

**Uniwersytet im. Adama Mickiewicza w Poznaniu**

**Wydział Filologii Polskiej i Klasycznej**

**Instytut Filologii Polskiej**

Dane Polskiej Bibliografii Literackiej jako dane badawcze.  
Perspektywa humanistyki cyfrowej i studium przypadku

Jakub Eichler

Rozprawa doktorska

promotor:  
prof. UAM dr hab. Konrad Dominas

promotor pomocniczy:  
dr Cezary Rosiński



Poznań 2023

# Spis Treści

<b>Wstęp</b> .....	<b>5</b>
<i>Digitalizacja literaturoznawstwa</i> .....	5
<i>Układ pracy i uwagi edytorskie</i> .....	16
<b>Rozdział 1. Humanistyka cyfrowa jako metoda – zarys i wyzwania</b> .....	<b>19</b>
<i>Zamieszkać na bazarze</i> .....	20
<i>Moda? Metoda? Metodologia? Szkoła?</i> .....	23
<i>Rys historyczny, rozwój i kontekst społeczny</i> .....	26
<i>Dane geograficzne i dotyczące położenia</i> .....	30
<i>Stan badań nad humanistyką cyfrową</i> .....	33
<i>Projekty cyfrowe</i> .....	40
<i>Globalizacja, dostępność, remediacja</i> .....	54
<i>Wątpliwości; jak nie działa humanistyka cyfrowa</i> .....	56
<i>Głosy krytyczne</i> .....	59
<i>Ku definicji humanistyki cyfrowej</i> .....	71
<i>Humanistyka cyfrowa w Polsce</i> .....	72
<i>Punkt wyjścia: projekty cyfrowe oparte na bibliografii</i> .....	74
<i>Uwagi końcowe</i> .....	76
<b>Rozdział 2. Otwarte dane badawcze</b> .....	<b>81</b>
<i>Idea ontologii versus Wielkie Wysypisko Wiadomości (z korporacyjnymi interesami w tle)</i> .....	81
<i>Stan badań i dobre praktyki w obszarze otwartych danych</i> .....	85
<i>Porządkowanie pojęć</i> .....	88
<i>Zasady FAIR w udostępnianiu danych</i> .....	92
<i>Obieg wymiany danych bibliograficznych</i> .....	95
<i>Uzgadnianie formatów</i> .....	96
<i>Bazy indeksowania czasopism a przypadek PBL</i> .....	97
<i>Otwarte licencje na udostępnianie danych</i> .....	103
<i>Dane wrażliwe</i> .....	107
<i>Plany zarządzania danymi badawczymi</i> .....	108
<i>Humanista cyfrowy kuratorem treści</i> .....	109
<i>Wiedza pozostaje rozproszona i domaga się nowych syntez</i> .....	109
<i>Wiedza na powierzchni</i> .....	110
<i>Wyzwania kwerendy w serwisach internetowych</i> .....	111
<i>Jaką narrację snuje sieć</i> .....	120

<i>Serwisy budowane społecznie</i> .....	121
<i>Kuracja treści</i> .....	126
<i>Dostęp</i> .....	128
<i>Coraz większe wolumeny danych</i> .....	128
<i>Futuryzm, czyli archiwistyka dla maszyn</i> .....	129
<i>Modele świadomości sieci</i> .....	132
<i>Uwagi końcowe</i> .....	134
<b>Rozdział 3. PBL jako przykład działalności przynależnej do humanistyki cyfrowej</b> .....	<b>136</b>
<i>PBL: od bibliografii do bibliograficznej bazy danych</i> .....	136
<i>Cele bibliografii</i> .....	136
<i>Rys historyczny i założenia</i> .....	137
<i>Pod czujnym okiem Cenzora</i> .....	139
<i>W stronę bazy danych</i> .....	143
<i>Bibliograf jako operator maszyny</i> .....	144
<i>PBL jako laboratorium wiedzy o kulturze</i> .....	145
<i>Nowy serwis PBL</i> .....	147
<i>Sieć instytucji związana z wytwarzaniem Bibliodata</i> .....	150
<i>PBL i inne podobne projekty na świecie</i> .....	153
<i>Porównanie z Biblioteką Kongresu</i> .....	154
<i>Baza retrospektywna i dążenie do aktualności</i> .....	157
<i>PBL cyfrowy: stan badań</i> .....	159
<i>Dla kogo powstaje PBL</i> .....	159
<i>Strategie i obecność PBL w serwisach społecznościowych</i> .....	174
<i>Połączone zbiory danych</i> .....	179
<i>Podsumowanie</i> .....	183
<b>Rozdział 4. Studium przypadku</b> .....	<b>185</b>
<i>Literatura na Świecie – hipertączya zza Żelaznej Kurtyny</i> .....	185
<i>Odbiorcy – między globalnością a lokalnością badań</i> .....	192
<i>Stan badań i inspiracje metodologiczne</i> .....	196
<i>Od Bibliodata do polisystemów</i> .....	204
<i>Wstępne przygotowanie techniczne</i> .....	206
<i>Analiza i przetwarzanie bibliografii LnŚ</i> .....	207
<i>Plan działań</i> .....	208
<i>Rozpoznanie i obróbka danych</i> .....	209
<i>Rozwój środków technicznych</i> .....	211
<i>Doświadczenia pracy z danymi</i> .....	212

<i>Pierwszy dataset - LnŚ</i> .....	216
<i>Praca z danymi PBL</i> .....	221
<i>Przetwarzanie danych Biblioteki Narodowej</i> .....	222
<i>MathFiction: baza akademicko-amatorska</i> .....	230
<i>Wizja połączonych zbioru danych</i> .....	236
<i>Wyzwania kompilacji źródeł</i> .....	237
<i>Charakterystyka datasetu</i> .....	239
<i>Obserwacje</i> .....	241
<i>Perspektywy i kolejne kroki</i> .....	242
<b>Podsumowanie</b> .....	<b>244</b>
<b>Zakończenie</b> .....	<b>246</b>
<b>Bibliografia</b> .....	<b>249</b>
<b>Spis ilustracji</b> .....	<b>270</b>
<b>Spis tabel</b> .....	<b>270</b>

# Wstęp

Praca koncentruje się na przedstawieniu zasobów Polskiej Bibliografii Literackiej jako źródła danych badawczych. PBL jest inicjatywą o korzeniach dokumentalistycznych, z których w XXI wieku wyrastają projekty humanistyki cyfrowej. Archiwistyczna praca dziesiątek (a niekiedy i setek) osób kontrybuuje obecnie do światowego ekosystemu wymiany danych w nauce jako dane bibliograficzne.

Niniejszy projekt stanowi też zapis prób mariażu wiedzy i doświadczeń literaturoznawcy z kompetencjami oraz praktykami sytuującymi się bliżej informatyki i informacji naukowej. Zamierzonym efektem końcowym było wytworzenie produktu, mieszczącego się w nurcie humanistyki cyfrowej, co wymagało podjęcia się próby opisu samego zjawiska *Digital Humanities*. Aby przybliżyć, jak mogą funkcjonować metody cyfrowe w literaturoznawstwie, chciałbym zaproponować osobom czytelniczym opis tekstu naukowego, który nigdy nie powstał. Początkiem będą realne interpretacje jednego utworu Marcina Świetlickiego, należące do współczesnej krytyki literackiej, pozostające w obrębie tradycyjnego literaturoznawstwa. Skupię się przede wszystkim na jednej interpretacji. Wychodzę od tradycyjnej polonistycznej interpretacji wiersza, aby odszukać i wskazać potencjał użycia narzędzi informatycznych. Następnie przenoszę refleksję na grunt ściśle związany z humanistyką cyfrową.

Digitalizacja literaturoznawstwa

Punktem odniesienia jest wiersz Marcina Świetlickiego pt. *McDonald's*. Główną interpretacją, do której się odwołuję, stanowi fragment zbioru szkiców krytycznych Pawła Próchniaka, *Promieniowanie tła*. Posiłkuję się także dla porównania spostrzeżeniami Mariana Stali i Wojciecha Bonowicza, o czym za chwilę. Moim głównym celem jest tutaj eksploracja nieznanego obszaru w interpretacji tekstu i wytworzenie pytań, które można zadać tekstowi, a następnie odpowiedzieć za pomocą cyfrowych narzędzi. Takie spekulatywne czytanie więcej ma nam powiedzieć o możliwościach literaturoznawstwa, wspieranego narzędziami cyfrowymi, niż o tym konkretnym przypadku wiersza. Wzorem dwóch interpretatorów, do których pracy się odnoszę, przytaczam ten tekst w całości:

Znajduję ślad twoich zębów w obcym mieście  
Znajduję ślad twoich zębów na swoim ramieniu.  
Znajduję ślad twoich zębów w lustrze.  
Czasami jestem hamburgerem.

Czasami jestem hamburgerem.  
Sterczy ze mnie sałata i musztarda cieknie.  
Czasami jestem podobny śmiertelnie  
do wszystkich innych hamburgerów.

Pierwsza warstwa: skóra.  
Druga warstwa: krew.  
Trzecia warstwa: kości.  
Czwarta warstwa: dusza.

A ślad  
twoich zębów  
jest najgłębiej,  
najgłębiej.

(94)<sup>1</sup>

---

<sup>1</sup> Cyt. za: Paweł Próchniak, „Czasami jestem”, w *Promieniowanie tła: szkice o wierszach i czytaniu*, Wydanie I (Wołowiec: Wydawnictwo Czarne, 2020), 25–37.

Marian Stala daje czytelnikom następującą propozycję interpretacyjną<sup>2</sup>: wiersz Świetlickiego to przede wszystkim wiersz miłosny<sup>3</sup>. Krytyk zauważa muzyczno-wokalny aspekt twórczości autora *Zimnych krajów*, a nawet czuje się w obowiązku rozróżniać, czy powyższy utwór jest bardziej wierszem, czy piosenką. Dla odbiorców Świetlickiego i jego grupy muzycznej Świetliki to pytanie zasadne np. z uwagi na kontekst poznania tego utworu lub najczęstszej formy kontaktu z nim. Dla Stali kwestia możliwego zrozumienia lub niezrozumienia „tego, co mówi poeta”. W odwołaniu do (konsumpcyjnej) kultury masowej interpretator widzi chęć nawiązania kontaktu z odbiorcą i zwrócenie uwagi na społeczne wyobrażenie aktu jedzenia hamburgerów. Dalej interpretator wykląda swój ogląd na centralny koncept tekstu: oto mówiący w wierszu mężczyzna porównuje siebie do hamburgera, „zjadanego” w akcie miłosnym przez kobietę, a „cały wywód wieńczy konkluzja, która jest miłosnym wyznaniem i prośbą”. Na ogólniejszym poziomie dla Stali interpretacja brzmi: „tym, co w człowieku najważniejsze – jest zdolność przyjmowania uczuć kogoś drugiego, sztuka bycia kochanym”. Pojawiają się jednak dwie komplikacje. Po pierwsze, krakowski badacz zwraca uwagę na fakt, że tylko ślady zębów i ugryzienie dają w wierszu znak obecności owej kobiety (może to uruchamiać konteksty kobiety fatalnej, ale Stala ogranicza takie zapędy do stwierdzenia niejasnej obawy). Druga wątpliwość wiąże się z innym obrazowaniem nałożonym na miłosne: wizją narcystyczną, wierszem o sobie, autoportretem przepełnionym słabością, co krytyk parafrazuje jako komunikat podmiotu lirycznego: „jestem ochłapem mięsa, który sprzedają jako smaczne danie”. Dla podkreślenia wagi tego napięcia, Stala przytacza cytaty z jeszcze dwóch utworów, aby pokazać, żywioły obecne w *McDonald's*: z jednej strony zwierzęce oddanie w miłosnym uniesieniu, z drugiej – zagubienie, pułapka i uczucie bycia zjadanym. Ta interpretacja kończy się prywatną refleksją Stali, że te utwory Świetlickiego, które wywierają na badaczu wrażenie, bawią go i przerażają.

Tymczasem Wojciech Bonowicz umieszcza ten sam wiersz w antologii wierszy religijnych Marcina Świetlickiego *Nieoczywiste*<sup>4</sup>. Poza odwołaniem do postaci Jonasza we wcześniej cytowanej interpretacji, nie mieliśmy sposobności zbadania tej perspektywy. Z pewnością

---

<sup>2</sup> Marian Stala, „Jedzenie hamburgerów”, w *Blisko wiersza: 30 interpretacji* (Kraków: Społeczny Instytut Wydawniczy Znak, 2013), 254–61. W dalszej części, odnosząc się do fragmentów interpretacji Stali w cudzysłowie, powołuję się zawsze na wyimki z tego właśnie tekstu.

<sup>3</sup> Do interpretacji tekstu *McDonald's* jako wiersza miłosnego odwoływał się też Krzysztof Hoffmann w dyskusji na spotkaniu z Wojciechem Bonowiczem i Marcinem Świetlickim w Poznaniu, towarzyszącym promocji zbioru *Nieoczywiste* w 2007 roku, gdzie wygłoszono referaty studenckie – archiwum własne (JE).

<sup>4</sup> Marcin Świetlicki, *Nieoczywiste: 77 wierszy religijnych Marcina Świetlickiego*, red. Wojciech Bonowicz, Seria Czarna (Kraków: Wydawnictwo EMG, 2007).

zwrot do odbiorcy, obecny w wierszu, można spotkać w utworach religijnych. Mają też one wątki miłosne, choć raczej nazwalibyśmy je uwielbieniem istoty wyższej i wybierając takie odczytanie – zrezygnujemy z erotycznego, proponowanego przez Stałę (lub syntetyzując, będzie trzeba się zdecydować na połączenie erotycznego z religijnym, z sugestią, że mamy przykład obrazowania ekstazy). Fakt, że w tekście Świetlickiego mamy do czynienia z opisem samoświadomości i w drodze poprzez kolejne warstwy najgłębszym poziomem staje się „dusza”, będzie najmocniejszym argumentem za taką próbą odczytania. Fakt, że wiersz jest 77 sugeruje, że trzeba było wybrać i ograniczyć się do zamkniętego zbioru. Czy mamy rozumieć, że właśnie te wiersze nadawały się najbardziej do tematycznego zestawienia? Nasuwa się też pytanie, na ile ważny był limit liczbowy i jakie niósł konsekwencje dla pominięcia lub włączenia konkretnych dzieł do tej książki<sup>5</sup>. Tu proszę czytelników o zapamiętanie następującej myśli: gest umieszczenia przez antologistę utworu w zbiorze (kompilacja) jest tożsamy z umieszczeniem rekordu w tematycznej bazie danych przez bibliografa lub dowolną inną osobę, tworzącą tego rodzaju wirtualny zbiór. Przyjęto jakiś zestaw kryteriów, które definiują zakres *kolekcji*. Umieszczenie *danej* w *kolekcji* jest wyborem i niesie konsekwencje dla odbioru tej poszczególnej informacji, jak i zagregowanej całości. W danych bibliograficznych często mamy do czynienia z przyporządkowaniem tematycznym dzieł. Określenie przynależności tekstu do wierszy religijnych (jak w tym wypadku), jak i każdej innej kategorii, jest wytworzeniem (meta)danej, związanej z tym konkretnym rekordem. To działanie umieszcza zarazem utwór w konfiguracji do pozostałych elementów preparowanego przez twórcę zbioru.

Ostatni głos o tym wierszu, który chciałbym przytoczyć, to esej Pawła Próchniaka *Czasami jestem*, z książki *Promieniowanie tła*<sup>6</sup>. Szkic został podzielony na 10 części o zróżnicowanej długości – otwiera go przytoczenie omawianego wiersza, a każda z numerowanych całości stanowi swoisty obraz krytycznej lektury lub refleksję czy uwagę na temat warsztatu egzegezy,

---

<sup>5</sup> Bonowicz odnosi się do zamysłu, choć nie do końca do samej liczby, we wstępie. Koncepcja zbioru jest tłumaczona anegdotą, która wyzwoliła intuicję: antologista, unawany za poetę religijnego, został zapytany przez uczestnika spotkania literackiego w Olkuszu, kogo jeszcze z żyjących polskich poetów by wskazał jako poetę religijnego wśród swoich rówieśników; Bonowicz miał bez wahania wskazać Świetlickiego, a im dłużej do tego pomysłu wracał, tym bardziej materializowała się książka ze zbiorem wierszy. Kryterium doboru było stworzenie opowieści o bohaterze podzielonej na cztery części. Wprowadzenie ma charakter dość manifestowy, z sugestią potrzeby wprowadzenia do debaty publicznej treści związanej z doświadczeniem religijnym. Zob. Świetlicki, 5-9.

<sup>6</sup> Próchniak, „Czasami jestem”. Pierwodruk: Paweł Próchniak, „Czasami jestem (czytając wiersz McDonald’s)”, w *Mistrz świata. Szkice o twórczości Marcina Świetlickiego*, red. Piotr Śliwiński (Poznań, 2011), 149–57. W przytoczeniu interpretacji Próchniaka wszystkie cytaty w cudzysłowie i cytatach blokowych pochodzą z tego tekstu.



nasuwające się w związku z tekstem Świetlickiego. Autor od początku akcentuje w utworze motyw *zranienia*. Interpretacja staje się też okazją do mówienia o interpretacji – komentarza, nazywającego wiersz partyturą i przywołania słów Rolanda Barthesa świadczących o czymś odwrotnym: o odkrywaniu „drzenia sensów” w utworach, co tak naprawdę odgrywałoby sam wiersz. Próchniak interesuje „fundamentalna nieoczywistość poezji”, która wyłania się ze swoistej przestrzeni niedookreśleń (upraszczając powiedzielibyśmy: z otwartości na interpretację). „I jeśli za sprawą wiersza wybrzmiewa coś, czego inaczej powiedzieć się nie da, jeśli w wierszu odnajdują głos rzeczy przejmujące i ważne, to jednocześnie ten sam wiersz prowadzi w nieznaną – poza horyzont wyobraźni, w stronę czegoś, co jest nie do pomyślenia i może już na zawsze takie pozostanie.”

Następnie interpretator przeprowadza rozbiór utworu. Nie wprost Próchniak odwołuje się tu także do znanej tektonicznej wykładni dzieła literackiego Romana Ingardena<sup>7</sup>. Autor przyrównuje wymienione przez Świetlickiego warstwy kolejnych części ciała, z których najbardziej wewnętrzną okazuje się dusza, do warstw znaczących w tekście. Daje to asumpt do zastosowania figury znanej w krytyce literackiej, w której wiersz opowiada nam o dziele literackim w ogóle, staje się wierszem o wierszu. Próchniak sygnalizuje, że zestawia „rudymenta”, aby przez model podstawowej lektury analitycznej (możemy powiedzieć: tego, co widać na powierzchni) dotrzeć do „porządku nieoczywistych znaczeń”.

Pisząc o strukturze czterech strof, interpretator zauważa zwornik budowy tekstu: „śląd” (po ugryzieniu). W budowie tekstu dostrzega kurczenie się, redukcję do ciszy. Próchniak zauważa:

W trzech pierwszych strofach poszczególne wersy są modulacyjnie domknięte, tworzą zwarte całości intonacyjne, będące – z jednym wyjątkiem – również całościami składniowymi. Ostatni czterowiersz to jedno zdanie – **rozbite, potrzaskane, złożone z powtórzeń, rozsypujące się, wypowiedane z wysiłkiem, na wydechu** (wyróżnienie moje – JE).

---

<sup>7</sup> Roman Ingarden, „Formy poznawania dzieła literackiego”, *Pamiętnik Literacki: czasopismo kwartalne poświęcone historii i krytyce literatury polskiej* 33, nr 1–4 (1936): 163–92. Przypomnijmy, że u Ingardena mamy podział na warstwy: a) brzmień słownych, b) znaczeniową, c) uschematyzowanych wyglądków (proponuję też rozumienie: wyobrażeń - JE), d) przedmiotów przedstawionych. Dla uporządkowania, u Świetlickiego warstwy nazwane *explicite* w wierszu to „Pierwsza warstwa: skóra./ Druga warstwa: krew./ Trzecia warstwa: kości./ Czwarta warstwa: dusza”. Próchniak mówi o: 1) postaci graficznej, 2) warstwie brzmieniowej, 3) semantyce i składni, 4) strukturze opisywanego świata – jest to częściowo zbieżne ze schematem Ingardenowskim.

Zauważmy tutaj, że praca analityczna odbywa się właściwie jednocześnie z interpretacyjną, tzn. przynajmniej już w gotowym tekście krytycznym interpretator przedstawia informacje z obu porządków: 1) przeglądu części składowych oraz ich wzajemnych relacji 2) konsekwencji dla obrazowania i subiektywnego nazywania funkcji tych elementów. Podobnie dzieje się w opisie aspektu brzmieniowego:

[O]bok rytmów powtórzeń i enumeracji słyhać też rytm trójzestrojowego wiersza tonicznego i rytmiczny oddech coraz krótszych fraz. te przenikające się takty prowadzą czytelnika przez kolejne warstwy wypowiedzanego świata. **Mówią o wewnętrznym pulsie egzystencji – skłóconym w sobie i zarazem mocnym** (wyróżnienie moje – JE).

To, co jednak interesujące w interpretacji Próchniaka dzieje się dalej w tej samej części: autor zwraca bowiem uwagę na celowe nagromadzenie głosek, które tworzą fonetyczne napięcie ze sobą. W dużym uproszczeniu, głoska „m” reprezentuje tutaj „materialną spoistość”, co kontrastują inicjalne ostre świsty głosek w słowach „śląd” i „zębów”. Z jednej strony mamy „płynną miękkość głosek opartych na rezonansie”, z drugiej „twarde, rezonujące, frykatywne” „r”. Za najmocniejszą frazę interpretator uważa tu wyliczenie kolejnych warstw w strofie trzeciej, a wcześniej podsuwa widoczną realizację opisywanego przez siebie efektu w drugiej („Sterczy ze mnie sałata i musztarda cieknie.”). Próchniak zwraca uwagę na to, że ta gra głosek milknie pod koniec tekstu.

W analizie składowych brzmienia wiersza, Próchniak przypisuje funkcje znaczeniowe lub związane z atmosferą. Dwie uwagi: po pierwsze, w odróżnieniu od Stali, autor *Wierszy na wietrze* nie przeprowadza rozróżnienia między tekstowością graficzną, a muzyczno-wokalną, zatem przyjmujemy, że jest to wariant opisu skoncentrowany na statycznym tekście drukowanym i wyświetlanym, w odróżnieniu do realizacji dźwiękowej. Po drugie: sposób organizacji tekstu możemy zatem opisać w odwołaniu do alfabetu fonetycznego, choć konkretne odczytania na głos i wykonania estradowe lub studyjne mogą się różnić między sobą<sup>8</sup>. Wynika z tego, że zgodnie z estetyczną funkcją nadmiaru nagromadzenia środków stylistycznych, możemy prześledzić dokonany przez Świetlickiego wybór głosek (słów, które je zawierają) i rozważyć: a) wpływ na brzmienie (czy rzeczywiście dobór środków wzmacnia

---

<sup>8</sup> Alternatywą wobec wszelkich badań muzyczności są także odwołania do anagramów de Saussure’a, mogłyby tu znaleźć zastosowanie. Por. Adam Dziadek, „Anagramy Ferdynanda de Saussure’a: historia pewnej rewolucji”, *Teksty Drugie : teoria literatury, krytyka, interpretacja* 71, nr 6 (2001): 109–25.

opisany przez Próchniaka efekt), b) alternatywy (czy aby zrealizować opisany efekt obecności załączków znaczeń w brzmieniu, dobrano najlepsze środki – np. jak słowo „musztarda” jest lepsze od słowa „keczup”, „majonez” czy wyrażenia „sos czosnkowy”), c) istotność wobec szerszego kontekstu (czy podana interpretacja zestroju głosek może być falsyfikowana na innych tekstach – np. co, jeśli podobne zestroje znajdziemy w innych wierszach tego autora, innych tekstach poetyckich; jak zaprezentuje się porównanie z tekstami niepoetyckimi).

Próchniak, podobnie jak Stala, używa innego tekstu Świetlickiego, by oświetlić dalszą część wywodu: przywołuję frazę „wyrazy, o które zahacza twój język” z wiersza bez tytułu z tomu *Zimne kraje*. Służy to podkreśleniu wagi użytych określeń – słów „węzłowych”, jak „McDonald’s”, „hamburger”, „lustro”, „śląd zębów” i wyrażenie „podobny śmiertelnie” – wiemy o nich, że nie pojawiły się przypadkowo w treści. Zwłaszcza to ostatnie określenie wywołuje zainteresowanie badacza. Próchniak pyta, co śmiertelność łączy z podobieństwem, „[c]o na skali intensyfikacji określeń oznacza »podobny śmiertelnie«?”. Zestawienie wszystkich pojęć wprowadza dynamikę: lustrzana symetria, symetria kochanków i obecność śmierci – być może jako niechybność, ale możliwe, że jako właściwość erotycznej relacji.

Pobrzmiwają tu echa wątków docierania do granic, eksces jak u Bataille’a: w innym miejscu ten sam interpretator zauważył na temat innego tekstu, że motyw negacji granic odsłania wątki suwerenności erotyzmu i jego erotycznej prawdy. Tak pisze o owej prawdzie erotycznej:

[J]ej osnową jest nagły rozlewający się nadmiar – wstrząsający, skandaliczny, graniczący bezpośrednio z występkiem, z wynaturzeniem, spokrewniony ze złem. I właśnie dlatego nurt prawdy erotycznej – nurt burzliwy i bezinteresowny – pracuje z siłą religijnego rytuału. Daje dostęp do »beźmiaru kryjącego się w miłości istoty śmiertelnej«. Zamyka w sobie żywą treść i dramatyzm egzystencji – jej gwałtowność, jej intensywność podszyta chłodem śmierci i czymś jeszcze bardziej zimnym, czymś rozpaczliwym, przesywającym do żywego, porażającym, czymś, co sprawia, że nasza miłość nieustannie zatracą się w materii nieporuszonej i martwej<sup>9</sup>.

Choć trzeba dodać, że gdy Próchniak omawia *Martwego* Bataille’a, to odnajduje w erotycznym ekscesie wzniosłość i *sacrum*, a omawiane u Świetlickiego obrazy śladu ugryzienia oraz

---

<sup>9</sup> Paweł Próchniak i Georges Bataille, „Bataille: słoneczna plama (na marginesie *Martwego*)”, w *Martwy* (Kraków, 2010), 43-44.

warstw „ja” wskazywałyby jednak na obecność granic (obrazowanie transgresji jest zgoła inne w każdym z tych tekstów).

Interpretacja Próchniaka kończy się dwojakim gestem. Najpierw przedstawieniem konkretnego obrazu w wierszu; następnie próbą otwarcia i sformułowania ogólniejszej zasady czy drogowskazu (interpretacja jest o wierszu, wiersz jest nie tylko o tym, o czym jest). Tym konkretnym obrazem jest mężczyzna w hotelowym pokoju patrzący na swoje odbicie w lustrze ze śladem zębów na ramieniu. Ta sytuacja uruchamia refleksję o duszy i tym, co skrywane najgłębiej. Ślad zębów nosi znamiona intymności i bólu. Próba dotarcia „najgłębiej” jest skwitowana milczeniem. Dla Próchniaka stanowi to asumpt do skonkludowania:

Jeśli prawda wiersza leży istotnie gdzieś po drugiej stronie słów, w tej obietnicy, którą składają nam słowa, jeśli kryje się w ich najgłębszej głębi, to zarazem przecież prawda ta jest całkowicie jawna – naoczna, w całej swej oczywistości zjawia się w ruchu lektury, w spojrzeniu ogarniającym kształt i formę zapisu, w podążaniu za porządkiem brzmień, za rytmem, za tokiem słów i obrazów, po ścieżkach, które obrysowują wypowiedaną rzeczywistość, stają się mapą wyobraźni zakorzenionej w tekście, uczepionej jego słów.

Każdy wiersz, także opisywany przez Próchniaka, jest więc potencjalnością, która otwiera się na lekturę. Aktualizuje się w momencie interpretowania, oferuje ciszę, następnie obrastając wrażeniami lektury, które nadają mu sensu.

Mamy tutaj zatem do czynienia z trzema diagnozami (miłosna, religijna i erotyczno-tanatyczna), częściowo w jakiejś mierze się wykluczającymi (wiersz miłosny czy religijny?), częściowo zbieżnymi (wiersz dotyczący spraw śmierci i miłości). Każdy z krytyków opiera swoje działanie interpretacyjne na własnych przesłankach i wywodzi z tego samego tekstu różne interpretacje. W moim odczuciu – każda z tych interpretacji jest uprawniona i każda z nich znajduje argumenty w tekście na poparcie wniosków. Podobnie jednak do adresata<sup>10</sup> słów podmiotu mówiącego w *McDonald's*, każdy interpretator pozostawia na tekście ślady własnego „ugryzienia” – właśnie przez arbitralne decyzje interpretatorskie (uwzględniające dobór przywołanych kontekstów), a także pomimo wspólnego punktu wyjścia w postaci tego

---

<sup>10</sup> Adresatki, osoby lub istoty, do której słowa są adresowane.

samemu wiersza. Tym, co może oferować humanistyka cyfrowa, jest potraktowanie tekstu jako zbioru danych, co sprowadzi tekst do pewnej obiektywizacji.

Po przekształceniu tekstu na zapis fonetyczny możemy sprawdzić tezę Próchniaka o częstym występowaniu głosek nosowych w sąsiedztwie „szeleszczących” (ale już nie, czy może się to kojarzyć ze ściekaniem musztardy po palcach i policzkach podczas konsumpcji hamburgera)<sup>11</sup>. Ustalimy frekwencję obu typów, ich procentowe występowanie w całym tekście. Następnie możemy tak opisany zbiór porównać z dowolnym korpusem: resztą tomu poetyckiego, wszystkimi wierszami kiedykolwiek napisanymi przez Świetlickiego, czy wreszcie zawartością książek poetyckich wydanych w tym samym roku lub dekadzie. Możemy także przyjąć klucz tematyczny i oprzeć korpus wyłącznie na wierszach dotyczących konsumpcji i konsumpcjonizmu, jedzenia i zwyczajów związanych z kulinariami. Warunkiem jest oczywiście przygotowanie korpusu z użyciem tej samej zasady, tj. konwersji treści na alfabet fonetyczny. Nie chodzi o to, czy sfalsyfikujemy lub potwierdzimy (prawdopodobnie słuszną) diagnozę Próchniaka na temat warstwy brzmieniowej wiersza. Takie narzędzie daje nam poparcie tezy w postaci twardych danych i czytelnej, zrozumiałej dla odbiorcy metodologii<sup>12</sup>.

Przeprowadza się też na tekstach badania stylometryczne, badając podobieństwo jednych utworów do drugich. Czy tekst *McDonald's* okaże się podobny pod względem zastosowanej leksyki, układu treści i budowania frazy do innych tekstów, które będziemy podejrzewać o tematykę religijną? Tego rodzaju interpretacja, idąc za gestem Bonowicza jako antologisty, opierałaby się np. na podkreśleniu wątków nieobecności (siły wyższej) lub piętna (zła, grzechu, winy), mogłaby także poprzez obrazowanie związane z ugryzieniem odwołać się do biblijnej historii Adama i Ewy (w zbiorze danych jest to niewidoczne wprost dla maszyny, która je przetwarza, ale uważny krytyk pamięta, że nawiązanie do tego motywu znajdziemy już w wierszu *Wstęp* z tomu *Zimne kraje*).

Z kolei rozwijając podejście Mariana Stali do tego tekstu, czy użyta leksyka i obrazy, które możemy otagować na potrzeby maszyny, czynią z wątku miłosnego ten dominujący

---

<sup>11</sup> Moglibyśmy za to dla przykładu porównać nagrania odgłosów konsumpcji i stopień ich podobieństwa brzmieniowego do wykonania utworu.

<sup>12</sup> Taka metoda z jednej strony usuwa „szum” – rzeczy nieweryfikowalne, pozostawiając informację w stanie czystym, z drugiej – wspiera ów „szum” (odczytanie afektywne, zaangażowane, żarliwe) za pomocą dostępnych obiektywizacji wywiedzionych z omawianego tekstu. Ten aspekt obiektywizmu służebnych metod cyfrowych będą się jeszcze starał zniuansować w dalszej części. Wszak spotkanie na styku praktyk tak od siebie odmiennych jak literaturoznawstwo i wytwarzanie oprogramowania (czy szerzej, inżynieria) musi prowadzić do napięć, zanim ustali się wspólny sposób działania.

w wierszu? W jakim stopniu? Czy możemy na korpusie wierszy wskazać szersze otoczenie tekstu, by uwypuklić, o jaki rodzaj tekstu miłosnego chodzi? Stala w swoim szkicu proponuje rozpięcie *McDonald's* między dwoma innymi tekstami: Świetlickiego wykonaniem własnym utworu Iggy'ego Popy *Now I Wanna Be Your Dog* oraz autorskim tekstem *Jonasz* – buduje to pewną dramaturgię interpretacji, tym bardziej, że *Jedzenie hamburgerów* kończy się afektywnym komentarzem krytyka. Wspiera to odbiór wiersza jako erotyku. Czy jeśli Stala powołuje się na aspekt „tego, co mówi poeta” i podaje swoje skojarzenie, to możemy szukać wszelkich możliwych wypowiedzi Marcina Świetlickiego (osoby) i po przeanalizowaniu ich maszynowo dotrzeć do stojącej za *McDonald's* intuicji, by z całym przekonaniem powiedzieć: „to jest to, co mówi poeta”?

Co, jeśli subiektywne odczucia interpretatorów i wybór konkretnej trajektorii odczytania kształtują przez lata percepcję czytelniczą tego wiersza, a poprzez porównanie z grupą podobnych utworów przestanie mieć taki oddźwięk, jak sugerują zawodowi czytelnicy? Dla mnie ów tekst jest tak szalenie wieloznaczny, że można podstawić bardzo wiele (ale prawdopodobnie nie: nieskończenie wiele) tematów i argumentować, że to one są głównym wątkiem wiersza. Co, jeśli *McDonald's* jest (dla nas) o sławie, nadwrażliwości, depresji, wiedzy, nałogu, ukąszeniu heglowskim lub przez kleszcza, traumie, władzy, wojnie, incydencie z udziałem groźnego zwierzęcia lub toksycznych wzorcach męskości?<sup>13</sup>

Jeśli krytycy w poszukiwaniu owej tematycznej osi odwołują się do szerszego dorobku autora, robią to w granicach pozwalających na zrozumienie tematu przez siebie i przez czytelnika. Maszynowe odczytanie uruchomiłoby takie porównanie na dużo większą skalę, rozszerzając percepcję jednych i drugich. W opisanych działaniach interpretacji i rozumienia utworu rysują się trzy perspektywy: blisko wiersza (Stala), dalej, w średnim dystansie (Próchniak), najdalej (Bonowicz). Proponowane zautomatyzowanie procesów analizy tekstu z odwołaniem

---

<sup>13</sup> Andrew Piper podaje bardzo ciekawy przykład emergentnych tematów: takich, które wymykają się weryfikacji zbioru tekstów przez człowieka. Analiza maszynowa poprzez klastrowanie grup tematycznych w przetwarzanych tekstach wybiera nieoczekiwane konfiguracje. Takim przykładem było badanie korpusu XIX-wiecznych niemieckich powieści, gdzie pojawił się temat (*topic*) nr 150, łączyący idee życia i śmierci, który okazał się być jednym z bardziej żywotnych w zbiorze: „My interest here is not with »life« in all of its nineteenth-century discursive forms. Rather, I am interested in the more limited question of how this topic, if we can call it that, organizes language in long narrative forms that are themselves very often about individual lives. What conceptual or metaphorical concerns does the topos of life and death carry with it through time? How does it establish new kinds of relationships between documents that we have not seen before?” – wytworzenie takiego nowego aparatu percepcji jest stawką jeśli nie niniejszego opracowania, to humanistyki cyfrowej w ogóle. Por. Andrew Piper, *Enumerations: data and literary study* (Chicago ; London: The University of Chicago Press, 2018), 68.

do korpusów innych dzieł byłoby perspektywą *dalej niż najdalej*. Może nawet w poszukiwaniu miejsca, o którym pisze Próchniak: poza horyzontem wyobraźni.

Krytyk literacki polega na wnikliwej lekturze wielu tekstów, sile swoich skojarzeń i przyjętej metodologii interpretacyjnej. Jest jednak istotą biologiczną o ograniczonych mocach poznawczych<sup>14</sup>. Maszynowa analiza tekstu opiera się na wprowadzonych danych<sup>15</sup> oraz przetwarzaniu informacji na skalę niedostępną człowiekowi. Intrygujące zatem wydaje się poszukiwanie takiej ilości i konfiguracji dostępnych nam informacji na dany temat, aby przebadać je i znaleźć zupełnie nieoczekiwany wynik. Wprowadzenie logiki „biznesowej” nakazuje zapytać: skoro interpretator najpierw określa „rudymenta” w pracy analitycznej, a później przechodzi do interpretacji, to czy możemy ten pierwszy etap zautomatyzować? Niesie to ciekawy walor (lub wpisana w proces „wadę fabryczną”) oddzielenia od siebie analizy i interpretacji. Jak przekonaliśmy się wcześniej, w tekście krytycznym mogą one zachodzić jednocześnie<sup>16</sup>.

Z perspektywy takiej obiektywizacji, polonistyczno-literaturoznawcza wiedza, niepodlegająca dyskusji, jaką posiadam o tekście *McDonald's*, to: autorstwo konkretnej osoby, co do którego badacze są zgodni; adresy bibliograficzne pierwodruku i kolejnych wydań wiersza; znane mi, przytoczone wyżej, teksty krytyczne, które odwołują się do tego dzieła, wraz z ich danymi bibliograficznymi; inne teksty tego samego autora, które także mogę wskazać pod kątem okoliczności wydania, jako dane bibliograficzne. Prawda, że jest to mało, w świetle interpretacji ukazanych wcześniej? Z jednej strony owszem, z drugiej – są to minima opisujące

---

<sup>14</sup> Porównuję tutaj dwa szkice Próchniaka, datowane z odstępem roku od siebie i poruszające tematy erotyczno-tanatyeczne, gdyż ich wątki wydały mi się mocno zbieżne. Maszynowa analiza *tekstów krytycznych* mogłaby zostać również nakierowana na poszukiwanie tego rodzaju prawidłowości (powracających figur, wątków, przykładów...). Inny taki przykład, który chciałbym podsunąć czytelnikom jako potencjalne ćwiczenie, to relacja działalności Andrzeja Sosnowskiego jako poety oraz jako tłumacza – można oczywiście wybrać dowolnego innego autora, parającego się obiema tymi pracami (Julia Fiedorczuk, Jacek Dehnel), choć z zastrzeżeniem, że najlepsi będą poeci często używający we własnych utworach zapożyczeń, sampli i *coverów* – stąd początkowy pomysł z Sosnowskim. Czy praca translatorska przechodzi do treści wierszy? Czy możemy wskazać istotne statystycznie transfery słów, zdań, znaczeń i tematów z przekładów do autorskiej poezji tłumacza? Nie jest to pytanie o oryginalność, a raczej o proces, psychologię twórczości i samoświadomość literacką.

<sup>15</sup> Dlatego tak istotna jest transparentna prezentacja metody tworzenia korpusu. Na etapie wprowadzania danych może się pojawić stronniczość – nieuwzględnienie istotnych elementów, nadreprezentacja jednego typu treści itp. Potrzebne jest rzetelne informowanie o przyjętym sposobie przygotowania danych wejściowych. Najlepszym wyjściem jest, poza samą prezentacją wyników, udostępnienie surowych danych, na których się pracowało.

<sup>16</sup> Choć nie będę utrzymywać, że jest tak w procesie myślowym krytyka – tekst zawierający interpretację jako gotowy produkt pracy interpretatora, stanowi już prezentację zastosowania narzędzi krytycznych do dzieła oraz przemyśleń egzegetyczno-aksjologicznych. Kompozycja szkicu krytycznego nie musi odtwarzać procesu dochodzenia do proponowanej interpretacji w sposób linearny.

system literacki (rozumiany jako produkcja i recepcja dzieł). W dalszej części pracy to właśnie nastawienie dominuje: spojrzenie na te elementy, co do których nie odbywa się negocjowanie sensu czy poszukiwanie niejasności, niedopowiedzeń. Również w przypadku cokolwiek efemerycznej i z definicji polisemicznej dziedziny poezji, to próba dotarcia do spraw, o których literaturoznawca może zaświadczyć z całą pewnością i które nie powinny być kontestowane w agonie (możemy delibrować nad tematyką wiersza, ale już nie w jakiej książce poetyckiej został wydany)<sup>17</sup>. Z nich oto obieramy punkt wyjścia do dociekań dążących w stronę uprawiania humanistyki cyfrowej, patrzącej na dane zgromadzone na temat kultury.

Czy to odziera interpretację literaturoznawczą ze swoistej magii, a – rozwijając myśl Waltera Benjamin – samemu dziełu jeszcze dobitniej odmawia aury?<sup>18</sup> W moim osobistym odczuciu tak się nie dzieje, gdyż operatorem narzędzi analizy, podobnie jak jej odbiorcą, pozostaje człowiek. Z perspektywy współistnienia różnych perspektyw badawczych możemy raczej mówić o komplementarności różnych odczytań literatury, jak i każdej innej dziedziny twórczości produkującej teksty kultury – w tym odczytań wspomaganym narzędziami cyfrowymi. Spróbuję zilustrować to paralelę metod informatyki i literaturoznawstwa w pierwszym rozdziale. Najpierw odniosę się jednak do budowy całości prezentowanego tekstu i charakterystyki stylistycznej wywodu.

Układ pracy i uwagi edytorskie

Pierwsze dwa rozdziały tekstu pełnią dwie funkcje: relacjonują poszukiwanie pola badawczego dla humanistyki cyfrowej i referują stan badań w dwóch potrzebnych dla niego zakresach: humanistyka cyfrowa jako całość oraz dane badawcze w humanistyce. Druga część, rozdziały

---

<sup>17</sup> O ile nastąpiłaby taka dyskusja, powinno ją zamknąć wyraźne wskazanie danych dotyczących wydania. W wypadku wiersza Świetlickiego wszystkie fakty systemu literackiego są znane. Można tu wszak zauważyć, że nie zawsze literaturoznawca, a w szczególności historyk literatury, dysponuje pełną informacją i w takich wypadkach tego rodzaju dyskusja (np. o pierwodruku, pierwszym ogłoszeniu tekstu lub prezentacji przed publicznością itd.) ma znaczenie.

<sup>18</sup> Dobrze znany przykład Benjaminowski z lat 30. XX wieku dotyczy dzieł malarskich, które występują w jednym egzemplarzu i tracą na powielaniu ich wizerunku w różnej formie za pomocą środków technicznych. Widzę tutaj paralelę ze względu na 1) sposób czytania tekstu kultury wywodzący się np. z tradycji hermeneutyki, w którym jako interpretatorzy odkrywamy ulotny sens dzieła (narzędzia cyfrowe ograniczałyby tu dowolność takich poszukiwań i umieszczałyby tekst w sferze bliższej faktom komunikacji, jak brzmienie zgłosek i analiza znaczeń oparta na frekwencji w dużym zbiorze) oraz 2) wpływ technologii na to, jak patrzymy na samo dzieło i je interpretujemy – w myśli Benjaminowskiej wartościowany negatywnie, ale dzięki humanistyce cyfrowej poddany ożywczej rewizji przez zestawienie zapoznanych treści ze statystyką, programowaniem (łatwość przeszukań i przetworzeń tekstu oraz zestawiania z innymi treściami) czy intermedialnością (inkorporowanie tekstów kultury z różnych porządków medialnych dla celów badawczych, a nawet na styku badań i działalności twórczej czy performatywnej). Por. Walter Benjamin, „Dzieło sztuki w epoce jego reprodukowalności technicznej”, w *Twórca jako wytwórca: eseje i rozprawy*, tłum. Robert Reszke (Warszawa: Wydawnictwo KR, 2011), 23–51.



trzeci i czwarty, zawężają zakres działania do minimalnego literaturoznawczego projektu cyfrowego z wykorzystaniem danych bibliograficznych, jako dowodu na użyteczność przedstawionych źródeł.

W pierwszym rozdziale omawiam ogólne założenia humanistyki cyfrowej wraz z jej krytyką i staram się opisać w relacji do istniejących dyscyplin. Odniesieniem są tutaj tradycyjna humanistyka, ze szczególnym uwzględnieniem literaturoznawstwa, jak i komponenty techniczne pochodzące głównie z informatyki.

W rozdziale drugim omawiam kwestie otwartych (meta)danych i ontologii w humanistyce. Zarysowuję też perspektywę korzystania z danych badawczych w naukach humanistycznych, z naciskiem na literaturoznawstwo. Na przykładzie recepcji twórczości wybranej autorki, Andrei Dworkin, ukazuję krajobraz rozproszenia danych dostępnych użytkownikom (w tym badaczom) i analizuję sposoby opisu z wykorzystaniem różnorodnych źródeł.

Trzecia część stanowi omówienie PBL jako bibliografii dziedzinowej, również w zarysie historycznym (przemiany), a docelowo jako inicjatywy humanistyki cyfrowej. Nakreślony zostaje kontekst tradycji autorskiej metody w tym pracy dokumentalistycznej, tła przemian metodologicznych, procesu digitalizacji zasobów oraz planów rozwoju.

Ostatnia część to dążenie do zbudowania *datasetu* z danych połączonych PBL oraz innych źródeł (zasoby Biblioteki Narodowej, bibliografia *Literatury na Świecie* autorstwa Ireny Bednarz, baza akademicko-społecznościowa *Math Fiction*). Przedmiotem tego zbioru danych jest trajektoria przekładów twórczości J. L. Borgesa w Polsce, ze szczególnym uwzględnieniem publikacji utworów tego autora w *Literaturze na świecie*.

W swoim założeniu formalnym praca jest połączeniem interdyscyplinarnego szkicu teoretycznoliterackiego z elementami raportowania oraz dokumentacji pracy projektowej i stanowi poszukiwanie formuły opowiadania za pomocą danych (*data storytelling*). Jako że pierwszy z tych dyskursów cechuje swoboda stylu, a drugi – skupienie na mierzalnych wartościach, najczęstszą figurą będzie tutaj (abstrakcyjny lub realnie istniejący) przykład zastosowania danej metody wraz z komentarzem. Dzięki temu narracja i interpretacje bliskie nauce o literaturze stanowią punkt wyjścia, a przetwarzanie informacji z użyciem narzędzi informatycznych stale dostarcza kontekstu dla dalszych dociekań. Jest to zatem połączenie dwóch żywiołów: snucia opowieści oraz zbierania i analizy faktów wyrażonych w danych. Dzięki temu informatyczne narzędzia mogą działać i jestem w stanie to opisać, a filologiczne nastawienie dostarcza lepszych środków do obrazowania tych procesów niż techniczny żargon.

Niniejszy projekt wytwarza pewien określony parametrami produkt. Jednocześnie, staram się pozostać blisko tematów i sposobów opisu przynależnych dziedzinie literaturoznawstwa, niejednokrotnie próbując przełożyć wprost użycie narzędzi informatycznych na badania teoretycznoliterackie. Spora część ma charakter spekulatywny i poszukujący: pewne zjawiska zostają opisane jako antycypacja kształtu uprawiania literaturoznawstwa w przyszłości.

W niniejszą rozprawę wpisana jest również dychotomia technologii i kultury: troska o dziedzictwo literatury i sztuki, wartości humanistyczne, spotykają się z komentarzami (również krytycznymi) do rozwoju narzędzi, jakimi posługujemy się na co dzień, dzięki rozwojowi internetu. Staram się skrótowo ukazać także konteksty skrajnie pozaliterackie, pochodzące z obszarów komunikacji, kultury mediów i zarządzania w przedsiębiorstwach technologicznych: ufam, że w ten sposób możliwe stanie się aktualne odniesienie do przedstawianych problemów i stawienie czoła realnym wyzwaniom intelektualnym, jak również kontrybucja do tego rodzaju humanistyki, który nie traci łączności komunikacyjnej ze światem pozanaukowym. Wprowadzenie niektórych wątków to także owoc mojej pracy zawodowej w branży handlu internetowego, gdzie punktem wspólnym do rozważań o bibliografii jest użycie baz danych i przetwarzanie informacji, które zawierają te zasoby. Paradoksalnie, w obu kontekstach znajdziemy miejsca wspólne, takie jak potrzeba rzetelnego opisu zawartości bazy, troska o możliwość łatwego znalezienia informacji przez odbiorcę, czy wreszcie łączenie ze sobą różnych źródeł, w celu uzyskania pełniejszego obrazu rzeczywistości. W kilku miejscach uznałem za stosowne poruszyć wątki zaangażowania społecznego humanistyki i jej wpływu na życie społeczne.

Starałem się używać inkluzywnego języka, stosując dostępne polszczyźnie narzędzia: osoby czytelnice napotkają w tekście użycie różnych form rodzajowych dla określeń osób i czynności. Cytaty prac w języku angielskim w miarę możliwości podaję w tłumaczeniu własnym; jeśli tak nie jest, zostały zaprezentowane dla oddania oryginalnego brzmienia.

Praca powstała jako owoc studiów w ramach programu interdyscyplinarnych studiów doktoranckich „Humanistyka Cyfrowa”, kierowanego przez Instytut Badań Literackich Polskiej Akademii Nauk oraz przy wsparciu Polsko-Japońskiej Akademii Technik Komputerowych jako partnera technicznego tego studium<sup>19</sup>.

---

<sup>19</sup> Studia realizowane są w ramach Programu Operacyjnego Wiedza Edukacja Rozwój dofinansowanego ze środków Europejskiego Funduszu Społecznego. Zob. <https://ibl.waw.pl/pl/edukacja/-interdyscyplinarne-studia-doktoranckie-humanistyk>

# Rozdział 1. Humanistyka cyfrowa jako metoda – zarys i wyzwania

Humanistyka cyfrowa bada zarówno obiekty materialne (np. kolekcje muzealne, druki), jak i niematerialne (modalność – sentyment wypowiedzi pisemnej, dane i metadane bibliograficzne)<sup>20</sup>. Zajmuje się analizą zjawisk, podaje hipotezy i propozycje ich interpretacji. Sama również, oprócz przetwarzania, *wytwarza* informacje i nową wiedzę. Połączenie narzędzi cyfrowych z warsztatem humanistycznym mają na celu trans- i interdyscyplinarne poszukiwania pod względem nowych ujęć znanych dotąd tematów badawczych, a nawet wynajdowania emergentnych zagadnień w toku prac<sup>21</sup>.

Następnie spojrzymy na sygnalizowane we wstępie spojrzenie na wytwarzanie oprogramowania jako proces nacechowany swoistą polifonią. Przejdziemy kolejno do charakterystyk humanistyki cyfrowej, uwzględniając konteksty historyczne, relacji dyscyplin oraz charakterystyki metodologicznej. W przywołanej literaturze skupiam się na takich przykładach, które bezpośrednio będą się odnosić do tekstów literackich. Ze względu na obszerność materiałów selekcja dotyczy takich źródeł, które inspirują do podjęcia działań nastawionych na rozwijanie zbiorów danych bibliograficznych. Zastanowimy się nad konsekwencjami tak uprawianego literaturoznawstwa, aby na końcu przejść do omówienia humanistyki cyfrowej pod kątem metodologicznym.

Posłużą do tego przykłady pozytywnych propozycji takich badań, ale również ich krytyka. Głosy sceptyczne pochodzą zarówno z pozycji tradycyjnie uprawianych nauk humanistycznych, jak i z wnętrza emergentnej metody. Dla dopełnienia obrazu *Digital*

---

<sup>20</sup> Należy pamiętać, że internet jest zależny od materialnej infrastruktury, to jest serwerów, centrów danych, sieci światłowodowej oraz satelitarnej itp., natomiast o niematerialności wymienionych zjawisk świadczy trudność wskazania ich w przestrzeni, co przybliża je do pojęć abstrakcyjnych i teoretycznych, jak np. gatunki literackie (także analizowane w badaniach humanistyki cyfrowej). Nawet jeśli wskażę konkretny adres bibliograficzny dzieła, to nie jest on materialnie przynależny do publikacji książkowej, w której się znalazł, w sensie celu, w jakim go używamy. Należy raczej do porządku informacji i (meta)danych. Analogicznie, sens i treść dzieł literackich oraz ich interpretacje nie są bezpośrednio związane z realizacją materialną w formie publikacji, choć jej typ może mieć wpływ na odbiór tekstu. Więcej o materialnym i środowiskowym aspekcie internetu: zob. Tatiana Schlossberg, *Inconspicuous consumption: the environmental impact you don't know you have*, First edition (New York: Grand Central Publishing, 2019). W przygotowaniu również książka Dietmara Offenhubera, dotycząca tematu materialności danych, zob. MIT Press, „Autographic Design”, <https://mitpress.mit.edu/9780262547024/autographic-design/>. Dostęp 2023-10-10

<sup>21</sup> Z tak ogólnym, a zarazem inkluzywnym ujęciem tematu przez Andrzeja Radomskiego spotkamy się w audycji *Głos nauki*, w odcinku poświęconym humanistyce cyfrowej: Agnieszka Stańczak, „Cyfrowa humanistyka”, *Głos Nauki*, <https://www.umcs.pl/pl/odcinki,20240,odc-5-cyfrowa-humanistyka,96829.chtm>. Dostęp 2023-07-15.

*Humanities* prezentują również kontekst tzw. nowych mediów oraz projekty polskie. Na zakończenie tego rozdziału, by zamknąć część związaną ze stanem badań, prezentuję istotne projekty stanowiące odniesienie do niniejszej pracy, a także wyzwania humanistyki cyfrowej jako metody i dobre praktyki w ich przewycięzaniu.

### Zamieszkać na bazarze

W kulturze rozwoju oprogramowania przeciwstawia się dwa modele pracy, nazwane przez Erica S. Raymonda katedrą i bazarem<sup>22</sup>. Autor pisał z perspektywy zwolennika otwartych licencji dla oprogramowania (temat otwartości i dostępności będzie jeszcze wielokrotnie wracał w kolejnych częściach niniejszych dociekań). Katedra opisuje wysiłek skierowany w określonym celu i poddany hierarchicznej strukturze zadań oraz odpowiedzialności, z wyraźnie zarysowanymi procesami projektowymi. Bazar to miejsce swobodnej wymiany idei, negocjacji, dynamicznie zmieniających się podejść do zadania, a także przestrzeń budująca społeczność wokół celu.

Przykładowym produktem-katedrą jest edytor tekstu *GNU Emacs*. W modelu wytwarzania kod źródłowy jest udostępniany na końcu, po wypuszczeniu produktu do użytku. Pomiędzy opracowywaniem nowych wersji dostęp do kodu źródłowego ma tylko zespół projektowy. Przykładowym produktem spod znaku bazaru jest system operacyjny Linux, stworzony głównie w czasie wolnym przez odpowiednio liczną grupę hakerów-entuzjastów, moderowanym przez Linusa Thorvaldsa, twórcę kodu źródłowego, od początku udostępnionego w sieci. W swojej książce Raymond oczywiście opowiada się za drugim (otwartym) podejściem, wskazując na efektywność w zwalczaniu pojawiających się błędów. Budowanie społeczności, którą motywuje wspólny cel, zwiększa liczbę oczu patrzących na projekt, a tym samym podnosi jakość jego wytworu<sup>23</sup>.

---

<sup>22</sup> Zob. Eric S. Raymond, *The cathedral and the bazaar: musings on Linux and Open Source by an accidental revolutionary*, Rev. ed (Beijing ; Cambridge, Mass: O'Reilly, 2001).

<sup>23</sup> Dodajmy, że współcześnie funkcjonuje pojęcie *playbor*, jako praca-zabawa, czyli zaprzęgnięcie grupy zapalonych fanów danej inicjatywy do wykonywania pracy, która wspiera projekt, najczęściej zupełnie nieodpłatnie. Ilustracją może być wykorzystywanie fanowskich modyfikacji gier, które studia projektowe wykorzystują jako narzędzie popularyzacji swoich tytułów lub rozwiązanie problemów technicznych, które pojawiły się w procesie produkcji. Rodzi to wątpliwości natury moralnej, jeśli pośrednio lub bezpośrednio zyskuje na tym wyłącznie wydawca, a zaangażowanie społeczności skupionej wokół projektu z jednej strony pozostaje niewynagrodzone w żaden sposób, a z drugiej pozostaje promowane. Motywacją osób wykonujących *playbor* może być np. chęć rozwoju umiejętności, środowiskowy prestiż, czy manipulowanie elementami gry (co Marcin M. Chojnacki nazywa metasprawczością).

Por. Kłosiński, Michał. *Hermeneutyka gier wideo: interpretacja, immersja, utopia = Hermeneutics of video games: interpretation, immersion, utopia*. Lupa obscura. Warszawa: Instytut Badań Literackich PAN. Wydawnictwo, 2018. Chojnacki, Marcin M. *Estetyka sprawczości w grach wideo*. Wydanie I. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2022.

Jako literaturoznawcy, mamy do czynienia ze swoistym „bazarem w domu”, teoria literatury nie jest bowiem monolitem o jasno wyznaczonych granicach tego, co wolno, a czego nie wolno robić z tekstem, kiedy się go opisuje w szczegółach, interpretuje i wartościuje. Poza określonym dziedzictwem, wywodzącym się z antycznej retoryki, czerpiącym z tradycji hermeneutycznej i interpretacji tekstów religijnych, a także bez wątplenia mocnych podstaw strukturalizmu wraz z jego rekonfiguracjami, redefinicjami i remiksami, które wszystkie posłużą jako pozytywistyczny wzorzec patrzenia na tekst, literaturoznawstwo stanowi w dużej mierze agon. Jest tego świadomy np. Jonathan Culler, wprowadzający początkującego czytelnika uniwersyteckiego w temat teorii literatury. Prezentuje on wiele metod interpretacji, nie dając żadnej primatu, ale raczej pokazując rozłożenie akcentów. Podejścia różnych szkół do tego samego tekstu mogą być diametralnie różne, co nie znaczy, że stanowią jakiś rodzaj gry o sumie zerowej:

Przy bardziej sformalizowanej interpretacji, dokonywanej w salach wykładowych, obowiązują różne reguły. Można rozpatrywać funkcje każdego poszczególnego elementu dzieła, jego relacje z innymi elementami, ostatecznie jednak interpretacja może wymagać gry w „o czym więc tak naprawdę jest ten utwór?”. Pytanie to nasuwa się nie z powodu niejasności tekstu; bardziej uzasadnione bywa w odniesieniu do utworów prostych niż złośliwie zawiłych. W tego rodzaju grze odpowiedź musi spełniać określone warunki: nie może być, na przykład, zbyt oczywista – musi być spekulacją. Powiedzieć, że „*Hamlet* jest sztuką o księciu Danii”, to odmówić udziału w grze. Możliwymi odpowiedziami są natomiast: „*Hamlet* opisuje zagładę elżbietańskiego porządku świata”, „*Hamlet* to sztuka o męskim lęku przed kobiecą seksualnością” albo „*Hamlet* mówi o tym, że nie można polegać na znakach”. To, co uważa się zwykle za „szkoły” krytyki literackiej bądź teoretyczne „podejścia” do literatury, to z punktu widzenia hermeneutyki skłonności do udzielania określonych odpowiedzi na pytanie „o czym” właściwie jest utwór: „o walce klas” (marksizm), „o możliwości ujednoczenia doświadczeń” (Nowa Krytyka), „o kompleksie Edypa” (psychoanaliza), „o powstrzymaniu sił wywrotowych” (neohistoryzm), „o asymetrii stosunków między płciami” (feminizm), „o autodekonstruktywnej

naturze tekstu” (dekonstrukcja), „o zmierzchu imperializmu” (teoria postkolonialna), „o matrycy heteroseksualnej” (badania literatury gejów i lesbijek).

Wymienione w nawiasach dyskursy teoretyczne nie są bynajmniej równoznaczne z metodami interpretacji: przede wszystkim podejmują próby wyjaśnienia zjawisk, które uznają za szczególnie istotne dla kultury i społeczeństwa. Wiele z tych teorii bada również mechanizm funkcjonowania literatury bądź dyskursu w ogóle, realizując tym samym założenia poetyki; będąc jednak odmianami hermeneutyki, teorie te prowadzą do szczególnego typu interpretacji, w których teksty uznaje się za wypowiedzi służące określonym celom. Przy grze w interpretację ważne nie jest podsuwanie trafnych odpowiedzi: jak pokazują podane przeze mnie przykłady, utrzymane w konwencji parodii niektóre odpowiedzi okazują się niejako z definicji łatwe do przewidzenia. Istotną rolę odgrywa to, jak się do nich dochodzi – sposób w jaki zawarte w tekście szczegóły zostaną wykorzystane podczas formułowania konkluzji<sup>24</sup>.

Culler podkreśla pewien tematyzm poszczególnych szkół, jednak wszystkie oświetlają przedmiot badań na swój sposób. Nie ma zatem w teorii literatury jednego głosu – kodu źródłowego katedry interpretacji, a raczej pozostaje polifonia rozproszonych podejść. Można podejrzewać, że metafora bazaru i odwołanie do informatycznego żargonu, gdzie stawiamy opozycję między wytwórcą a użytkownikiem jako współtwórcą, spodobałyby się pragmatystom, dla których każde odczytanie tekstu jest jego użyciem.

Jak widzimy, zajmowanie różnych pozycji krytycznych i interpretacyjnych, a także negocjowanie opisu przedmiotu badań, są żywiołem teorii literatury od dawna. Przyjęty sposób uprawiania literaturoznawstwa pozostawia pole dla eksperymentu, również z użyciem badań na danych, aby poszerzyć kontekst analizy dzieła. Zamiast agonu i konkurujących ze sobą „o rację ostateczną” szkół teoretycznoliterackich, widziałbym oparty na wymianie pomysłów i informacji zwrotnej bazar – pełne wrzawy miejsce, gdzie każdemu przyświeca idea, że może lepiej współtworzyć swoją dziedzinę.

---

<sup>24</sup> Jonathan D. Culler, *Teoria literatury. Bardzo krótkie wprowadzenie*, tłum. Maria Bassaj, Wydanie pierwsze (Warszawa: Prószyński i S-ka, 1998).

Moda? Metoda? Metodologia? Szkoła?

Patrząc na wymienione przez Cullera przykłady podejść do tekstów, nie jesteśmy w stanie ekstrapolować podobnych założeń na to, co nazywamy humanistyką cyfrową. Nie jest ona szkołą, która udziela wciąż tej samej odpowiedzi na pytania o swój przedmiot albo filtruje rzeczywistość pod względem istotnych dla siebie kryteriów. Nie jest w rozumieniu Cullera teorią jako taką, ponieważ nie podsuwa wyjaśnień otaczających nas zjawisk, tak jak czynią to najbardziej znane teorie literackie. Nie tworzy jednolitego filtra czy siatki pojęciowej, przez które patrzymy na świat. Jest nierozzerwalnie związana z metodami wypracowanymi na gruncie humanistyki, ale rozbudowuje własne strategie opisu czy metody badawcze.

Tym, co jednak wspólne między projektami humanistyki cyfrowej a teorią, jest dążenie do poszerzenia naszego pola widzenia i oglądu na rzeczywistość. Obie są sposobami wytwarzania dodatkowej aparatury poznawczej. Jest to szczególnie ważne w zmieniającym się świecie mediów cyfrowych. Odwołam się do dłuższej wypowiedzi Hansa Ulricha Gumbrechta ze wstępu do zbioru esejów, poświęconych rozumieniu-wytwarzaniu sensu (*making sense*):

What is at stake, then, is the invention of a new epistemology capable of theorizing and analyzing ways of sense making that perhaps no longer include effects of „meaning” and „reference”. The first step in this direction would be to recognize that, at least so far, we simply do not know what interfaces such as those between TV and psychic systems, or between computers and social systems, look like. As soon as we are able to develop some understanding of such interfaces, the question will arise of how we can and should use our „insight” in this respect. Systematically speaking, „sense making” will continue to be our field of reference, but perhaps the same will no longer be the case with meaning, representation and reference. In any event, the phenomena with which we are going to deal will probably look very different from those that we used to associate with the whole horizon of „sense making” – so different that it is now far too early to think about what the relationships between such phenomena (our future „objects” of research) and our discourses might look like; too early also to say how our dealing with new modes of sense making will make sense. Such questions are definitely not – not yet – the topic of the essays presented in this volume. But I assume that the specific

perspectives from which they analyze and reconstruct different literary and nonliterary modalities of sense making in Western history prove themselves to be an appropriate point of departure for such an agenda. But if „point of departure” as a metaphor implies an idea of the goals to reach and the ways leading to these goals, it might even be too much to say that we are now at a point of departure. What we experience these days is rather the bare inviability of certain concepts and assumptions on which our profession used to be grounded. Our notion of „meaning” still works quite well with texts – but textuality is no longer the predominant form in which knowledge circulates. If we want to maintain the traditional equivalence between „history” and „historiography”, we have to admit that history depends as much on an author's subjective intuitions as any other narrative – and it can then hardly serve as a stable ground for any political claims. Finally, what the New York Times used to call „the upheaval in Eastern Europe” should make us think whether the division between the private sphere and the public sphere, as well as rationality as a normative orientation of decision making, is still available as a basis for our pathetic concept of „politics.” Anyone who raises such issues, at least in Germany, is likely to be labeled a „neoconservative”. But it might turn out that the only obvious reference of the adjective „conservative” is not to raise them, that the most dangerous of all fictions is to behave as if we could still rely on such traditions as meaning, history, and politics<sup>25</sup>.

Słowa te zaczerpnąłem z książki z 1992 roku. Jestem ciekaw, jak badacz sformułowałby tę myśl dzisiaj<sup>26</sup>. Pod względem celu produkcja czy wytwarzanie sensu dla badacza ze Stanford

---

<sup>25</sup> Hans Ulrich Gumbrecht, *Making Sense in Life and Literature*, Theory and History of Literature, v. 79 (Minneapolis: University of Minnesota Press, 1992), 12-13. Zastosowano nieznaczne zmiany w interpunkcji.

<sup>26</sup>Badacz poruszył część wątków w wystąpieniu na temat przyszłości humanistyki w 2016 roku, zob. *Hans Ulrich Gumbrecht (Stanford) delivers Annual Prokhorov Lecture HQ*, 2016, <https://youtu.be/88lj4C8fakM?si=CMTE4psUFtsuurlS>. Dostęp 2023-09-16. Wystąpienie, dotyczące przyszłości humanistyki, wskazuje np. na jej rolę w zapewnianiu, że ośrodki akademickie są ośrodkami intelektualnymi (budowanie kultury intelektualnej, wymiana myśli – ale także stymulowanie wielości perspektyw). Trzy kierunki-rekomendacje programowe, które wymienia dla humanistyki to: ryzykowne myślenie (badanie hipotez), kontemplacja (w przeciwieństwie do tego, co opisuje się jako *research*, badania, zwł. empiryczne), wyobrażenia (związana ze śmiałością badawczą). W powyższej dyskusji, odwołując się do monopolu silnych ośrodków badawczych i tradycji średniowiecznej kultury akademickiej wskazuje między innymi na kwestię



są istotne, problematyczne staje się natomiast przewidzenie trajektorii zmian technologicznych. Odwołuję się do – wydawać by się mogło – przepastnej różnicy w czasie: szczególnie w Polsce od momentu, gdy Gumbrecht ową myśl sformułował do lat dwudziestych XXI nastąpił znaczny przeskok w wymianie wiedzy na każdym poziomie w społeczeństwie (prywatnym, akademickim, medialnym). Internet, będący w centrum omawianych przeze mnie zjawisk, odpowiada za znaczące przyspieszenie owych przemian. Sam odwołuję się w przypisie nie do publikowanych notatek, skryptów, czy wyimków z wystąpień Gumbrechta w formie drukowanej, ale do rejestracji wideo całego wykładu tego autora, dostępnej nieprzerwanie<sup>27</sup> dla wszystkich użytkowników w sieci.

„Sens” oderwany od „znaczenia” i „referencji” – czy byłby zatem „porozumieniem”? Wypowiedź Gumbrechta przykuwa oczywiście uwagę cyfrowego humanisty ze względu na wzmiankę o przemianach interfejsów, stosowanych przez użytkowników do kontaktu z informacją. Autor wskazuje na aspekt aktualizacji gromadzenia wiedzy i opisu rzeczywistości wraz z tymi zmianami. Tym bardziej dramatycznie wybrzmiewa tutaj uwaga o „wymieraniu” pojęć, którymi posługuje się humanistyka w jej tradycyjnej formie. Tym bardziej zasługuje na uznanie próba ocalania ich – a przynajmniej ich sensu.

Zakończenie cytatu przenosi te uwagi na wymiar polityczny. Poszukiwanie znaczenia i sensu, wysiłek hermeneutyki, a także epistemologiczne dociekania, nie mogą wszakże zostać w pełni oderwane od swojego kontekstu społecznego. Jak zobaczymy w kolejnych przykładach, zajmowanie się dziedzictwem „literatury świata”, zbiorem literatury powszechnej, może być pod względem przetwarzania danych wyzwaniem, związanym z globalizacją. Gumbrecht odnosi się zbiorczo do przemian w regionie, do którego należy również Polska. W części pracy poświęconej PBL jako projektowi cyfrowemu, także pojawi się ten kontekst – formy opisu kultury, archiwizacji dziedzictwa i gromadzenia danych nie są wolne od wpływu władzy, polityki, a nawet systemowego zarządzania nauką i badaniami. Współcześnie możemy

---

swoistego recyklingu idei znanych w kulturze od wieków, ale wymagających dostosowania do współczesności. Z kolei w książce z 2014, poświęconej doświadczeniu czasu, wskazuje wyraźnie na hipermedialny charakter współczesnego doświadczenia, wszechobecność mediów i techniki użytkowej. Stwarza to poznawcze symultaniczności, które odwodzą od – prafrazując ustalenia badacza – przeżywania namacalnej rzeczywistości (Gumbrecht czyni też wiele uwag, wskazujących na uprzedzenia do praktyk wspieranych technologią, zarówno w życiu codziennym, jak i uprawianiu nauki – owe uprzedzenia są jak najbardziej świadome i niekoniernie skrajne, tzn. autor traktuje je jako doświadczenie pokoleniowe). Zob. Hans Ulrich Gumbrecht, *Our Broad Present: Time and Contemporary Culture* (New York: Columbia University Press, 2014).

<sup>27</sup> Z zastrzeżeniem, że owa ciągłość oznacza nieograniczony dostęp w czasie, z wyłączeniem sytuacji zaprzestania działalności serwisu hostującego lub usunięcia materiału przez osobę zarządzającą kontem w tym serwisie. Ten temat będzie powracać, np. ilekroć napotkamy rozróżnienia między prywatnymi i publicznymi źródłami.

zaktualizować te spostrzeżenia o jeszcze jeden czynnik, czyli interesy korporacyjne i wzajemne relacje z akademią, tzn. wpływ biznesu na naukę i nauki na biznes, np. w obszarze transferu wiedzy<sup>28</sup>. Narzędzia cyfrowych humanistów są dostarczane przez globalne przedsiębiorstwa nastawione na maksymalizację zysku, a cała infrastruktura internetowa stanowi rozproszoną sieć wpływów takich przedsiębiorstw. Dochodzi do tego kwestia otwartej nauki, otwartego dostępu do informacji i dostępności danych – część repozytoriów również pozostaje komercyjna. Jest to nieunikniona zależność współczesnego humanisty cyfrowego od biznesu.

Można też powiedzieć, że (w pewnym sensie: instytucjonalna) humanistyka cyfrowa jako organizacja wiedzy naukowej rozwija się równoległe z oddolnymi projektami społecznościowymi opartymi na wiedzy, ulepszanymi z pragmatycznych pobudek (hobby, nauka szkolna i akademicka, wymiana informacji w grupie). Poruszam tę kwestię przy okazji opisu treści tworzonych przez użytkowników w rozdziale poświęconym otwartym danym.

Rys historyczny, rozwój i kontekst społeczny

Za swego rodzaju mit założycielski humanistyki cyfrowej, ze względu na ścisłą współpracę z IBM jako partnerem technologicznym, uznaje się prace zakonnika Roberto Busy (*Corpus Tomisticum*, konkordancja dzieł Tomasza z Akwinu), rozwijane jeszcze w latach 40. XX wieku, w celu określenia charakterystycznych cech językowych i uformowania korpusu tekstów, z uwzględnieniem tekstów „satelickich” (przypisywanych niegdyś autorowi oraz jego niepełnych prac)<sup>29</sup>. Mamy też inne odniesienia; w przywołanej już audycji *Głos Nauki*<sup>30</sup> Radomski wymienia przynajmniej dwie inicjatywy: badania tzw. kliometrystów, Conrada

---

<sup>28</sup> Nie jest to przedmiotem pracy i nie czuję się kompetentny by wypowiadać się na temat transferu wiedzy z filologii do biznesu, ale zaryzykuję hipotezę, że więcej takiego „zastrzyku wynalazczości” z nauki do biznesu daje językoznawstwo niż literaturoznawstwo. Mam na myśli wkład w rozwój przetwarzania języka naturalnego (NLP), tłumaczenie maszynowe, wielkie modele językowe (nazywane dziś sztuczną inteligencją), analiza i generowanie tekstu oraz inne pokrewne rozwiązania. Miejsc wspólnych literaturoznawstwa i przedsiębiorczości np. opartej na technologii szukałbym w obszarze metadanych, blisko związanym z tematem niniejszej rozprawy – społecznym tagowaniem kultury, ontologiami danych dotyczących kultury, agregacją treści według klucza tematycznego i podobnych. Odwołuję się do tych kwestii w rozdziale związanym z otwartymi połączonymi danymi, natomiast warto wspomnieć o tym, że serwisy społecznościowe związane z kulturą muszą się odwoływać do pewnych klasyfikacji ustalonych na gruncie humanistyki (literaturoznawstwo, filmoznawstwo), np. przy podziałach gatunkowych utworów. Takie rozwiązania polegają na kompetencjach kierunkowo wykształconych ekspertów tych dziedzin. Widzę w tym przede wszystkim pracę kapitału społecznego: transfer wiedzy i umiejętności zaangażowanych osób. Osobnym tematem jest zbiór kompetencji związany z tworzeniem światów i fabuł np. na potrzeby gier wideo, por. Michał Kłosiński, *Hermeneutyka gier wideo: interpretacja, immersja, utopia = Hermeneutics of video games: interpretation, immersion, utopia*, *Lupa obscura* (Warszawa: Instytut Badań Literackich PAN. Wydawnictwo, 2018). Piotr Kubiński, *Gry wideo: zarys poetyki* (Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, 2016).

<sup>29</sup> Susan Schreibman, Raymond George Siemens, i John Unsworth, red., *A companion to digital humanities*, Blackwell companions to literature and culture 26 (Malden, MA: Blackwell Pub, 2004).

<sup>30</sup> Stańczak.

i Meyera, które pomogły wprowadzić w ekonomii podejście oparte na matematyce (czyli na danych) zamiast dotychczasowego narracyjnego<sup>31</sup> oraz powstanie na uniwersytecie w Cambridge jednostki Literary and Linguistic Computing Centre (LLCC) (1964), obecnie działającej w ramach współczesnego Cambridge Digital Humanities<sup>32</sup>. Kolejny etap rozwoju przynosi upowszechnione użycie komputerów osobistych w latach 80. XX wieku, również w nauce. Następnym przełomem był rozwój internetu od lat 90.: umożliwiło to lepszą komunikację i publikację badań (współpraca międzynarodowa), rozwinęły się archiwa cyfrowe. Od tamtego czasu rozwija się intensywnie nowy przedmiot badań: dane cyfrowe. W XXI wieku głównym obszarem są duże zbiory danych (*big data*) i technologie, takie jak uczenie maszynowe (nazywane sztuczną inteligencją, AI).

Przed wyłonieniem się *Digital Humanities* używane było także pojęcie *Humanities Computing*. Wskazuje to na akcent przesunięty w stronę strony technicznej, obliczeniowej. Współczesny rozwój technologii związanych z wymianą informacji i komunikacją (ICT, *Information and Communications Technology*) to impuls napędzający zmiany w obszarze nauk o mediach. Owe przemiany wpływają także na humanistykę i to w różny sposób: kształtują nowe sposoby odbioru i nadawania w kulturze, a za tym idzie pojawienie się nowych przedmiotów badań. Nowe narzędzia oznaczają również nowe metody prowadzenia tych badań.

Humanistyka cyfrowa rozwija się w kontekście tzw. nowych mediów. Te z kolei czerpią z idei, które pojawiły się dużo wcześniej w kulturze. Można powiedzieć, że rozwiązania, które obserwujemy jako współczesny internet, są wpisane w przekaz kulturowy i w nim zanurzone (choć wydawałoby się, że przedstawia się je jako wytworzone przez konkretne przedsiębiorstwa za zamkniętymi drzwiami, w izolacji od społecznych praktyk kulturowych). Badacze zwrócili uwagę na antycypacyjny charakter literatury, filozofii i sztuki przy opisie zmian technologicznych, jak chociażby w zbiorze *New Media Reader*, w którym wyróżniono cztery grupy tematyczne: I Złożoność, zmiana i nieokreśloność; II Media kolektywne, media osobiste; III Projektowanie, aktywność i akcja; IV rewolucja, opór i start Sieci – podejmowane tematy i formy przekazywania wiedzy oraz opowieści są bardzo różnorodne (proza, esej filozoficzny, publicystyka, artykuł naukowy, komiks)<sup>33</sup>. Ta polifoniczność sposobów

---

<sup>31</sup> Zob. Michael Hauptert, „The Impact of Cliometrics on Economics and History”, *Revue d'économie politique* 127, nr 6 (2017): 1059, <https://doi.org/10.3917/redp.276.1059>. Owe badania wpłynęły także na historię ekonomii i pokrewne dziedziny.

<sup>32</sup> Por. Literary and Linguistic Computing Centre (LLCC), <https://www.repository.cam.ac.uk/handle/1810/131666>. Dostęp 2023-09-16. Obecna witryna związana z humanistyką cyfrową: Cambridge Digital Humanities, <https://www.cdh.cam.ac.uk/>. Dostęp 2023-09-16.

<sup>33</sup> Noah Wardrip-Fruin i Nick Montfort, red., *The New Media Reader* (Cambridge, Mass: MIT Press, 2003).

opowiadania odzwierciedla również charakter infosfery, z którą mamy współcześnie do czynienia w sieci. Lev Manovich nakreśla we wstępie pole nowych mediów „od Borgesa do HTML”<sup>34</sup>, uzmysławiając czytelnikowi istnienie pewnych odpowiedników-korespondencji między projektami sztuki a rozwojem technologii (HTML jako „ogród o rozwidlających się ścieżkach”). Cyberkultura to dla autora *Języka nowych mediów* po prostu przeniesienie kultury do wnętrza technologii jako wehikułu. Narzędzia komputerowe stanowią tu platformę dystrybucji treści kulturowych (i kulturalnych). Tym samym, kultura staje się danymi, a na wyższym od pojedynczej informacji poziomie następuje mieszanie się konwencji (konwencje kultury przenikają konwencje technologii). Manovich podaje tu przykład „przełamania” pojedynczości obrazu – w sieci ilustracja, obrazek, zdjęcie, jest elementem znaczącym, ale poprzez nadanie większej liczby jakości czy cech np. „klikalność”, czyli zdolność do wprowadzania interakcji, ztraca status pojedynczego pola (zwłaszcza w przypadku rozbicia tych właściwości na różne obszary tego samego obrazu). Nowomediálne strategie estetyczne odtwarzają też znane wcześniej tropy – także w warstwie krytyki<sup>35</sup>. W odwołaniu do algorytmizacji, nowe medium (dodajmy do słów Manovicha – każde nowe medium, będące usprawnieniem poprzednich) pozwala na szybsze wykonanie zadań, które wcześniej wymagało „pracy ręcznej”<sup>36</sup>. Tu zauważmy, że ów przeskok zaobserwujemy między spisem książki z autopsji w celu dodania do bibliografii, a masowym zarządzaniem rekordami bibliograficznymi w bazach danych (rzędu setek, tysięcy i milionów takich pojedynczych zapisów). Manovich określa także nowe media jako spełnienie zamierzeń dwudziestowiecznej awangardy (wracam do jego ustaleń na ten temat z innego tekstu w dalszej części), a także owocem rozprzestrzeniania się podobnych idei w obszarach sztuki i informatyki (*computing*). Media cyfrowe jawią się także jako spełnienie snów kontrkultury lat 60. XX wieku.

Mariaż kultury z technologią jest zatem bezsprzeczny. Istotny jest społeczny kontekst – jeśli narzędzia masowej analizy danych wykorzystuje się w takich dziedzinach jak agitacja polityczna, sterowanie opinią publiczną poprzez media, agresywny marketing i dezinformacja – z ryzykiem niepowodzenia, ale ostatecznie – osiągając wynik – to dlaczego nie spróbować

---

<sup>34</sup> Wardrip-Fruin i Montfort, 13-25.

<sup>35</sup> „In other words, at some point photography, telephones, cinema, and television each were »new media«” - Wardrip-Fruin i Montfort, 19. Podobnie kiedy w niniejszej pracy posługuję się ustaleniami Manovicha, to ze świadomością, że obecnie nowym medium są np. duże modele językowe, a nie zakupy w internecie.

<sup>36</sup> Zapisuję tłumaczenie *manual work* w cudzysłowie, ponieważ nie musi to być praca *stricte* ręczna – nowe media i technologie jak najbardziej odciążają pracownika przy zadaniach umysłowych, zwłaszcza w społeczeństwie informacyjnym.

w humanistyce? Wymaga to pewnej badawczej śmiałości i pogodzenia się z mniej spektakularnymi rezultatami, ale i tak daje nagrodę w postaci nowej informacji o przedmiocie.

Mimo aktualności ustaleń co do charakterystyki mediów cyfrowych, współczesny krajobraz wymiany informacji zmienił się od czasu swoistego technoentuzjazmu sprzed np. 2010 roku. Innymi słowy, „nowe media”, o których pisał Manovich, okrzyki i stworzyły pole dla niestandardowych komunikatów (zamiast „czystej” wymiany wiedzy i informacji, jak zdaje się sugerować porównywanie do modernizmu, współcześnie spotykamy się z przekazami ironicznymi, surrealistycznymi, ale też niebezpiecznymi poznawczo, np. propagandą – dość wspomnieć popularność pojęcia „post-prawda”<sup>37</sup>). Silna propozycja humanistyki, oferującej rzetelne źródła wiedzy, może być odpowiedzią na tego rodzaju kryzysy komunikacji i jest częścią społecznej odpowiedzialności nauki, stąd przytaczam kilka wątków i źródeł na ten temat. By ująć to łagodnie – w porównaniu do entuzjastycznego stanu nowych mediów, opisywanego u Manovicha, jako przestrzeń eksperymentu, równości i kreatywności, lata 20. XXI wieku nie są już w internecie tak jednoznacznie „hipisowskie”.

Wśród przemian historycznych o dużym wpływie społecznym można wymienić protesty Occupy Wall Street (2011) i inne późniejsze rozruchy w USA (ruch Black Lives Matter, szczególnie demonstracje i walki w 2020 oraz szturm na Kapitol po ogłoszeniu wyników wyborów prezydenckich 2021 i przegranej Donalda Trumpa), wyjście Wielkiej Brytanii z UE (określane jako *Brexit*, 2016-2020), pandemię COVID-19 (2019, obecnie jako endemia), inwazję Federacji Rosyjskiej na Ukrainę (2022). Część z tych działań uwzględnia paniczne zachowania uczestników komunikacji w internecie, kampanie dezinformacyjne i ataki hybrydowe z wykorzystaniem infosfery, zwłaszcza internetu i w szczególności mediów społecznościowych.

Christopher Wylie z perspektywy uczestnika (i współwinnego) opowiada o wpływie na *Brexit* i wybory prezydenckie w USA w 2016, jaki wywarły algorytmy dobierające wyświetlanie treści w mediach społecznościowych na określone grupy użytkowników<sup>38</sup>. Shoshana Zuboff przestrzega przed oddaniem nadmiernej władzy w ręce przedsiębiorstw o globalnym zasięgu

---

<sup>37</sup> Artur Bartkiewicz, „Oxford Dictionaries uznały «post-prawdę» za słowo 2016 roku”, *Rzeczpospolita*, 2016-11-16, <https://www.rp.pl/spoleczenstwo/art3137761-oxford-dictionaries-uznaly-post-prawde-za-slowo-2016-roku>. Dostęp 2023-09-16.

<sup>38</sup> Christopher Wylie, *Mindf\*ck: Cambridge Analytica, czyli jak popsuć demokrację*, tłum. Michał Strąków (Kraków: Insignis, 2020).

i o mechanizmach inwigilacji w późnym kapitalizmie<sup>39</sup>. Peter Pomerantsev opisuje jak działają propagandyści w internecie, związani z Kremlm, poprzez *trolling*, czyli rozprowadzając fałszywe informacje i podważając społeczne zaufanie do instytucji, jak i źródeł wiedzy (w tym naukowej)<sup>40</sup>. *Trolle Putina* to studium relacja fińskiej dziennikarki Jessiki Aro, która opowiada o doświadczeniu prześladowania przez *trolli* związanych z rosyjskimi wpływami, w związku z jej śledztwami na temat tej działalności<sup>41</sup>. Wszystkie powyższe pozycje łączą diagnozy kryzysu zaufania do mediów, dlatego tak ważne jest budowanie infrastruktury wiedzy, które pomogą nie tylko w badaniach naukowych, ale również usprawnią ustalanie faktów wszystkim użytkownikom internetu (*fact-checking*)<sup>42</sup>. Osobnym problemem może być „remonopolizacja” sieci, tzn. oddanie zarządzania ruchem i infrastrukturą we władanie kilku największych korporacji technologicznych<sup>43</sup>.

Piszę o tym dlatego, że w pewnym sensie wiele pól działania, które tradycyjnie kojarzymy z humanistyką, zostało przejętych przez siły rynku i tzw. *bad actors*. Dotyka to publicystykę, krytykę kulturalną (literacką, ale też np. filmową i muzyczną), edukację, uczestnictwo w kulturze, czy nawet powszechne zdolności do autorefleksji i krytycznego myślenia. Poruszam te wątki w części związanej z krytyką humanistyki cyfrowej. Istnieje jednak pozytywna strona i chciałbym zaprezentować przykład możliwego działania, które nie jest kontrowersyjne i przynosi pozytywny efekt. Na zasadzie kontrastu chciałbym sięgnąć po inną gałąź ludzkiej aktywności, której obecność w kulturze, podobnie jak w przypadku literatury, możemy prześledzić aż do starożytnej Grecji (i jeszcze wcześniej).

Dane geograficzne i dotyczące położenia

Sport stanowi dziedzinę, w której użycie danych wspiera istniejące już metody pracy. Nie jest niczym nowym zastosowanie dietetyki czy indywidualnych planów treningowych opartych na danych. Współczesne wysiłki w tej dziedzinie rozwijają się jednak również na inne sposoby,

---

<sup>39</sup> Shoshana Zuboff, *Wiek kapitalizmu inwigilacji: walka o przyszłość ludzkości na nowej granicy władzy*, tłum. Alicja Unterschuetz, Wydanie I (Poznań: Zysk i S-ka Wydawnictwo, 2020).

<sup>40</sup> Peter Pomerantsev, *To nie jest propaganda: przygody na wojnie z rzeczywistością*, tłum. Aleksandra Paszkowska, Wydanie pierwsze (Warszawa: Wydawnictwo Krytyki Politycznej, 2020).

<sup>41</sup> Jessikka Aro, *Trolle Putina*, tłum. Marta Laskowska, Wydanie I (Kraków: Wydawnictwo Sine Qua Non, 2020).

<sup>42</sup> W publicystyce polskiej porusza te wątki Piotr Fortuna. Na temat dwóch książek wymienionych jako ostatnie zob. Fortuna Piotr, „Więcej niż propaganda”, *Dwutygodnik* 299, nr 1 (2021), <https://www.dwutygodnik.com/artukul/9329-wiecej-niz-propaganda.html>. Dostęp 2023-09-16.

<sup>43</sup> W tym gronie wymienia się Alphabet (dawniej Google), Amazon, Meta Platforms (dawniej Facebook), Apple. Czasami również dodaje się Microsoft. Mówi się o wielkiej czwórce lub piątce, pod zbiorczą nazwą GAFA lub GAFAM. Zob. np. Scott Galloway, *The Four: The Hidden DNA of Amazon, Apple, Facebook and Google* (London: Bantam Press, 2017). Wyd. pol. Scott Galloway, *Four Wielka Czwórka: ukryte DNA: Amazon, Apple, Facebook i Google*, tłum. Jolanta Kubiak, Wydanie I (Poznań: Dom Wydawniczy Rebis, 2018).

na przykład z użyciem danych geolokalizacji w dyscyplinach drużynowych (sportowcy noszą lokalizatory podczas sesji treningowych, niekiedy jest to też stosowane w trakcie rozgrywek). Dzięki temu powstaje mierzalna informacja zwrotna na temat przebytego dystansu, pozycji w czasie i relacji do pozostałych osób z drużyny. Podstawowymi celami sztabów szkoleniowych są osiągnięcie przewagi nad konkurującymi drużynami oraz zmniejszenie ryzyka kontuzji. Przykłady takich informacji to: przebyty dystans, prędkość i przyspieszenie zawodnika, kadencja biegu, liczba kontaktów (kolizji) i metabolizm (szacowane wydatkowanie energii na podstawie wysiłku). Dane dotyczące pozycji mają w założeniu wspierać obiektywizowane podejmowanie decyzji: ocenę formy graczy i zachowań związanych z realizacją strategii trenera (np. w formie mapy termicznej zajmowanych pozycji na boisku)<sup>44</sup>. By powołać się na konkretny przykład dyscypliny: w rozgrywkach piłki nożnej, klubowych i międzynarodowych, wykorzystuje się dane lokalizacji graczy, aby w czasie rzeczywistym sprawdzać realizację planu taktycznego drużyny oraz w razie konieczności dokonywać zmian zawodników, a także poza spotkaniami monitorować obciążenie treningowe<sup>45</sup>. W piłce nożnej bardzo wcześnie zaczęto stosować pomiary lokalizacji. Co warto podkreślić, w każdej dyscyplinie parametry będą inne: długość sprintów warunkują reguły gry oraz powierzchnia boiska; interwały wysiłku i spoczynku są inne w zależności od czasu rozgrywek, w tym także regulaminowych przerw. Przepisy poszczególnych dyscyplin wpływają na biomechanikę; mają ponadto realny wpływ na odczyt danych z urządzeń noszonych przez sportowców<sup>46</sup>.

Interesujący jest dla mnie ów aspekt obiektywizacji podejmowania decyzji, tzn. moment, w którym możemy się zgodzić, że dane prezentują jakiś faktyczny stan i wspomagają planowanie oraz działanie. W taki właśnie sposób odczytuję dążenia czy pretensje humanistyki cyfrowej i metod wykorzystania danych w literaturoznawstwie. Statystyczna analiza danych czy *Data Science*<sup>47</sup> może być nauką pomocniczą dla literaturoznawstwa, natomiast wymagać

---

<sup>44</sup> Torres-Ronda L. et. al., „Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis”, *Sports Medicine - Open* 8, nr 1 (grudzień 2022): 15 <https://sportsmedicine-open.springeropen.com/articles/10.1186/s40798-022-00408-z>.

<sup>45</sup> Goal.com, *Footballers don't 'wear bras' - sporting reasons for under-shirt clothing explained*, <https://www.goal.com/en/news/footballers-dont-wear-bras-sporting-reasons-under-shirt-clothing-explained/1aakl5v6271f814s624c5ws52t>. Dostęp: 2023-09-16.

<sup>46</sup> Torres-Ronda L. et al.

<sup>47</sup> W polszczyźnie proponowanymi ekwiwalentami *Data Science* są „danetyka” i „danologia” (dla zwięzłości np. wobec sformułowania „nauka o danych”). Funkcjonuje także pojęcie „inżynieria danych”, dla odzwierciedlenia części związanej z przetwarzaniem i analizą danych. Osobiście skłaniam się do pozostawienia terminu w angielskiej wersji i w kolejnych częściach pracy przeważnie stosuję to rozwiązanie. Por. Robert Ryciak,

to będzie adaptacji do konkretnych zagadnień. Nie możemy się spodziewać tak łatwego przełożenia mierzalnych wskaźników na wnioski jak to ma miejsce w sporcie (specyfika przedsięwzięć sportowych bardzo też odbiega od humanistyki, biorąc pod uwagę aspekty rywalizacji, charakterystyki czynności oraz typ danych, na których się pracuje). To, co jednak wspólne w obu przypadkach, to element decyzji, czy będziemy ufać danym, czy intuicji. Wyobraźmy sobie teraz humanistę cyfrowego jako trenera drużyny. Dane przedstawiają jakiś obraz sytuacji (np. wspierają decyzję o posadzeniu ważnego zawodnika na ławce albo świadczą o nieistotności utworu literackiego do uwzględnienia w dalszych krokach analizy). Osoba decyzyjna ma jednak swoją wizję (zaufanie do zawodnika lub przeświadczenie o ważności utworu): jak postąpić w takiej sytuacji i jak się odnieść do tego, wydawałoby się, zobiiektywizowanego obrazu sytuacji: zaufać przyjętej przez siebie metodzie, czy zrewidować jej parametry, otwierając się na wnioski wyciągnięte z wyników pomiaru?

W humanistyce związanej mocniej z interpretacją, tej „blisko tekstu”, bliskim skojarzeniem do wyżej zaprezentowanych metod będzie geopoetyka, czyli jak podaje w najbardziej znanym polskim wprowadzeniu do tego zagadnienia Elżbieta Rybicka „orientacja, której celem jest badanie interakcji pomiędzy twórczością literacką a przestrzenią geograficzną”<sup>48</sup>. Co charakterystyczne, mimo zaawansowanej „danetyzacji” (*datafication*) dziedziny, jaką jest geografia, w przywołanej pracy znajdziemy przede wszystkim konteksty antropologiczne i badań kulturowych, bardziej mowa tu o doświadczaniu czy opisywaniu przestrzeni w dziełach literackich.

Z kolei wśród prac humanistów cyfrowych spotkamy się z położeniem akcentu na wszelkie mierzalne aspekty interpretacji położenia geograficznego. Odwołaniem są systemy informacji geograficznej (GIS, *geographic information systems*), czyli systemy informacji ściśle związane z danymi położenia i lokalizacji. Tego rodzaju informacje, poza swoim typowym zastosowaniem geodezyjnym, militarnym czy planistycznym, mogą służyć do innych celów. Na przykład jako surowe dane mogą być przetwarzane jako odniesienia w badaniach humanistycznych<sup>49</sup>. Spotkamy takie podejście w zbiorze informacji o humanistyce cyfrowej

---

„Danetyka, czyli o polskim tłumaczeniu Data Science”, 2019-10-03, <https://www.kodolamacz.pl/blog/data-science-po-polsku/>. Dostęp 2023-09-16.

<sup>48</sup> Elżbieta Rybicka, *Geopoetyka. Przestrzeń i miejsce we współczesnych teoriach i praktykach literackich* (S.l.: Universitas, 2014), 10.

<sup>49</sup> Por. David J. Bodenhamer, John Corrigan, i Trevor M. Harris, red., *The spatial humanities: GIS and the future of humanities scholarship*, Spatial humanities (Bloomington: Indiana University Press, 2010). W dalszej części tego rozdziału nawiązuję do prac, które analizują miejsca wydania dzieł. Niektóre z tych badań odwołują się wprost do GIS.



University of Exeter<sup>50</sup>, zbiorze narzędzi, które rekomendują biblioteki University of Texas<sup>51</sup>, czy wstępie do humanistyki cyfrowej Temple University<sup>52</sup>. Poza środowiskiem akademickim, firmy oferujące wykorzystanie komercyjne GIS również wskazują konteksty humanistyki, np. w formie scenariuszy lekcji i tematycznych kolekcji do wykorzystania w badaniach<sup>53</sup>. Przykładem polskiego projektu łączącego refleksję dotyczącą przestrzeni, geograficzne dane (np. wizualizacje kartograficzne) jest Nowa Panorama Literatury Polskiej<sup>54</sup>.

Jak widać, rysują się dwa porządki uprawiania literaturoznawstwa interdyscyplinarnego, idącego w kierunku geografii<sup>55</sup>. Pierwsza jego odmiana będzie poszukiwać toposów i wątków w tradycji, odwoływać się do kulturowych kodów oraz praktyk, pozostając przy tradycyjnym zestawie narzędzi, związanych z interpretacją. Druga może kod potraktować na cyfrową modłę i korzysta z danych lokalizacji, by wytwarzać z nich naukową narrację lub wspierać formułowanie wniosków.

Stan badań nad humanistyką cyfrową

Zwymiarowanie w obszarze kultury przynosi dyskurs nowych mediów, o którym już wspomnieliśmy wcześniej na zasadzie przenikania tekstów kultury do tworzenia technologii. Mamy wszakże do czynienia z procesem odwrotnym, czyli „ucyfrowieniem” dyskursu o sprawach humanistycznych.

---

<sup>50</sup> Zob. „Working with Geospatial Data - Digital Humanities – Introduction” w *LibGuides at University of Exeter*, <https://libguides.exeter.ac.uk/digitalhumanities/geospatial/>. Dostęp: 2023-09-16.

<sup>51</sup> Zob. „GIS & Mapping - Digital Humanities Tools and Resources” w *LibGuides at University of Texas at Austin*, <https://guides.lib.utexas.edu/digitalhumanities/gis-mapping>. Dostęp: 2023-09-16.

<sup>52</sup> Zob. „GIS Mapping - Introducing the Digital Humanities” w *Research Guides at Temple University*, <https://guides.temple.edu/c.php?g=78518&p=506481>. Dostęp: 2023-09-16.

<sup>53</sup> Zob. „GIS for Humanities” w *Esri*, <https://learn.arcgis.com/en/paths/gis-for-humanities/>. Dostęp: 2023-09-16.

<sup>54</sup> Zob. „Nowa Panorama Literatury Polskiej”, <https://nplp.pl/>. Dostęp: 2023-09-16. Na stronie znajdziemy projekty ściśle związane z konkretnymi tematami przestrzennymi, jak atlas miejsc istotnych dla polskiego romantyzmu (co dodaje wizualnego odbiorcy kontekstu dla wyobrażenia sobie przepływu idei między ośrodkami oraz podróży romantyków) czy w mniejszej skali: multimedialne plany Warszawy, które ukazują miasto w odniesieniu do życia i twórczości Bolesława Prusa (trzy wersje przedstawienia graficznego zawierają ważne lokacje dla: Głowackiego-osoby, Prusa-twórcy i Wokulskiego-postaci *Lalki*).

<sup>55</sup> Nieco komplikując ten obraz, możemy jeszcze wskazać wątki sytuacjonistyczne: psychogeografii i dryfu, np. u Guya Deborda, co przenosi akcent w stronę aktywności, a nawet aktywizmu w doświadczaniu przestrzeni – nie jest to jednak ściśle związane z dalszymi wywodami, stąd zaledwie sygnał o tym możliwym odgałęzieniu tematycznym. Por. np. Paweł Krzaczkowski i Mateusz Kwaterko, red., *Przewodnik dla dryfujących: antologia sytuacjonistycznych tekstów o mieście*, Bęc. Seria Kieszonkowa (Warszawa: Fundacja Nowej Kultury Bęc Zmiana : Austriackie Forum Kultury, 2016).

Charakterystyki nowych mediów wyjaśnia Manovich w swojej klasycznej już pracy *Język nowych mediów*<sup>56</sup>. Formułuje przy tym zbiór prawidłowości, jakie zachodzą ze względu na przemiany w dominującym modelu wymiany informacji. Pierwszą istotną cechą jest wyrażenie informacji numeryczne (dzięki temu może być przeliczana). Obiekt nowych mediów przez możliwość obróbki algorytmicznej zyskuje zdolność do programowania go. Kolejna to modułowość, czyli zwiększenie operatywności zbiorów danych, ponieważ stają się elementami „układanki” dla twórcy, odbiorcy i badacza. Manovich używa też pojęcia struktury „fraktalnej” – wspomniane moduły daje się łączyć i rozłączać bez utraty sensu. Możemy dyskutować nad bezstratnością (znaczenia) w tym wypadku, ale jest to pytanie o kwestię gradacji, a sama zasada sformułowana przez badacza działa. Taką modułową czy fraktalną budowę ma też sieć WWW, złożona z wielu oddzielnych elementów. Autor opisuje internet jako nieciągły i niehierarchiczny. Kolejny ważny aspekt nowych mediów stanowi automatyzacja. Manovich wskazuje, że możemy usunąć dzięki niej intencjonalność. Dotknęliśmy tego aspektu, mówiąc o splocie analizy i interpretacji dzieła w tekstach krytycznych: prawdopodobnie nie są to działania rozłączne i krokowe, jak mogłyby sugerować strukturalistyczne metody odczytywania tekstów. Z kolei automatyzacja oznacza zawsze posługiwanie się procedurą, co zapewnia jej również przewidywalność. Użycie narzędzi cyfrowych stwarza szansę na redukcję „szumu” i zachowanie sygnału (pożądaną informację) – w działaniach związanych z humanistyką realizuje się to np. jako autokorekta (z pożytkiem, ale nieraz szkodą przy nazwach własnych i intencjonalnych niestandardowych wariacjach zapisu). Automaty wytwarzają, edytują, a także udostępniają obiekty cyfrowe i pomagają je filtrować (wyszukiwanie w obrębie danego systemu lub podsystemu). Media cyfrowe opisuje również zmienność (wariacyjność) i zastąpienie stałych zmiennymi (*variables*). Manovich wskazuje tutaj na powszechną praktykę w tworzeniu kodu programistycznego (używa się zmiennych zamiast stałych wszędzie tam, gdzie to możliwe)<sup>57</sup>. Z perspektywy wdrożenia i zastosowania, odbywa się to przy jednoczesnym przeniesieniu ciężaru kognitywnego na użytkownika (odbiorcę) przy wyszukiwaniu, gromadzeniu i przetwarzaniu informacji. Manovich podaje tu przykład zakupów i przeglądania listingu produktów przez klientów zamiast bycia obsługiwanymi przez personel tradycyjnego sklepu, ale ten sam mechanizm

---

<sup>56</sup> W oryginale Lev Manovich, *The Language of New Media*, 8. print, A Leonardo Book (Cambridge, Mass.: MIT Press, 2001).

<sup>57</sup> Moje minimalne doświadczenie z programowaniem w języku Python potwierdza tę informację – w ramach mentoringu technicznego, prowadzonego w związku z projektem doktoranckim, stałe wprowadzałem jako wartości testowe, dla sprawdzenia działania kodu w kontrolowany sposób. Docelowo w takim miejscu miała się pojawić zmienna.

zachodzi, gdy pomyślimy o poszukiwaniu danych na określony temat humanistyczny ze wsparciem ekspertów i samodzielnie, przy użyciu sieci: taki przykład przytaczam w kolejnym rozdziale. Ostatnią cechą mediów cyfrowych stanowi transkodowanie, czyli wpływ mediów cyfrowych na kulturę. Manovich przewidywał, że cywilizacja zmierza od kultury mediów do kultury oprogramowania. Nowe media są rozpowszechniane, tworzone i odbierane za pomocą komputerów; jak zauważa Manovich, skutkuje to zastosowaniem logiki związanej z komputerami do tych procesów, w przeciwieństwie do mediów tradycyjnych. Z czasem dyskursy i logiki obu porządków, tradycyjnej humanistyki i komputerowej inżynierii, będą się przenikać – ta prognoza Manovicha właściwie może zostać uznana za spełnioną współcześnie<sup>58</sup>.

Istotne z naszego punktu widzenia są także cechy wariacyjności, które wymienia badacz, dlatego wróć do tego wątku, by je przytoczyć: 1) przechowywanie zawartości w relacyjnej bazie danych, 2) rozdzielenie poziomu danych od interfejsu (ale nadal z potencjałem możliwości dostępu do danych bezpośrednio lub za pośrednictwem GUI, *Graphic User Interface*), 3) wykorzystanie przez medium dopasowania informacji o użytkowniku do zmiany struktury, elementów medium, wyświetlania treści, 4) interaktywny schemat wyboru (logika rozgałęzień drzewa), 5) hipermedia (w tym hipertekst), pozwalające na przechodzenie między treściami bez z góry założonego „prawidłowego” porządku i kierunku, 6) uaktualnienia i poprawki – media cyfrowe stale się zmieniają i podlegają aktualizacjom, 7) skalowalność: rozumiana tu np. jako dopasowanie wyświetlanej grafiki do wydajności komputera albo wczytywanie tylko początkowych rekordów kwerendy w GUI jako przykładu, z możliwością pobrania całości jako pliku.

Według tego autora możemy także mówić o estetyce baz danych w wymiarze takiej bazy jako formy symbolicznej<sup>59</sup>. Baza danych to samodzielna forma kulturowa. Manovich podkreśla obecność zasady narzucenia komputerowej ontologii na rzeczywistość: mamy do czynienia ze strukturami danych i algorytmami. Do algorytmu da się sprowadzić każdą czynność, a każda rzecz wymagająca opisu może zostać zgodnie z logiką bazodanową jako obiekt. Pozornie zdradza to dynamikę pasywności i aktywności, ale obiekty muszą zostać aktywnie wytworzone

---

<sup>58</sup> Mam na myśli popularność i powszechny użytek smartfonów oraz innych urządzeń osobistych, które wykonują zadania komputerów, użycie mediów cyfrowych w edukacji oraz codziennym życiu, a nawet powszechne zastosowanie narzędzi cyfrowych w kwestiach osobistych finansów, ubezpieczeń, ochrony zdrowia, życia obywatelskiego i spraw urzędowych.

<sup>59</sup> Lev Manovich, „Database as Symbolic Form”, w *Database Aesthetics. Art in the Age of Information Overflow*, red. Victoria Vesna, t. 20, Electronic Mediations, b.d., 39–60, dostęp 1 marzec 2023. Dostosowany przedruk jako część rozdziału piątego w Manovich, *The Language of New Media*.

i są stale przetwarzane, a algorytm ma też cechy stałości, gdy weźmiemy pod uwagę pozostawienie kodu, by działał, a ten będzie działał cały czas w ten sam, przewidywalny sposób. Bazy wymagają także indeksowania metadanymi. Tu Manovich przekazuje interesujące porównanie literackie. Biorąc pod uwagę, że każdy obiekt ma swoje dane opisowe, spełniła się w rzeczywistości literacka przepowiednia Borgesa o mapie wielkości terytorium, które opisuje. Na tym jednak nie koniec, ponieważ logikę bazodanową Manovich przykłada do narracyjności. Fabuła może być opisana umownie jako algorytm. Baza danych natomiast stoi w centrum działalności kreatywnej<sup>60</sup> ale nie może zostać opisana w parze z narracyjnością, fabułą (pomimo istnienia „interaktywnych” fabuł jako konstelacji punktów do przejścia w wybranej konfiguracji, sama przedustawna tabela zawierająca te punkty nie stanowi jeszcze narracji). Pod względem semiotycznym, zawartości baz danych składają się z warstw, z których każda jest opracowywana rozłącznie względem pozostałych (i może na nią wpływać inny aktor czy instancja, np. masowa zmiana metadanych na zbiorze zdjęć przez kod programistyczny i edycje warstwy graficznej przez człowieka). Relacje elementów baz mają charakter syntagmatyczny i paradygmatyczny<sup>61</sup>. Z jednej strony niosą znaczenie w sekwencji, z drugiej – mają przypisane klasy i właściwości, przez co rozpoznajemy typy. Elementy syntagmatyczne są eksplicytne, a paradygmatyczne implicytne: widzimy realizację z użyciem danego elementu; znamy alternatywy, których nie zastosowano w ich miejsce oraz abstrakty, które je opisują. Manovich widzi odwrócenie porządku w przypadku nowych mediów, gdzie baza jako paradygmat staje się czymś materialnym, a narracja (syntagmat) dematerializuje się. Interaktywne interfejsy prezentują swym użytkownikom możliwości wielu wyborów naraz, i to w więcej niż jednym wymiarze<sup>62</sup>.

---

<sup>60</sup> Bazy zdjęć, filmów i innych materiałów powszechnie stosuje się w działalności komercyjnej w mediach, komunikacji i reklamie (częściowo przekłada się to na produkcję kultury masowej). By odwołać się do adekwatnego przykładu humanistycznego oraz lokalnego, w ramach konsorcjum DARIAH i przy udziale uczestnika z UAM w Poznaniu (Konrad Dominas) rozwija się projekt E-Theseus Map. Jego celem jest stworzenie narzędzia webowego do wizualizacji i mapowania relacji zachodzących między postaciami, motywami i wątkami mitu na przykładzie podania o Tezeuszu. W dalszej kolejności ma to pozwolić na rozszerzenie części mapy o nowe konteksty odbioru: religijne, literackie, psychologiczne, językoznawcze itp. Samo narzędzie ma możliwość stać się platformą także dla innych tematów badawczych opartych na fabule. Dzięki takim rozwiązaniom badacze i entuzjaści mogą śledzić wersjonowanie opowieści, czyli patrzeć na fabuły w bazodanowy sposób. E-Theseus Map to przykład infrastruktury, która to zapewnia.

<sup>61</sup> W tym miejscu Manovich odwołuje się do ustaleń de Saussure’a oraz późniejszych komentarzy Barthesa, czyli z jednej strony ogólnych praw działania języka, a z drugiej poststrukturalistycznego rozszerzenia tych prawidłości na inne systemy znaków.

<sup>62</sup> Wiedzą o tym wszyscy, którzy korzystają z serwisów strumieniowania wideo i spędzili więcej czasu na wyborze treści niż oglądaniu jej.

On the level of the whole object, the user is made aware that he or she is following one possible trajectory among many others. In other words, the user is selecting one trajectory from the paradigm of all trajectories that are defined<sup>63</sup>.

Manovich podkreśla dosłowność interaktywności, którą oferują media cyfrowe. Doświadczenie użytkownika zostaje porównane do syntagmatycznej sekwencji, związanej z ograniczonymi wyborami, w granicach paradygmatu. Badacz widzi w tym wpływ kina jako medium, gdzie wydarza się jedna rzecz naraz – w przeciwieństwie do bardziej przestrzennych form, gdzie mieliśmy do czynienia z większą intensywnością symultaniczności. Opisuując coś, co nazywa „kompleksem bazy danych”, Manovich proponuje spojrzenie na narrację i bazę danych jako rywalizujące ze sobą tryby wyobraźni<sup>64</sup>. Tym, co będzie dla nas istotne w kontekście powyższych ustaleń, z myślą o bibliograficznych bazach danych, jest interakcja człowiek-komputer w sposobie docierania do informacji i wzajemna gra obu elementów wyobraźni, o których pisze Manovich.

O naukowej popularności świadczy wydanie latach 2004 i 2015 dwóch przewodników po *Digital Humanities*<sup>65</sup>. W pierwszym zapoznamy się z historycznym rozwojem tytułowej metody, a także jej aplikacjami dla różnych dyscyplin (tak różnorodnych jak archeologia i leksykografia). Omówione zostają także reguły komputerowe w odniesieniu do humanistyki (podejście stosunkowo zbieżne do prezentowanego przez Manovicha). W rozdziale o zastosowaniach czytelnicy pierwszego przewodnika mogą szukać inspiracji i zapoznać się ze studiami przypadków. Osobną część poświęcono wytwarzaniu, rozpowszechnianiu oraz archiwizacji danych – te czynności są często punktem wspólnym nauk humanistycznych i technicznych. Drugi przewodnik, już bardziej zniuansowany, porusza konkretne zagadnienia problemowe – obowiązuje w nim klucz grupowania tematów (infrastruktury, kreacja, analiza, rozpowszechnianie i rozwój *Digital Humanities*), co świadczyłoby o tym, że w odstępnie prawie

---

<sup>63</sup> Manovich, „Database as Symbolic Form”, 49.

<sup>64</sup> Siłąc się na lokalny przykład spoza tekstu badacza, poezję Eugeniusza Tkaczyszyna-Dyckiego wyobrażam sobie programowo jako bazę danych, a u Kiry Pietrek widziałbym zbiór narracji. Z zastrzeżeniem, że każda książka poetycka może być rozumiana jako bazodanowa kompilacja.

<sup>65</sup> Schreibman, Siemens, i Unsworth, *A companion to digital humanities*; Susan Schreibman, Ray Siemens, i John Unsworth, red., *A New Companion to Digital Humanities*, 1. wyd. (Wiley, 2015), <https://doi.org/10.1002/9781118680605>. Pierwszy z nich dostępny w całości *online*: <https://companions.digitalhumanities.org/DH/>.

dziesięciolecia przestała być potrzebna mapa nowego obszaru, a bardziej użyteczny stał się model targów, w rodzaju humanistyczno-cyfrowego „Expo” projektów<sup>66</sup>.

W dokumencie z marca 2009 roku zatytułowanym *The Promise of Digital Humanities*<sup>67</sup>, co jest wczesnym użyciem terminu, pada propozycja zdefiniowania obszaru badawczego i charakterystyki. Ów *whitepaper* przygotowano na Uniwersytecie Kalifornijskim w Los Angeles, w tonie sugerującym wyraźnie chęć zaznaczenia pozycji jednostki i ubiegania się o środki dla dalszego rozwoju pojawiającej się (emergentnej) dyscypliny, ze względu na dotychczasowe realizacje na tym polu. Nie chciałbym się tu skupiać na tym elemencie promocyjno-retorycznym, o którym warto po prostu pamiętać, ponieważ bardziej istotne w odniesieniu do naszego tematu są tu próby opisanie samej humanistyki cyfrowej w tamtym miejscu i czasie.

Samo *Digital Humanities* zostaje określone jako termin parasolowy, obejmujący praktyki tworzenia, stosowania i interpretowania nowych technologii cyfrowych oraz informacyjnych. Autorzy zwracają uwagę na to, jak pojawiające się rozwiązania techniczne wpływają na wiele dziedzin humanistycznych pod względem sposobów prowadzenia badań i proponują to nowe określenie jako naturalny rozwój humanistyki tradycyjnej – w żadnym wypadku porzucenie już wypracowanych narzędzi. Z jednej strony jest to więc diagnoza i propozycja odpowiedzi na zmieniający się świat, z drugiej – podkreślenie zachowania tożsamości już istniejących dyscyplin naukowych.

Charakterystyki wymienione przez autorów to interdyscyplinarność, kolaboratywność (oparcie na zespołowości), zaangażowanie społeczne (impakt; autorzy szczególnie podkreślają też jej związek z instytucjami zachowującymi dziedzictwo kulturowe), globalność (ze względu na łatwość wymiany informacji w internecie), aktualność i istotność (w wymiarze pedagogicznym m.in. przygotowanie osób uczących się na wyzwania współczesnego świata, zanurzonego w informacji cyfrowej).

Widać w tym komunikacie apel do rozwijania współpracy międzywydziałowej i międzyuczelnianej. Siłą rzeczy jest to dokument skupiony na szansach związanych z funkcjonowaniem uniwersytetu jako instytucji, podkreśleniu roli i osiągnięć UCLA, a także

---

<sup>66</sup>Z perspektywy naszego tematu najważniejsza byłaby część dotycząca infrastruktury, a w nim tekst Jennifer Edmond o infrastrukturze i współpracy oraz artykuł poświęcony danym połączonym w części poświęconej analizie, zob. Schreibman, Siemens, i Unsworth, *A New Companion to Digital Humanities*, 54-65, 251-273.

<sup>67</sup>Todd Pressner i Chris Johanson, „The Promise of Digital Humanities”, 1 marzec 2009. W dalszej części opracowania odwołuję się do całości tego tekstu.

pozyskiwaniem finansowania, zarówno ze środków publicznych, jak i z sektora prywatnego. Na liście przykładowych projektów UCLA znajdziemy m.in. Encyklopedię Egiptologii, trójwymiarowy model Forum Romanum, czy wirtualna rekonstrukcja 3D katedry w Santiago de Compostela. Przenikają się tutaj kompetencje i współpraca m.in. historyków, lingwistów, informatyków, a nawet urbanistów i architektów.

Jako kluczowe obszary rozwoju humanistyki cyfrowej twórcy dokumentu wymieniają: 1) strategie zatrudniania, 2) pozyskiwanie funduszy grantowych, 3) dedykowane przestrzenie i zasoby informatyczne (w tym zasoby ludzkie), 4) kadrę zarządzającą grantami, 5) zespoły odpowiedzialne za archiwizowanie, digitalizację i publikację danych, wytworzonych w badaniach *Digital Humanities*, z zachowaniem standardów, metadanych i dostępności danych, 6) wsparcie instytucyjne w formie wewnątrzuczelnianych grantów skierowanych do wielu wydziałów oraz wspierających projekty humanistyki cyfrowej na wczesnym etapie rozwoju (inkubacja).

Przedostatni punkt jest szczególnie ważny dla kolejnych naszych dociekań. Wpisuje się on bowiem w ogólną troskę humanistyki o zbieranie doświadczeń oraz zachowanie ciągłości dostępu do informacji w działalności badawczej i edukacyjnej. Realizuje też potrzebę dostępności nauki i komunikowania wyników badań na zewnątrz uczelni.

We wszystkich punktach wyraźnie rysuje się, poza mocno podkreślonym wątkiem finansowania nauki<sup>68</sup>, potrzeba budowania infrastruktury dla nowego typu prowadzenia badań. Taką infrastrukturę tworzą nie tylko miejsca i narzędzia, ale również odpowiednio przeszkoleni ludzie, dokonujący transferu wiedzy. Silny nacisk został położony na wytwarzanie wiedzy i publikację badań. Pod względem dydaktyki ważnym i aktualnym po dziś dzień spostrzeżeniem pozostaje konieczność budowania u osób uczących się tzw. *media literacy* czyli świadomości świata mediów. Współczesne pokolenia studentów uczelni wyższych to tzw. *digital natives*, osoby wychowane wśród urządzeń elektronicznych z dostępem

---

<sup>68</sup> Dwie uwagi odnośnie aspektu finansowania. Warto też pamiętać o kontekście szkolnictwa wyższego w USA i systemu jego finansowania, który różni się od polskiego czy europejskiego. Odwołania do prywatnych sponsorów i zewnętrznych źródeł dofinansowania prawdopodobnie są częścią tamtejszej kultury akademickiej. Z drugiej strony, kwestia znalezienia dodatkowego budżetu na badania akurat w obszarze humanistyki cyfrowej wiąże się z realnym kosztem narzędziowym, który może wykraczać poza planowane wydatki danej jednostki, prowadzącej badania. Mam tu na myśli współpracę z informatykami, opłacanie infrastruktury w sieci (przechowywanie danych), narzędzia (komputery, licencje na oprogramowanie) i inne koszty, które mogą okazać się konieczne (np. prace utrzymaniowe czy naprawy).

do internetu. Tym prawdziwsza diagnoza badaczy z UCLA, skłaniająca do łączenia tradycji badań z komponentami cyfrowymi.

### Projekty cyfrowe

Humanistyka cyfrowa przede wszystkim realizuje się w projektach. Mogą to być duże inicjatywy grantowe, obejmujące szersze infrastruktury między wieloma ośrodkami (tak jak w konsorcjach DARIAH<sup>69</sup> i CLARIN<sup>70</sup>), ale również mniejsze, powstające w małych grupach badawczych lub jako inicjatywy indywidualne. Nawet w przypadku tych ostatnich, aspekt kolaboracyjny jest bardzo istotny. Humanistyki cyfrowej nie uprawia się samotnie, „dla siebie”, ani „sobie a muzom”. Oczywiście impakt i wymiar społeczny „niecyfrowej” czy tradycyjnej humanistyki także jest istotny, ale można w niej silić się na większe syntezy i tematy badawcze w pojedynkę. W przypadku humanistyki cyfrowej tak się nie dzieje, nad danym zagadnieniem najczęściej pracują grupy o różnych kompetencjach i każda osoba wnosi wkład z innej domeny. Gdyby przyszło nam porównywać do dyscyplin sportowych, to w humanistyce cyfrowej więcej byłoby konkurencji drużynowych, a tradycyjnych dziedzinach – indywidualnych.

Nie ma tu jednolitego tematu, przedmiotu badań dla całości zjawiska, które nazywamy „humanistyką cyfrową”. Badane zagadnienie może pochodzić z jednej z dziedzin humanistycznych, może też sytuować się na przecięciu różnych dziedzin. Projekty mogą łączyć się modułowo w większe programy, co zwiększa ich złożoność i różnorodność w obrębie tej samej inicjatywy.

Do badań humanistycznych przenika dyskurs biznesowy, z dziedzin zarządzania i IT. Coraz bardziej popularne stają się metodyki *Scrum* i *Agile* (projektowanie zwinne). Humanisłci

---

<sup>69</sup> DARIAH, czyli Digital Research Infrastructure for the Arts and Humanities to „pan-europejska infrastruktura dla badaczy humanistyki i sztuki”. Jej misją jest umożliwienie zdigitalizowanych badań, a także łączenie osób badających zagadnienia związane z tymi obszarami nauki, jak również sieciowa dystrybucja wiedzy, idei narzędzi i metod. Przy użyciu nowoczesnych narzędzi, organizacja dąży do wyniesienia badań na poziom europejski: wspiera współpracę między dyscyplinami oraz krajami. DARIAH powstało w 2014 roku jako European Research Infrastructure Consortium (ERIC). Wśród rozwijanych w obrębie konsorcjum inicjatyw na pewno warto w kontekście PBL wymienić grupę Bibliographical Data, którą kierują Vojtěch Malínek, dyrektor Czeskiej Bibliografii Literackiej, oraz Tomasz Umerle, zastępca kierownika Pracowni Bibliografii Bieżącej w Poznaniu. Zob. „DARIAH | Digital Research Infrastructure for the Arts and Humanities”, <https://www.dariah.eu>. Dostęp: 2023-09-23.

<sup>70</sup> CLARIN, czyli Common Language Resources and Technology Infrastructure (założona w 2012 jako CLARIN ERIC) to infrastruktura, służąca do badań nad językiem naturalnym i efektywniejszej wymiany danych związanej z tymi badaniami. W swojej strategii kieruje się zasadami otwartej nauki oraz FAIR. Wśród centrów na całym świecie działa polski oddział tej organizacji, z Politechniką Wrocławską jako liderem projektu i przy współpracy sześciu innych ośrodków. Zob. "Homepage | CLARIN ERIC", <https://www.clarin.eu/>. Dostęp: 2023-09-16. oraz „CLARIN-PL”, <https://clarin-pl.eu>. Dostęp: 2023-09-16.



cyfrowi, również w Polsce, chętnie posługują się tą terminologią. Pracują w *sprintach*, czyli krótkich interwałach, np. dwutygodniowych, po których jest dostarczany gotowy produkt, rozumiany jako minimalny efekt: rzecz gotowa do użycia czy poznania przez odbiorcę i w jakiejś mierze skończona (mająca wytyczone granice, określony zakres). W dalszej kolejności skaluje się i rozwija te owoce pracy badaczy w większe rozwiązania, w sposób modułowy, jak również udostępnia dla innych. Przeciwnieństwem *Agile* jest podejście *Waterfall*, czyli kaskada: stałe wytwarzanie takiej samej wartości jak np. produkcja jednego typu produktu w fabryce. W doświadczeniu humanisty akademickiego także spotkamy te podejścia. *Agile* może być zarządzanie wydaniem jednego numeru czasopisma w trybie projektowym (projekt jest inicjatywą skończoną), a *Waterfall* można przyrównać do pracy dydaktycznej, opartej na programach nauczania ze wskaźnikami efektów kształcenia (obserwujemy wyraźny cykl roczny czy semestralny w edukacji kolejnych roczników studentów i praca ta wydaje się nie mieć końca). Mowa o wytwarzanych funkcjonalnościach (*features*). Praca odbywa się *iteracyjnie*. Humanisci cyfrowi kreślają konieczność wytworzenia funkcjonalności np. zgodnie z zasadami priorytetyzacji *MoSCoW* (*Must have, Should have, Could have, Will not do* – od celu niezbędnie koniecznego i nienegocjowalnego, poprzez efekty możliwe, ale nieobowiązkowe, do odrzuconych w danym działaniu, ale możliwych do realizacji w innych okolicznościach).

Nie dziwi to w sytuacji, kiedy sam rozwój nauki często odwołuje się do zarządzania projektami, a umiejętności, znajomość metodyk i certyfikaty stanowią wymagania pracy przy niektórych inicjatywach, w tym grantowych (np. PRINCE2 czy PMBOK jako metody i certyfikaty w obszarze zarządzania projektami). Nie wydaje się, by było to wyłącznie modą czy nowinką, skoro praca projektowa stanowi w świecie akademickim często obecne rozwiązanie systemowe.

Wiąże się to na pewno z pracą zespołową. Tak jak wskazują badacze z UCLA, kolaboratywny aspekt projektów cyfrowych jest nieodzowny. To nie tylko w języku opisu inicjatyw i adaptowaniu metod zarządzania, ale także w podziale zadań i wykonawstwie widać konieczność interdyscyplinarnego podejścia. Humanista cyfrowy z zapleczem wiedzy kierunków humanistycznych jak językoznawstwo i literaturoznawstwo może pełnić rolę kierownika projektu czy głównego użytkownika, zlecającego zadania do osoby (lub jednostki) odpowiedzialnej za część techniczną. Zdarzają się też coraz częściej sytuacje transferu kompetencji i humanisci sięgają po narzędzia informatyczne. Z jednej strony po to, by samemu wytwarzać komponenty projektów, z drugiej – aby ułatwić porozumiewanie się z pozostałymi

członkami zespołu i uzmysłowić sobie wachlarz możliwości technicznych. Działa tutaj także metaforyka laboratorium,<sup>71</sup> do której wracam przy okazji krytyki zjawiska *Digital Humanities*.

To, co badaczowi podpowiada intuicja, badania z użyciem narzędzi cyfrowych mogą wesprzeć obiektywizacją w postaci pracy na danych. Zostały już omówione dwa wzajemnie się napędzające w humanistyce cyfrowej dyskursy: literaturoznawczy i projektowo-techniczny. Ten drugi może też przybrać bardziej naukowy charakter, z odwołaniem do matematyki i fizyki. Chciałbym zaprezentować pracę badacza działającego na pograniczu dwóch dziedzin, w której widzę impuls emancypacji humanistyki cyfrowej jako bardziej technicznego narzędzia.

Nieprzypadkowo, aby w ogóle zacząć refleksję nad semiotyką w humanistyce cyfrowej, Michael Filimowicz odnosi się do filozofii Jacques'a Derridy<sup>72</sup>. Koncepcje francuskiego badacza na stałe wrosły w główny nurt badań literackich i stanowią istotny punkt odniesienia dla humanistyki w ogóle<sup>73</sup>. Filimowicz skupia się na próbie „wynalezienia na nowo” myślenia o koncepcji różnicy i swoistego rozminowania krytycznego stosunku cybernetyki, obu dokonanych przez Derridę w *O gramatologii*<sup>74</sup>. Zacznę od drugiej kwestii, która wydaje się ciekawsza: oto francuski autor identyfikuje całokształt badań cybernetycznych jako domenę podległą pismu: „Obszarem pisma, bez względu na to, czy ma ono zasadnicze granice, czy nie, będzie wreszcie cały obszar objęty *programem* cybernetycznym. Jeśli przyjmiemy, że teoria cybernetyki może wchłonać wszelkie pojęcia metafizyczne – nawet takie, jak dusza, życie, wartość, wybór, pamięć – które jeszcze niedawno służyły przeciwstawianiu człowiekowi maszyny, to będzie ona musiała zachować pojęcie pisma, śladu, gramu czy grafemu do chwili, gdy ujawni się również jej historyczno-metafizyczna przynależność”. We fragmencie dotyczącym opozycji człowiek-maszyna Derrida wprowadza przypis, który brzmi: „Wiadomo, że na przykład Wiener pozostawia »semantyce« uznane przezeń za zbyt grube i zbyt ogólne przeciwieństwo między żywym i nieżywym itd., mimo to do określania maszyn wciąż używa wyrażeń takich jak »organy zmysłów«, »organy poruszające« itp.”.

---

<sup>71</sup> Por. Bruno Latour i Steve Woolgar, *Laboratory life: the construction of scientific facts* (Princeton, N.J.: Princeton University Press, 1986). Do metaforyki laboratorium wracam w części związanej z krytyką *Digital Humanities*.

<sup>72</sup> M. Filimowicz, *Sysemics [w:] Information Semiotics for Digital Humanities: Towards a New Conception of Quantity and Quality in Culture*, Vancouver 2020, s. 7-21. W dalszej części przytaczam główne tezy wyводу.

<sup>73</sup> W innym miejscu omawiam analizę zawartości czasopisma „Teksty drugie” autorstwa Macieja Maryła, której wyniki, dotyczące wystąpień nazwisk twórców, potwierdzają tę obserwację.

<sup>74</sup> J. Derrida, *O gramatologii*, przeł. B. Banasiak, Warszawa 1999.

Filimowicz widzi w takim postawieniu sprawy kilka problemów. Po pierwsze, totalizujący gest odrzucenia całej szkoły czy nurtu badań przy powołaniu się na jedno, niedokładne (niezidentyfikowane w konkretny sposób) źródło. Po drugie, zawłaszczenie dyskursu cybernetycznego, poprzez sklasyfikowanie go jako części piśmiennictwa, dzięki czemu może stanowić „teren” filozofii Derridy. Po trzecie, projektowane zawłaszczenie przez cybernetykę, „w jej obecnym wcieleniu jako teorii systemów i informatyki”<sup>75</sup>, zainteresowań filozofów (Filimowicz obok Derridy wymienia Heideggera) zagadnieniami metafizycznymi. Po czwarte, amerykański badacz sugeruje swoistą zazdrość i napięcie Derridy, związane z tym, że jakaś inna dziedzina zbliża się do opisanego jako pierwsza problemów, które nurtują filozofa. Filimowicz podaje jako przykłady dekonstrukcję binarnych opozycji człowieka i zwierzęcia oraz umysłu i maszyny. Jednocześnie amerykański badacz zauważa, że teoria systemów postuluje wspólne własności organizacji dla każdego systemu, aby mógł być uznawany za system. Tymczasem u Derridy klasyfikacja całej teorii systemów jako części piśmiennictwa niejako ma służyć poddaniu jej kontroli za pomocą narzędzi hermeneutycznych, by przeprowadzić na niej procedurę dekonstrukcji - co nie zajmuje wcale teoretyków systemów, będących zbyt zajętymi budowaniem, by obchodziło ich to burzenie <sup>76</sup> . Po piąte i ostatnie, Filimowicz wprost oskarża francuskiego filozofa o sofizmat rozszerzenia, gdy ten pisze, że cybernetyka będzie musiała zachować pojęcie pisma (i pozostałe wymienione pojęcia – dop. JE). Według amerykańskiego badacza nie ma takiego roszczenia ze strony Wienera i podobnych mu teoretyków, a nawet wręcz przeciwnie. Cybernetyka, przekonuje Filimowicz, interesuje się wieloma więcej sposobami opisu zjawisk niż filozofia, wśród nich: językiem naturalnym, syntezą mowy, schematami elektrycznymi, wizualizacją informacji, telemetrią, matematyką, algorytmami, kodem komputerowym, czujnikami audiowizualnymi itp. Na obronę podejścia spod znaku dekonstrukcji warto dodać w tym miejscu, że filozofia, zwłaszcza filozofia nauki może się zajmować wspomnianymi sposobami komunikacji. Jest jednak niewątpliwie przypisaniem adwersarzowi konkretnej intencji stwierdzenie, że nie obejdzie się bez pisma, gdy nie było to przez niego w ogóle postulowane.

W ten sposób, w oparciu o konkretne cytowania, Derrida zostaje odczarowany – nie jako wierny filozofii podejrzeń demaskator, ale raczej czempion stylu, dostosowujący fakty

---

<sup>75</sup> Filimowicz, 11 (tłumaczenie własne).

<sup>76</sup> Dosłownie, i nie bez złośliwości, pisze Filimowicz: „Derrida needs to inscribe systems theory as a writing system, so that philosophy can wrest »command and control« of cybernetics strictly on hermeneutic grounds. And because systems theorists write, philosophy can attempt to claim the semiotic ground out from under them, through deconstruction, an activity that is probably of little interest to systems theorists, since they are absorbed with tasks such as building robots and other technical feedback systems”. – M. Filimowicz, 11-12.

do swego wywodu. Jest to dla mnie ważne nie o tyle, by wytykać błędy autorowi inspirującemu kolejne pokolenia literaturoznawców, co raczej do obserwacji, że nasza dyscyplina wymaga w niektórych wypadkach „dekonstruowania dekonstruktorów”. Podejmując wyzwanie Filimowicza, możemy cofnąć się o krok, zaobserwować swoiste uprzedzenie ze strony humanistyki i zapytać o jego zasadność. Jest to także dobre ćwiczenie perspektywy do współpracy w interdyscyplinarnych zespołach badawczych.

Odwołajmy się teraz do drugiej kwestii, czyli wymyślenia na nowo dyskursu semiotycznego i problemu różnicy w humanistyce cyfrowej (do czego było nam potrzebne oczyszczenie atmosfery, gdy zaobserwowaliśmy jak wpływowa teoria literatury spogląda na metodologie wynikające wprost z refleksji nad postępem technicznym). Filimowicz, obok krytyki Derridy, podaje następującą koncepcję pozytywną: punktem wyjścia badacza było wszak poszukiwanie kategorii zbieżnej z *différance*, lecz pozbawionej odwołań do metafizyki (nie)obecności.

Celem Filimowicza jest opisanie spektrum między tym, co mierzalne, a tym co jakościowe, w odbiorze sygnału (dźwiękowego, optycznego)<sup>77</sup>. Głównym punktem odniesienia w tych badaniach jest przekaz audiowizualny. W tym celu badacz wyróżnia pięć etapów przyswajania informacji – sygnału, odwołując się do tradycji fenomenologicznej. Dzięki temu projekt Filimowicza ma pogodzić tradycyjną humanistykę, skupioną na znaku, znaczeniu i interpretacji, ze spojrzeniem inżynierskim (oprócz działalności naukowej autor zajmuje się produkcją dźwiękową i muzyczną).

Poszukując własnego modelu semiotyki informacji, Filimowicz proponuje oddzielanie perceptów od afektów. Te pierwsze są świadome, drugie pozostają poza odbiorem jednostkowym. Można sparafrazować (żywię nadzieję, że bez uszczerbku sensu), iż autor przedstawia tutaj kontinuum poznawcze takich sygnałów: od biologicznie i technicznie nierejestrowalnych fenomenów, przez te dające się zaobserwować przez człowieka, a następnie takie, które są możliwe do odbioru ze wsparciem narzędziowym, co wykracza poza możliwości organizmu biologicznego.

Poziomy rozumienia przedstawiam w formie tabeli zgodnie ze układem w cytowanej pracy, który dodatkowo sytuuje sygnały opisowe przez Filimowicza wobec koncepcji Gilberta Simondona i Rona Ihde. Pierwsza dotyczy jednostkowego poznania i indywidualizacji, a zatem podziału na osobowy odbiór rzeczywistości; kontekst technologiczny wspiera swego rodzaju

---

<sup>77</sup> Filimowicz posługuje się pojęciami *quanta* i *qualia*, czyli kwanty i kwalia/qualia – jako podstawowe jednostki w odniesieniu do ilości i jakości sygnału – poddane opisowi kwantytatywnemu i kwalitatywnemu.

wykroczenie poza ludzką kondycję dzięki technologii (poziom transindywidualny), co otwiera konteksty trans- oraz posthumanizmu. Druga łączy fenomenologię z filozofią technologii i ukazuje poziomy obserwacji jednostkowej w ujęciu skali tej obserwacji; rolą technologii jest wsparcie kultury i rozszerzenie zdolności poznawczych jednostki. Propozycja Filimowicza ma być odpowiedzią na metodę schizoanalizy Gilles’a Deleuze’a Félix’a Guattariego i wyraźnie separować wartości kwantytatywne (S1-S2) od kwalitatywnych (S4-S5), z poziomem S3 mediującym na ich granicy (autor podkreśla, że granice między nimi są płynne, ale postuluje podejście systemowe do badania perceptów i afektów)<sup>78</sup>.

*Tabela 1. Porównanie klasyfikacji odbioru sygnałów przez obserwatora w ujęciu Filimowicza*

Signal	Level	Simondon	Ihde	Comments
S1	Physical	Preindividual		
S2	Organic	Transindividual		„Ground” of Affects
S3	Phenomenal	Transindividual	Microperception	„Ground” of Percepts (technically extended)
S4	Phenomenological	Individual	Micro/Macro	
S5	Formal	Individual	Macroperception	

Tak przygotowana rama pojęciowa służy w dalszej części pracy jako odwołanie dla humanisty, aby zakorzenił refleksje dotyczące znaczenia w problematyce doświadczania fenomenów aparaturą zmysłową oraz rejestrowania ich aparaturą techniczną. Filimowicz konsekwentnie w swojej pracy będzie podkreślał różnice między mierzalnymi zjawiskami i ich interpretacjami. Dzięki temu jest jasne, w którym momencie badacz wskazuje na aspekty ściśle inżynierskie, a kiedy na hermeneutyczną część pracy. Autor łączy tutaj zarówno odniesienia do fenomenologii i tradycji semiotyki, jak i wiedzę ściśle techniczną, np. z zakresu akustyki i inżynierii dźwięku. Przykład tego badania ilustruje tendencję humanistyki cyfrowej

<sup>78</sup> Filimowicz, 103.

do opierania się na minimalnych jednostkach nienegocjowalnych<sup>79</sup>. W przypadku danych bibliograficznych także będą one poszukiwane do pobierania, łączenia ze sobą, przetwarzania i prezentowania informacji. Kontekst fenomenologiczny nie będzie istotny. Akcent przesunie się w stronę świadomych decyzji zespołów i badaczy przy pracy na materiale bibliograficznym (czyli raczej będzie mowa o dobrych praktykach i politykach zarządzania danymi oraz rejestrowaniu faktów na drodze przyjętej metody, niż rozważań nad fenomenologicznym aspektem postrzegania np. egzemplarza książki spisywanego z autopsji i pracę intencjonalnej świadomości nad tym aktem).

Warto jednakże wrócić na chwilę do wspomnianej już koncepcji Romana Ingardena i odczytania wiersza Świetlickiego przez Próchniaka. Pewne koncepcje, takie jak szukanie „warstw” są stale obecne w literaturoznawstwie, a humanistyka cyfrowa, szukająca nowych narzędzi, tylko aktualizuje zastosowanie teorii. Przesuwają się akcenty, zmienia się kontekst. Jak zauważymy na przykładzie prac Franco Morettiego, jest także napięcie w pracach nurtu *Digital Humanities* pomiędzy naukami przyrodniczymi, humanistycznymi i informatyką (przyrodniczy element może być inspiracją albo jak u Filimowicza jednym ze źródeł modelu, w tym wypadku w związku z badaniami akustycznymi).

Ważną częścią badań cyfrowych w humanistyce jest także wizualizacja danych i szerzej: prezentacja wyników. Materiały wizualne często pełnią tu dużo ważniejszą rolę niż w tradycyjnych badaniach literaturoznawczych. Istotną pozycją opisującą to zagadnienie jest książka Johanny Drucker o humanistycznej wizualizacji i wyświetlaniu<sup>80</sup>. Na gruncie polskich badań zajęła się tym tematem Maja Starakiewicz z Jakubem Woynarowskim, tworząc przystępny przewodnik komunikacji wizualnej w humanistyce (ich publikacja zasługuje na uwagę ze względu na klarowność informacji oraz doświadczenie obojga autorów w projektowaniu graficznym)<sup>81</sup>. Poniżej prezentuję badania, w których wizualizacja jest istotna ze względu na wolumen przetwarzanych danych.

Bardzo rozpoznawalną pozycję w odniesieniu do humanistyki cyfrowej stanowi *Distant reading* Franco Morettiego, zarówno jako propozycja metodologiczna, jak i lekturowa<sup>82</sup>.

---

<sup>79</sup> Co tłumaczy też Filimowiczowską obronę poprzez atak na takie szkoły czy podejścia humanistyczne, które nakazują wątpić w możliwości poznawcze obserwatora i poszukiwać wewnętrznych sprzeczności w dyskursie.

<sup>80</sup> Johanna Drucker, *Visualization and interpretation: humanistic approaches to display* (Cambridge, Massachusetts: The MIT Press, 2020).

<sup>81</sup> Maja Starakiewicz i Jakub Woynarowski, *Model i metafora: komunikacja wizualna w humanistyce*, Wyd. 1 (Kraków: Korporacja „Ha!art”, 2019).

<sup>82</sup> Franco Moretti, *Distant reading* (London ; New York: Verso, 2013).

W książce o tym tytule badacz proponuje metodę „czytania z oddali”, aby więcej zobaczyć w procesie przemian literatury, w krótkich szkicach oświetla różne nieraz problemy, ale zawsze ustawiając się w owej znacznej odległości od tekstów.

Jest to oczywiście nawiązanie do metody *close reading* (uwważnego czytania) spod znaku formalizmu amerykańskiego Nowej Krytyki – przekorne, może nawet nieco parodystyczne w zamiarze, ale jednocześnie trafne poza tym kontekstem. Tam, gdzie formaliści amerykańscy proponują lekturę zorientowaną na jednostkowość i autonomię dzieła, bez zewnętrznych kontekstów (politycznych, ideologicznych, społecznych, historycznych), aby odkryć pewien „tekst sam w sobie”, twórca *Distant reading* poszerza owe konteksty do maksymalnego oddalenia od dzieła i bazując na podstawowych informacjach, doszukuje się generalizacji. To może posłużyć do wskazywania trajektorii zmian oraz tendencji przemian gatunkowych.

Przedmiot badań jest na wskroś literaturoznawczy: mamy do czynienia z analizami historii i przemian powieści europejskiej. Zakres czasowy i geograficzny nakreślono bardzo ambitnie, znajdziemy odwołania od Novalisa po Kafkę i Joyce’a. W klasycznej pracy, opartej na analizie, interpretacji i wartościowaniu dzieła literackiego, czy – mówiąc wprost – związanej z lekturą, spodziewalibyśmy się ograniczenia przedmiotu badań do wybranych przykładów, dla lepszego panowania nad tym materiałem. Moretti od początku wskazuje obszary problemowe, takie jak archipelag tradycji narodowych w tworzeniu powieści jako gatunku, ale też swoisty izomorfizm i paralele w obrębie systemu literatury. Widzimy jednak, że spacjotemporalne kryteria doboru przykładów dążą do globalnego ujęcia zjawiska.

Odwołując się do *Weltliteratur* czy *World literature*, Moretti ma także świadomość imperialnego czy kolonialnego aspektu takich starań. Jest to bardzo samoświadomy zabieg i w tym miejscu warto wspomnieć, że owszem, przygotowanie materiału zawsze jest subiektywnym wyborem (uwzględnienie danych źródeł w *datasecie* ma podobną wagę co np. wybór haseł w tematycznej encyklopedii o ograniczonej objętości albo wierszy do antologii – ustala się hierarchia) – ale w sytuacji globalnej wymiany danych możemy taki globalistyczny czy kolonialny aspekt mitygować, co rozwijam w kolejnych rozdziałach. Wątki specyfiki regionu i kultury dotyczą także przedmiotu badań Morettiego, który przywołuje przykłady badań powieści w różnych częściach świata. Prawidłowością staje się swoisty kompromis między obcą formą a lokalną kulturą dla samego gatunku, dlatego włoski autor podkreśla, że powieść może być fenomenem zachodnioeuropejskim w swojej prototypowej formie. W przypadku wychodzenia poza ramy swojego pierwotnego otoczenia kulturowego,

ów gatunek wymaga zabiegów adaptacyjnych. Podobnie jest z przekładami literatur zagranicznych na języki kultur o różnej tradycji i odmiennym podziale gatunkowym pisarstwa.

Na zasadzie opozycji wobec „ważnego czytania” Nowej Krytyki, niezamierzona gra słów w przekładzie na polski czyniłaby z metody Morettiego lekturę „nieważną”, jednak mamy świadomość, że w grę będzie wchodzić „bliska” i „daleka” lektura. Możemy się także posłużyć pojęciami dystansu i *dystansowania się* od tekstu. Przyjrzyjmy się, jakie metodologiczne konsekwencje ma oddalenie soczewki obserwacji dla badań literaturoznawcy, oddając mu głos. W poniższym fragmencie Moretti tłumaczy założenie metody i odnosi się do wcześniej cytowanego hasła Marca Blocha na temat historycznych syntez – „lata analizy dla dnia syntezy”<sup>83</sup> (zachowuję oryginalną pisownię wyróżnień):

Now, if we take this model seriously, the study of world literature will somehow have to reproduce this ‘page’ – which is to say: this relationship between analysis and synthesis – for the literary field. But in this case, literary history will quickly become a very different from what it is now: it will become ‘second hand’: a patchwork of other people’s research, *without a single direct textual reading*. Still ambitious, and actually even more so than before (world literature!); but the ambition is now directly proportional to *the distance from the text*: the more ambitious the project, the greater must the distance be<sup>84</sup>.

Antycypując tutaj krytykę takiego podejścia, Moretti wskazuje na istotny problem *close reading*: ubóstwo kanonu, którym się posługujemy w takich badaniach (obok Nowej Krytyki autor wymienia też dekonstrukcję jako jedno z podejść „blisko tekstu”). Argument byłby więc taki: jeśli chcesz, badaczko, poznać tendencje na większej skali czasowej i przy większym zasięgu geograficzno-kulturowym, nie będziesz w stanie korzystać z uważnej lektury, ponieważ przekroczy to Twoje zdolności poznawcze, a przez to ograniczysz się do zbyt małego obszaru (i dopowiedzmy dalej: nie dostrzeżesz szerszej perspektywy badanego problemu). Następnie Moretti podkreśla, że dystans staje się wręcz warunkiem osiągnięcia nowej wiedzy. Rodzi to naturalny opór, skoro autor *Distant reading* wyraźnie podkreśla, iż badawczy dystans od literackiego materiału, który mamy tu na myśli, powiększa się tak bardzo, że tracimy uważność poziomu odczytania pojedynczego tekstu. Przekonujemy się, że włoski badacz

---

<sup>83</sup> Czyli ‘years of analysis for a day of synthesis’ Cyt. za: Franco Moretti, *Distant reading* (London ; New York: Verso, 2013), s. 47.

<sup>84</sup>Moretti, 48.



proponuje uprawiania literaturoznawstwa bez czytania literatury. W obliczu tej demaskacji pozostaje tylko brawura:

And if between the very small and the very large, the text itself disappears, well, it is one of those cases when one can justifiably say, Less is more. If we want to understand the system in its entirety, we must accept losing something. We always pay a price for theoretical knowledge: reality is infinitely rich; concepts are abstract, are poor. But it's precisely this 'poverty' that makes it possible to handle them, and therefore to know. This is why less is actually more<sup>85</sup>.

Jeśli tracimy z oczu „drzewo” tekstu na rzecz „lasu” korpusu, należy sobie zadać pytanie, czy to jeszcze badanie literatury<sup>86</sup>. Implikowane ryzyko metod masowego przetwarzania danych w literaturoznawstwie, polegające na wytwarzaniu „falsyfikatów” na podstawie niepełnych informacji (tak jak wspomniane braki źródłowe, zrodzone przez imperialne obarczenie – *bias*, niedostępność danych itp.) możemy mitygować poprzez dokładne anotowanie korpusu, na którym odbywa się praca badawcza i sporządzanie dokumentacji projektu przy przetwarzaniu informacji. Dzięki temu umożliwia się krytykę i badania porównawcze, a także odtwarzanie wyników. Przez porównanie do nauk przyrodniczych, możemy odtworzyć jak propozycja Morettiego ma szansę zadziałać. Jesteśmy sobie w stanie wyobrazić swego rodzaju sztafetę badawczą: laborant sporządzający opis składu chemicznego w próbce wody nie musi znać widoku jeziora, z którego ta woda pochodzi – ale jezioro widział technik, który pobrał próbkę i dostarczył ją do laboratorium. Tego rodzaju klastry, oparte na kompetencjach i zadaniach, działają już w humanistyce – podział pracy na odcinki może jednak wciąż stanowić *novum* w interpretowaniu tekstów, a zatem uprawianiu historii i teorii literatury, jakie znamy na wydziałach humanistycznych.

---

<sup>85</sup> Moretti, 49. Autor powołuje się także w przypisie na Maxa Webera mówiącego o obiektywizmie nauk społecznych, aby sparafrazować główną ideę; przytaczam ten fragment w niezmienionej formie: „(...)to quote Weber again: 'Concepts are primarily analytical instruments for the intellectual mastery of empirical data'. ('Objectivity in Social Science and Social Policy', p. 106). Inevitably, the larger the field one wants to study, the greater the needs for abstract 'instruments' capable of mastering empirical reality". Tamże.

<sup>86</sup> Na marginesie, drzewo dosłownie pojawia się jako metafora, gdy Moretti przywołuje filogenetyczne typologie i zestawia je z drugą metaforą – falą. Czyni tu uwagę krytyczną wobec komparatystyki, która powinna być radykalna i modelem fali oraz przepływów sprzeciwiać się sztywnemu modelowi typologicznego czy taksonomicznego drzewa, dzielącego literaturę na zbiór literatur narodowych, w sposób niemal genealogiczny (tylko taka komparatystyka liczyłaby się dla autora *Distant reading*). Wróćmy do tego rodzaju metafor przy innym omawianym tekście Morettiego.

Interesujące, że część cyfrowa i wizualizacyjna badań Morettiego pojawia się stopniowo, „przesącza” się w pracy. Zapewne wynika to z konieczności oswojenia odbiorców z proponowaną, jednak radykalną metodą. Pierwszym narzędziem staje się algorytmizacja i wizualizacja, za pomocą „ewolucyjnego” drzewa, przemian gatunku prozy detektywistycznej. Popularność Conan Doyle’a względem innych współczesnych mu autorów<sup>87</sup> Moretti stara się wytłumaczyć obecnością wskazówek w fabule takiej fikcji. Ten zabieg, dziś doskonale znany, miał być jak w teorii ewolucji pewną emergentną cechą, dającą przetrwanie. Utwory opowiadające o wskazówkach w śledztwie miały większą szansę na popularność. Moretti dzieli też obecność tego elementu na kolejne opozycje nieobecność i obecność wskazówek, a dalej na wskazówki konieczne, widoczne i możliwe do odszyfrowania (przez czytelnika). Proponowana tutaj analiza jest frekwencyjna i uwzględnia oznaczanie (nadawanie markerów) poszczególnym tekstom. Moretti przywołuje wszakże te dociekania, jak i badanie szerszego trendu na zbiorze 108. utworów z magazynu *Strand*, jako tradycyjną pracę filologiczną (własną, ale też na zasadzie współpracy z grupą studentów i asystentką). Nie ma tu jeszcze śladu metod cyfrowych, poza wizualizacją na wykresie, a raczej intuicja w stronę wchodzenia na wyższą skalę badań. To, co zaskoczyło włoskiego badacza w wynikach, to utrzymywanie się utworów bez nowego środka rozwoju fabuły obok tych, które ten środek zaadaptowały. Moretti podaje możliwe hipotezy (pierwsza: autorzy utworów bez wskazówek w fabule szukają alternatywnych środków; druga: ci autorzy są już uformowani pod względem stylu i nie stosują nowego środka, musi ich wymienić nowe pokolenie – oba założenia mogą współistnieć ze sobą). Dodaje jednak także trzecie: fikcja detektywistyczna jest połączeniem akcji i suspense, działa na dwóch tempach. Opowiada o zbrodni (dynamizm), ale odtwarza śledztwo (retardacja). Badacz rzuca też mimochodem: „I have no data for this hypothesis (and someone else will have to do the reading this time)”<sup>88</sup>.

W kolejnych szkicach aspekt wizualizacji danych się zagęszcza. Moretti przedstawia zagregowane dane dotyczące popularności filmów, wyprodukowanych w Hollywood, za granicą, formułując wnioski dotyczące wpływu ich gatunków na rozpowszechnianie się i odbiór w innych krajach. Praca interpretacyjna jest właściwie odniesieniem do danych statystycznych i geograficznych, popartym filologiczną wiedzą o gatunkach tekstów kultury. Grząski to grunt z perspektywy tradycyjnie rozumianych badań (np. ograniczonych do recepcji

---

<sup>87</sup> Wcześniej mowa w książce o „rzeźni literatury”, czyli masowym niweczeniu pamięci o niektórych utworach – w czytelniczej świadomości po latach pozostają tylko najpopularniejsze. Dziś najczęściej specjaliści jak Moretti mogą wymienić rynkowych konkurentów dla autora przygód Sherlocka Holmesa i tytuły ich dzieł.

<sup>88</sup> Moretti, *Distant reading*, 83.

w jednym kraju i bazujących na czytaniu recenzji), ale znajduję tu dowód na to, że kompetencja literaturoznawcza czy kulturoznawcza będą nieodzowne w tego typu badaniach, ponieważ dane nie będą przedstawiać znaczenia bez właściwej prezentacji oraz interpretacji badacza.

W tekście poświęconym krytyce książki opisanej poniżej, Moretti rozwija myśl, że jego analizy należą do porządku eksplikacji, wyjaśniania, bardziej niż interpretacji. Mają więcej wspólnego z ciągami przyczynowo-skutkowymi niż teleologicznym dociekaniem sensu. Co dla nas istotne, badacz przystaje na zarzut, że jego prace zawierają element interpretacyjny (o czym szerzej przy okazji krytyki), ale zwraca uwagę na dynamikę wzajemnego napędzania się między tymi porządkami. Zapożyczając się u Cullera, powtarza słowa o wystarczającej liczbie interpretacji i dodaje: w obliczu istnienia w obszarze interpretacji „wszystkiego co można było powiedzieć (moja parafraza – JE), trzeba wytworzyć nowe ciągi przyczynowo-skutkowe dla kolejnych rzeczy, które można powiedzieć”.

W szkicu poświęconym historii i teorii powieści, na podstawie materialnych dowodów Moretti stara się badać jej rozwój jako fenomen ściśle europejski. Nawiązuje do narzędzi komparatystycznych, przekonując, dlaczego nie obserwujemy rozwoju bogatej tradycji powieści chińskiej w XVIII w. i estetycznego zwrotu w Europie. Parametry tego badania są dość uproszczone, mowa o długości utworu, skali produkcji literackiej. Przywołane są też głosy ówczesnych krytyków i pisarzy. Moretti wskazuje na kryteria rynkowe dla rozwoju form estetycznych, tzn. ściśle przełożenie form dystrybucji i odbioru na formy estetyczne (zwiększenie konsumpcji przy osłabieniu uważności-skupienia prowadzą do wytworzenia prostszych podgatunków popularnych).

Zaprezentowana zostaje również analiza stylu na imponującym korpusie 7000 tytułów (niedosięglym, dodajmy, dla typowego kursu semestralnego czy rocznego w dydaktyce akademickiej, do lektury „uważnej” i „bliskiej”). Dane wejściowe są znów proste, np. liczba wydanych powieści, długość utworu, długość tytułu. Wnioskowania formułowane przez Morettiego idą w stronę interpretacji konsekwencji długości tytułu dla semantyki i odbioru dzieła (np. tytuł jako streszczenie fabuły vs wyłącznie nazwa własna w tytule vs tytuł jako abstrakcyjne pojęcie). Odważniejsze wnioski wiążą się z pokazaniem jak poszerzanie się rynku powieści koreluje ze skróceniem tytułów utworów. Pod względem semantycznym, jak niektóre pomniejsze gatunki (powieść gotycka) zdradzają tendencje swego nazewnictwa (nazwy miejsc: emblematem staje się tu *Zamczysko w Otranto*).

Ostatni fragment stanowi prezentację analizy sieciowej (m. in. na tekstach chińskich, choć najbliższy rodzimemu czytelnikowi będzie *Hamlet*) w odczytaniu dzieła literackiego. Poprzez wykreślenie grafu – siatki powiązań między bohaterami dzieła (kto z kim rozmawia lub wchodzi w interakcję, jak często), możemy określić mocne i silne węzły takiej sieci. Niektóre postaci mogą nam się jawić w innym świetle, niż początkowo by to wynikało z tradycyjnej interpretacji. Jak zauważa Moretti, może to też pokazywać strukturalne różnice w kreowaniu protagonistów w literaturach Wschodu i Zachodu.

Interesującym walorem *Distant reading* (książki), poza prezentowanym pomysłem – motywem przewodnim – jest odautorski komentarz do każdego eseju pomieszczonego w zbiorze, tworzący swego rodzaju dziennik doświadczeń badacza. Moretti komunikuje nie tylko napotkane fakty dotyczące literatury i swoje spostrzeżenia, ale prowadzi czytelnika we wstępach do tych szkiców, jakby udostępniał luźne notatki laboratoryjnego eksperymentu obok głównego raportu z badań.

Przedstawiam drugą istotną dla tematu niniejszych rozważań książkę Morettiego (a zarazem ważną dla humanistyki cyfrowej w ogóle i konieczną do odnotowania), *Graphs, Maps, Trees: Abstract Models for a Literary History*<sup>89</sup>, asynchronicznie: mimo wcześniejszej daty wydania, jest dużo „czystsza” pod względem metody, oparta na spójnych metaforach i conceptach, jak i pozbawiona wspomnianych didaskaliów czy paratekstów, nie licząc posłowania Alberto Piazzę. Tutaj Moretti prezentuje adaptację koncepcji innych nauk do literaturoznawstwa, opierając się na masowej analizie danych. Będą to kolejno: teoria grafów, zaczerpnięta z metod kwantytatywnych historii, idea mapy przynależna do domeny geografii oraz drzewiaste struktury, służące wcześniej do obrazowania teorii ewolucji.

W przypadku grafów czytelnik otrzymuje propozycję prześledzenia historii powieści lat 1740-1900 jako gatunku, ale wyrażonej w podziale na około 40 podgatunków. Te małe gatunki, jak np. powieść militarna, gotycka czy *Bildungsroman*, charakteryzuje krótki, często jednopokoleniowy żywot. W tej wielości i zmienności form, badacz poszukuje jednolitego wzorca. Tym, co stara się podkreślić, jest efemeryczność form cząstkowych, składających się na pozornie monolityczne zjawisko, które zwykło się nazywać „powieścią”.

W części związanej z mapami, Moretti poszukuje nowych odczytań poprzez weryfikację relacji przestrzennych w interpretowanych dziełach. Najwięcej miejsca zajmują analizy *Our Village*

---

<sup>89</sup> Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History*, Paperback ed (London: Verso, 2007).

Mary Mitford (ukazanie lokalnej społeczności w swoistych kręgach wpływu i relacji, które da się wyrazić za pomocą najpierw mapy, a później diagramu, uwzględniającego stosunki w społeczności lokalnej, opisywanej w utworze. Znajdziemy także w formie dygresji interesujące zestawienie miejsc zamieszkania bohaterów utworów, których akcja dzieje się w Paryżu i ich miłości. Moretti powołuje się na ten przykład ze swojego innego ekstu i jego krytykę<sup>90</sup>. W tym miejscu dokonuje się także przeskok pojęciowy od mapy do diagramu, aby uwypuklić różnice w opisie orientacji przestrzennej w geografii i kreślenie relacji opisywanych przez geometrię.

Co do drzew, ważną rolę spełnia pojęcie odgałęzień. Przykład widzimy w *Distant Reading*, gdy mowa o opowieściach detektywistycznych. Jedne formy literackie rozwijają się, inne zanikają. Co dla nas najistotniejsze, Moretti wskazuje na powiązanie rozwoju gatunkowego z danymi przestrzennymi – stąd powiązanie z genetyką. Podkreśla ten fakt Piazza w posłowie, przyrównując tu zmiany pola literackiego do mutacji w genetyce. Dla kolejnych dociekań warto zachować tę myśl w kontekście przekładu i przekazu międzykulturowego. Moretti odwołuje się wprost do ewolucji gatunków w piśmiennictwie Darwina i Dawkinsa<sup>91</sup>. Przywołane zostają wątki klasowości i dyskursu klasowego w ogóle, jako przykład czynników różnicujących styl i treść powieści. Stawką jest tu wskazanie drobnych różnic stylu w obrębie jednego przypisania gatunkowego.

Moretti nazywa podejście, realizowane zbiorczo w tych trzech obszarach, materialistyczną koncepcją formy. Antycypując pytanie o odniesienie powyższego do Marksowskiej teorii, wskazuje dwójaki aspekt: jest mu bliskie pojęcie formy jako siły czy energii (co wiązałoby *Graphs, Maps, Trees* z taką odmianą hermeneutyki podejrzeń). Z drugiej strony jednak Moretti kategorycznie stwierdza, że nie ma jednej ramy pojęciowej, która w badanym przez niego obszarze wyjaśniałaby wszystko (co wyraźnie wskazuje na promowanie podejścia przynajmniej interdyscyplinarnego, na pewno uwzględniającego wiele perspektyw badawczych).

---

<sup>90</sup> Mowa o Franco Moretti, *Atlas of the European Novel: 1800-1900* (London New York: Verso, 1998). Autor przytacza zastrzeżenia do tego tekstu zgłoszone przez Claudio Cerretiego.

<sup>91</sup> Tradycyjne podejście humanistyczne w rozwijaniu myśli Darwina prezentuje Krzysztof Uniłowski, wskazując na zjawisko drobnych zmian w ewolucji gatunku oraz rolę czynnika zewnętrznego w selekcji. Ta ostatnia, w aspekcie hodowli, powiązana zostaje z rolami w systemie literackim, które wpływają na kształt gatunkowy dzieł (recenzenci, krytycy). Zmiany gatunkowe podlegają w swoich przemianach tak wewnętrznym, jak i zewnętrznym wpływom. Zob. Krzysztof Uniłowski, „Czy gatunki i konwencje ewoluują (dzisiaj)?”, *Teksty Drugie*, nr 3 (2011): 52–63.

Matthew Jockers także proponuje badania na korpusach tekstów „z lotu ptaka” (odwołując się też wprost do ustaleń Morettiego). Swoją metodę nazywa makroanalizą<sup>92</sup>. Podobnie jak poprzednio przedstawiany autor, Jockers bada zbiór powieści. Analizowany przez niego korpus stanowi literatura XIX-wieczna z amerykańska, angielska, irlandzka i szkocka. Dzięki zastosowaniu metod *big data*, Jockers wskazuje najważniejsze motywy i tendencje, pochodzące bezpośrednio z masowej analizy treści tekstów. Szczególną uwagę chciałbym zwrócić na obecność osobnego rozdziału dotyczącego metadanych (czyli danych na temat danych), co jest istotnym wyróżnieniem. Bez opisu metadanych, skuteczne badania na taką skalę nie mogłyby się wydarzyć. Konkretny przykład literaturoznawczego „śledztwa” to u Jockersa przemiany tożsamości irlandzkich imigrantów w USA końca XIX wieku – i to właśnie w oparciu o dane *na temat* utworów, a nie interpretacje. W rozdziale 8., dotyczącym motywów (*themes*), autor porównuje koncepcje i sądy wyrażane przez krytyków i badaczy literatury z danymi otrzymanymi w frekwencyjnej analizie stylistyki tekstu, dokonanej narzędziami komputerowymi. W omówionych źródłach robi wrażenie agregacja aż pięciuset tematów z korpusu 3346 książek. Na wyróżnienie zasługuje praktyka „wystawienia” danych do oglądu przez czytelników – komponent cyfrowy towarzyszy książce, odbiorcy sami mogą dokonywać przekształceń<sup>93</sup>. Właściwie zaciera to też granice między autorami a odbiorcami w modelu rozprzestrzeniania się wiedzy: takie działanie wytwarza międzynarodowe sieci kontrybutorów, na podobieństwo wspomnianej kultury technologicznego bazaru.

Podobne strategie obiera Katherine Bode, analizując piśmiennictwo w rodzimej Australii<sup>94</sup>. W jej wypadku będzie to korpus danych bibliograficznych związany z australijskimi powieściami, a stawką rewizje ustaleń historycznoliterackich. Za pomocą badań na danych stara się z jednej strony oświetlić dotychczasowe badania nową wiedzą, wydobytą nowymi narzędziami, a z drugiej – zaprezentować wartościowe metody cyfrowe do pracy na zagregowanych danych, czyli niejako wyeksponować właśnie ów warsztat.

Globalizacja, dostępność, remediacja

Warto rozważyć, jak takie szeroko zakrojone badania rezonują z postmodernistycznym przeświadczeniem o braku możliwości dla stworzenia wielkich syntez, czy swego rodzaju podejrzliwością wobec totalizujących metodologii. Mam na myśli skupienie na detalu,

---

<sup>92</sup> Matthew Lee Jockers, *Macroanalysis: digital methods and literary history*, Topics in the digital humanities (Urbana: University of Illinois Press, 2013).

<sup>93</sup> Zob. "Macroanalysis" w *Matthew L. Jockers*, <https://www.matthewjockers.net/macroanalysisbook>. Dostęp 2023-09-16.

<sup>94</sup> Katherine Bode, *Reading by numbers: recalibrating the literary field* (New York: Anthem Press, 2012).

wytyczenie niewielkiego, ścisłego, ale dającego się opanować zakresu pracy krytyka czy badacza. W przytoczonych wcześniej badaniach ten detal się zmienia – opisywana skala jest szeroka, ale wpływ badacza na materiał czy też kontakt z przedmiotem się ograniczają i to one zamieniają się w ów szczegół, możliwy do opanowania. Kontrolę sprawujemy nad parametrami maszynowej analizy i opisu, a nie doбором wąskiego zakresu przedmiotowego.

Wspomniany już aspekt globalizacji i dobór źródeł, który można krytykować za skoncentrowanie przede wszystkim na wybranych krajach europejskich i zasobach amerykańskich (czy jak w przypadku Bode – australijskich). Nie posądzam badaczy o złą wolę – raczej o korzystanie z takich zbiorów, które były im dostępne (można to przyrównać do tradycyjnych kwerend w poszukiwaniu fizycznych egzemplarzy). Widziałbym w tym potrzebę wyeksponowania danych dotyczących innych kultur i regionów, aby zwalczać stronniczość podejścia (*bias*). Dotyczy to również zasobów związanych z polską kulturą.

Jak zostało to pokazane wcześniej, ekosystem mediów cyfrowych wymaga pracy (pobierania, przetwarzania, publikowania) na cyfrowych obiektach. Nie wszystkie źródła dla uprawiania literaturoznawstwa w takiej formie są dostępne. Część wymaga dostosowania, także w formie remediacji, czyli przekształcenia zapisu w jednym medium na drugie<sup>95</sup>. Mam tu na myśli szczególnie budowanie zasobów cyfrowych na podstawie piśmiennictwa dostępnego obecnie jedynie w formie drukowanej. Ten wątek powróci przy opisie PBL (kontekst przemian technologicznych i zmian metody pracy). Podstawowa kwestia brzmi: aby uniknąć globalizacyjnej stronniczości w prezentacji danych, zasoby muszą być powszechnie i łatwo dostępne dla badaczy; tylko tak można zniwelować sięganie po „dyżurne” przykłady i promować widoczność własnych języków, kultur i literatur, spoza tych najczęściej opisywanych i szczególnie spoza anglojęzycznego obszaru wymiany informacji naukowej.

Dostrzegam dwojaki problem stojący przed badaczem, który chce użyć makroanaliz czy opisanego wcześniej *Distant reading* do szeroko zakrojonych badań literackich. Po pierwsze, jak daleko możemy się oddalić od tekstu, aby wciąż powiedzieć o nim coś sensownego? Pozostaje tu aktualny dylemat przedłożony we wcześniej przywoływanym cytacie z Gumbrechta – nowe narzędzia odmieniają nasze poznanie przedmiotu. Po drugie, klątwa tego rodzaju badania może też polegać na tym, że wystawi się na krytykę niezależnie od rezultatu. Po „czytaniu z dystansu”, jeśli potwierdzimy to, co jakkolwiek inny badacz

---

<sup>95</sup> J. David Bolter i Richard Arthur Grusin, *Remediation: Understanding New Media* (Cambridge (Mass.): MIT Press, 2000).

napisał o dziele, stworzymy być może *communitas*, będziemy należeć do wspólnoty interpretacyjnej. Gorszy rezultat czeka nas, gdy konkluzje staną w opozycji do przyjętego sądu. Krytyka wniosku przeradza się w krytykę metody. Gdzie tkwi w takim razie wartość tych badań? Przekonajmy się w ogniu krytyki.

Wątpliwości; jak nie działa humanistyka cyfrowa

Największą obawą przed stawianiem odważnych pytań jest ryzyko potwierdzenia dawno już znanych faktów; udowodnienie tez, które tradycja badań literackich zdążyła zarchiwizować w burzliwych dyskusjach, zakończonych konsensusem.

Efekt pracy na danych literackich może nie być oszałamiający, ale jego wymierność niesie zupełnie inne implikacje. Daje to nam informację innego rodzaju: pewną i konkretną, wolną od modalności i intuicji sądu badacza. Jedynym czynnikiem, który to zmienia, jest możliwość podważenia sposobu, w jaki dane przygotowano – a to z kolei daje sposobność umieszczenia informacji w innym kontekście (np. poszerzenia o dodatkowy zbiór) i ponownego przeliczenia. Drugim istotnym aspektem jest potwierdzenie „liczbą” tego, co badacz ogłasza za pomocą osądu intuicji.

Trudno też obronić takie *nihil novi*, gdy dowód został już przedłożony. Pisze o tym Andrew Piper w formie akademickiej anegdoty (tłumaczenie własne – JE):

Dla każdego, kto kiedykolwiek prezentował metody obliczeniowe, reakcja „Ale czy nie wiedzieliśmy tego już wcześniej?” odciska piętno na świadomości. Nic bardziej nie wywołuje poczucia zawodu... i frustracji. Zawsze mam ochotę odpowiedzieć: cóż, wierzyliśmy, że to wiemy, a teraz posiadamy więcej adekwatnych dowodów, by w to wierzyć jeszcze mocniej. Ale to odrobinę przegadane<sup>96</sup>.

W cytowanym tekście Piper ucieka się nawet do przykładu zaprezentowania matrycy lub wykresu bez uzupełnionych danych, aby odbiorcy najpierw pomyśleli, jakiego wyniku się spodziewają i spróbowali odgadnąć rezultat. Dopiero później wyjawiane są ustalenia takich badań. Relacjonując zastosowanie takiego triku ze słuchaczami jednego z jego wystąpień na żywo w Berkeley, Piper dzieli się kilkoma istotnymi spstrzeżeniami<sup>97</sup>. Po pierwsze,

---

<sup>96</sup> Andrew Piper, *Did we already know that?* w *.txtlab @ mcgill*, <https://txtlab.org/2018/09/did-we-already-know-that/> Dostęp: 2023-09-23.

<sup>97</sup> Przytaczam wnioski za: Piper. Eksperyment autora dotyczył zbioru danych o zmianach stylu poetów w różnych okresach życia, m.in. pod względem skomplikowania leksyki. Por. Andrew Piper, *Ennumerations*, Chicago 2018, 147-177.



nie mamy tak mocnych przekonań na temat zjawisk o dużej skali, jak to relacjonujemy po fakcie. Badacz zauważył też jednak na podstawie odpowiedzi zebranych z sali, że uśrednione odpowiedzi całej grupy dość dobrze przewidywały, jaki może być wynik badania. Według Pipera dowodzi to tego, że większa skala ma przewagę nad pojedynczą refleksją. Zależy to oczywiście od charakteru badania, metodologii, a także kompetencji „ankietowanych” (mimo wszystko, badacz prawdopodobnie wygłaszał swoje wystąpienie do osób obeznanych w dziedzinie).

To podejście Pipera przywodzi na myśl, jak niemal 40 lat wcześniej Stanley Fish kontestował rozumienie kanonu i formułowanie systemów wiedzy<sup>98</sup>. O ile Fish zwraca uwagę na istnienie pewnych wspólnot interpretacyjnych, ustalających razem pewien poziom wiedzy (redukując poglądy pragmatystów: osiąga się w ten sposób własną grupową „prawdę”), o tyle Piper proponuje wzbogacenie konsensusu o mierzalny i sprawdzalny doświadczalnie czynnik.

Możemy powiedzieć, że nie dowiedzieliśmy się tego, co już wiedzieliśmy, a raczej, że innymi metodami dochodzimy do tych samych lub podobnych wniosków. Najciekawsze zatem powinny być takie miejsca, gdzie oparte na danych badania odbiegają nieco od ogólnie przyjętych interpretacji. Daje to szansę, by uzupełnić wiedzę lub zastanowić się nad poprzednimi ustaleniami. Jeszcze jednym argumentem za wykorzystaniem danych w badaniach literaturoznawczych jest istniejące już zanurzenie literaturoznawstwa w rzeczywistości danowej. Wspominałem o tym przy okazji kontekstu społecznego i medialnego.

Wracam teraz do problemu zarysowanego odnośnie szkicu Próchniaka: nie jest tak, że zwrot cyfrowy daje nowe spojrzenie na analizę, interpretację i wartościowanie dzieła literackiego (to znaczy: nakazuje zerwać z dotychczasową wiedzą i prowadzić te działania na nową modłę). Uważam, że nurt humanistyki cyfrowej oferuje kontynuację badań, zgodnie z założeniami metod literaturoznawstwa, ale posługuje się przy tym unowocześnionymi narzędziami. Co więcej, sądzę, iż tradycyjne literaturoznawstwo, zwłaszcza takie, które nie ukrywa strukturalistycznego rodowodu, już od dawna zadaje pytania, na które humanistyka cyfrowa chce obecnie odpowiadać – z tą różnicą, że możemy udzielać na nie odpowiedzi z dużo większą dozą precyzji.

---

<sup>98</sup> Por. Stanley Eugene Fish, *Interpretacja, retoryka, polityka: eseje wybrane*. red. Andrzej Szahaj. tłum. Krzysztof Abriszewski. Wyd. 2. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, 2008.

Moim koronnym przykładem na taki stan rzeczy niech będzie studium o Kafce Martina Walsera<sup>99</sup>. Badacz dąży do ścisłości swojej metody, aby była naukowa, dosłownie klasyfikuje części mowy, liczy części mowy użyte w tekście *Procesu* i na podstawie danych liczbowych wysnuwa interpretacje oraz sądy na temat dzieła autora *Kolonii karnej*. W podsumowaniu pracy objaśnia zamysł tego dzieła, ustawiając się w kontrze do badaczy zorientowanych na hermeneutyczne poszukiwania:

Również nauka o literaturze może być tak ścisła, aby móc we właściwym sensie używać określeń „prawdziwy” i „fałszywy”. (...) Memu nauczycielowi, Friedrichowi Beissnerowi zawdzięczam, że nie wpadłem chyba na całkowite bezdroże. Bez jego pomocy może i ja uprawiałbym spekulacje na temat tego, czy w Kafkowskim gmachu więcej jest z sądu czy z łaski. A tak zabrałem się do liczenia jego drzwi i okien, i ustalania, jakie postacie noszą cylindry, jakie zaś brody, i dlaczego Kafka tak często używa słowa „choć”. O sądzie i łasce zmienia się niekiedy poglądy dość szybko, podczas gdy to, co tylko starannie zarejestrowało się zgodnie z przynależną do tego funkcją, potrafi dłużej zachować swe znaczenie. Chyba, że rejestrator grubo się przeliczył. Ale to można przecież sprawdzić<sup>100</sup>.

Walser pisze te słowa jeszcze w latach sześćdziesiątych, przed pojawieniem się internetu, przed masowymi statystycznymi analizami tekstu i zanim Andrew Piper zaprezentuje wyniki swoich badań na połączonych korpusach powieści anglo- i niemieckojęzycznej. Zmieniła się tylko skala: co Walser był w stanie osiągnąć w pojedynkę, mrówczą pracą, naocznym zliczaniem słów w zaledwie jednej powieści, Piper (oczywiście angażując zespół i używając narzędzi komputerowych) mnoży na wiele więcej tytułów. Pragnę w ten sposób podkreślić, że impuls kwantyfikujący w literaturoznawstwie istniał dużo wcześniej od „mody” na humanistykę cyfrową czy „zwrotu cyfrowego”.

Dlatego upatruję w humanistyce cyfrowej bardziej szansę, niżli przemijającą modę; daje ona możliwość stawiania tych samych kwestii w bardziej globalny sposób, uogólnienia perspektywy w patrzeniu na problem, łączy odległe obszary, a także – odwołując się do metafory koła hermeneutycznego – uruchamia proces mediacji między ludzkim

---

<sup>99</sup> Martin Walser, *Opis formy - studium o Kafce*, tłum. Edmund Misiółek, Biblioteka Krytyki Współczesnej (Warszawa: Państwowy Instytut Wydawniczy, 1972).

<sup>100</sup> Walser, 128-129.

i maszynowym rozumieniem (przetwarzaniem) informacji. Między wyjaśnianiem i rozumieniem, między ogółem i szczegółem, następuje ruch od pracy na danych do ludzkiego procesu myślowego i z powrotem – z każdym obrotem koła, dowiadujemy się więcej.

Głosy krytyczne

Humanistyka cyfrowa spotyka się z dwojaką krytyką: z jednej strony z pozycji tradycyjnego uprawiania znanych nam do tej pory dyscyplin, z drugiej z pewnego rodzaju drugą falą, tzn. opiniami pochodzącymi z wnętrza tego ruchu.

Najpierw chciałbym się zająć kilkoma punktami z tekstu Ryszarda Nycza, zawierającego uwagi do tzw. nowej humanistyki<sup>101</sup>. Humanistyka cyfrowa jest w tym tekście jednym z kilku przejawów owej nowości, obok humanistyki zaangażowanej, kognitywnej, posthumanizmu, i badań opartych na sztuce (badania artystyczne, humanistyka artystyczna, *art based research*). Celem Nycza jest wskazanie pewnych trudności, nieporozumień i niezgodności, związanych z nowymi formami uprawiania humanistyki. Skupimy się tutaj oczywiście na uwagach i refutacjach związanych z humanistyką cyfrową, zarysowując kontekst, gdzie to potrzebne lub inspirujące dalsze dociekania.

Nycz rozpoczyna od odniesienia się do (domniemanego) kryzysu humanistyki: granice dyscyplin zatarły się, wszyscy naukowcy funkcjonują w mniej lub bardziej otwartej (współdzielonej) przestrzeni, pod nadzorem opinii publicznej. Możemy przyjąć, że inne nauki (np. techniczne) są w publicznym dyskursie wartościowane wyżej (choć autor *Tekstowego świata* nie wskazuje ich wprost). Skutkuje to swoistym „kryzysem tożsamości” nauk humanistycznych i swego rodzaju kompleksem niższości, związanym z (nie)przydatnością tych dyscyplin, w porównaniu z innymi dziedzinami. Krakowski badacz formułuje tu dwie uwagi. Pierwsza: humanistyka zawsze reagowała na tego rodzaju przemiany i niejednokrotnie je wywoływała<sup>102</sup>. Druga: odwołując się do *Koszmaru partycypacji* Markusa Miessena,

---

<sup>101</sup> Ryszard Nycz, „Nowa humanistyka w Polsce: kilka bardzo subiektywnych obserwacji, koniektur, refutacji”, *Teksty Drugie* 1 (2017): 18–39, <https://doi.org/10.18318/td.2017.1.2>. Korzystałem też z tłumaczenia Davida Schaufflera publikowanego w periodyku *Er(r)go*: por. Ryszard Nycz, „The New Humanities in Poland: A Few Subjective Observations, Conjectures, and Criticisms”, tłum. David Schauffler, *Er(r)go. Teoria - Literatura - Kultura*, nr 43 (2021-12-30): 315–38, <https://doi.org/10.31261/errgo.11711>. W dalszej części odwołuję się do polskiej wersji tekstu.

<sup>102</sup> Nycz zostawia uzupełnienie tej luki jako ćwiczenie dla czytelnika, zaproponujemy zatem w losowej kolejności: w dyskursie popularnonaukowym wynalazek Gutenberga idzie w parze z Reformacją. Egipskie piramidy miały funkcję sakralną. Dziedzictwo kulturalne i historyczne jest powszechnie stosowanym kodem nazewnictwa miejsc w przestrzeni publicznej, budującym tożsamość i stanowiącym punkt odniesienia (znajdowania drogi). Współcześnie, przynajmniej od powojnia, jeśli nie wcześniej, ważnym wątkiem jest relacja spekulatywnej fikcji i fantastyki naukowej, a nawet popkultury względem rozwoju samej nauki. Teksty kultury tych obszarów

Nycz pisze, że zadaniem humanistyki pozostaje kultywowanie samoświadomości, wrażliwości i kreatywności, a zatem politycznie korzystnym wydaje się przechwycenie narzędzi tej części nauki i wpędzenie humanistów w swoiste poczucie winy za ten stan rzeczy. Z tej sytuacji wyłaniają się trzy wyzwania, przed którymi stoją wszyscy chętni, aby humanistykę uprawiać: użyteczność (wartości utylitarne, impakt, wpływ społeczny), poziom innowacyjności oraz profesjonalizm dyscyplin.

Co do rangi merytorycznej humanistyki w Polsce: w tekście zostaje zarysowana ciekawa opozycja strategii uprawiania nauki w odniesieniu do globalizacji czy stanu międzynarodowej wymiany w badaniach. Z jednej strony możemy próbować odwoływać się do wzorców rodzimych (określiłbym to jako strategię emancypacji), z drugiej – próbować się dostosować do międzynarodowych standardów (roboczo przyjmijmy, że byłaby to strategia inkluzji). Prześmiewczo, choć adekwatnie, Nycz punktuje przeszkody dla obu skrajności:

(...) nie widać np., by istniała rdzenna poetyka piastowska bądź zostały odkryte jagiellońskie studia kulturowe” oraz „najbardziej nawet profesjonalne siły i środki rzucone na uruchomienie anglojęzycznych (prawdziwie, nie anglopodobnych) czasopism, powiedzmy, Uniwersytetu Posthumanistycznego w Lęborku czy serii wydawniczej Wyższej Szkoły Krajobrazu w Ojcowie nie doprowadzą bynajmniej (w najmniejszym nawet stopniu) do automatycznego podniesienia się h-indeksu publikacji ich pracowników w światowym obiegu<sup>103</sup>.

Pierwsza ścieżka naznaczona jest resentymentem i swego rodzaju ksenofobią, druga – niemalże donkiszoteryjnym (Menardowskim?) porywem wobec nieubłaganych praw globalizacji w nauce i dominacji tego, co Nycz nazywa „fabrykami” wiedzy i teorii. To te ośrodki mają według badacza niewyrażony wprost monopol na dyktowanie (niepisanych często) reguł, jakim należy się podporządkować. Istotna to uwaga, zważywszy na potrzebę widoczności badań oraz problemów i tematów z rodzimego kręgu kulturowego. W jakimś sensie jest to aktualność powtarzania „nie gęsi”: Polacy mają humanistykę, kulturę, literaturę – i nowoczesne metody badań... Z jednej strony, bez takiej repetycji nie można budować marki i rozpoznawalności. Jest to również związane z troską o dziedzictwo (obecność samych tematów badawczych, związanych z polską kulturą, a nawet samego języka na arenie

---

dosłownie wychowują nowe pokolenia naukowców. Wybrane przykłady pokazują nam wzajemny wpływ nauk o duchu i troski o zachowanie kultury z rozwojem technologii.

<sup>103</sup> Nycz, 21.

międzynarodowej). Z drugiej – potrzeba uznania praktyk obcych wtedy, kiedy służą takiej promocji i są po prostu skuteczne – w przeciwieństwie do wymyślania od nowa czegoś, co może uchodzić za standard. Dodatkowo, użycie znanych i uniwersalnych („przechodnich”) narzędzi oraz kompetencji zwiększa szanse na współpracę z ośrodkami na całym świecie, co przekłada się na ten „emancypacyjny” cel. Nycz zaznacza paradoksy i wyzwania obu podejść, zatem idąc za jego myślą, potrzeba pewnej hybrydy: „drogi środka”, osadzonej w tradycji, ale ze świadomością trendów oraz kontekstu globalnego. Autor *Tekstowego świata* mówi wprost o dynamice centrum i peryferii, do czego przyjdzie nam jeszcze wrócić, rozważając wybrane badania w celu zbudowania *datasetu*, który mógłby posłużyć międzynarodowej grupie odbiorczej.

W kwestii profesjonalizmu dyscypliny, Nycz wskazuje na pułapki „uzurpacji” nowych podejść do humanistyki. Zbija jednak ten argument, wskazując na stałe przemiany pola badań. Druga wątpliwość zdaje się tu poważniejsza, przytaczam w całości: „programy nowej humanistyki to rzeczywiste zagrożenie dla zachowania owego profesjonalizmu: negują one bowiem stabilność i określoność przedmiotu badania, odrębność i jedność metody oraz w ogóle możliwość zbudowania teorii w nowoczesnym rozumieniu naukowej humanistyki – czyli jako usystematyzowanego zespołu ogólnych twierdzeń, wieńczących ostatecznie proces opracowania kompleksu orzeczeń szczegółowych”. Tu dygresja: nawet niniejsze rozważania miejscami będą wyłamywać się z ram ściśle pojętego literaturoznawstwa, odwołując się do nauk o informacji, bibliologii i informatyki. Te części pozostają jednak komplementarne: oświetlają one główny przedmiot badań, jakim pozostaje literatura (system piśmiennictwa).

Co do samej humanistyki cyfrowej Nycz odnosi się z ironią do mody na „laboratoria” (co odczytuję przede wszystkim jako konwencji nazewniczej i resentyment). Badacz formułuje trzy wątpliwości. Pierwsza związana jest z tworzeniem „ryzykownej rywalizacji metod jakościowych oraz ilościowych”<sup>104</sup>. Argument Morettiego zostaje odwrócony: tam gdzie twórca metody *distant reading* wykazuje ograniczenia poznawcze pojedynczego badacza w przyswojeniu materiału badań literackich, możliwe do przewyżnienia za pomocą technologii, Nycz przypomina, że nie taki jest cel humanisty:

To spostrzeżenie jest wiele warte; ma w sobie coś z odkrycia wielkiego przeoczenia. Kiedy jednak damy już upust swemu zachwytowi dla bystrości autora i przewyżnimy moment zawstydzenia własną

---

<sup>104</sup> Nycz, 19.

ślepotą, powinniśmy zadać sobie pytanie, czy rzeczywiście są to rywalizujące ze sobą strategie badawcze i czy aby na pewno jedno warto zastępować drugim. Scjentyistyczne dążenie do posiadania wiedzy pełnej i neutralnego (pozbawionego wartościowania) jej opisu nigdy nie było przecież pierwszoplanowym celem humanistyki... Nie rozwijam tego wątku szerzej; zauważę tylko, że niezależnie od tych wątpliwości są to programy, które pozwalają zadawać korpusowi przedmiotowemu inne niż dotąd pytania – i uzyskiwać inne, istotne, nowe, niekiedy fascynująco inspirujące odpowiedzi<sup>105</sup>.

Druga wątpliwość dotyczy cokolwiek pozorowanego obiektywizmu dyscypliny. Sam, odwołując się wcześniej do interpretacji wiersza Świetlickiego, korzystam z tego dyskursu opierania się na danych, który wyraźnie w humanistyce cyfrowej się zaznacza: oto będziemy w stanie pracować na „czystej” informacji, otrzymując wyniki pozbawione badawczej stronniczości (*bias*) i zaprezentujemy je w neutralny sposób, aby pozbawić odbiorcę wątpliwości – wszak dane powinny wyrażać idee jednoznacznie<sup>106</sup>. Pierwsza krytyczna myśl, która się tutaj rodzi to opór humanisty, dla którego wątpliwość i krytyczna refleksja są podstawowym żywiołem pracy (mniej może w przypadku tej części językoznawstwa, której bliżej do nauk ścisłych i preskrytywizmu, ale w opartym na *interpretacjach* literaturoznawstwie szczególnie). Nycz zaś podchodzi do tego problemu następująco: przede wszystkim metoda nie jest przezroczysta – każda taka deklaracja byłaby raczej uzurpacją: „Mit ten maskuje przecież bardzo daleko posunięte procesy selekcji, schematyzacji oraz homogenizacji włączanego w badania »materiału«”<sup>107</sup>.

Zapytajmy jednak również: do jakich tekstów odwołuje się Nycz? Nie kwestionując słuszności uwag, dla porządku odnotujmy odniesienia. Wspomniana zostaje recenzja ważnego zbioru *Digital Humanities*, autorstwa Nikolausa Fogle’a – a zatem pośrednio także omawiana tam antologia. Przywołany zostaje też Franco Moretti, nieco emblematycznie, podstawiony jako

---

<sup>105</sup> Nycz, 25.

<sup>106</sup> A jednak, antycypując dalsze dociekania, od razu musimy wtrącić za Cezarym Rosińskim, że „nie ma niewinnych danych” – selekcja jest interpretacją, zob. „Nie ma niewinnych danych”, *OPERAS-PL* (blog), 2021-12-09, <https://operas.pl/2021/12/09/nie-ma-niewinnych-danych/>. Por. również zastosowanie (m. in. Gadamerowskiej) hermeneutyki do opisu architektury informacji jako opartej na przed-sądach przez Marcina Roszkowskiego: Marcin Roszkowski, „Kaźda reprezentacja jest interpretacją – w stronę hermeneutycznej koncepcji architektury informacji”, *Zagadnienia Informacji Naukowej - Studia Informacyjne* 57, nr 2(114) (2018-12-05), <https://doi.org/10.36702/zin.455>.

<sup>107</sup> Nycz, 25.

synekdocha wyłaniającej się dyscypliny – jego teksty autobiograficzne, opisujące metodę działania autora *Distant Reading*<sup>108</sup>. Badacz odwołał się także do tekstu Urszuli Pawlickiej, publikowanego w tym samym numerze *Tekstów Drugich*, do którego wrócę w dalszej części. Jak widać lista nie jest wyczerpująca, ale tworzy odpowiedni przekrój dla formułowanych uwag (autor zaznacza, że tworzy bardziej krajobraz przemian humanistyki, nie pochyla się zatem nad niuansami każdego z przywoływanych trendów).

Krytykę Nycza określiłbym jako upomnienie się o *humanistykę* w humanistyce cyfrowej – tak, aby nie stała się statystyką, informatyką czy innym, niepodobnym już do badań nad kulturą tworem metodologicznym. Konkluzja artykułu, w efektowny, retoryczny sposób akcentuje stałość przemian w obrębie humanistyki. Sam wątek cyfrowy określony zostaje wcześniej jako część kultury partycypacji – analizowane przez Nycza trendy w znacznym stopniu zawierają element uczestnictwa i szeroko pojętej inkluzywności (tak pod względem metodologicznym, jak i społecznym).

Na konkretnym przykładzie, związanym z metodami, można prześledzić krytykę humanistyki cyfrowej, jaką prowadzi Maciej Eder, analizując badania stylometryczne<sup>109</sup>. Aby dać kontekst obliczeniowych badań nad literaturą, autor odwołuje się najpierw do stylometrii. Jednym z najbardziej rozpalających umysły badaczy oraz czytelników problemów jest tutaj atrybucja autorska, skądinąd, możliwa do sprowadzenia do „zero-jedynkowego stanu”: na podstawie badań z użyciem metod ilościowych potwierdza się autorstwo analizowanego tekstu lub nie. Ćwiczenie, które proponuje Eder, polega na wyskalowaniu metod takich badań do pracy nad większymi korpusami:

Skoro możliwe jest rozpoznawanie grup utworów napisanych przez jednego autora, to wolno zasadnie sądzić, że w obrębie dużych korpusów będzie możliwe odnajdywanie podobieństw między całymi grupami tekstów czy to zbliżonych pod względem tradycji literackiej, czy to zbliżonych czasem powstania, czy wreszcie wykazujących mniej lub bardziej subtelne zależności intertekstualne<sup>110</sup>.

---

<sup>108</sup> Zob. F. Moretti *Przypuszczenia na temat literatury światowej oraz Więcej przypuszczeń*, przeł. P. Czaplński, „Teksty Drugie” 2014 nr 4. Cyt. za: Nycz.

<sup>109</sup> Maciej Eder, „Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii”, *IBL PAN, call no. P.I.2524*, nr 2 (2014), <https://doi.org/10.18318/td>.

<sup>110</sup> Eder, 92.

Eder posługuje się metaforą „odcisku palca” autora, by zobrazować problem określania atrybucji. Odnosi się jednak do szerszych kryteriów stylu. Wskazuje na prawidłowości, które aplikują się tak do indywidualów, jak do grup, twierdząc, że pewne czynniki kulturowe da się zaobserwować w językowych nawykach stylu. Badacz osobno wskazuje też na rolę osoby autorskiej jako łatwą do określenia w takim badaniu. „Odcisk palca” staje się czymś dużo bardziej efemerycznym, biorąc pod uwagę kontekst socjokulturowy, tradycje, sieci wzajemnych nawiązań itp. Eder krytykuje badania stylometryczne z kilku powodów. Jednym z nich jest analiza materiału, w której przypisuje się autorstwo na podstawie najczęściej występujących słów. Eder wskazuje tu na pierwszą pułapkę w postaci całości synsemantycznych (np. spójniki). Badacz nie dodaje jednak, że najczęściej występujące słowa pozbawione autonomicznego znaczenia zostaną odfiltrowane w badaniu jako tzw. *stop words*, co jest powszechną praktyką. Eder twierdzi, że takie badanie pomija cechy stylistyczne; możemy argumentować, że raczej pozostanie to kwestią kalibracji użytych narzędzi.

Obraz tego, jak niewiele współczesna stylometria ma wspólnego z tradycyjnie rozumianym stylem, dać mogą powstające w ostatnich latach prace na temat atrybucji publikowane przez badaczy technologii informacji, w których do rozróżniania autorów z powodzeniem stosuje się zbitki dwóch lub więcej wyrazów, relacje składniowe między wyrazami, a nawet najczęstsze zbitki liter (np. dwóch albo trzech)<sup>111</sup>

Przywołany zostaje przykład, który rzeczywiście nie daje się sensownie odczytać przy użyciu tak sformułowanej zasady. Kolejną kwestią wzbudzającą uzasadniony niepokój, jest dobór korpusu porównawczego przy metodzie badań podobieństwa tekstów.

Mówiąc obrazowo i z zamierzoną przesadą: jeśli przy dowodzeniu autorstwa *Erotyków* przypisywanych Sępowi Szarzyńskiemu zostanie użyty korpus, powiedzmy, dzieł Mickiewicza, Lechonia i Szymborskiej, to jako najbardziej prawdopodobny autor zostanie wskazany Mickiewicz – bądź co bądź bliższy stylem XVI-wiecznemu autorowi *Erotyków* niż Lechoń albo Szymborska. Ta charakterystyczna przypadłość metod bazujących na mierze odległości będzie miała istotne konsekwencje nie tylko w atrybucji autorskiej, lecz także w próbach zastosowania stylometrii do badania szeroko rozumianych

---

<sup>111</sup> Eder, 94.



relacji między tekstami literackimi: jakość wyniku eksperymentu będzie bardzo mocno zależna od jakości (oraz zawartości) korpusu porównawczego<sup>112</sup>.

Możemy się tutaj z całą pewnością zgodzić z ostatnim zdaniem. Pozostaje wątpliwość, czy rzeczywiście występują aż tak duże problemy z przygotowaniem korpusów<sup>113</sup>. Następnie Eder przechodzi do eksperymentu z określeniem tendencji na masową skalę: analiz podobieństwa między 28. powieściami XIX i XX wieku. Porównanie dotyczy najbliższego sąsiada, na zbiorze najczęstszych słów i graficznie zostaje przedstawione jako drzewo z parami najbliższych tekstów obok siebie. Eder po zmianie parametru badania z 100 na 300 słów rejestruje odmienne wyniki i konkluduje: „[w] jaki sposób badacz ma wiedzieć, które parametry dają prawdziwy obraz korpusu, a które prowadzą na stylometryczne manowce? Jak ma wierzyć w obiektywizm wyników, kiedy te przeczą sobie nawzajem?”<sup>114</sup>. Badacz dostrzega ryzyka, zwłaszcza przy pochopnym analizowaniu w ten sposób dużych zbiorów danych.

W drugiej części tekstu Eder bierze na warsztat teorię grafów i analizę sieciową, aby ten sam zbiór opisać jako siatkę relacji. Tutaj głównym zarzutem, który wydaje mi się ciekawy, jest brak informacji dla badacza, tzn. niewielki pożytek z przedstawienia wizualnego informacji. Graf sieciowych powiązań niekoniecznie komunikuje zależności, które da się ująć w formie wniosku. Można odnieść wrażenie, że w tym konkretnym przypadku decyduje o tym niewielka próbka, co Eder sam podsuwa czytelnikowi: zamiast 28 powieści lepsze byłoby kilkadziesiąt lub kilkaset.

Głównym problemem, poruszonym przez Edera, jest swoisty „kompleks obiektywizmu”, a zatem podobny temat, do którego odwoływał się Nycz. Z jednej strony, te same dane dają te same wyniki w określonych warunkach. Z drugiej – następuje jakaś interpretacja tych wyników, co skazuje nas na subiektywność. Przygotowanie danych, jak będziemy później dowodzić, także może być interpretacją. Możemy też wskazać, że to po stronie badacza jest dobór parametrów badania i jego uzasadnienie. W swoich własnych przykładach, podanych przekornie, Eder wymienia też propozycje częściowych rozwiązań (uśrednienie dendrogramu w pierwszym badaniu, dobór algorytmu i próbki w drugim eksperymencie).

---

<sup>112</sup> Eder, „Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii”.

<sup>113</sup> Tu uwaga w związku z kolejnymi częściami: zadaniem humanistycznych baz danych jest właśnie dostarczenie odpowiednich informacji, aby tego typu badania były rzetelne.

<sup>114</sup> Eder, „Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii”.

Należy zaznaczyć, że mimo wyraźnego uzasadnienia tych obaw, już w obrębie przywoływanych tekstów, jak i w praktyce badaczy stosujących stylometrię powyższe ryzyka są znane, a co za tym idzie – mitygowane. W niektórych miejscach Eder celowo podaje też przykład użycia danego narzędzia, w formie eksperymentu (myślowego lub technicznego), do zagadnienia, które niekoniecznie da się w ten sposób badać. Bardziej byłaby to więc lista wskazówek, czego nie należy robić, zbiór dobrych praktyk *à rebours*, które krytyk zostawia nam ku przestrodze.

Do opisu łączenia humanistyki z technologią używane są niekiedy określenia: zwrot laboratoryjny<sup>115</sup>, obliczeniowy<sup>116</sup> i cyfrowy<sup>117</sup>. Urszula Pawlicka zadaje pytanie o zasadność, jak pisze, „powrotu” do laboratorium – skoro taka metodologiczna próba nie jest pierwszą, a za pierwszym razem się to nie udało (wyróżnienia oryginalne w tekście – JE):

[k]onsekwencje przekroczenia progu laboratorium są ogromne, wiążą się bowiem z modyfikacjami w procesach i metodach prowadzenia badań naukowych. Laboratorium doprowadza do następujących przesunięć w badaniach humanistycznych: od *opisu* do *produkcji*, od *metody* do *narzędzi*, od *tekstu* do *artefaktu* oraz od *teorii* do *praktyki*. Laboratorium przeorientowuje dotychczasowe założenia humanistyki i przepisuje jej słownik na nowo. W efekcie mamy do czynienia nie z tekstami, ale z *danymi*; nie z pracami, ale *projektami*; nie z teorią, ale z *praktykowaniem teorii*; nie z analizą kulturową, ale z *produkcją narzędzi* do analizy, a raczej *analizy kulturowej*. Laboratorium wiąże się zatem z całkowitą przebudową humanistyki, która zapożycza kategorie z innych nauk, by stać się „nauką”<sup>118</sup>.

Autorka upatruje tutaj taktycznego zwrotu w odpowiedzi na kryzys i powoływanie takich „labów” humanistycznych jako strategię insyтуcjonalną (co nakazuje patrzeć nieco krytyczniej

---

<sup>115</sup> Urszula Pawlicka-Deger, „The Laboratory Turn: Exploring Discourses, Landscapes, and Models of Humanities Labs”, *Digital Humanities Quarterly* 014, nr 3 (2020-09-25).

<sup>116</sup> David Berry, „The Computational Turn: Thinking about the Digital Humanities”, *Culture Machine* 12 (2011-02-18), <https://sro.sussex.ac.uk/id/eprint/49813/>.

<sup>117</sup> Deutsche Forschungsgemeinschaft, „Digitaler Wandel in den Wissenschaften”, 2020-10-28, <https://doi.org/10.5281/ZENODO.4191345>.

<sup>118</sup> Urszula Pawlicka, „Humanistyka: pracownia, centrum czy laboratorium?”, *Teksty Drugie* 1 (2017): 314–33, <https://doi.org/10.18318/td.2017.1.26>. W przytoczonym fragmencie odwołanie w przypisie do: Lev Manovich, „The Science of Culture? Social Computing, Digital Humanities, and Cultural Analytics” (2015), <http://manovich.net/index.php/projects/cultural-analytics-social-computing>.

na takie teksty jak przywoływany zbiór „obietnic” z UCLA). Dla nakreślenia tła, badaczka proponuje cofnąć się do historii humanistyki i prześledzić, ile to zwrotów ma ona za sobą: antypozytywistyczny (owo wyjście z „pierwszego” laboratorium), zwrot strukturalny, przełom poststrukturalny, zwrot kulturowy, zwrot pragmatyczny, zwrot performatywny i zainteresowanie nową socjologią, w tym socjologią nauki, wreszcie swoisty zwrot technologiczny. Pawlicka wskazuje tu m.in. humanistykę cyfrową i pokrewne względem niej filozofię technologii, archeologię mediów, studia nad platformami (*platform studies*) czy studia nad oprogramowaniem (*code studies*). Już samo namnożenie tych zwrotów wskazuje w uproszczeniu na dwie powracające tendencje: swoistą sinusoidę zwracania się to ku strukturze i teorii, to znowu ku dynamizmowi i praktyce. Pawlicka zwraca też uwagę na tożsamość dyscyplin idiograficznych w konfrontacji z rozumieniem naukowości wyłącznie jako podejściem nomotetycznym.

Taktyczne zbliżenie się humanistyki do nauki jest odpowiedzią na neoliberalny porządek, który obejmuje uniwersytet, pisze Pawlicka, mimo, a może: z powodu konfliktu wartości między tymi dwiema ideami. Rynek okazuje się tutaj silniejszy, co autorka dostrzega w ekonomizacji i komercjalizacji nauki.

Zwrot taktyczny w humanistyce oznacza zatem rozwój takich kierunków, które wpiszą się w narzucony przez neoliberalną edukację model nauki opartej na kryteriach rynkowych. Celem jest zmodyfikowanie i rozwój humanistyki, która udowodni, że tak jak nauka, potrafi być empiryczna i praktyczna. W takich warunkach powstaje humanistyka cyfrowa<sup>119</sup>.

Autorka nie ma wiele dobrego do powiedzenia o opisywanej tu metodzie: referując prace Grusina<sup>120</sup> i Raley<sup>121</sup> wskazuje na kryzysowe otoczenie *Digital Humanities* („laby” pojawiają się po kryzysie finansowym w 2008 r.) oraz tworzenie napięć z tradycyjną humanistyką poprzez wprowadzanie pragmatyzmu i mierzalności do badań (co może być dobre

---

<sup>119</sup> Pawlicka, „Humanistyka”, 325.

<sup>120</sup> Richard Grusin, „The Dark Side of Digital Humanities: Dispatches from Two Recent MLA Conventions”, *Differences* 25, nr 1 (2014-05-01): 79–92, <https://doi.org/10.1215/10407391-2420009>.

<sup>121</sup> Rita Raley, „Digital Humanities for the Next Five Minutes”, *Differences* 25, nr 1 (2014-05-01): 26–45, <https://doi.org/10.1215/10407391-2419991>.

z administracyjnego punktu widzenia, ale toksyczne dla rozwoju tradycyjnych humanistycznych modeli metodologicznych)<sup>122</sup>.

Aby zilustrować różne podejścia – strategie, Pawlicka posługuje się metaforami sposobów pracy humanisty: pracowni, centrum i laboratorium. Pracownia byłaby odgrózeniem się od świata zewnętrznego w uprawianiu nauki – dwa pozostałe podejścia mają to zmienić. Centrum to strategia otwarcia na świat, pod względem zadań:

(...) budowanie kolekcji i archiwów cyfrowych, dostarczanie narzędzi technologicznych, obsługiwane informacyjnych portali uczelni, koordynacja cyfrowych projektów (...), organizowanie szkoleń, warsztatów i konferencji, oferowanie konsultacji, badanie wpływu technologii cyfrowych na prowadzenie badań naukowych, popularyzacja włączenia narzędzi cyfrowych do badań i nauczania (...) rozwijanie zaawansowanych form nauczania cyfrowego, wykorzystywanie nowych technologii w badaniach literackich, budowanie pomostu między humanistyką, nauką i informatyką, między kadrami i studentami oraz między uniwersytetem i szerszą społecznością<sup>123</sup>

Może się wszakże centrum jawić jako swego rodzaju dysponentem takiej wymiany i rozwoju, narzędziem w jakiejś mierze kontrolującym (centralne jest kierownicze). Pawlicka opisuje trzeci porządek w formie życzeniowej jako pewną syntezę. Laboratorium standaryzuje, dokumentuje i eksperymentuje, ale jest także miejscem spotkań i budowania wspólnoty naukowej. Obok rytuałów zadaniowych wytwarza także te społeczne. Pawlicka formułuje zatem na koniec propozycję pozytywną, (również) w odwołaniu do społecznego zaangażowania humanistyki:

W laboratorium humanista, „tradycyjny” i „cyfrowy”, powinien usiąść wraz z programistą, technologiem czy designerem, by wspólnie dyskutować o problemach rozpatrywanych transdyscyplinarnie, jak np. o etyce technologii cyfrowych, mediach empatycznych, archiwach cyfrowych, kształtowaniu pamięci cyfrowej, narzędziach

---

<sup>122</sup> Upatrywałbym przykładu takiego dyskursu np. w walce o „pole” jaką prezentuje krytyka Derridy zaprezentowana przez Filimowicza.

<sup>123</sup> Pawlicka, „Humanistyka”, 329.

cyfrowych dyskryminujących ze względu na płeć czy o aplikacjach naruszających ludzką intymność<sup>124</sup>.

Z tekstu Pawlickiej płynie zatem ważna przestroga, by nie fetyszyzować naukowości oraz zdobyczy techniki. Drugim wątkiem jest efemeryczność – choć tutaj możemy argumentować, że całe zespoły poświęciły swój czas na rozwój tego typu badań (co już jest pozostawieniem śladu w rozwoju humanistyki jako takiej) i próbować odwrócić argument historyczny autorki: przekonamy się o losach humanistyki cyfrowej przy kolejnym zwrocie, skoro zmiana wpisuje się w nieunikniony sposób w przemiany naszych dziedzin.

Dla dalszych rozważań w kontekście PBL można też przyjąć, że *Digital Humanities* dobrze odpowiada na potrzeby instytucjonalne i metodologiczne tego projektu i że sama działalność bibliograficzna (dokumentalistyczna, archiwistyczna) pozostaje na tyle blisko techniki, że zastosowanie metod cyfrowych jest nie tylko rozwojowe czy inspirujące, ale konieczne zadaniowo.

Pojawia się także głos krytyczny w obrębie samego nurtu humanistyki cyfrowej, wyrażony przez Katherine Bode, jako krytyczna nota do podejścia *distant reading* i makroanaliz<sup>125</sup>. Jak zostało to wspomniane, ta badaczka jak najbardziej używa analiz statystycznych do opisu literatury. Krytyka, którą wyraża, ma jednak aspekt metodologiczny: Morettiemu wypomina nadużywanie słowa „fakty”, tymczasem dane to jeszcze nie fakty. Za niewystarczające uznaje wskazywanie bibliografii cyfrowych (nazwanie źródła), bez uwzględnienia zmienności tych baz – preferowane jest publikowanie *datasetów* lub precyzyjniejszych odniesień. Autorka przestrzega także przed stosowaniem współczesnych kategorii do danych historycznych, np. ignorowaniu warunków upowszechniania się wiedzy we wcześniejszych stuleciach. Niesie to ryzyko interpretacji danych. Widzimy tu zatem krytykę wyższego poziomu złożoności; nie są kontestowane podstawy (wszak Bode sama korzysta z metod cyfrowych), ale metody. Pytanie nie brzmi „dlaczego?” ale „jak?”. Według australijskiej badaczki ważniejsze od wyłonienia się nowego sposobu opisu dawnych problemów jest pojawienie się nowych przedmiotów badań przy zastosowaniu nowych narzędzi<sup>126</sup>. To ważny głos, ukazujący rolę

---

<sup>124</sup> Pawlicka, „Humanistyka”, 332-333.

<sup>125</sup> Katherine Bode, „The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”, *Modern Language Quarterly* 78, nr 1 (1 marzec 2017): 77–106, <https://doi.org/10.1215/00267929-3699787>.

<sup>126</sup> Co byłoby ciekawą odpowiedzią na słowa z przytaczanego artykułu Edera: „Tak samo jak teleskop jest jedynie narzędziem, użytecznym instrumentem, dzięki któremu widać więcej niż nieuzbrojonym okiem, tak samo statystyka może się stać użytecznym narzędziem, które pozwoli na analizę wielkich korpusów: pytania

badacza *Digital Humanities*, pracującego na danych, jako kuratora treści, a także apel o bardziej transparentny proces. Za najważniejszą myśl uważam spostrzeżenie Bode, że publikacja danych nie jest konkluzją, ale raczej zaproszeniem do dalszych dociekań.

W środowisku biznesowym, ale też w codziennie dostępnej komunikacji medialnej (wszędzie tam, gdzie statystyka jest nauką pomocniczą), możemy napotkać przykłady manipulacji danymi – dopasowanie faktów do liczb lub na odwrót. Na podobny sentyment zdaje się zwracać uwagę Eder, mówiąc o subiektywizmie badacza. Sama Bode podkreśla odpowiedzialność badawczą, niemniej pokusa lub ryzyko manipulacji i przeoczeń w prezentowaniu danych to realne ryzyka, dlatego tak ważne jest udostępnianie dokumentacji metody badań i sporządzanie opisu efektów.

Możemy też sformułować następujący zarzut: po co uczestniczyć w tradycyjnym obiegu informacji naukowej i prowadzić badania literaturoznawcze, jeśli za cały proces wytwarzania wiedzy odpowiadają narzędzia cyfrowe? Czy potrzebujemy drukowanych książek, które mówią o tym, że literaturoznawca przeprowadził analizy statystyczne na danych dotyczących powieści, gdzie prezentowane są wykresy?<sup>127</sup>

Warto wspomnieć, że pierwsze próby sformułowania kanonu dla internetowej wymiany w obrębie literackości w sieci nie mogły się uwolnić od medium drukowanego: wydany w 2003 *Liternet.pl*<sup>128</sup> jest publikacją książkową i przeglądany po dwudziestu latach wprawia w zdziwienie jako fizyczna książka, która traktuje o praktykach internetowych. Funkcjonuje do dzisiaj portal pod tą samą nazwą i adresem, gdzie zawartość stanowią odniesienia do utworów literackich. W publikacji pod redakcją Piotra Mareckiego znajdziemy natomiast teksty o hipertekstowości, społecznościach w sieci, blogach, net art (sztuce internetowej). Eseje i teksty krytyczne uzupełniają także wywiady i teksty o charakterze artystycznym, a jednym z dysonansowych elementów jest umieszczenie hiperłączy w bibliografiach (zdradzają brak bezpośredniego połączenia tych mediów ze sobą).

---

stawiane literaturze pozostają przecież te same”. Eder, „Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii”, 96.

<sup>127</sup> Niebezpieczeństwem jest także problem datafikacji (*datafication*), czyli zamieniania w dane coraz większej liczby aspektów codziennego życia. Przykład sportowy, przytoczony wcześniej, jest zasadny; mamy także pozytywne skutki analizy statystycznej w zarządzaniu miastami (transport publiczny, oświetlenie) czy codzienną aktywnością fizyczną (prywatne urządzenia biometryczne, promujące ruch w ciągu dnia itp.). Jednak zbyt proste, pochopne przenoszenie schematów obliczeniowych w obszar nauk humanistycznych może nieść ryzyko.

<sup>128</sup> Piotr Marecki, red., *Liternet.pl*, Wyd. 1 (Kraków: Rabid, 2003).

Od tamtego czasu zmieniło się jednak bardzo wiele i tak jak Liternet znalazł miejsce w sieci, tak i pozostałe impulsy krytyczne oraz twórcze zadomowiły się w cyberprzestrzeni. Na uwagę w przypadku badań nad hipertekstem na pewno zasługuje chociażby portal Techsty<sup>129</sup>. Mamy także przykład badań historii i archiwizacji polskiej sieci prowadzonych przez Marcina Wilkowskiego, wyrażanych w pracy naukowej (mającej swoje określone rygory), publicystyce i formie blogowej<sup>130</sup>.

W odwołaniu do części cyfrowej i drukowanej, ukuła się formuła przedstawiania raportu z prac i publikowania wniosków, co może być wymogiem instytucjonalnym. Publikacje książkowe wskazują jednak na miejsca w sieci i odwrotnie. Ruch z medium drukowanego jest też łatwiejszy dzięki projektowaniu publikacji z użyciem np. kodów QR, które po zeskanowaniu urządzeniem mobilnym odsyłają do adresu w internecie. Rozwój serwisów w sieci sprawił, że niekiedy są samowystarczalne i nie wymagają legitymizacji za pomocą ukazania się tych samych treści drukiem. Powszechną praktyką jest podział na część książkową, zapoznającą czytelniczki z ideami badań i ich tokiem, a także komponent cyfrowy (surowe dane w repozytorium, wizualizacja, strona towarzysząca badaniom), widzimy to w takich pracach jak *Enumerations* Andrew Pipera.

Ku definicji humanistyki cyfrowej

Wielość perspektyw, którą przedstawiam, metodologiczna nieokreśloność (mówimy bardziej o sposobie prowadzenia badań, nie o szkole badawczej)<sup>131</sup>, tworzą nieostry obraz obszaru badawczego<sup>132</sup>. Określeń i prób zdefiniowania zjawiska było wiele, zwraca na to uwagę Maciej

---

<sup>129</sup> Zob. "Techsty - literatura i nowe media", <http://techsty.art.pl>. Dostęp: 2023-09-23.

<sup>130</sup> Zob. "Marcin Wilkowski", <https://wilkowski.org>. Dostęp: 2023-09-23. Strona zawiera także bibliografię prac naukowych autora.

<sup>131</sup> Porównując różne definicje, Maryl zauważa też: „[k]luczowe jest tu uznanie humanistyki cyfrowej nie tyle za konkretną dyscyplinę czy metodę, ile zespół różnych praktyk badawczych, wykorzystujących cyfrową (pierwotnie bądź wtórnie) formę przedmiotu badań”. Zob. Maciej Maryl, „Kim są polscy humaniści cyfrowi?”, *Teksty Drugie* 1 (2017): 286–300, <https://doi.org/10.18318/td.2017.1.24>.

<sup>132</sup> Bardzo zniuansowany obraz wyłania się także z ustaleń Anny Kamińskiej, która w próbach ustalenia definicji prezentuje analizę dyskursu wokół *Digital Humanities* w piśmiennictwie polskim i amerykańskim. Upraszczając, w USA więcej pojawia się głosów, które prezentują humanistykę cyfrową jako nową dyscyplinę czy dziedzinę, wykorzystującą nowe technologie; w Polsce mówi się o nowych narzędziach i sposobach prowadzenia badań w obrębie dotychczasowych dyscyplin i opisuje *Digital Humanities* jako nurt lub model badań. W obu przypadkach trudno o jednoznaczną definicję, zakres pojęcia jest szeroki i rozmyty. Wśród cech, oba środowiska wymieniają cyfryzację badań, w szczególności analizy materiału, otwartość i pracę w zespołach. Tylko polscy badacze odwoływali się także do sytuacji humanistyki jako kryzysu i humanistyki cyfrowej jako odpowiedzi na ten kryzys. Z kolei amerykańscy badacze prezentowali bardziej pragmatyczne podejście, nie tylko jak Polacy wymieniając ogólne możliwości, ale też wskazując konkretne narzędzia i zastosowania. Kamińska jako punkt wspólny podkreśla zatem użycie narzędzi komputerowych do opracowywania materiału humanistyki, sugestią *Digital Humanities* jako kolejnego etapu rozwoju badań tradycyjnych i swego rodzaju promocją wykorzystania narzędzi cyfrowych w badaniach oraz dowartościowania humanistyki (poprawa

Maryl<sup>133</sup>. W swojej pracy zdecydowałem się posłużyć bardzo użyteczną definicją humanistyki cyfrowej, zaprezentowaną przez tego autora, zredukowaną do absolutnego minimum: badania humanistyki cyfrowej to takie badania humanistyczne, w których element cyfrowy (komputerowy, obliczeniowy) jest niezbędny do rozwiązania danego problemu badawczego<sup>134</sup>. I jeszcze krócej: projekt humanistyki cyfrowej to w takim razie *działanie humanistyczne z obowiązkowym komponentem cyfrowym*.

### Humanistyka cyfrowa w Polsce

Maryl przedstawia przekrojowy opis projektów, działań i inicjatyw w artykule *Kim są polscy humaniści cyfrowi?* – z niego dowiadujemy się przekroju dyscyplinarnym, używanych narzędziach czy statusie zawodowym badaczy<sup>135</sup>. W interpretacyjnej części napotkamy dwa tropy, charakteryzujące tę grupę w 2017 roku: 1) niewielkie użycie infrastruktur bazodanowych i przetwarzania danych przez respondowanych 2) pod względem celu stosowania metod cyfrowych, na pierwszym miejscu pojawiły się podstawowe czynności badawcze: odkrywanie, zbieranie i tworzenie materiałów badawczych. Widzę to jako spojrzenie tych respondentów na emergentną metodę jako szansę usprawniania badań. W odwołaniu do drugiego punktu Maryl, korzystając z typologii Tary McPherson, mówi że jest to profil humanistów „komputerowych” i „blogujących” bardziej niż „multimodalnych”<sup>136</sup>.

---

wizerunku, także w związku z pozyskiwaniem dofinansowania). Autorka konkluduje: „warto zatem spojrzeć na humanistykę cyfrową jako na projekt czy inicjatywę popularyzującą wykorzystanie nowych technologii w badaniach humanistycznych oraz promującą określone wartości w odniesieniu do pracy naukowej tej grupy naukowców” (usunięto wyróżnienie oryg. – JE). Zob. Anna Kamińska, „Cyfrowa humanistyka – koncepcja, kierunki i stan rozwoju oraz powiązania z informatologią” (Doktorat, Warszawa, Uniwersytet Warszawski, 2020), Doktoraty Wydziału Dziennikarstwa Informatyki i Bibliologii (Archiwum), <https://depotuw.ceon.pl/handle/item/3853>.

<sup>133</sup> Maryl, Maciej, „F5: odświeżanie filologii”, *Maciej Maryl F5 : odświeżanie filologii Teksty Drugie : teoria literatury, krytyka, interpretacja* 146, nr 2 (2014): 9–20.

<sup>134</sup> Dłuższa wersja tej krótkiej definicji brzmi: „Humanistyka cyfrowa to taki prąd w humanistyce, który polega - w dużym uogólnieniu - na wykorzystaniu narzędzi cyfrowych w badaniach. Chodzi o badania humanistyczne, dla których komponent informatyczny jest centralny, kluczowy i niezbywalny; o badania, których nie dałoby się przeprowadzić bez komputerów i oprogramowania (lub wymagałoby niewspółmiernie wielkich, wieloletnich nakładów pracy dużego zespołu badawczego); o traktowanie materiałów kulturowych (tekstów, obrazów, nagrań...) jako źródeł danych; o wykorzystanie ogromnych zasobów kulturowych, które możemy przetwarzać dzięki digitalizacji (pomyślcie tylko o tych wszystkich książkach, zdjęciach, obrazach, słownikach, bazach danych – już zdigitalizowanych i czekających na digitalizację); o zestawianie tych materiałów i wyszukiwanie związków czy relacji między nimi (proszę sobie wyobrazić porównanie języka WSZYSTKICH polskich powieści)”. Maryl, Maciej, „Co to jest humanistyka cyfrowa – odpowiedź na ankietę Czasu Kultury”, *Czas Kultury* 185, nr 10 (2015): 10–11.

<sup>135</sup> Maryl, „Kim są polscy humaniści cyfrowi?”. Jak zaznacza autor, badanie ankietowe nie jest reprezentatywne - było natomiast otwarte i promowane w środowisku.

<sup>136</sup> Te typy to „humaniści komputerowi (*computing humanists*), kładący nacisk na projekty infrastrukturalne (archiwizację, digitalizację, zachowanie dziedzictwa), humaniści blogujący (*blogging humanists*), zainteresowani nowoczesnymi możliwościami publikacji i komunikacji naukowej, oraz najnowszy typ humanistów multimodalnych (*multimodal humanists*), którzy łączą zainteresowanie bazami danych,



Autor formułuje końcowy wniosek: cyfrowych narzędzi używa się w polskich badaniach na wczesnym etapie procesu badawczego, a stopniowo zakres ich stosowania będzie się poszerzać. Maryl wskazuje jednak cechy otoczenia badań, które mają wpływ na dalszy rozwój humanistyki cyfrowej i stanowią trudności do przezwyciężenia: 1) rozproszenie (polskie struktury konsolidują się z dziesięcioletnim opóźnieniem względem USA, różne ośrodki działały niezależnie), 2) zarządzanie (brak spójnej polityki cyfrowej dla trzech resortów, których humanistyka cyfrowa dotyczy), 3) pieniądze (niskie wydatki na naukę, kosztochłonność projektów, brak dobrej propozycji grantowej dla tego typu badań).

Głównym źródłem wiedzy o polu badawczym jest raport *Humanistyczne projekty cyfrowe w Polsce*, który Maryl sporządził wspólnie z Marcinem Werłą, w związku z tworzącymi się strukturami konsorcjum DARIAH (szerzej o tej organizacji w dalszej części pracy)<sup>137</sup>. Przeprowadzono badanie ankietowe w środowisku naukowym, do którego w ramach odpowiedzi zgłoszono prawie 80 projektów. Badacze zwracają uwagę na wielość i różnorodność: „[p]odkreśla to tylko fakt, iż cyfrowość nie jest wcale nowym paradygmatem badań, tylko narzędziem do realizacji przeróżnych celów badawczych często odmiennych dyscyplin”<sup>138</sup>. Duży odzew środowiska naukowego każe zaś autorom raportu zauważyć potrzebę istnienia tego rodzaju bazy wiedzy dla rodzimej humanistyki cyfrowej. Ich opracowanie daje szeroki przekrój działań, w które angażowali się i angażują nadal rodzimi badacze. Odbiorczynie mogą poznać nazwy i rodzaje podmiotów realizujących, status poszczególnych projektów, źródło finansowania według typu, słowa kluczowe, typy i efekty tych inicjatyw. Polska Bibliografia Literacka jest tu określona mianem laboratorium wiedzy o współczesnej kulturze polskiej (tak nazywa się grant związany z bazą bibliograficzną; uwzględniał modernizację bazy danych i remediację roczników bibliografii ogłoszonej drukiem)<sup>139</sup>. Siłą rzeczy cały raport prezentuje model kwestionariusza (autorzy zdecydowali się na minimalne poprawki, znaczna większość treści to odpowiedzi respondentów ankiety); daje wszakże obraz panoramy projektów cyfrowych realizowanych w kraju.

---

narzędziami badawczymi czy nowymi formami współpracy, wykorzystując jednocześnie potencjał mediów wizualnych i audialnych” Maryl, 290.

<sup>137</sup> Marcin Werla i Maciej Maryl, „Humanistyczne projekty cyfrowe w Polsce” (Poznań-Warszawa, 2014), <https://rcin.org.pl/ibl/publication/84544>.

<sup>138</sup> Werla i Maryl.

<sup>139</sup> Szerzej odwołuję się do tego tematu i innych projektów w obrębie PBL w rozdziale trzecim.

Punkt wyjścia: projekty cyfrowe oparte na bibliografii

Szczególne miejsce dla naszych rozważań zajmują dwa badania, ze względu na zastosowane metody, a także przyjęte założenia dotyczące informacji na wejściu i na wyjściu (*input – output*). W praktyce pokazują one wykorzystanie danych bibliograficznych do formułowania wniosków, a także dają wyobrażenie o potencjalnych grupach odbiorczych dla opracowania *datasetu*, czym zajmuję się w rozdziale czwartym.

W pierwszym z badań, Jan van der Meer przeanalizował dane związane z polskim piśmiennictwem XVIII wieku, epoki stanisławowskiej (1764-1795)<sup>140</sup>. Celem było opisanie systemu literackiego, pod względem produkcji, wymiany, krytyki, mediacji i recepcji. Na dostępne dane składają się głównie informacje opublikowane w bibliografii *Nowy Korbut* (tomy 4-6) oraz odwołania do wyspecjalizowanych badań nad epoką, reprezentujących tradycyjne metody opisu historycznoliterackiego. Założonym wynikiem jest tutaj podanie szacowanej liczby uczestników takiego systemu literackiego: producentów, mediatorów, post-producentów (krytyków), a także szacunkowa liczba odbiorców. Przytaczając tło i wyjaśniając odmienny od wybieranych przez pozostałych badaczy okres badań, van der Meer podaje charakterystykę epoki, skupiając się na danych liczbowych tam, gdzie to możliwe: a) pojawienie się nowych gatunków, b) instytucjonalizacja krytyki literackiej w czasopiśmie, c) wzrost liczby księgarń, ich profesjonalizacja i promocja, d) wzrost liczby drukarni i zmniejszenie wśród nich proporcji zakładów związanych z Jezuitymi i organizacjami kościelnymi na rzecz świeckich mieszkańców Warszawy, e) powiększenie się publiczności czytelniczej i zwiększenie różnorodności tej grupy.

Dla określenia produkcji literackiej van der Meer redukuje początkową liczbę 533 producentów literackich w systemie, których rejestruje *Nowy Korbut* dla epoki oświecenia, do tych aktywnych w analizowanym okresie, otrzymując 215 tylko w epoce stanisławowskiej (wykluczeni zostają autorzy aktywni poza ramami czasowymi oraz publikujący w innych dziedzinach niż literatura piękna). Były to osoby, które wydały przynajmniej jedno dzieło. Przy opracowaniu danych widać prawidłowości, jak np. związki z trzema silnymi ośrodkami (dwór królewski w Warszawie, Puławy Czartoryskich, inteligencja warszawska). Badacz zauważa też stosunkowo rozwiniętą „międzynarodowość” wyrażoną w wielojęzyczności publikacji. Estymowana liczba aktywnych pisarzy, uwzględniając anonimowych autorów,

---

<sup>140</sup> Jan IJ. van der Meer, „The Literary System of the Stanisław Age in Poland: Its Participants”, *Russian Literature* 49, nr 4 (maj 2001): 449–88, [https://doi.org/10.1016/S0304-3479\(01\)80047-8](https://doi.org/10.1016/S0304-3479(01)80047-8).

to około setka wydających przynajmniej kilka dzieł rocznie w różnych gatunkach i 150-200 takich, którzy byli mniej aktywni (pojedyncze przekłady, publikacje poezji itp.).

Analogicznie postępuje van der Meer z danymi dotyczącymi krytyki literackiej (np. taka obserwacja na temat krytyków-recenzentów: 18-20 krytyków, wyłącznie mężczyzn, żaden nie zajmował się tylko krytyką – nie była to ówczesnie profesja). Analizowane są dane materialne, jak nakłady, liczba wydawanych tytułów, informacja o liczbie zakładów drukarskich, księgarni itp. Autor bierze też pod uwagę działalność bibliotek i czytelników, pod kątem oceny recepcji i życia literackiego. Nakłady czasopism służą tu do szacunkowej oceny skali ich czytelniczego odbioru.

Van der Meer transparentnie zaświadcza o limitach swojego podejścia, np. nie uwzględnia w badaniu niemieckojęzycznej części mieszczańskiej w północno-zachodniej Polsce. Stoi za tym intencja badań nad polszczyzną i językami rozumianymi w równym stopniu (stąd francuski został uwzględniony). Pominięte zostały też aspekty patronatu i klienteli: prominentne postaci miały wpływ na system, szczególnie producentów i recenzentów. W końcowych wnioskach jest jednak w stanie uznać na podstawie uzyskanych wartości liczbowych, że system literacki, mimo że nie był szeroki, da się opisać jako samoorganizujący się, relatywnie autonomiczny, produkujący polemiki w sieciach autorów i krytyków.

Drugim ważnym badaniem, wykorzystującym dane bibliograficzne, są studia nad English Short-Title Catalogue (ESTC) wytworzonego w British Library, które wprost mogą przekładać się na ustalenia w dziedzinach historii i historii literatury; autorami tej analizy są Leo Lahti, Niko Ilomäki i Mikko Tolonen<sup>141</sup>. Zakres geograficzno-kulturowy tworzą Wielka Brytania i USA, a czasowy to lata 1470-1800. Danymi wejściowymi są określenia czasu i miejsca publikacji, formaty (wielkość), tytuły, autorzy i ich lata życia. Pozwalają one jednak po odpowiednim przygotowaniu i przetworzeniu rozprawiać na temat dostępnej wiedzy o systemie literackim – dodajmy, w zupełnie inny sposób, niż robiłby to historyk literatury, starając się stworzyć syntezę wiedzy na temat piśmiennictwa epoki i opierając się na danych z autopsji oraz własnej refleksji<sup>142</sup>. Dzięki analizie formatów publikacji autorzy proponują

---

<sup>141</sup> „[L]ibrary catalogues can provide a valuable data resource for historical studies. We show how the analysis can be performed within a comprehensive quantitative framework, following the best open-science practices of transparency, reproducibility and code sharing”. Leo Lahti, Niko Ilomäki, i Mikko Tolonen, „A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470-1800”, *LIBER Quarterly: The Journal of the Association of European Research Libraries* 25, nr 2 (2015-12-04): 89, <https://doi.org/10.18352/lq.10112>.

<sup>142</sup> Mam tu na myśli np. tzw. małe i duże „cegły”, przewodniki po literaturze pisane w formie eseistycznej, głównie do użytku edukacyjnego na studiach wyższych, które porządkują najważniejsze zjawiska historycznoliterackie w ułożeniu umownej periodyzacji w podziale na epoki. Jednym z dobrze znanych

np. refleksję nad zmianami konsumpcji papieru i zapotrzebowania systemu wytwórczego literatury na ten surowiec (sprawa, której nadają znaczenie, a może umknąć refleksji skupionej na treści dzieł, dorobku twórców i świadectwach recepcji). Z kolei liczba edycji (wznowienia) pozwala określić z przybliżeniem liczbę „sprzedanych” kopii (z zachowaniem proporcji gdy przykładamy miary współczesnych rynków książki do realiów historycznych) a także dają liczbowe podstawy do opisywania historii rozwoju narzędzi drukarskich do efektywniejszego kopiowania (zwiększania nakładów). Wśród uwag końcowych, autorzy wskazują potencjalne nowe obszary badawcze dla takiego ujęcia, np. oświecenia i europejskiej Republiki listów. Wyrażone zostaje też oczekiwanie, że instytucje będą dzielić się danymi na potrzeby takiego typu badań<sup>143</sup>.

Co znaczące, w obu badaniach z właściwie rudymenarnych informacji na temat tekstów otrzymujemy zbiór nowych wniosków. Wystarczy zmiana skali postrzegania problemu, zadanie odpowiednich pytań badawczych i analiza dostępnych informacji. Stwarza to pole dla prowadzenia otwartych i reprodukowalnych badań ilościowych systemu literackiego<sup>144</sup>. Podkreślona zostaje także (szczególnie w drugim tekście) transparentność udostępnienia danych i metod badawczych, nie tylko dla odtworzenia wyników, ale również aktualizacji na wypadek pojawienia się nowych informacji.

#### Uwagi końcowe

Podsumowując, humanistyka cyfrowa tworzy projekty i strategie, ale nie jest metodologią. Wymagałoby wytworzenia spójnego systemu, na miarę innych dużych prądów w humanistyce, takich jak strukturalizm, psychoanaliza czy dekonstrukcjonizm. Nie brakuje tu istotnych nazwisk oraz prac, ale na pewno nie uświadczymy w dokonaniach cyfrowych humanistów spójności metody, która mogłaby nam pozwolić nazywać omawiane zjawisko szkołą czy metodą badawczą. Digitalne narzędzia tworzą raczej doraźne konfiguracje, które odpowiadają na zarysowanie przez badacza problemu i pomagają opisać zagadnienie oraz przedstawić rozwiązania.

---

przykładów są prace Jerzego Ziomek dot. renesansu: Jerzy Ziomek, *Renesans*, Wyd. 11, Wielka historia literatury polskiej (Warszawa: Wydawnictwo Naukowe PWN, 2012); Jerzy Ziomek, *Literatura Odrodzenia*, Wyd. 3 popr. i uzup. Dzieje Literatury Polskiej : synteza uniwersytecka (Warszawa: Wydaw. Naukowe PWN, 1999). Przy tworzeniu takiego omówienia, *computing*, „operacje liczbowe” zachodzą mentalnie, co jest subiektywne. W przyszłości tego rodzaju syntetyczny opis może być wspomagany maszynowo.

<sup>143</sup> „We anticipate that as libraries and other data holders give researchers and the general public full access to their data resources, there will be a rapidly increasing demand for open source ecosystems of data analysis in this field, the application of which we have demonstrated in this article”. Lahti, Ilomäki, i Tolonen, „A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470-1800”, 108.

<sup>144</sup> Lahti, Ilomäki, i Tolonen.

Co to oznacza dla uprawiania literaturoznawstwa? Wspólnym punktem, który wyłania się z przywoływanych wcześniej prac, jest potraktowanie materiału literackiego (zawartości dzieł lub danych i metadanych na ich temat, na przykład bibliograficznych) jako punktów danych, zasilających maszyny obliczeniowe, w celu przedstawiania analiz oraz wizualizacji. Wkład kompetencji historyczno- i teoretycznoliterackich nadal jest istotny: wymienione prace badawcze opierały się na silnej podstawie studiów dotyczących literatury i traktowały narzędzia filologiczne jako główny komponent przy opisie danych oraz formułowaniu wniosków. Cyfrowi humaniści, których dokonania zaprezentowałem, w większości kształcili się najpierw w humanistycznych dziedzinach, a cyfrowy „narzędziownik” zdobywali i rozwijali w trakcie studiów, badań i prowadzenia projektów interdyscyplinarnych.

Chris Houghton wymienia trzy wyzwania stojące przed humanistyką cyfrową: 1) Analiza dużych zbiorów danych wymaga (dostępu do – JE) danych; 2) Hostowanie i utrzymanie dużych zbiorów danych w ramach działalności instytucji 3) Wytwarzanie narzędzi technicznych do prowadzenia analiz<sup>145</sup>. Proponuje także swym czytelnikom sposoby na poradzenie sobie z tymi problemami.

Przede wszystkim autor zauważa rolę instytucji tworzących archiwa: „There has been a sea of change in humanities studies in recent years. As part of the first wave, librarians and researchers have sought to harness the newfound power of digital archives to not only collect, but just as importantly, to curate, clean and analyze large sets of data”<sup>146</sup>. Houghton zauważa również na wstępie, że wyjściowa sytuacja dla badań *Digital Humanities* ujawniła rozłączność umiejętności pomiędzy tymi niezbędnymi do opracowywania problemów badawczych merytorycznie oraz tymi, które służą do czysto technicznego wsparcia w tym procesie. W jego przykładzie: bardzo mało osób wykazuje jednocześnie zainteresowanie osiemnastowieczną powieścią gotycką *oraz* posiada zdolności programistyczne, pozwalające na zbieranie, harmonizację i analizę danych na ten temat.

Autor odwołuje się, jak inni wcześniej, do słowa „obietnica” w kontekście humanistyki cyfrowej: tym razem ze względu na dostępność informacji. Jak zauważa Houghton, przygotowanie danych w bazach humanistycznych jest już dużo lepsze; oprócz tego można korzystać z narzędzi kwerend w wielu źródłach jednocześnie. Przekłada się to na skrócenie

---

<sup>145</sup> Chris Houghton, „Guest Post — Three Challenges (and Solutions) to Expand Digital Humanities”, *The Scholarly Kitchen* (blog), 2022-09-26. <https://scholarlykitchen.sspnet.org/2022/09/26/guest-post-three-challenges-and-solutions-to-expand-digital-humanities/>.

<sup>146</sup> Houghton.

czasu potrzebnego na dotarcie do informacji i uwolnienie sił i czasu na analizę, opis oraz wysnuwanie wniosków. Autor doprecyzowuje cel: „We should, however, never think of *Digital Humanities* as a replacement for traditional methods of enquiry, but instead as a way of enhancing them and uncovering new possibilities”<sup>147</sup>.

Pierwsza trudność: by analizować duże zbiory danych, trzeba mieć dane. Pozorna tautologia zdradza swój sens, gdy rozważymy dostępność źródeł, na których nam zależy. Dane mogą nie być wytworzone. Mogą istnieć, ale w ograniczonym dostępie. Może być tak, że mają określony format, wymagający zastosowania wyłącznie określonych narzędzi. Wytwarzanie danych to także łączenie i harmonizacja już istniejących informacji, jak również oferowanie możliwości połączenia z innymi danymi. Houghton przy tym punkcie podkreśla, jak ważne jest utrzymywanie archiwów: upewnianie się, że kolekcje zasobów wirtualnych są aktualne. Oprócz tego aspektu istotna jest troska o szeroki zakres tematyczny tych danych, aby dało się opracowywać na ich podstawie różne tematy badawcze. Konkretnym przykładem jest stworzenie przez biblioteki uniwersyteckie możliwości dostępu do wielu źródeł połączonych.

Z drugim problemem wiąże się konieczność utrzymania dostępu do danych w otoczeniu instytucyjnym. Projektom z zakresu humanistyki cyfrowej bardzo często grozi efemeryczność, poprzez ich rozproszenie czasowe i domenowe (np. różne miejsca w sieci, hostujące zasoby czy określony czas aktywności strony projektu, po którym przestaje być dostępna). Dlatego zawczasu warto przewidzieć te trudności i pomyśleć o trwaniu i dostępności wytworów cyfrowych. Dobrą praktyką staje się więc utrzymywanie projektów np. na podstronach nadrzędnych instytucji, a nie w niezależnych miejscach w sieci. Houghton podpowiada również, że instytucje powinny organizować a) dostęp, w tym subskrypcje, do zasobów (np. korpusy, bazy czasopism), b) narzędzia materialne (*hardware*) i niematerialne (oprogramowanie, licencje na obliczenia w chmurze) oraz c) zabezpieczenia zasobów i narzędzi, jak również wszelkiego *hardware*, związanego z ryzykami w projektach (niepowołany dostęp, awarie, czasowy brak dostępu, błąd ludzki i inne).

Trzecią kwestią do rozwikłania pozostaje połączenie kompetencji cyfrowych i humanistycznych. Houghton wspomina o zmniejszaniu luki kompetencji poprzez użycie specjalnie przystosowanych środowisk, które nie wymagają kodowania, np. multiwyszukiwarek i filtrowań (możemy dodać tu narzędzia typu *no code*, czyli takie, które nie wymagają umiejętności programowania, a oferują efekty dzięki sterowaniu za pomocą

---

<sup>147</sup> Houghton.

interfejsów graficznych<sup>148</sup>). Oszczędza to znacznie czas na zbieranie, kurację, czyszczenie i analizę danych.

Wśród uwag końcowych Houghton wskazuje szansę na otwarciu badań cyfrowych dla większej części kadry uczelnianej oraz studentów. Zauważa otwarcie na nowe obszary badawcze. Widzi także okazję wykształcenia kadr cyfrowych humanistów o umiejętnościach technicznych przydatnych w innych kontekstach (*transferable technical skills*), co sprzyja tak instytucjom, jak uczącym się osobom. Z jednej strony obniża się zatem próg wejścia do pracy z danymi, z drugiej zaś przetwarzanie danych stymuluje rozwój nauki.

Jak to zostało zasygnalizowane przy okazji dwóch ostatnich przytoczonych badań (system literacki epoki stanisławowskiej van der Meera oraz badania korpusu ESTC autorów z Uniwersytetów w Wageningen i Helsinkach), dane bibliograficzne stanowią bardzo interesujący obszar: z podstawowych informacji możemy wywodzić wnioski i opisy znaczące dla narracji naukowych, które dotyczą systemu literackiego. Dostępne „na wejściu” fakty zdają się być zaledwie rudymenarne, ale osadzone w kontekście istniejących ustaleń – oświetlają zastane prace w zupełnie nowy sposób. Dzięki użyciu narzędzi technicznych, związanych z informatyką oraz informacją naukową, można zobjektywizować optykę badań humanistycznych – także w szeroko pojętym literaturoznawstwie. Otwiera to pole dla *Bibliographic Data Science*, czyli badań (bardziej) reprodukowalnych, zespołowych i usieciowionych (zarówno poprzez infrastrukturę techniczną, jak i kadry)<sup>149</sup>.

---

<sup>148</sup> Houghton wymienia Gale, HathiTrust Digital Library, Google Books oraz ProQuest jako narzędzia wyszukiwania i selekcji danych. Innym przykładem jest Orange, narzędzie do uczenia maszynowego i wizualizacji danych, opublikowane w wolnym dostępie na stronie <https://orangedatamining.com/>. Zob. też Junting Huang, „Programming Without Coding: Orange for Digital Humanities”. 2023-07-15. <https://medium.com/@juntinghuang/programming-without-coding-orange-for-digital-humanities-39b60078cd27>. Dostęp 2023-09-16. Autor tego tekstu co prawda posiada obie grupy umiejętności (informatyczne oraz literaturoznawcze), ale zachęca filologów do rozwijania zdolności informatycznych, prezentując Orange jako dobry wstęp do nauki. Zauważa także paradoksy tego, jak istotne jest opowiadanie historii w prezentacji danych i jak w komponencie cyfrowym praca dzieli się na dwie części: programowanie to logiczne rozwiązanie problemu, a kodowanie jest wytworzeniem implementacji tego procesu w skodyfikowanym zbiorze komend. Wśród innych narzędzi, których używają humaniści cyfrowi, możemy wymienić takie jak Voyant Tools, RapidMiner, Gephi, StoryMaps czy Omeka. Jest to zależne od potrzeb w projekcie i zadania do wykonania przy użyciu tych aplikacji. Por. <https://liberalarts.tamu.edu/codhr/resources/dh-tools-and-software/>.

<sup>149</sup> Wśród infrastruktur związanych z tym nurtem można wymienić fińską grupę Helsinki Computational History Group (zob. <https://www.helsinki.fi/en/researchgroups/computational-history>), grupę roboczą DARIAH Bibliographical Data (zob. <https://www.dariah.eu/activities/working-groups/bibliographical-data-bibliodata/>; lista publikacji: [https://zenodo.org/communities/bibliodatawg\\_dariah?q=&l=list&p=1&s=10&sort=newest](https://zenodo.org/communities/bibliodatawg_dariah?q=&l=list&p=1&s=10&sort=newest)); na polskim gruncie zagadnienie omawia także Marcin Roszkowski, por. Marcin Roszkowski, „Bibliographic Data Science – konceptualizacja obszaru badawczego”, *Przegląd Biblioteczny* 90, nr 1 (2022-10-31): 5–26, <https://doi.org/10.36702/pb.914>.

Ten temat analityki opartej na danych bibliograficznych zostanie jeszcze rozwinięty w kolejnych rozdziałach. W następnej części pracy przejdziemy do zdefiniowania, czym są dane badawcze i przyjrzymy się, jak można tworzyć ontologie dla humanistyki – a może przede wszystkim, choć nie wprost, postaramy się dowieść, dlaczego to może być istotne w dyskursie literaturoznawczym przyszłości.



## Rozdział 2. Otwarte dane badawcze

Idea ontologii *versus* Wielkie Wysypisko Wiadomości (z korporacyjnymi interesami w tle)

Pod koniec XX wieku Clifford Stoll nie szczędził sceptycyzmu dla *novum*, jakim w świecie mediów masowych jawił się internet. W tekście opublikowanym w 1995 w Newsweeku, zatytułowanym *Why the Web Won't Be Nirvana*, autor wyśmiewa jako mrzonki pomysły w rodzaju powszechności pracy zdalnej, bibliotek interaktywnych czy szkolnych lekcji multimedialnych<sup>150</sup>.

Zakupy przez internet Stoll uznaje za niewypał, ze względu na brak bezpiecznych płatności i rzeczywistego kontaktu klienta ze sprzedawcą. Podobne słowa przypisuje się także Stanisławowi Lemowi, który, przewidziawszy w prozie niektóre wynalazki techniczne na długo przed ich powstaniem, jeszcze za życia miał okazję obserwować rozwój międzynarodowej sieci. Autor *Bajek robotów* miał powiedzieć, że butów nie da się kupić przez internet (ponieważ wymaga to przymierzania)<sup>151</sup>. Okazuje się, że rynek obuwniczy zaadaptował się do tej okoliczności, np. za pomocą mechanizmu ułatwionego zwrotu nietrafionego zakupu.

Konkluzją artykułu Stolla jest zbyt duża przewaga bezpośredniego ludzkiego kontaktu nad tym wirtualnym, aby sieć mogła stać się „krajem marzeń”. Nie tylko handel miał się według niego zatrzymać na rzeczywistości dostępnej fizycznie (cieleśnie). Sytuacje, które wymienia, to przyjacielskie rozmowy przy kawie, uczestnictwo w wydarzeniach kulturalnych, jak koncerty, a także wirtualna rzeczywistość i (cyber)seks – wszystkie nie mające sobie równych „w realu” względem internetowych substytutów.

---

<sup>150</sup> Clifford Stoll, „Why the Web Won't Be Nirvana”, Newsweek, 1995-02-26, <https://www.newsweek.com/clifford-stoll-why-web-wont-be-nirvana-185306>.

<sup>151</sup> Tomasz Borejza, „Stanisław Lem przewidział, co internet zrobi z ludzkim mózgiem”, *Krowoderska.pl* (blog), 2020-05-26, <https://krowoderska.pl/stanislaw-lem-przewidzial-co-internet-zrobi-z-ludzkim-mozgiem/>. Dostęp 2023-09-16. W cytowanym tekście pojawia się również ciekawa wzmianka, dotycząca braku „centrum” internetu pod względem nadawania i tworzenia treści: wzrasta szybkość przesyłu informacji, ale nie zawsze jej jakość. Niekoniecznie także muszą występować mechanizmy kontroli tejże. Borejza cytuje również sformułowany przez Lema problem etyczny, wywiedziony z twórczości Josepha Conrada: nowe technologie dają nowe narzędzia, ale nie zmieniają natury człowieka w którym „tkwi kanibal”. Dla naszych rozważań jest to o tyle istotne, że odnosi się wprost do krytycznej analizy treści, weryfikacji prawdziwości komunikatów i poszukiwania wiedzy.

Wraz z rozwojem sieci wygląda na to, że kolejne opisane przez Stolla obszary ustępują wirtualizacji doświadczeń. Wykorzystanie internetu nie musi być w nich dominujące, ale jest znaczące i zauważalnie zmienia ludzkie interakcje. Doświadczenie pandemii COVID-19 pokazało zaś dobitnie, że znaczna część populacji światowej może przez jakiś czas pozwolić sobie na dobrowolny *lockdown* i jak najbardziej substytuować przy pomocy internetu życie w społeczeństwie – zamieniając się w sieć pojedynczych wysp z dostępem do internetu.

Dziś można śmiać się z prób przewidywania przyszłości przez Stolla. Mało tego, złośliwcy przypominają jego artykuł na rocznicę jego publikacji, porównując niepewny los papierowego wydania *Newsweeka* do sukcesu internetowych portali informacyjnych współczesności<sup>152</sup>. Jednym z czynników uniemożliwiających dostrzeżenie powstającej przyszłości na pewno był zawrotny rozwój infrastruktury, co upowszechniło korzystanie z sieci. Dla Stolla *web* musiał być czymś elitarnym, znanym garstce osób, najczęściej korzystających z niego w kontekście zawodowym czy akademickim. Tymczasem w drugiej dekadzie XXI dostęp do internetu żartobliwie określa się jako jedną z podstawowych potrzeb człowieka.

Ale to nie tylko swoisty antywebowy luddyzm przemawia w pisarstwie Stolla. Autor ma wszak wykształcenie techniczne (jest astronomem) i pracował jako administrator systemów w Lawrence Berkeley Laboratory. W tej ostatniej roli, w latach osiemdziesiątych, brał udział w schwytaniu niemieckiego hakera, pracującego dla KGB, co opisał w książce *Kukulcze jajo*<sup>153</sup>. Felieton z *Newsweeka*, który chybił celu, otwiera wyznanie, że autor spędził w sieci już dwadzieścia lat (w 1995!) i owszem, ma z tym związane pozytywne wrażenia, lecz nie oszołomiły go na tyle, by kibicował upowszechnieniu tego wynalazku jako części życia nas wszystkich. W książce *Krzemowe Remedium* (w oryginale – *Silicon Snake Oil* – „Krzemowa maść na szczury”?) Stoll wypowiada się w podobnym tonie: uznaje internet za wysypisko informacyjnych śmieci<sup>154</sup>.

Jednym z powodów sceptycyzmu Stolla wydaje się być zbyt bliska perspektywa: korzystał z dobrodziejstw internetu u jego zarania, w codziennej pracy naukowej i w ramach grup

---

<sup>152</sup> Jon Levine, „That Time in 1995 Newsweek Predicted This ‘Internet’ Thing Wouldn’t Catch On”, 2018-02-06, <https://www.thewrap.com/newsweek-predicted-internet-wouldnt-catch-on-1995/>. Dostęp 2023-09-16.

<sup>153</sup> Clifford Stoll, *Kukulcze jajo*, tłum. Tomasz Hornowski (Poznań: Dom Wydawniczy Rebis, 1998); por. Clifford Stoll, „Stalking the Wily Hacker”, *Communications of the ACM* 31, nr 5 (maj 1988): 484–97, <https://doi.org/10.1145/42411.42412>.

<sup>154</sup> Clifford Stoll, *Krzemowe remedium. Garść rozważań na temat infostrady*, tłum. Tomasz Hornowski (Poznań: Dom Wydawniczy Rebis, 2000).

dyskusyjnych. Dlatego jego akademickie oczekiwanie rzetelnej wiedzy spotyka się tutaj z początkowym chaosem sieci, tak pod względem komunikacyjnym, jak technologicznym:

What the Internet hucksters won't tell you is tht the Internet is one big ocean of unedited data, without any pretense of completeness. Lacking editors, reviewers or critics, the Internet has become a wasteland of unfiltered data. You don't know what to ignore and what's worth reading. Logged onto the World Wide Web, I hunt for the date of the Battle of Trafalgar. Hundreds of files show up, and it takes 15 minutes to unravel them—one's a biography written by an eighth grader, the second is a computer game that doesn't work and the third is an image of a London monument. None answers my question, and my search is periodically interrupted by messages like, „Too many connections, try again later.”<sup>155</sup>

Na przełomie XX i XXI rzeczywiście takie projekty jak Wikipedia i pierwsze sieci społecznościowe czy fora mogły nie budzić zaufania badaczy akademickich. Z czasem jednak ta kolektywna kuracja treści weszła na wyższy poziom, wraz ze wzrostem skali: im więcej zaangażowanych, tym większa weryfikowalność informacji – choć trzeba zaznaczyć, że działa w internecie dynamika grupowa i międzygrupowa<sup>156</sup>.

Innym krytykiem społecznego tworzenia treści w internecie stał się Andrew Keen, w swojej książce atakując tytułowy „Kult amatora”<sup>157</sup>. Keen wytyka niekompetencję osobom posługującym się internetem i aspirującym do roli ekspertów. Autor posługuje się przy tym m.in. określeniem *egocasting*, dla wyrażenia narcystycznego sposobu komunikacji w sieci. Znamy go dziś z roli *influencerów* jako przekaźników treści dla odbiorcy, czy to recenzentów, czy społecznych dziennikarzy lub po prostu wykonawców w procesie reklamy. Tam, gdzie Stoll wytyka niekompetencję amatorskim twórcom treści w internecie, Keen idzie dalej i wskazuje na ich swoisty narcyzm. W humanistyce podobieństwa takiego dyscyplinującego

---

<sup>155</sup> Stoll, „Why the Web Won't Be Nirvana”. Zachowałem pisownię oryginalną.

<sup>156</sup> Na temat społecznego konstruowania wiedzy por. Amy S. Bruckman, *Should You Believe Wikipedia?: Online Communities and the Construction of Knowledge*, 1. wyd. (Cambridge University Press, 2022), <https://doi.org/10.1017/9781108780704>.

<sup>157</sup> Por. Keen, Andrew. *The cult of the amateur: how today's internet is killing our culture*. 1st ed. New York: Doubleday/Currency, 2007. Korzystałem z recenzji Jakuba Nowaka, por. Jakub Nowak, „Andrew Keen. 2007. Kult amatora. Jak Internet niszczy kulturę. Warszawa: Wydawnictwa Akademickie i Profesjonalne, ss. 198 Bohdan Jung (red.). 2010. *Wokół mediów ery Web 2.0*. Warszawa: Wydawnictwa Akademickie i Profesjonalne, ss. 263”, *Nowe Media*, b.d.

podejścia znajdziemy (z zachowaniem wszelkich proporcji) w pobłażliwym opisie komunikacji kulturalnej mediów masowych autorstwa George'a Steinera<sup>158</sup>, gdy pisze, że szersza publiczność jest skazana na bycie kierowanymi przez „mandarynów” ze względu na złożoność kultury – a odnosi to pojęcie do krytyków literackich, także związanych z akademią – i to określenie przyłgnęło też do samego autora *Gramatyk tworzenia*<sup>159</sup>. U Steinera następuje podkreślenie skrajności opozycji między tworzeniem kanonu a demokratyzacją kultury. Podobnych wątków, inspirujących ten temat, doszukamy się też zresztą w komentarzach do kultury popularnej autorów związanych ze szkołą frankfurcką<sup>160</sup>. Da się przy tym zauważyć jak różne są od siebie podejścia: konserwatywnej humanistyki, odwołującej się do konieczności wykształcania ekspertów, technoentuzjazmu, oddającego rolę

---

<sup>158</sup> Ilustruje to fragment: „Zwyczaj i wartości o przemożnej wadze w dzisiejszym społeczeństwie konsumpcyjnym Zachodu są odwrotnością wartości przeważających w wymagowanym społeczeństwie bezpośrednim (cytuję szkic *Wtórne miasto*, część książki *Rzeczywiste obecności*, oryginalnie wydanej w 1989; po trzydziestu latach diagnoza wydaje się podobna, również w Polsce łatwiej dostrzec skutki rozwoju kultury masowej późnego kapitalizmu – dop. JE). Dominuje wtórność i pasożytnictwo. Wyedukowanemu społeczeństwu stręczy się miliony słów drukowanych, przedstawianych, nadawanych, wyświetlanych: o książkach, do których ludzie nigdy nawet nie zajrzą, o muzyce, której nigdy nie wysłuchają, o dziełach sztuki, których nie zaszczyca choćby spojrzeniem. W powietrzu unosi się nieustające brzęczenie komentarza estetycznego, gotowych sądów, fabrycznie pakowanych ocen, przypieczętowanych bullą nieomyślności. (...) Biblioteka przeciętnego college'u czy uniwersytetu subskrybuje mniej więcej trzy do czterech tysięcy periodyków humanistycznych. Wiele tysięcy towarzystw naukowych, stowarzyszeń (...) publikuje zeszyty i wydania zbiorowe, dokumentując rozmaite pasje badawcze w mniej lub bardziej oficjalnym i usystematyzowanym formacie. (...) Żaden komputer nie zinwentaryzuje wiadomości bieżących, dorocznych memorabiliów, zapisków z kolejnych sabatów, celebrujących taki czy inny rytuał kanonicznej eksplikacji. (...) Faulkner był przedmiotem tysiąca uczonych artykułów i prac (...) zanim powstał periodyk i biuletyn w całości poświęcony jego dziełu. Pas transmisyjny wypływa kolejne komentarze o twórczości Ezry Pounda czy Samuela Becketta. Mandaryńskie szaleństwo wtórnego dyskursu zatruwa myśl i wrażliwość. (...) Poważna sztuka, muzyka literatura nie są *interesujące* w takim sensie, w jakim musi być dziennikarstwo. (...) roszczenia dziennikarstwa wobec dzieła i życia człowieka mają charakter totalitarny (...) Środki masowego przekazu zatrudniają doradców małżeńskich i astrologów. Dlaczego nie miałyby zatrudniać krytyków sztuki i recenzentów muzycznych?” Zob. George Steiner, *Rzeczywiste obecności: czy w tym, co mówimy, coś w ogóle jest?*, tłum. Olga Kubińska, Wydanie drugie zmienione (Gdańsk: słowo/obraz terytoria, 2023), 22-24. Rozstrzelony druk zamieniłem w zapisie na kursywę.

<sup>159</sup> Por. Maya Jaggi, „George and his dragons”, *Guardian*, 2021-03-17.

<https://www.theguardian.com/books/2001/mar/17/arts.highereducation>. Dostęp 2023-09-16. Zob. też Robert Boyers, „George Steiner as Cultural Critic: Confronting America”, *New England Review* 15, nr 2 (1993): 136–56.

<sup>160</sup> Mam na myśli teorię m.in. przemysłu kulturowego i fetyszizmu towarowego. O aktualności teorii krytycznej szkoły frankfurckiej w późniejszych badaniach oraz wobec kultury masowej późnego kapitalizmu: por. Marcin Hylewski i Tomasz Burdzik, „Teoria krytyczna szkoły frankfurckiej jako krytyka kultury masowej”, 2014. <https://doi.org/10.6084/M9.FIGSHARE.1050567>. Przytaczam ostatnie zdanie tej pracy: „[teoria krytyczna] w ogromnym stopniu przyczyniła się do rozwoju interdyscyplinarnych badań nad kulturą masową. Wydaje się, że jej znaczący wpływ na myśl współczesną wynika z ciągłej aktualności zjawisk, które uczynili przedmiotem swojej analizy: kultury masowej i hegemonii kapitalizmu (w szerokim rozumieniu, obejmującym nie tylko ekonomię, ale i kulturę, politykę oraz sferę stosunków społecznych)”. Zob. też Andrzej Szahaj, *Teoria krytyczna szkoły frankfurckiej: wprowadzenie*, Współczesna Myśl Humanistyczna. Polskie Spojrzenia (Warszawa: Wydawnictwa Akademickie i Profesjonalne. Grupa Kapitałowa WSiP, 2008).

prowordrów właściwie każdemu i nikomu, a także publicystyki, zazdrosnej o pozycję względem mas<sup>161</sup>.

Typowe słowa krytyki dotyczą zatem kilku obszarów, w tym: partycypacji osób o różnym poziomie wiedzy w tworzeniu treści na dany temat, stroniczości doboru i przedstawiania informacji, połączonej ze swoistym narcyzmem (interpretowanym jako cecha komunikacji w sieci w ogóle), niestabilności infrastruktury i efemeryczności webowych źródeł. Te ostatnie argumenty, dotyczące technologii zdają się już być nieaktualne.

Stan badań i dobre praktyki w obszarze otwartych danych

W tej części pracy odwołuję się do dynamiki tworzenia i przetwarzania danych – między trzema instancjami: zespołami eksperckimi, użytkownikami sieci, a także maszynami. Przedstawiam standardy, rekomendacje i dobre praktyki, których użycie zapewnia widoczność w sieci. Ze względu na zakres badań, skupiam się na wymianie informacji w obrębie badań naukowych, z okazjonalnymi wtrąceniami tematów związanych z biznesem i sieciami społecznościowymi, tam gdzie to niezbędne lub gdzie otwiera to nowe pole refleksji.

Najpierw chciałbym zająć się najbardziej niepewnym terminem, związanym ze społecznym tworzeniem treści. Tim O'Reilly ukuł pojęcie *Web 2.0*, opisując w ten sposób usługi i serwisy angażujące społeczności użytkowników w tworzenie struktur wiedzy. Jest to swoisty pojęciowy trik, który miał przekonać odbiorców do tego, że oto spotykają się z „nową, lepszą siecią”. Zanim przejdę do wyjaśnienia, zastrzegam, że ustalenia O'Reilly'ego wybrane zostały ze względu na przystępność prezentacji.

W opublikowanym w 2007 tekście<sup>162</sup> przedstawia model „starej” sieci, odchodzącej do lamusa, oraz emergentnych serwisów, które wyprą poprzednie rozwiązania. Kluczem tego sukcesu były „sprytniejsze” modele biznesowe, zorientowane na użytkownika. Z tych opozycji między starą a nową siecią wybieram najbardziej istotne dla reprezentacji systemów wiedzy w internecie. Britannica Online ma tutaj zostać zastąpiona przez Wikipedię (jest to oczywiście bardziej przykład typu serwisu wypierającego inny niż rzeczywista zmiana organizacyjna). Strony osobiste znane z sieci 1.0 wybiera blogowanie. Spekulacja nazwami domen (wykupywanie i rezerwowanie adresów w sieci, które mogą okazać się zyskowne, ze względu

---

<sup>161</sup> W nieco prostszym ujęciu: niewykluczone, że widać tu hierarchię trzech poziomów kompetencji: między związanym zarówno z publicystyką, jak i akademią Steinerem, publicystą oraz przedsiębiorcą Keenem i wreszcie amatorami w sieci, którzy oddolnie rozbudowują swoje małe systemy i syntezy wiedzy. W każdej relacji poziomu wyższego do niższego napotkamy swego rodzaju resentyment lub krytykę.

<sup>162</sup> Tim O'Reilly, „What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software”, *Communications & Strategies* 65, nr 1st quarter 2007 (2007): 17–37.

na nazwę) ustępuje miejsca optymalizacji wyszukiwania (SEO, search engine optimization). Statystyki wyświetleń stron ma wyprzeć miara kosztu za kliknięcie. Screen scraping (przetworzenie i przygotowanie przez maszynę danych z innego programu w formie danych wyjściowych dla człowieka) zastępują usługi sieciowe. Publikowanie treści ma tutaj ustąpić partycypacji użytkowników w ich tworzeniu. Systemy zarządzania treścią (CMS, *content management systems*) zastępują narzędzia wiki (encyklopedie społeczne, często korzystające z takiej infrastruktury jak Wikipedia<sup>163</sup>). Katalogi (taksonomia) mają się zmienić w tagowanie (folksonomia). Z kolei „lepkość” treści (*stickiness*), czyli zdolność serwisu do przyciągnięcia i zatrzymania użytkownika na dłuższy czas wizyty, zostaje zastąpiona syndykacją (*syndication*), czyli przechowywaniem treści w jednym miejscu i wysyłaniem do wielu miejsc (stron) naraz, w oparciu o metryki korzystania z nich. Nie wszystkie z opisanych powyżej narzędzi, metod i praktyk doczekały się takiej pełnej transformacji. Zmiana, która zaszła, w wielu przypadkach (serwisy oparte na wiedzy, w tym encyklopedyczne, wykup domen, opozycja społecznych i eksperckich sposobów wytwarzania i opisywania zasobów wiedzy). Patrząc jednak na nie, możemy zauważyć, że częściowo żyjemy w przyszłości, którą opisał O’Reilly – używamy *Web 2.0*.

Jak przekonuje Marcin Wilkowski<sup>164</sup>, korzystanie z tego określenia niesie za sobą problemy co do chronologii i zakresu pojęcia. Istniały projekty o podanych przez O’Reilly’ego cechach jeszcze przed powstaniem tych, które opisywał jako *Web 2.0*, a intencje tego opisu miały charakter raczej biznesowy. Wilkowski dosłownie wskazuje, że działalność O’Reilly’ego w promocji nowego pojęcia pozwoliła przenieść koszty tworzenia treści z właścicieli portali na użytkowników (wystarczyło dostarczyć infrastrukturę, by ci obsłużyli siebie i innych sami za pomocą jej narzędzi). Aspekt demokratyzacji sieci przez korporacje zarządzające treściami pozostaje więc pod znakiem zapytania, a samo określenie bardziej tworzy rozpoznawalną markę parasolową, niż opis rzeczywistości (ze względu na wspomniany już anachronizm i sugestywne odwołanie do wersjonowania, które nie ma odzwierciedlenia w standaryzacji i technologii, a jedynie marketingowo i PR-owo da się je uzasadnić). Używam tego pojęcia zbiorczo i umownie, wykorzystując jego rozpoznawalność, natomiast nie bez zastrzeżeń. Tam, gdzie to możliwe, poza odwoływaniem się do porządku *Web 2.0* będę też mówić o treściach wytwarzanych przez użytkowników, serwisach i mediach społecznościowych, budowanych

---

<sup>163</sup> Jako przykład działającego do dziś projektu można wymienić zbiór fanowskich baz wiedzy, Fandom: zob. „Fandom”, <https://www.fandom.com/>. Dostęp: 2023-09-23.

<sup>164</sup> Marcin Wilkowski, „Nie potrzebujesz Web 2.0 w historii webu”, *Web96* (blog), 18 marzec 2023, <https://web96.pl/2023/03/18/web-2-0.html>. Dostęp 2023-09-16.

społecznie, czy społecznej sieci/*Webie*. Rama pojęciowa cech owych serwisów, którą przedstawił O'Reilly, wydaje się funkcjonalnie i poznawczo obiecująca, natomiast intencja, która stoi za prezentacją tych informacji (zachęcanie inwestorów wspierania internetu po tzw. kryzysie dot-comów w 2001), pozostaje wątpliwa dla badacza i warto o tym pamiętać.

Z kolei Manovich w 2002 roku porównał rewolucję nowych mediów do ruchów awangardowych w sztuce początku XX wieku, nazywając awangardę „oprogramowaniem” (*software*)<sup>165</sup>. Interfejsy i rozwiązania „wcieliły” i wprowadziły w rzeczywistość idee wizjonerów, które wyprzedzały swoje czasy pod kątem masowego wykorzystania w reprodukcji technicznej (choć aspekt wcielenia zapisujemy w cudzysłowie, wszak mowa o wirtualnych zasobach). Narzędzia cyfrowe umożliwiają kreowanie dowolnych i interaktywnych asamblaży, opartych na informacjach – tekście i materiałach graficznych. Wspólną cechą z ruchami awangardowymi jest też nacisk na kreatywne wykorzystanie typografii i to, co dzisiaj nazywamy identyfikacją wizualną, czyli zbiór cech graficznych, które powodują rozpoznawalność marki<sup>166</sup>. Ten wątek jest o tyle ciekawy, że moglibyśmy też mówić o programowaniu poprzez rozpoznawalność, tj. „wdrukowywaniu” odbiorcom obecności określonych sygnałów, dla wywoływania skojarzeń i odpowiedniej reakcji opartej na percepcji – tak jak to się dzieje, gdy w zupełnie obcym miejscu osoba rozpoznaje logotyp międzynarodowej sieci sklepów czy lokali *fast food* i rozważa skorzystanie z ich usług ze względu na znajomość tego symbolu. Wykraczamy tu jednak za daleko w stronę mechanizmów psychologii reklamy, więc podobnie jak w przypadku wiersza Świetlickiego, zawierającego lokowanie nazwy baru szybkiej obsługi, wypada zawrócić do głównego wątku<sup>167</sup>.

---

<sup>165</sup> Manovich, Lev. „Avant-garde as Software”. *Artnodes 0*, nr 2 (2002-12-01). <https://doi.org/10.7238/a.v0i2.681>.

<sup>166</sup> Marką może być w tym kontekście serwis internetowy, linia produktów danej firmy, przedsiębiorstwo, osoba (marka osobista danego autora książki czy celebryty), partia polityczna itp.

<sup>167</sup> Nie sposób wskazać w tym miejscu całej debaty na ten temat, ale pozostawiam garść odniesień związanych z estetyką. Badacze z dziedzin humanistycznych analizują związki sztuki z reklamą, jak „przechwytywanie” przez kapitalizm zjawisk ze świata twórczości artystycznej, relacje i przenikanie się tych dwóch światów w biografiami artystów, a wreszcie i nowe kategorie estetyczne, „oswajające” awangardę i odbierające jej subwersywne narzędzia. Możemy wskazać np. kontrast odważnych instalacji Kurta Schwittersa wobec jego projektów reklamowych, por. tekst opisujący wystawę poświęconą temu tematowi: José Juan Barba, „Kurt Schwitters. Avant-Garde and Advertising”, *Metalocus*, 2014-11-27, <https://www.metalocus.es/en/news/kurt-schwitters-avant-garde-and-advertising>. Relacje niezależności sztuki i zależności reklamy wobec rynku opisuje Jackson Lears, na sposób, który nazwałbym preposterystycznie pre-postmodernistycznym, tzn. ryzykownie optymistycznym co do autonomii sztuki (odniesienie do koncepcji Mieke Bal jest w tej aliteracji jest przypadkowe i niezamierzone). Zob. Jackson Lears, „Uneasy Courtship: Modern Art and Modern Advertising”, *American Quarterly* 39, nr 1 (1987): 133, <https://doi.org/10.2307/2712634>. Nowe kategorie estetyczne we współczesności (*zany, interesting, cute*) proponuje Sianne Ngai, w pewnym momencie przypisując w swojej książce kategorię *cute* (czegoś uroczonego, słodkiego) właśnie awangardzie. Por. Sianne Ngai, *Our Aesthetic*

W niniejszej pracy zajmujemy się jednakże informacjami, które są budowane przez tekst. Strona wizualna prezentowania informacji jest wtórna, tzn. nie zależy np. od typografii: krój pisma przy prezentowaniu bibliografii można dowolnie zmieniać, bez wpływu na cel, jaki ona spełnia. Mówimy też o danych, czyli swoistej abstrakcji, która posiada swoje ukonkretyzowane realizacje – o informacji, która podlega przechowywaniu, publikowaniu i wymianie w obrębie systemów.

### Porządkowanie pojęć

Zanim przejdziemy do tematu bibliograficznych danych badawczych, potrzeba zdefiniować najważniejsze pojęcia i określić stosunek humanistyki do danych na trzech głównych poziomach. Wyróżniam tu 1) ogólne zastosowanie w nauce, 2) szczegółowe w obrębie badań związanych z literaturą, a także 3) typy przetwarzanych informacji. Przede wszystkim, kiedy mowa w niniejszej pracy o danych badawczych, to takie dane są rozumiane jako efekt pracy naukowej, prezentowany szerszej społeczności w ramach dyscypliny. Takie udostępnienie oznacza możliwość dalszych przekształceń w kolejnych pracach, niekoniecznie związanych z oryginalną publikacją, czy wytworzeniem początkowego *datasetu*. Dane badawcze komunikują bezpośrednio wyniki opracowania przez twórców ich agregacji, a pośrednio również metodologię ich sporządzenia, w formie dokumentacji i metadanych.

Tego typu badania mogą posłużyć do określania autorstwa (lub prawdopodobieństwa tegoż), charakterystyki wydźwięku danej wypowiedzi, przemian stylu, grupowania tematów dzieł, jak i badań komparatystycznych. Widzieliśmy przykłady tego typu zastosowań metod obliczeniowych we wcześniej przywołanych pracach Bode, Jockersa, Morettiego i Pipera, różnorodnych, ale sięgających po stylometrię i pokrewne analizy.

### Dane badawcze

Najprostsza definicja operacyjna: dane badawcze to wszystko, co powstaje w wyniku badań<sup>168</sup>. Przyjmujemy dla potrzeb niniejszej pracy, że to wszelkie informacje (dane i metadane), które są zbierane, wytwarzane i przetwarzane w toku prac badawczych, a także publikowane jako wynik tychże prac. W ten sposób obieg się zamyka i raz opublikowane dane służą jako materiał do kolejnych badań.

---

*Categories: Zany, Cute, Interesting* (Cambridge, Massachusetts London, England: Harvard University Press, 2015).

<sup>168</sup> Wystąpienie Bożeny Bednarek-Michalskiej „Zarządzanie danymi badawczymi – informacje ogólne i plany zarządzania danymi” <https://www.youtube.com/watch?v=iopHvHKFD-A> (dostęp: 2020-03-01).



Cele udostępniania danych to m.in.: 1) transparentność – jeśli dane powstają w procesie publikacji wyników, ale nie są częścią ogłaszanych wniosków, to dostęp do takich częściowych informacji zwiększa wiarygodność badań; 2) możliwość powtórzenia badań przez innych naukowców – zgodnie z dobrymi praktykami społeczności naukowej, czyni to badania falsyfikowalnymi i reprodukowalnymi, wszędzie, gdzie jest taka możliwość<sup>169</sup>; 3) efektywność kosztów: nie duplikuje się badań dzięki korzystaniu z cudzych danych – wpływa to pozytywnie na skalowalność badań, kiedy każda kolejna inicjatywa naukowa związana z danym przedmiotem nadbudowuje zbiór wiedzy i rozwija możliwy do eksploracji zasób danych; 4) kontakty międzynarodowe – dzięki rozwiniętym środkom komunikacji w internecie, fakt publikacji (dobrze opisanych) danych pozwala dotrzeć do nich innym badaczom, z jakimi naukowiec-wytwórca oryginalnego *datasetu* nie miał do tej pory łączności, wyłącznie dzięki wspólnemu zainteresowaniu badawczemu; 5) popularyzacja badań w społeczeństwie – dotyczy to szczególnie takich form publikacji i przedstawiania badań, które angażują odbiorcę, zawierają element popularyzatorski, a nawet rozrywkowy (będą to bardziej takie części projektów, które stanowią osobną stronę internetową lub inny rodzaj „doświadczenia” w sieci, z aspektem narracyjnym, wyraźnie skierowane do odbiorców niezwiązanych z pracą naukową – w przeciwieństwie do publikowania danych w surowej formie, wymagającej dodatkowej pracy dla ich przeglądania, zrozumienia, interpretacji, obróbki oraz łączenia z innymi informacjami w kontekście).

Dane badawcze w literaturoznawstwie

Jeśli zostało powiedziane o „wszystkim co powstaje podczas badań”, moglibyśmy Skupimy się tutaj jednak W kontekście literaturoznawstwa możemy mówić o wielu kontekstach występowania takich inicjatyw. Przytoczę kilka najbardziej znanych. Dane badawcze w literaturoznawstwie chyba najbardziej są kojarzone ze stylometrią i analizą sentymentu. Jest to badanie rozpoznawalne, wykorzystujące korpusy tekstów. Łączy ono kompetencje językoznawcze z literaturoznawczymi (aby badać retorykę przekazu, trzeba wcześniej ujednolicienia analizy, np. opisu części mowy z podziałem na funkcje i wagi), a efekty są tyleż łatwe do zrozumienia, nawet w popularyzacji nauki, co intrygujące badawczo <sup>170</sup> .

---

<sup>169</sup> Możemy wyrazić uzasadnione wątpliwości w kwestii badań tradycyjnej humanistyki, która nie wytwarza danych; wymienione tutaj spostrzeżenia dotyczą pracy na danych badawczych, zatem znajdują zastosowanie do humanistyki cyfrowej i wszelkich badań inter- i transdyscyplinarnych, które wykracza poza typowe humanistyczne narzędzia, np. literaturoznawcze, włączające dane badawcze do swojego zakresu.

<sup>170</sup> Tutaj uwaga związana z relacjami humanistyki cyfrowej i genetyki tekstów: przygotowując opis stylometryczny, humanista cyfrowy posługuje się pewnym „ostatecznym” tekstem, tzn. jedną i jednolitą formą. Tymczasem jak wiemy np. z prac P-M. de Biasiego, żywiołem genetyki tekstów jest śledzenie przekształceń rękopisów pisarzy i innych „półproduktów” gotowego dzieła. O ile narzędzia cyfrowe zasadniczo umożliwiają

Mamy również do czynienia z makroanalizami i *distant reading*, co przedstawiłem w poprzednim rozdziale. Ciekawym kierunkiem są także badania związane z edytorstwem cyfrowym czy łączące aspekty historii literatury, edytorstwa i wizualizacji związanej z wersjonowaniem tekstu – jak badania Marka Troszyńskiego nad *Samuelem Zborowskim*<sup>171</sup>. Innym przykładem jest użycie danych i metadanych związanych z archiwizacją dorobku dziedzictwa i kultury, jak dane z dziedziny bibliografii.

Dane bibliograficzne

Dane bibliograficzne (również jako *bibliodata*) będą tu rozumiane jako bibliografia wyrażona poprzez dane, w tym dane badawcze. Statyczny wymiar bibliograficznego adresu i charakterystyki konkretnej publikacji zostaje zamieniony na zmediatyzowaną konstelację informacji, udostępnionych do przekształceń, zgodnie z przywoływanymi już wcześniej spostrzeżeniami Manovicha na temat modułowego charakteru zawartości *Webu*. Możemy też argumentować, że każda bibliografia stanowi opis za pomocą danych.

Dane połączone

Dane połączone to nie tylko *datasets* złożone z więcej niż jednego źródła. Przynajmniej jako *Linked Data* rozumie się takie dane, które można na zasadzie semantycznych połączeń parować ze względu na wspólne właściwości. Potrzebne są do tego rozwinięte formy opisu metadaneowego. Główne standardy promuje w związku z tym zagadnieniem organizacja W3C<sup>172</sup>. Dzięki zastosowaniu stałych identyfikatorów DOI („miejsca” zasobu w sieci), tagowania i metadanych, uniwersalnych identyfikatorów (jednoznacznych przypisań cechy np. kontroli autorytatywnej), słowników kontrolowanych i pracy nad konwersją formatów danych,

---

analizę wersjonowania tekstu i porównania takiej pracy w toku, zazwyczaj w badaniu stylometrycznym czy analizie sentymentu będziemy mieć do czynienia ze standaryzacją wszystkich części pod względem reguł zapisu i włączaniem do korpusu określonej stałej wersji danego tekstu. Możliwa jest digitalizacja kolejnych etapów powstawania dzieła, jednak każdorazowo wymagałaby prawdopodobnie stosownego oznaczenia przez osobę przygotowującą tego rodzaju zbiór – *dataset*. Wspominam o tym, ponieważ rysuje się ciekawy obszar badań, inspirowany metodami wypracowanymi na gruncie genetyki tekstów, nierzadko nawet znajdującej miejsca wspólne z psychoanalizą w badaniach literackich, prowadzący nas w stronę „genetyki tekstu cyfrowego”. Nasuwa się wszakże refleksja o przemianach we współczesnym powstawaniu tekstów literackich i szerzej tekstów kultury: nierzadko nie pozostawiają one po sobie śladu i jest to charakterystyczne dla używanych mediów. Ów „półprodukt”, badany przez genetyka tekstu, stale się sam nadpisuje, kiedy pisarz tworzy w obrębie jednego pliku tekstowego, w przeciwieństwie do np. XIX- i XX-wiecznej literatury, pomieszczonej w brulionach i notatkach autorów, dostępnych w archiwach. Taka „cyfrowa genetyka tekstów” stanęłaby przed zadaniem śledzenia kolejnych wersji tekstu już w zupełnie innej rzeczywistości medialnej. Por. Pierre-Marc de Biasi, *Genetyka tekstów*, tłum. Filip Kwiatek i Maria Prussak, *Filologia XXI 5* (Warszawa: Instytut Badań Literackich PAN. Wydawnictwo, 2015).

<sup>171</sup>Marek Troszyński i Juliusz Słowacki, *Alchemia rękopisu „Samuel Zborowski” Juliusza Słowackiego*, *Filologia XXI 11* (Warszawa: Instytut Badań Literackich PAN Wydawnictwo, 2017).

<sup>172</sup> Zob. „Semantic Web Standards”, [https://www.w3.org/2001/sw/wiki/Main\\_Page](https://www.w3.org/2001/sw/wiki/Main_Page). Dostęp: 2023-09-16.

możliwe jest przejście od zbierania danych do ich efektywnej analizy<sup>173</sup>. Używa się modelowania metadanych za pomocą grafów i tzw. „trójek” RDF (dla polonistów, podmiot-orzeczenie-dopełnienie, dla osób zorientowanych technicznie podmiot-predykat-obiekt).

Otwarte dane

Zasoby otwarte rozumiemy tu jako dostępne dla wszystkich<sup>174</sup>. Są to powszechne informacje, z ograniczonym właścicielstwem lub w pełni uwolnione – znajdują się w domenie publicznej. Otwartość danych ma związek z otwartą nauką. Użyteczną klasyfikacją będzie tutaj rozróżnienie na złotą i zieloną drogę otwartego dostępu do prac naukowych<sup>175</sup>. Złota oznacza pierwsze wydanie tekstu w wolnym dostępie. Zielona droga to wydanie w dostępie zamkniętym i wykorzystanie prawa autora do ponownej publikacji (samo-archiwizacji), jako *preprint* lub *post-print* (materiały sprzed oddania do druku lub recenzji oraz takie, które udostępnia się jako otwarte już po publikacji, ale mogą mieć drobne zmiany, np. formatowania).

Otwarte dane połączone

Na drodze syntezy dwóch ostatnich typów otrzymujemy z otwarcia *Linked Data: Linked Open Data*. Jest to koncepcja również związana z siecią semantyczną, otwarte dane stosują schemat RDF; użytkownicy mogą uzyskać dostęp do danych, a aspekt otwartości oznacza możliwość modyfikacji treści, bez wpływu na oryginalne źródło. Stwarza to globalny zasób wymiany informacji w całym internecie. Podstawową przesłanką do tworzenia takich systemów jest przeświadczenie, że dane mają tym większą wartość, im bardziej da się je łączyć z innymi zbiorami<sup>176</sup>.

Metadane

W największym skrócie, metadane są danymi na temat danych. Mogą dotyczyć użytych klasyfikacji i formatów czy proveniencji informacji. Przekazują informacje o okolicznościach zapisu i wersjach. Są to także odniesienia do treści materiałów, np. jako słowa kluczowe i klasyfikacje tematyczne, związane z architekturą informacji w serwisach i przyjętymi standardami opisu. Jak argumentuje Maciej Maryl, metadane są także swego rodzaju

---

<sup>173</sup> Eero Hyvönen, „Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery”, red. Pascal Hitzler i Krzysztof Janowicz, *Semantic Web* 11, nr 1 (2020-01-31): 187–93, <https://doi.org/10.3233/SW-190386>.

<sup>174</sup> Kontekst otwartości danych jest szerszy, por. „Otwarte dane i zasoby” w *Ministerstwo Cyfryzacji - Portal Gov.pl*, <https://www.gov.pl/web/cyfryzacja/otwarte-dane-publiczne>. Dostęp: 2023-09-23.

<sup>175</sup> Zob. „open-access.network: Open Access Green and Gold”, <https://open-access.network/en/information/open-access-primers/green-and-gold>. Dostęp: 2023-09-23.

<sup>176</sup> Zob. „Linked Open Data - W3C eGovernment Wiki”, [https://www.w3.org/egov/wiki/Linked\\_Open\\_Data](https://www.w3.org/egov/wiki/Linked_Open_Data). Dostęp: 2023-09-23.

subkodem, tworzącym opowieści, również prywatne<sup>177</sup> - czy to jako identyfikator sieci, z którą się łączymy, czy wskazując na formy organizacji wiedzy (prywatne i publiczne). Zapewniają kontekst informacji.

Dane dotyczące cytowań

Same dane dotyczące cytowań prac, nawet pozbawione treści, stanowią użyteczny zbiór informacji. Jak argumentują Silvio Peroni i jego współpracownicy, cytowania powinny zawsze być publikowane w otwartym dostępie i uznane za *commons* – dobro publiczne<sup>178</sup>. Stworzyli taki zasób na podstawie informacji dostępnych w serwisie PubMed i opublikowali w wolnym dostępie jako otwarte dane połączone. W ten sposób inni badacze, wydawcy i instytucje mogą korzystać ze zbioru (używać, rozbudowywać go), w dowolnym celu, bez zastrzeżeń prawnych.

Zespół humanistów cyfrowych z Bolonii opisał także temat cytowań pod względem technologii <sup>179</sup> . Porównano narzędzia ekstrakcji informacji bibliograficznych z pełnotekstowych plików PDF z artykułami naukowymi. Jest to przykład działania na dostępnym materiale, które oferuje ciekawe efekty. Cytowania są oczywiście wykorzystywane w mierzeniu wpływu i badaniach bibliometrycznych.

Peroni i David Shotton stoją na czele Open Citations<sup>180</sup>, organizacji *non-profit*, której celem jest udostępnianie danych bibliograficznych jako LOD dla lepszego przepływu informacji w komunikacji naukowej, co przekłada się na realizację celów otwartej nauki.

Zasady FAIR w udostępnianiu danych

Ramą, którą się posłużyć przy opisie są zasady *FAIR Data (FAIR DATA Principles)*. Zaleca je m.in. Komisja Europejska dla programu badań i innowacji Horizon 2020<sup>181</sup>. Akronim określa cztery podstawowe czynniki otwartości danych: *Findable*, czyli możliwe do odszukania; *Accessible*, to znaczy dostępne; *Interoperable* – interoperowalne (dające się przetwarzać);

---

<sup>177</sup> Maciej Maryl, „Subnarracje metadanych”, *Teksty Drugie*, nr 3 (2014): 7–11.

<sup>178</sup> Peroni, Silvio, Alexander Dutton, Tanya Gray, i David Shotton. „Setting Our Bibliographic References Free: Towards Open Citation Data”. *Journal of Documentation* 71, nr 2 (9 marzec 2015): 253–77. <https://doi.org/10.1108/JD-12-2013-0166>.

<sup>179</sup> Alessia Cioffi i Silvio Peroni, „Structured References from PDF Articles: Assessing the Tools for Bibliographic Reference Extraction and Parsing”, w *Linking Theory and Practice of Digital Libraries*, red. Gianmaria Silvello i in., t. 13541, Lecture Notes in Computer Science (Cham: Springer International Publishing, 2022), 425–32, [https://doi.org/10.1007/978-3-031-16802-4\\_42](https://doi.org/10.1007/978-3-031-16802-4_42).

<sup>180</sup> Zob. „OpenCitations – Home”, <https://opencitations.net/>. Dostęp: 2023-09-23.

<sup>181</sup> Trwającego w latach 2014-2020 i kontynuowanego obecnie przez inicjatywę Horizon Europe do 2027.

*Reusable* (również jako: *Re-usable*) – dające się wykorzystać ponownie. Grupa, z Markiem D. Wilkinsonem jako głównym autorem, opublikowała te wskazówki w 2016<sup>182</sup>.

Z jednej strony nie jest to sformalizowany zbiór reguł, który twórcy narzucają jako badawczy kanon – z drugiej, opracowano te zasady na podstawie wieloletnich doświadczeń i z myślą o zrównoważonym rozwoju, a pożytek z korzystania z *FAIR* na szeroką skalę legitymizuje wspomniane użycie przez Komisję Europejską. Warto też podkreślić, że twórcy tego podejścia budowali na poprzednich doświadczeniach i rozwijali inne koncepcje. *FAIR* odnosi się do rozwiązań wypracowanych na gruncie Concept Web Alliance, *JDDCP* (*Joint Declaration of Data Citation Principles*), *DSA* (*Data Seal of Approval*). Dążenie ku spójnemu, dostępnemu szeroko i dogodnemu w użyciu dla wszystkich ekosystemu wymiany danych naukowych uwzględnia inkluzywne podejście również do metod opisu tejże wymiany.

Głównym problemem, który ma rozwiązać jest istotna potrzeba wymiany i ponownego użycia danych badawczych, co wymaga polepszenia infrastruktury dla nich. Stąd pojawienie się spójnych i mierzalnych wskaźników. Zarządzanie danymi (*Data Management*) w sprawny sposób daje wsparcie nowych badań i nowych odkryć, dzięki przyspieszeniu procesu wytwarzania wiedzy. Istotny staje się kontekst maszyn, które będą odczytywać te dane.

Przejdę teraz do szczegółowego opisu tych cech na podstawie listy proponowanej przez twórców *FAIR*. Każda cecha posiada cztery charakterystyki (trzy w przypadku interoperacyjności). Poniżej tłumaczenie punktów:

Aby dane dało się odszukać (*Findable*):

F1. (meta)dane mają przypisane globalnie unikatowe i stałe identyfikatory

F2. dane są opisane przez bogate metadane (zdefiniowane w R1 poniżej)

F3. metadane w czytelny i wyraźny sposób zawierają identyfikator danych, do których się odnoszą

F4. (meta)dane są rejestrowane lub indeksowane w zasobie pozwalającym na przeszukiwanie

---

<sup>182</sup> Mark D. Wilkinson i in., „The FAIR Guiding Principles for Scientific Data Management and Stewardship”, *Scientific Data* 3, nr 1 (grudzień 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

Aby były dostępne (*Accessible*):

A1. (meta)dane dają się odnaleźć na podstawie swojego identyfikatora przy użyciu standardowego protokołu komunikacji

A1.1 protokół jest otwarty, darmowy i uniwersalnie dostępny do implementacji

A1.2 protokół pozwala na procedurę autoryzacji, gdy to konieczne

A2. metadane są dostępne, nawet kiedy dane już nie są dostępne

Aby były interoperowalne (*Interoperable*):

I1. (meta)dane są opisane przy użyciu dostępnego (*accessible*), udostępnionego (*shared*) i nadającego się do szerokich zastosowań języka dla reprezentacji wiedzy.

I2. (meta)dane używają słowników, które stosują się do zasad *FAIR*

I3. (meta)dane zawierają kwalifikowane odniesienia do innych (meta)danych

Aby były możliwe do ponownego użycia (*Reusable*):

R1. (meta)dane<sup>183</sup> są bogato opisane poprzez wiele dokładnych i adekwatnych (*relevant*) atrybutów

R1.1. (meta)dane są oddane do użytku z czytelną i dostępną licencją użycia danych

R1.2 (meta)dane są połączone ze szczegółową proveniencją

---

<sup>183</sup> Oryginalna publikacja w tym miejscu czyni wyjątek i zamiast konsekwentnego użycia „(meta)data” stosuje zapis „meta(data)”, co pozwoliłem sobie poprawić.

R1.3. (meta)dane spełniają standardy społeczności, adekwatne (*relevant*) dla danej dziedziny<sup>184</sup>.

Obieg wymiany danych bibliograficznych

W humanistyce ontologie pomagają wejść na wyższy poziom organizacji wiedzy pod względem klasyfikacji i przyporządkowań tematycznych. Coraz istotniejsze będą narzędzia wykorzystujące metadane w celu wyszukiwania, identyfikacji i klasyfikacji cyfrowych obiektów (możemy to porównać do ustawiania przybliżenia w urządzeniach optycznych – od generalizacji w metadanych do informacji w danych). Przykładem takiego projektu, związanego z rodzimą humanistyką cyfrową jest platforma GoTriple, agregująca metadane. Przydatną cechą jest agnostyczny językowo sposób porównywania informacji (zapewniają to uporządkowane metadane i stosowanie uniwersalnych formatów)<sup>185</sup>.

Raport twórców tego projektu, grupy roboczej TRIPLE (Transforming Research through Innovative Practices for Linked interdisciplinary Exploration )<sup>186</sup> zawiera dobre praktyki w kontekście metadanych jako danych badawczych. Główne wskazówki dotyczą: stałych identyfikatorów, słów kluczowych, abstraktów i cytowań. Pod względem stałych identyfikatorów wskazuje na różnorodność źródeł i potrzebę konsolidacji właśnie za ich pomocą (muszą być możliwe do odczytania przez maszyny). Kolejnym aspektem jest wielojęzyczność (obecna w międzynarodowym środowisku naukowym), a także różnorodność, nazwana *bibliodiversity*: różne organizacje tworzą i przetwarzają dane, robiąc to według własnych metodologii i kryteriów. Stąd powstające ontologie powinny uwzględniać łączliwość z tymi systemami. Występuje także potrzeba agregacji mniejszych źródeł i kompleksowe podejście do całości. Wiąże się to także z rozproszeniem tych „miejsc” produkcji i przechowywania danych produkowanych przez humanistyczne projekty: poziom analizy powinien uwzględniać tę wielość.

Z kolei ustalenia na temat krajobrazu danych bibliograficznych grupy roboczej DARIAH<sup>187</sup> ukazuje dwie ważne opozycje w podejściu do obiegu *bibliodata*: publicznie-prywatni interesariusze oraz wytwarzanie-użycie danych. Wśród publicznych podmiotów możemy

---

<sup>184</sup> Wilkinson i in., „The FAIR Guiding Principles for Scientific Data Management and Stewardship”.

<sup>185</sup> Dostępna pod adresem <https://gotriple.eu/>.

<sup>186</sup> Umerle, Tomasz i in., „TRIPLE Deliverable 8.5: Guidelines on the Research Data in the Humanities”, 2023-01-20, <https://doi.org/10.5281/ZENODO.7568873>.

<sup>187</sup> Tolonen, Mikko i Umerle, Tomasz, „An Analysis of the Current Bibliographical Data Landscape in the Humanities. A Case for the Joint Bibliodata Agendas of Public Stakeholders - Video Presentation”, 2022-10-11, <https://doi.org/10.5281/ZENODO.7118150>.

wymienić instytucje dokumentujące dziedzictwo czy uniwersytety. Prywatnymi są w tym kontekście np. przedsiębiorstwa zajmujące się utrzymywaniem komercyjnych bibliograficznych baz danych.

Different business and data-ownership models create a natural division between public and private sector. Despite this division, these two types of stakeholders are deeply interdependent. At the same time, they are subject to internal divisions: this includes the split between libraries, archives, and museums (LAM) and research institutions in the public sector and the competition between smaller companies and the consolidating forces of corporations in the private one<sup>188</sup>.

Twórcy raportu pokazują nam trzy tendencje: zwiększony napływ *bibliodata* (więcej publikacji i metadanych); automatyzacja przetwarzania tych danych (wytwarzanie, czyszczenie, wzbogacanie i łączenie, udostępnianie); badania oparte na danych (rozwój narzędzi otwiera możliwości takiego ich zastosowania). Diagnozą jest tutaj wspólny interes publicznych interesariuszy w kultywowaniu budowania sieci swobodnego dostępu do *bibliodata* i wymiany wiedzy, rekomendacją – dalsze działania w kierunku użycia danych bibliograficznych jako materiału do poszukiwań wniosków na temat społeczeństwa, kultury czy historii.

Uzgadnianie formatów

Dane PBL przez długi czas cechował własny format danych (*proprietary data format*). Baza wyrasta z metodyki opracowywania monografii, a ta była zrodzona konkretnymi potrzebami i zadaniami, stojącymi przez PBB – nie wspominając o realiach historycznych i rozwoju projektu przed cyfryzacją. Użycie własnego formatu danych (*proprietary data format*) nie jest zalecane przez specjalistów zarządzania danymi<sup>189</sup>. Tak w PBL, jak i w innych projektach widać jednak tendencje do łączenia różnych zbiorów.

Użycie powszechnego formatu zwiększa interoperacyjność danych<sup>190</sup>, co jest elementem standardu FAIR. Dzięki zastosowaniu takiej metody przedstawienia i przechowywania

---

<sup>188</sup> Tolonen, Mikko i Umerle, Tomasz.

<sup>189</sup> W zbiorze dobrych praktyk autorstwa zespołu Bibliotek Uniwersytetu Stanforda odradza się stosowanie własnego formatu danych: „Best practices for file formats | Stanford Libraries”, dostęp 2019-11-13, <https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats>.

<sup>190</sup> Zob. Wojciech Fenrich, „Selekcja-i-przygotowanie-danych-badawczych-do-udostępnienia.pdf”, Uniwersytet Warszawski Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, 2019-12-03, <https://drodb.icm.edu.pl/wp-content/uploads/2019/10/Selekcja-i-przygotowanie-danych-badawczych-do-udost%C4%99pnienia.pdf>.



zawartości, aby maszyny mogły je odczytać, rozszerza się zakres użycia owej zawartości. Dane jako przedmiot analiz stają się możliwe do odczytania przez popularne i niewymagające wykupu licencji narzędzia oraz można łączyć te dane z innymi zbiorami, w celu przeprowadzania dokładniejszych porównań.

Ten aspekt interoperacyjności jest właściwie kluczowy, aby baza danych, gromadząca dziedzictwo kulturowe, mogła zaistnieć na szerszej scenie. Przykładowo, jeśli użyty zostanie format MARC 21 lub Dublin Core, jest większa szansa, że przedstawiciele innych instytucji i ośrodków badawczych wykorzystają te dane (ze względu na popularność zastosowania tych dwóch formatów). Powszechne formaty danych ułatwiają także komunikowanie na zewnątrz ze względu na możliwość przetwarzania danych przez maszynę.

Przypadkami, w których warto rozważyć użycie własnego formatu danych lub połączenia formatów otwartego i własnego, są ryzyka utraty informacji, np. na skutek konwersji. Kiedy używa się własnego formatu danych, istotne staje się zapewnienie odpowiedniej dokumentacji. Kluczowymi informacjami w niej zawartymi są nazwa i producent oprogramowania oraz wersja programu użytego do utworzenia danych. Dokumentacja powinna być łatwo dostępna, np. jako plik Readme.txt. W podanym przykładzie nazwa jest konwencjonalna i intuicyjnie rozpoznawalna jako instrukcja, a rozszerzenie pliku uniwersalnie dostępne do użycia w najpopularniejszych systemach operacyjnych.

Dokumentacja służy odbiorcom danych, dostarczając instrukcji postępowania. Pomaga także wdrażać nowych pracowników oraz zapewnia przechowanie wiedzy obecnych specjalistów, rozwijających bazę. Dzięki niej zespół ma pewność odtwarzania rezultatów oraz możliwość wyjaśniania, jak utworzono dane.

Bazy indeksowania czasopism a przypadek PBL

Aneta Drabek opracowała poradnik dla wydawców czasopism; nakreśliła w nim kontekst tego, jak ważna jest widoczność badań naukowych<sup>191</sup> i ukazuje wagę bibliograficznych baz danych

---

<sup>191</sup> „Najważniejszym celem działań każdego uczonego jest nie tylko przeprowadzenie badań, lecz także wprowadzenie wynikających z nich koncepcji i rozwiązań metodycznych do międzynarodowego obiegu informacji, tak aby inni mieli możliwość się z nimi zapoznać. To wprowadzenie odbywa się przede wszystkim poprzez opublikowanie wyników, czyli zaprezentowanie ich w czasopismach naukowych lub monografiach, czy też przedstawienie na konferencjach. Dopiero wtedy możemy mówić o prawdziwym włączeniu w obieg naukowy. Jednakże jest to dopiero pierwszy etap drogi, która prowadzi od autora do czytelnika. Sukces w działalności badawczej polega na tym, że publikowane prace są czytane, a wyniki twórczo wykorzystywane przez innych naukowców. Tak więc autorowi powinno zależeć nie tylko na tym, aby jego tekst w ogóle się ukazał, ale też żeby miał szansę dotrzeć do osób, które mogłyby być zainteresowane opublikowanymi w nim treściami”. Aneta Drabek, „Indeksowanie czasopism w referencyjnych bazach danych”, 2018, <https://doi.org/10.6084/M9.FIGSHARE.5683972>, 5.

w dostępie do informacji, wskazując na wielość dostępnych tytułów publikacji naukowych<sup>192</sup>. Pod względem metadanych, autorka porusza kwestię istotności stałych identyfikatorów – w tym kontekście dwoma najważniejszymi, które Drabek zauważa jako pojawiające się coraz częściej, będą identyfikator indywidualny autora, czyli ORCID (Open Researcher and Contributor ID) oraz DOI – *Digital Object Identifier* (adres w internecie, który pozwala dotrzeć do konkretnej publikacji).

Autorka przedstawia następujące korzyści z rejestracji czasopisma w referencyjnej bazie danych<sup>193</sup>:

- wyszukiwanie literatury naukowej – nie tylko jako forma przygotowania do badań, gdzie atut stanowi przeglądanie treści pod kątem tematu, ale także pozyskania metadanych do bibliografii własnej pracy badawczej;
- ocena czasopisma i jego rekomendacja – w przypadku niektórych baz zgłoszenie do niej tytułu wiąże się z wieloetapową weryfikacją w odwołaniu do eksperckich standardów, a włączenie czasopisma do nich zwiększa prestiż (Drabek podaje Web of Science Core Collection i Scopus jako przykłady takich źródeł bibliograficznych);
- wiarygodność danych – profesjonalne przygotowanie zawartości baz najczęściej rozkłada się na kilka etapów, a specjaliści wprowadzają dane z autopsji; autorka zwraca też uwagę na stały dostęp do rekordów jako wartość (odbiorca może powracać do sprawdzonych informacji i ma gwarancję dostępności, ponieważ nie usuwa się zapisów w bazach);
- cytowania – umieszczenie w bazach bibliografii załącznikowych artykułów, co daje wgląd w odwołania i podstawę wyliczeń współczynników impaktu (dodajmy, że poza wskaźnikami oceny nauki, pozwala to również na analizy sieciowe między artykułami, związane z przedmiotem badań, a nie działalnością naukową autorów tekstów), jak i odkrywa dla badaczy potencjalnie nowe, interesujące publikacje do uwzględnienia we własnych zbiorach źródeł;

---

<sup>192</sup> „(...)zalew naukowych publikacji spowodował, że coraz większą rolę zaczynają odgrywać źródła, które ułatwiają wyszukiwanie literatury na dany temat, ale także spełniają rolę sita, zbierając tylko wartościowe publikacje naukowe i pomijając jednocześnie te nienaukowe czy też słabej jakości. To sprawia, że poszukujący informacji może szybko i sprawnie znaleźć potrzebne mu dane. Taką rolę pełnią we współczesnej nauce bibliograficzne bazy danych”. Cyt. za: Drabek, 5. Dalej autorka zwraca też uwagę na rolę bibliograficznych baz danych jako banków informacji, służących do ewaluacji nauki.

<sup>193</sup> Drabek, s. 9-16.

- rozbudowane rekordy bibliograficzne – „Wiele baz danych przygotowuje szczegółowe informacje o każdej indeksowanej publikacji, często uzupełniając je o elementy, których nie ma w czasopiśmie (np. słowa kluczowe, hasła tematyczne, klasyfikację, język publikacji, abstrakt, afiliację, linki do pełnych tekstów i inne), które znacząco zwiększają możliwości wyszukiwawcze. Wyniki są bardziej precyzyjne, co może wpłynąć na szybsze dotarcie do bardziej adekwatnych tekstów” pisze autorka poradnika<sup>194</sup>;
- pełne teksty w bazach danych – niektóre źródła oferują dostęp do całej treści, np. w formie plików do pobrania (może to być również odczyt *online*); Drabek zwraca tu ponownie uwagę na wiarygodność rozwiązania (pewność pobrania przez odbiorcę odpowiedniej wersji tekstu); widoczność w internecie przekłada się też tutaj na szanse samego autora danej publikacji na zacytowanie treści; z kolei z perspektywy wydawcy czasopisma jest to często dobre rozwiązanie, dzięki otrzymaniu wsparcia przy utrzymaniu dostępu pełnotekstowego przez podmiot zarządzający bazą, opis metadanych, wsparciu technicznemu, a także możliwej retrodigitalizacji zbiorów dostępnych wcześniej wyłącznie w formie papierowej;
- informacja o czasopiśmie – istotne gdy czasopismo nie posiada swojej strony w internecie lub jakiegokolwiek oficjalnego miejsca w sieci, które zwiększa jego widoczność; może także służyć jako uzupełnienie informacji, jeśli dana publikacja funkcjonuje w internecie, ale zyskuje w ten sposób dodatkowy punkt dostępu;
- prezentacja w internecie – profil czasopisma w bibliograficznej bazie danych może zawierać dane do kontaktu, informacje o wydawcy, punktacji i wskaźnikach bibliometrycznych, czy wspomniane wcześniej strony internetowe pisma – Drabek podkreśla, że jednolity układ takich baz może sprzyjać łatwiejszemu znalezieniu pożądaných informacji (strony poszczególnych czasopism mogą się różnić między sobą, a interfejs bazy bibliograficznej przekazuje te same informacje w stały sposób); oprócz aspektu wymiany wiedzy mamy tutaj także do czynienia z walorem promocyjnym dla danego tytułu;
- lepsza widoczność czasopisma w internecie – przy tym punkcie Drabek odwołuje się do optymalizacji wyszukiwarek internetowych w kontekście naukowym (akademickie pozycjonowanie, *Academic SEO*): praktyka wielu użytkowników sprowadza się do użycia wyszukiwarek naukowych, szczególnie cieszącego się największą

---

<sup>194</sup> Drabek, s. 11.

popularnością Google Scholar – umieszczenie danych w bazach bibliograficznych bezpośrednio przekłada się na dostęp tą metodą i większą ekspozycję odbiorcy na treści;

- statystyki – informacja zwrotna z takiego serwisu może wspierać decyzje redakcji i wydawcy pisma; przykładem może być liczba pobrań lub wyświetleń artykułu; autorka poradnika zauważa, że dzięki śledzeniu tych informacji można przejść do „zbadania aktualnych trendów w nauce, mogą także pomóc w kształtowaniu polityki naukowej danego periodyku, podejmowaniu decyzji o zaproszeniu do współpracy wybranych autorów lub też o przekształceniu czasopisma (połączeniu, rozdzieleniu na serie tematyczne) czy nawet zmianie profilu tematycznego”<sup>195</sup>;
- promocja autorów – dzięki generowaniu wykazu publikacji, baza bibliograficzna może promować dorobek badacza (tu dodajmy także korzyści dla promowania określonego tematu czy twórcy – również w zakresie przedmiotu badań ma to znaczenie);
- promocja instytucji – dane o afiliacji autorów przekładają się na wkład do rankingów uczelni;
- wzajemna rekomendacja – bazy zawierają listy współpracujących czasopism, a strony publikacji w konkretnym miejscu odnoszą się do indeksowania w bazach;
- możliwość wykonywania wielowymiarowych analiz – Drabek pisze: „Bibliograficzne bazy danych stały się podstawą większości badań biblio- i nauko-metrycznych. Dzięki zgromadzeniu tysięcy czy nawet milionów rekordów możliwe stało się prognozowanie rozwoju nauki, rozpoznawanie wzorów publikowania w dziedzinach naukowych, ocena jednostek, badaczy czy czasopism”<sup>196</sup> – do tego punktu warto nadmienić, że, zgodnie z przyświecającym nam celem, te same dane mogą zostać użyte do badań w dziedzinach humanistycznych, tzn. posłużyć również do refleksji nad systemem literackim i innych inicjatyw naukowych; jest to domeną projektów humanistyki cyfrowej, która stara się objąć wspomniany wolumen danych i formułować wnioski na podstawie jego analiz;
- Punktacja czasopism naukowych – indeksowanie w bazach bibliograficznych przekłada się na prestiż i wyższą punktację czasopism (ten aspekt nie będzie istotny dla dalszych rozważań);

---

<sup>195</sup> Drabek, s. 14.

<sup>196</sup> Drabek, s. 15.

- Integracja z multiwyszukiwarkami – narzędzia, które przekazują zapytanie do wielu źródeł naraz stają się nieodzowne w pracy badawczej; multiwyszukiwarki prezentują ujednolicony wynik ze zbiorczej kwerendy połączonych źródeł; przy tym punkcie Drabek zauważa: „Czasopismo nieindeksowane w żadnej bazie pozostaje faktycznie poza obiegiem naukowym, a potencjalni zainteresowani (zwłaszcza ci, którzy korzystają tylko z baz danych i bibliografii) mogą nigdy się o nim nie dowiedzieć. Wydawcy muszą dbać o swój wizerunek na rynku, co ułatwiają im bazy danych, ponieważ indeksując czasopisma, stają się doskonałym i darmowym miejscem promocji. Obecność danego periodyku w bazie danych to ważny element promocji samego czasopisma, ale także autorów w nim piszących, treści oraz samego wydawcy. Warto podjąć trud, aby jego zawartość oraz najważniejsze informacje znalazły się w renomowanych źródłach. Włożona w przygotowanie odpowiednich danych praca na pewno przyniesie pożądane efekty w postaci zwiększenia widoczności opublikowanych tekstów, poszerzenia grupy potencjalnych odbiorców, jak również podwyższenia szansy na uzyskanie cytowań”<sup>197</sup> – jak w przypadku poprzednich punktów, ową widoczność w świecie nauki chcielibyśmy zastąpić optyką dostępności bibliograficznych danych badawczych, tzn. podkreślić ten sam cel cząstkowy (otwartość, możliwość odnalezienia i przetwarzania informacji), ale zaakcentować inny cel końcowy (na dobrą sprawę dane stanowią cel sam w sobie i nie musimy się odnosić do treści).

Rozwijając ostatni punkt: oznacza to dla nas, że nie liczy się „głębokość” ruchu od danych do konkretnego źródła (od pojedynczego zapisu, przez treść, aż do powołania się na konkretny tekst naukowy), ale raczej zgromadzenie jak najlepiej opisanego metadanymi i jak najszerszego zbioru informacji o *tekście*, który następnie możemy dowolnie przetwarzać.

Część poradnika związana z podstawowymi informacjami na temat baz danych opisuje także bazy bibliograficzne. PBL pojawia się tutaj wśród innych źródeł, a wybrane zostają omówione w aneksie<sup>198</sup>. Projekt PBB IBL nie wszedł akurat do tej grupy opisanej bardziej szczegółowo, jak można się domyślać z uwagi na profil tekstu: nie jest to *stricte* baza mająca na celu

---

<sup>197</sup> Drabek, s.16.

<sup>198</sup> „Wśród baz bibliograficznych można wymienić: Philosopher’s Index, MLA International Bibliography, Polską Bibliografię Literacką, Bibliografię Historii Polski, Polską Bibliografię Prawniczą, Polską Bibliografię Lekarską, BazHum, iSybisław, AGRO, Biological Abstracts, GeoRef, MathSciNet, Inspec, DBLP Computer Science Bibliography i wiele innych.” Drabek, s.20-21.

zwiększenie widoczności publikacji naukowych w bieżącym obiegu, a raczej zasób dokumentujący wybrany zakres piśmiennictwa i produkcji kulturalnej w sposób ciągły.

W odwołaniu do wymienionych powyżej korzyści, warto pamiętać o tej różnicy w zakresie działania PBL od bibliograficznych baz skoncentrowanych na wymianie informacji w nauce. Z pewnością to, co projekt IBL ma do zaoferowania, to możliwość wyszukania literatury na dany temat (zarówno w kluczu haseł przedmiotowych, jak i autorskim), wiarygodne i rzetelne dane, opracowywane z autopsji przez zespół PBB (a w ostatnich latach ze wsparciem narzędzi cyfrowych), stałość dostępu do źródła w formie cyfrowej. Atutem są także rozbudowane opisy bibliograficzne, szczególnie w obszarze adnotacji, które odwołują się do treści. Ważne są w tym aspekcie także klasyfikacje wytworzone przez zespół PBB. Na pewno nie możemy w przypadku PBL mówić o dostępie pełnotekstowym czy aspektach promocji, w tym wzajemnych rekomendacji z „dostawcami” źródeł.

Jest za to kilka kwestii, które możemy przeanalizować jako drogi rozwoju dla PBL. Pod względem widoczności czy prezentacji czasopism i informacji o nich, serwis bibliograficzny rozwijany przez PBB w istocie dodaje kolejne miejsce w sieci, ale, nie ma potwierdzonych informacji, by zwiększał znacząco zasięg dotarcia źródeł bieżących do odbiorcy. Ma za to duży udział w aspekcie dokumentalistycznym. PBL nie udostępnia również na stronie statystyk, choć mogłoby to być interesujące, zarówno dla wydawców i innych uczestników rynku wydawniczego, jak i dla odbiorców akademickich bądź prywatnych. W dalszej części pracy, ściśle związanej z działalnością PBL, powołuję się na taki raport z użyciem narzędzia Google Analytics. Szerszy dostęp do niektórych zawartych w nim danych byłby na pewno ciekawy dla osób korzystających z bazy. Dostęp do danych to cecha modernizowanej strony PBL. To ona będzie umożliwiać również analizy wielowymiarowe. Nowy serwis zawiera takie odniesienia do strony IBL, wskazuje partnerów: Centrum Humanistyki Cyfrowej i Poznańskie Centrum Superkomputerowo-Sieciowe. Szerzej odnoszę się do kwestii rozwoju w rozdziale kolejnym.

Poza zakresem działalności PBL będą takie aspekty jak ocena i punktacja czasopism, informacje o cytowaniach (mimo bogatych informacjach o temacie dzieła, którego dotyczy zapis). Nie dotarłem do informacji o uwzględnieniu bazy bibliograficznej rozwijanej przez PBB w multiwyszukiwarkach. Pojawia się jako rekomendowane źródło na stronach uczelni wyższych wśród zasobów wiedzy. Dane PBL są także częścią kilku serwisów, w tym międzynarodowych, opartych na łączeniu danych z wielu zbiorów, co ułatwia wyszukiwanie w nich wszystkich jednocześnie.

Otwarte licencje na udostępnianie danych

Najpopularniejszym narzędziem oznaczania licencji jest zbiór *Creative Commons*<sup>199</sup>. Nie są to części przepisów, ani organizacja prawnicza, tylko inicjatywa społeczna. Każde użycie licencji CC jest jednak deklaracją twórcy na temat dozwolonego użytku, na jaki się godzi, w świetle praw autorskich i wykorzystania. To popularne narzędzie w działalności kreatywnej (muzyka, literatura, film, inne), ale także w działalności naukowej i aktywistycznej. Istnieją specjalne wyszukiwarki treści do ponownego użycia na wolnej licencji, a także ustawienia zaawansowane najważniejszych wyszukiwarek internetowych, które uwzględniają tylko takie wyniki.

Dla danych badawczych najlepsze są takie licencje, które są najbardziej otwarte, aby inni naukowcy mogli swobodnie z nich korzystać. Oczywiście można też wybierać licencje np. z obowiązkową atrybucją autorstwa (CC-BY), dla uznania wysiłków twórcy *datasetu* i w trosce o cytowania.

W lipcu 2015 roku Biblioteka Narodowa Niemiec udostępniła swoje rekordy na wolnej licencji o najbardziej otwartym charakterze, przekazując te informacje do domeny publicznej<sup>200</sup>. Zawartość obejmowała około 11 milionów danych kontroli autorytatywnej i ponad 14 milionów zapisów bibliograficznych. Dane udostępniono w powszechnie dostępnych serwisach, także do zautomatyzowanego pobierania, po uprzednim zalogowaniu do tych serwisów (użytek jest darmowy). Wymienione zostały usługi Data Shop, SRU i OAI. Intencją tej zmiany było umożliwienie korzystania z bardziej masowych rozwiązań przy użyciu metadanych bibliograficznych, zamiast pojedynczych wyszukiwań za pomocą interfejsu. Odbiorcy mogą łatwiej konfigurować zapytania, aby uzyskiwać pożądane informacje.

Władze DNB miały trojaki powody takiej zmiany modelu biznesowego. Z perspektywy polityczno-ekonomicznej, oceniono tę inicjatywę jako korzystną przez statutową działalność na rzecz dobra publicznego. Otwartość była już wcześniej cechą działania Biblioteki Narodowej Niemiec, a zadeklarowanie licencji CC0 1.0 umożliwiło dostosowanie się do tego, jak działa wymiana informacji w internecie. Dzięki temu możliwe jest budowanie zasobów danych połączonych i sieci semantycznych. Poza realizowaniem swojej misji, organizatorzy

---

<sup>199</sup> Zob. „Creative Commons”, <https://creativecommons.org/>. Dostęp: 2023-09-16.

<sup>200</sup> Meyer-Heß A., Rupp J., *What happens if you publish the National Bibliography under a CC0 license? – Experiences of the German National Library (DNB)*, dokument dostępny online: <http://library.ifla.org/1414/1/210-meyer-en.pdf> (dostęp 2020-01-04). Komunikat prasowy dostępny pod adresem: <https://www.dnb.de/EN/Ueber-uns/Presse/ArchivPM2015/metadatenCC0.html>.

przedsięwzięcia liczyli także na promocję, tzn. zwiększenie liczby odwiedzin i użytkowników korzystających z serwisu.

Pojawiły się także powody związane z danymi. Już wcześniej dane były one dostępne; pobierano natomiast opłatę związaną z ponownym ich użyciem, argumentowaną wewnętrznym zapewnieniem kontroli jakości (standard katalogowania). W miarę rozwoju bazy danych, część informacji zaczęła być wszakże pozyskiwana z zewnętrznych źródeł, a nie wytwarzana w ramach działalności DNB. Obecnie biblioteka zapewnia normalizację i standaryzację w ramach formatu MARC 21 do ponownego nieodpłatnego wykorzystania. Opłaty były także związane z pracą po stronie zespołu, nie istniało zautomatyzowane rozwiązanie. Otwarty i bezkosztowy dostęp idzie tutaj w parze z oddaniem do użytku narzędzi samoobsługowych (*self-service*), tak aby to użytkownik konfigurował import danych ze źródła. Trzecim czynnikiem były realia rynku: sprzedaż metadanych stanowiła istotne zasilenie budżetu, ale w DNB spodziewano się, że ten strumień przychodu zaniknie wraz z rozwojem technologii. Działanie obliczono zatem na stopniową adaptację do przewidywanej sytuacji. Dane uwalniano stopniowo w latach 2010-2015, również w formie *Linked Open Data*, monitorując straty przychodu, z największym i ostatecznym otwarciem zasobów 1. lipca 2015.

Spowodowało to znaczący wzrost zainteresowania, mierzony tworzeniem nowych kont dostępu (czterokrotnie więcej). Reakcje dzieliły się na dwie grupy: niedowierzanie i szybką adaptację. Część użytkowników pytała DNB, czy dobrze rozumieją informację o uwolnieniu danych. Znaczna grupa pośród nich pobierała dane z serwera FTP (odpłatnie) i zdecydowała się na model bezkosztowy. Drugi typ reakcji reprezentowały przede wszystkim zespoły wytwarzające oprogramowanie i rozwijające wyszukiwarki internetowe. Wśród nich dominowała gotowość techniczna na dzień uruchomienia nowego typu dostępu i podjęto akcje już wcześniej, aby pozostało wtedy tylko zgłosić się po dane dostępne, a później pobierać zapisy bibliograficzne.

Po roku DNB zebrała doświadczenia. Wspomniany znaczny wzrost liczby użytkowników, głównie z krajów niemieckojęzycznych, stworzył okazję do tworzenia społeczności odbiorczej, która przekazuje informację zwrotną na temat swoich oczekiwań co do danych i dostępu do nich. Są to instytucje (np. biblioteki publiczne), firmy (np. księgarnie) i osoby prywatne. Zmieniła się gwarancja jakości: DNB używa zewnętrznych źródeł i nie może zapewnić poziomu dostępnego przy tworzeniu zasobów wewnątrznie. Udostępnienie danych spotkało się zasadniczo z dobrym przyjęciem.



W podsumowaniu autorzy deklarują implementację systemu zarządzania relacjami z klientem (CRM, *Customer Relation Management*), dla lepszego zaopiekowania potrzeb odbiorców. Innym usprawnieniem jest dodanie narzędzi analityki użycia zawartości oraz wprowadzenie badań ankietowych wśród użytkowników. Autorzy podkreślają też, że z jednej strony zwiększyła się widoczność i istotność bazy DNB w świecie, z drugiej – spadły przychody do budżetu. Możemy wszakże zauważyć, że w pewnym sensie znikła usługa objęta płatnością (gwarancja standardu przygotowania materiału). Korzyść dla ekosystemu danych na świecie wydaje się być niewątpliwa, zwłaszcza w kontekście użycia *bibliodata* w celach badawczych.

Takie obserwacje potwierdza raport – ekspertyza zespołu ICM, powstały pod kierownictwem Marka Niezgódki<sup>201</sup>. Jest to publikacja odwołująca się do doświadczeń międzynarodowych i przyjmująca poziom państwowy zarządzania za główny cel analizy. Zebrane zostały wnioski w oparciu o informację zwrotną z instytucji z USA, Niemiec, Wlk. Brytanii, Holandii, Danii, Szwecji, Norwegii, Finlandii, Francji, Australii, Brazylii i ChRL. Kluczowe wnioski i rekomendacje to:

- Otwarty dostęp na poziomie rządowym najczęściej jest traktowany jako narzędzie zarządzania wiedzą, jej lepszego rozpowszechniania i wykorzystania oraz zwiększenia widoczności krajowych badań.
- Projekt implementacji typu top-down jest bardziej efektywny (OD wprowadzany przez fundatorów, uniwersytety, agendy rządowe).
- Rząd powinien wspierać otwarty dostęp na poziomie prawnym i finansowym.
- Dobrze rozpoznany rozwiązaniem wprowadzania OD jest powołanie specjalnej komisji (przy rządzie, parlamencie, ministerstwie nauki lub radzie badawczej), która powinna zbadać skutki wprowadzenia otwartego dostępu oraz zaproponować ścieżkę wprowadzenia OD.
- Infrastruktura techniczna odgrywa kluczową rolę. Zwykle to ona rozwijana jest w pierwszej kolejności (Niemcy, Australia, Dania). Pod uwagę należy wziąć interoperacyjność infrastruktury.
- Zmiany powinny być wprowadzane stopniowo (czas na społeczne konsultacje, oswojenie się z ideą). W niektórych przypadkach OD było

---

<sup>201</sup> Marek Niezgódka i in., „Wdrożenie i promocja otwartego dostępu do treści naukowych i edukacyjnych. Praktyki światowe a specyfika polska. Przewidywane koszty, narzędzia, zalety i wady” (ICM, 2011), <http://depot.ceon.pl/handle/123456789/1545>.

wprowadzane w instytucji jednego typu (np. Narodowa Biblioteka Naukowa w Chinach) lub dla jednego typu zasobów (np. program Cream of Science w Holandii, digitalizacja naukowych czasopism humanistycznych we Francji).

- Większość krajów rozwija zarówno zieloną drogę (repozytoria umożliwiające samoarchiwizację), jak i złotą drogę (dedykowanie środków na publikowanie w czasopismach OD). Płacenie za publikowanie w czasopismach OD jest bardziej wspierane w krajach, gdzie rynek wydawnictw naukowych jest dobrze rozwinięty i jest przedmiotem zainteresowania zagranicznych naukowców.
- Uwzględnić należy aktorów różnego typu (fundatorzy, naukowcy, przedsiębiorcy, wydawcy, agencje rządowe, społeczeństwo, instytucje edukacyjne).
- Najważniejsze trudności we wprowadzaniu OD to: opór tradycyjnych wydawców (Wielka Brytania, USA, Australia), prawa własności intelektualnej (Niemcy, Wielka Brytania), brak wsparcia ze strony społeczności naukowej (Chiny, Brazylia), brak działań koordynujących.
- Digitalizacja czasopism może zostać użyta jako dźwignia do wprowadzenia OD (Brazylia, Francja).
- Potrzebne są działania na poziomie krajowym i międzynarodowym<sup>202</sup>.

Jak widać, systemowe wsparcie otwartości danych sprzyja cyrkulacji wiedzy i staje się nieodzownym elementem globalnych przemian.

Gdzie najlepiej udostępnić dane? – pyta Bożena Bednarek-Michalska. I podaje rekomendację: w dobrym towarzystwie <sup>203</sup>. W repozytorium zaufanym, prestiżowym, posiadającym odpowiednie procedury. Najlepiej repozytorium międzynarodowym. Wracam jeszcze do poranika Drabek. Spora część tej pracy zostaje poświęcona temu, jak dodać teksty do baz danych. Prezentowane są kryteria dla Scopus, Web of Science Core Collection, DOAJ oraz ERIH Plus. Jak widzieliśmy w przypadku rekomendacji z tego tekstu, element prestiżu i eksperckiej preselekcji czasopism może niestety stać w sprzeczności z otwartością takich

---

<sup>202</sup> Niezgódka i in.

<sup>203</sup> Zob. Bożena Bednarek-Michalska, „Zarządzanie danymi badawczymi – informacje ogólne i plany zarządzania danymi” <https://www.youtube.com/watch?v=iopHvHKFD-A>. Dostęp: 2020-03-01. (fragm. ok. 37. minuty)

danych z perspektywy czytelniczej. Na szczęście wiele serwisów z komercyjnym dostępem, jak JSTOR czy Scopus, również zawiera częściowe rozwiązania wolnego dostępu do treści: niektóre wynikają z decyzji autorskich o licencji, inne to edytorska decyzja właścicieli bazy o np. uruchomieniu „podglądu” w darmowej wersji. Będzie to odpowiadało modelowi hybrydowemu. Można także robić użytek z danych bibliograficznych i danych do cytowania (co było wspomniane przy okazji badań Peroniego i innych i Open Citations).

### Dane wrażliwe

Dane osób stają się coraz ważniejszym kontekstem przetwarzania informacji w internecie, co każdy użytkownik sieci może poznać, odwiedzając strony i wpływając na ustawienia prywatności i ciasteczek (pliki *cookies*, służące m.in. do zapamiętywania odwiedzin, analityki, tworzenia profilu odwiedzającego i rekomendacji wewnątrz serwisu, np. optymalizacji wyświetlania reklam). Ważną kwestią przy zarządzaniu danymi badawczymi jest zachowanie poufności tam, gdzie to konieczne.

W związku z regulacjami ochrony danych osobowych (znane jako RODO, GDPR), każdy podmiot, zbierający i przechowujący dane, ma obowiązek respektować prawa osób, których te dane dotyczą. Są to między innymi prawa do modyfikacji, aktualizacji, czy usuwania danych z bazy. Osoba, której dane dotyczą może egzekwować prawo do zapomnienia, czyli wniesić o kompletne usunięcie informacji o sobie z danego zbioru. Z jednej strony mamy do czynienia z prawnym obowiązkiem administratora danych i prawami jednostki do rozporządzania danymi na swój temat – z drugiej, stwarza to potencjalne ryzyko dla wysiłków archiwistycznych i dokumentalistycznych. By pozwolić sobie na *reductio ad absurdum*, gdyby ogłoszono w Polsce wielki strajk pisarzy i poetów, skutkujący wniesieniem o usunięcie ich danych osobowych, takich jak imię i nazwisko oraz biogram z baz bibliograficznych, czy zespoły dokumentalistyczne musiałyby wymazać wszelką gromadzoną wiedzę na temat tych osób?

W przypadku Polskiej Bibliografii Literackiej troska o dane osobowe jest zasadna. Dane osobowe to takie, które jednoznacznie mogą wskazać konkretną osobę. Nie są nimi np. imię i nazwisko, ponieważ można domniemywać, że istnieje więcej osób, które się tak nazywają. Ale już imię i nazwisko w połączeniu z danymi teleadresowymi – tak. Przenosząc to na grunt pracy bibliograficznej, imię i nazwisko autora stanowią dane dotyczące osoby. W połączeniu z tytułem publikacji – są to dane wskazujące konkretną osobę.

Przy uznaniu takich danych za dane osobowe, objęte mechanizmem wyrażania zgód i będące przedmiotem przetwarzania, możliwość egzekwowania prawa do zapomnienia powodowałaby utratę danych badawczych. Jednakże sposób operowania danymi jest zgoła inny. Pracownicy Pracowni Bibliografii Bieżącej sporządzają zapisy z autopsji, mając kontakt z fizycznymi egzemplarzami, pochodzącymi z bibliotek. Informacja zawarta w danych pochodzących od wydawcy, nie stanowi zbioru danych osobowych. Dane o zapisie (bibliograficzne) stają się publicznie dostępne w momencie wydania konkretnego tytułu. Oznacza to, że przywołani polscy pisarze i poeci powinni liczyć się z gromadzeniem i przetwarzaniem danych na temat ich *twórczości*, a dane nazywające osobę przestają być danymi osobowymi, gdy mówimy o materialnym czy niematerialnym fakcie publikacji dzieła. W tym wypadku trudno powiedzieć, czy jest to pomnik trwalszy niż ze spiżu, ale na pewno trwalszy niż prawo do bycia zapomnianym w świetle przepisów RODO. W znacznej większości przypadków spodziewam się, że jest to w interesie tych osób autorskich, aby być rozpoznawalnymi.

Plany zarządzania danymi badawczymi

Obok FAIR Data, w europejskich inicjatywach badawczych są promowane *Data Management Plans*, zbiory charakterystyk projektów, które same w sobie tworzą metadane, opis inicjatyw. Plan zarządzania danymi zawiera sekcje obejmujące kilka najważniejszych obszarów<sup>204</sup>: opis danych i metadanych, charakterystyka zbierania ich oraz ponownego wykorzystania; dokumentacja oraz aspekty kontroli jakości; przechowywanie oraz tworzenie kopii zapasowych (*backups*) – w tym bezpieczeństwo i ochrona danych wrażliwych, o ile występują; podstawy prawne i odniesienia do przepisów (jak wspomniane RODO, ale też inne, np. dotyczące praw autorskich czy niezbędnych licencji); udostępnianie i długotrwałe przechowywanie danych (w repozytoriach i archiwach – często jest to wymogiem instytucji czy grantodawców, aby dostęp był zapewniony w podanym okresie).

Istnieje wiele wzorów (*templates*), a nawet narzędzi do tworzenia takich planów zarządzania danymi, z niewielkimi odstępstwami pod względem układu treści, nazewnictwa pól itp. Zaletą niektórych z nich są interaktywne formularze, do wypełnienia w interfejsie internetowym, przez który później wydaje się polecenie wygenerowania pliku lub wydruku gotowego

---

<sup>204</sup> Wytyczne Narodowego Centrum Nauki - plan zarządzania danymi  
[https://www.ncn.gov.pl/sites/default/files/pliki/regulaminy/wytyczne\\_zarzadzanie\\_danymi.pdf](https://www.ncn.gov.pl/sites/default/files/pliki/regulaminy/wytyczne_zarzadzanie_danymi.pdf) (dostęp: 2020-03-01). Por. Science Europe, „Practical Guide to the International Alignment of Research Data Management - Extended Edition”, 2021-01-27, <https://doi.org/10.5281/ZENODO.4915861>.

dokumentu<sup>205</sup>. Wiele gotowych wzorów odnosi się wprost do FAIR Data lub, przez to, że te standardy są tak adekwatne do zarządzania danymi badawczymi, zawiera te same punkty, nie odwołując się bezpośrednio do tego zbioru pryncypiów.

Badacze wypełniają Podczas poszukiwań własnych na potrzeby projektu danowego, zauważyłem, że w praktyce badawczej część tych pól może pozostawać pusta tzn. nie aplikuje się do prezentowanego pola badawczego czy opisywanego projektu. Rzadziej będzie to zapewne motywowane świadomym wyborem o niewprowadzeniu informacji.

Humanista cyfrowy kuratorem treści

Jak zauważa Katherine Bode<sup>206</sup>, przygotowujący zestawy danych stają przed zadaniem kuratora treści, podobnie do redaktorów książek. Zbiór danych (kolekcja, dataset) wyraża pewną ideę i poprzez dobór zawartości reprezentuje zestaw kryteriów, jakimi posługiwali się jego twórcy. W refleksji krytycznej na temat humanistyki cyfrowej widzieliśmy, jak ważny to aspekt – także w przypadku danych połączonych wybór odpowiedniego skrawka z szerszego ekosystemu będzie mieć znaczenie.

Aby dopełnić tej części pracy, po zreferowaniu dobrych praktyk, chciałbym odwołać się do przykładu, który zilustruje, jak chaotyczne może być badanie wybranego zagadnienia dotyczącego kultury na podstawie informacji dostępnych w sieci. Mam nadzieję dowieść, jak potrzebna jest interwencja, związana z ontologiami, dla otwartego dostępu do rzetelnej informacji, popartej wiedzą specjalistyczną.

Wiedza pozostaje rozproszona i domaga się nowych syntez

Współczesny użytkownik *webu*, zdobywający wiedzę i angażujący się w dyskursy krytyczne (społeczne, kulturalne, akademickie), poznaje swój temat dociekań z wielu źródeł niemalże jednocześnie. Uniwersytecka lub szkolna dyskusja w ramach zajęć, nieformalna wymiana zdań na żywo i na platformach społecznościowych, wydarzenia kulturalne oraz tradycyjne źródła lekturowe przeplatają się z kontaktem ze sztuką i fikcją w czasie wolnym.

Chciałbym zobrazować przenikanie się wielu rozproszonych źródeł, za których powstaniem stoją zróżnicowane, często sprzeczne, interesy i intencje. Każde z nich może stanowić punkt

---

<sup>205</sup> Takich narzędzi jest wiele i kwerenda w sieci może pokazać bardziej przydatne, w zależności od potrzeb badawczych. Większość serwisów, które odwiedzałem, oferuje przechowywanie DMP, a zatem wymaga stworzenia konta lub zalogowania się np. danymi konta ORCID. Zob. np. „DMPTool”, <https://dmptool.org/>. Dostęp: 2023-09-16.

<sup>206</sup> „As with a traditional scholarly edition, the composition of a curated data set will reflect what the editor (or editorial team) perceives as most relevant to understanding a particular literary system”. Bode, „The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”.

wejścia dla osoby zapoznającej się z danym tematem, jak i splot pośrodku wędrówki – odsyłacz do innych źródeł.

Weźmy konkretny przykład: prace Andrei Dworkin, autorki związanej z drugą falą feminizmu. Wybieram ten przypadek jako złożony poznawczo, badawczo intrygujący, a jednocześnie nie do końca obecny na gruncie polskich dyskusji i nadający się do dalszych badań. Prace Dworkin wpisują się w nadal żywy nurt myślowy i prowokują debaty. Jak się jednak przekonamy, rodzimy czytelnik może napotkać wyzwania, próbując dotrzeć do rzetelnych źródeł wiedzy na temat jej poglądów i wypowiedzi.

Wiedza na powierzchni

Zainteresowanie tą osobą i jej twórczością może mieć różne źródła, widzę trzy główne. Pierwsze to edukacja: akademickie poszukiwania (np. związane z *gender studies*, feminizmem, a nawet, jak się przekonamy, historią prawa) lub praca projektowa związana z jakimś zagadnieniem. Drugie to amatorska chęć poznania tej osoby pod wpływem zapoznania się z jakąś fragmentaryczną informacją, taką jak przeczytanie cytatu jej autorstwa, obejrzenie fragmentu filmu z jej wypowiedzią, czy też wyszukiwanie nazwiska, ponieważ pojawiło się w dyskusji na żywo lub w mediach. Trzeci obszar scharakteryzowałbym jako motywowany politycznie i ideowo, tzn. wynikający z chęci zrozumienia dynamiki i historii ruchu feministycznego, śledzenie myśli i losów jego najważniejszych przedstawicielek i przedstawicieli. Każdy z tych typów zainteresowania może także przybrać formę polemiczną, to znaczy znajduje motywację bardziej w niezgodzie niż zainteresowaniu lub aprobacie. Takich scenariuszy może być więcej, mogą się także wzajemnie przenikać (biorąc chociażby pod uwagę tradycję feminizmu akademickiego, czyli połączenie motywacji pierwszej i trzeciej).

Na wstępie ograniczę się tutaj do minimalnego zarysowania kontekstu działalności Dworkin. W Polsce jej postać może interesować osoby badające historię feminizmu naukowo, jak i zainteresowane tym ruchem społecznym i naukowym. Ze względu na konkretny moment i miejsce jej działalności (głównie Stany Zjednoczone lat 60. XX wieku do początku XXI wieku, publikacje książkowe ok. 1974-2002), nie wszystkie tematy tego pisarstwa będą adekwatne dzisiaj. Można założyć, że to teksty mocno osadzone w kontekście społeczno-politycznym epoki, w której pisarka działała najprężniej.

Autorka angażowała się mocno w działalność publicystyczną, stąd być może warte uwagi byłyby pod względem stylu. Poruszała tematy dziedziny prawodawstwa, związane z przemocą seksualną i prostytutką. Krytykowała też pornografię jako przemysł legitymizujący kulturowo

patriarchalne normy i normalizujący zachowania przemocowe. Była współautorką propozycji reformy systemu prawnego, która pozwoliłaby pozwać twórców treści pornograficznych przez osoby, które poczułyby się nimi w jakiś sposób skrzywdzone. W czasach studenckich brała udział w protestach przeciwko wojnie w Wietnamie. Możliwa zatem jej obecność także w dyskursie dotyczącym legislacji, praw obywatelskich oraz spraw karnych. Dworkin była ofiarą przemocy seksualnej (zarówno jako dziecko, jak i osoba dorosła) i w niektórych tekstach wręcz opowiadała się za odwetowym użyciem przemocy wobec sprawców tego typu nadużyć. Ze względu na radykalizm poglądów niewykluczone, że jej wypowiedzi publiczne mogłyby stanowić dobry punkt odniesienia do badań dyskursu publicznego, zasad argumentacji, analizy sentymentu w debacie oraz w studiach związanych z szeroko pojętym dziennikarstwem i naukami politologicznymi, np. pod kątem jej retoryki – także, a być może zwłaszcza, w wymiarze historycznym. Wreszcie, jako osoba mająca żydowskie pochodzenie, w części z prac odwołuje się do Zagłady, jak i do doświadczeń Żydów oraz państwa Izrael. Ostatni znaczący moment w historii, gdzie Dworkin poznajemy jako komentatorkę zdarzeń, to sprawa Clinton-Lewinsky (oskarżenie prezydenta USA o molestowanie seksualne przez jego pracownicę w późnych latach 90.).

Wszystkie powyższe informacje na jej temat czerpałem najpierw z angielskiej<sup>207</sup>, następnie polskiej<sup>208</sup> wersji strony poświęconej jej na Wikipedii, a także z biograficznego filmu dokumentalnego pt. *My name is Andrea*<sup>209</sup>, zawierającego elementy fikcyjnego przedstawienia zdarzeń z jej życia, jak np. spotkania z przedstawicielem *Beat Generation*, Allenem Ginsbergiem. Polska Wikipedia wymienia wybrane publikacje książkowe z zaznaczeniem, że „dotychczas nie było polskich edycji”<sup>210</sup>. Uznaję te źródła za łatwo dostępne dla polskiego odbiorcy posługującego się internetem i najprawdopodobniejsze punkty startu w typowej sieciowej kwerendzie kogoś zainteresowanego tematem.

Wyzwania kwerendy w serwisach internetowych

Przechodząc do jej obecności w danych bibliograficznych, Andrea Dworkin funkcjonuje w polskich serwisach typu *Web 2.0*, zajmujących się rynkiem wydawniczym. Pierwszym, który sprawdziłem, jest Lubimy Czytać, portal rozwijany przez użytkowników. Platforma społecznościowa należy do grupy kapitałowej Znak, a więc jest związana z jednym

---

<sup>207</sup> „Andrea Dworkin” w *Wikipedia*, [https://en.wikipedia.org/wiki/Andrea\\_Dworkin](https://en.wikipedia.org/wiki/Andrea_Dworkin). Dostęp: 2023-05-28.

<sup>208</sup> „Andrea Dworkin” w *Wikipedia*, [https://pl.wikipedia.org/wiki/Andrea\\_Dworkin](https://pl.wikipedia.org/wiki/Andrea_Dworkin) dostęp: 2023-05-28.

<sup>209</sup> Strona internetowa projektu: <https://www.mynamesandreamovie.com/> dostęp: 2023-06-01.

<sup>210</sup> „Andrea Dworkin” w *Wikipedia*, [https://pl.wikipedia.org/wiki/Andrea\\_Dworkin](https://pl.wikipedia.org/wiki/Andrea_Dworkin) dostęp: 2023-05-28.

z wydawców na rynku książki w Polsce<sup>211</sup>. Po wyszukiwaniu, strona autorki zawiera trzy wpisy i wszystkie są pozycjami anglojęzycznymi<sup>212</sup>.

Biblionetka odnotowuje na podstronie poświęconej autorce<sup>213</sup> jedną publikację, wprowadzoną przez użytkownika, ale jest ona zbiorem szkiców i widać tutaj dużo więcej postaci. Tym jednym tytułem są *Prognozy: Trzydziestu myślicieli o przyszłości*, gdzie znajdziemy również m.in. teksty Noama Chomsky’ego, Umberta Eco czy Lynn Margulis. Mało rozwinięty profil feministycznej działaczki świadczy prawdopodobnie o tym, że użytkownik bardzo skrupulatnie dodał wszystkie nazwiska autorów obecnych w i właściwie dlatego w bazie Biblionetki w ogóle pojawiło się nazwisko autorki *Pornography*.

Inny istotny serwis społecznościowy związany z rynkiem wydawniczym, nakanapie.pl, nie odnotowuje żadnych publikacji Dworkin. Nie powinna dziwić nieobecność dzieł czy mała reprezentacja, skoro brak polskich tłumaczeń.

Oprócz tego, koncepcja tych serwisów zakłada przede wszystkim monetyzację poprzez linki afiliacyjne do internetowych księgarni i zysk z reklam. Dlatego najlepiej będą opisane publikacje popularne (kryterium liczby czytelników ponad estetyczną oceną autorytetów, np. krytyki literackiej) oraz nowości wydawnicze (w tym przypadku istotny będzie aspekt promocji i obecność tych dzieł w bieżącym życiu literackim). Teksty Dworkin nie wpisują się w owe tendencje, które byłyby motywatorem do uzupełnienia informacji o autorce i jej dziełach. Widzimy zatem, że zbiory tworzone społecznościowo są względem siebie rozłączne i myśląc o stworzeniu korpusu lub bibliografii tematycznej skorzystalibyśmy na połączeniu informacji w nich zawartych.

Wspomniany już film *My name is Andrea* został wyświetlony podczas festiwalu Tofifest w 2023 jako polska premiera<sup>214</sup>. Ta projekcja to jeden ze wspomnianych przeze mnie punktów wejścia odbiorcy w temat twórczości Dworkin – w mojej ocenie jeden z bardziej ryzykownych

---

<sup>211</sup> Wspominam o tym ze względu na prywatno-komercyjny aspekt. Z perspektywy użytkownika serwis może służyć do budowania prywatnej biblioteczki i eksploracji lekturowej, ale już dla właścicieli będzie narzędziem monetyzowanym przez reklamy oraz dostarczającym widoczności własnym tytułom. Działa to także w systemie implementacji Lubimy czytać jako źródła ocen użytkowników do sklepów internetowych. Na stronie internetowej wydawnictwa Znak odnajdziemy informację na ten temat w formie elementu graficznego. Zob. „Wydawnictwo Znak” <https://www.znak.com.pl/>. Dostęp: 2023-09-16.

<sup>212</sup> Zob. „Andrea Dworkin” w *Lubimy czytać*, <https://lubimyczytac.pl/autor/52950/andrea-dworkin>. Dostęp: 2023-07-15.

<sup>213</sup> Zob. „Dworkin Andrea” w *Biblionetka*, <https://www.biblionetka.pl/author.aspx?id=118819>. Dostęp: 2023-07-15.

<sup>214</sup> Zob. „My name is Andrea” w *Tofifest* [https://tofifest.mojfestiwal.pl/movies/ZsAEisu6jE\\_8LcCjpbMzSA](https://tofifest.mojfestiwal.pl/movies/ZsAEisu6jE_8LcCjpbMzSA). Dostęp: 2023-09-23.



interpretacyjnie. Nie chcę pobłażliwie demonizować braku aparatu krytycznego u (potencjalnych) widzów czy sugerować szczególny wymóg kompetencji do odbioru takiego tekstu kultury. Film wyświetlono w sekcji „Niepokorne”, która na różne sposoby obrazowała walkę kobiet o równe prawa (m. in. *The Conductor* o dyrygentce Marin Alsop i *Dream's Gate* o kobiecych oddziałach w Rożawie). Organizatorzy festiwalu jako punkt odniesienia dla widza współczesnego proponują porównanie z ruchem #MeToo<sup>215</sup>, co pozwala dać wyobrażenie przemian historycznych i jednocześnie podkreślić nieustającą aktualność omawianych w dziele problemów.

Warto jednak zaznaczyć, iż jak każdy współczesny film dokumentalny, jest to utwór świadomy własnego zapośredniczenia. Zawiera też elementy performatywne i afektywne, tzn. fragmenty fabularyzowane. Nie tylko referuje zdarzenia, ale też odwołuje się do emocji widzów. Za inne ryzykowne media uważam materiały wideo na platformach internetowych, np. klipy z występów Dworkin lub wycinki jej wypowiedzi w materiałach o charakterze publicystyki społecznej. Pod tym względem bibliografia i tekst, główne obszary poszukiwań niniejszej pracy, raczej referują w rzeczowy sposób i przynależą do porządku powolnej lektury (*slow reading*).

Dzięki ekranizacji życiorysu i biografii pisarskiej Dworkin, otwiera się przed nami horyzont poszukiwań w bazach danych związanych z filmem. Filmweb, prężnie działająca polska internetowa baza kinematografii, posiada podstrony dotyczące Dworkin<sup>216</sup> i filmu *My name is Andrea*<sup>217</sup>. O autorce dowiemy się niewiele: lata życia, powiązanie z filmem dokumentalnym i ciekawostki dotyczące stanu cywilnego (dwa małżeństwa). W tworzeniu tej podstrony brała udział jedna osoba. Profil filmu prezentuje opis tytułu i obsadę, jeszcze we wrześniu 2023 kontrybuowały w tworzeniu strony dwie osoby (użytkownicy portalu inni niż osoba tworząca profil Dworkin).

Internet Movie Database (IMDb) jest międzynarodowym odpowiednikiem Filmwebu i dostarcza ciekawej filmografii. Znajdujemy oczywiście wpis związany z filmem

---

<sup>215</sup>Tamże. Porównanie z #MeToo dobrze zestawia spójność intencji działalności Dworkin z tą współczesną akcją, której celem było zasygnalizowanie, że wiele kobiet współcześnie doświadcza molestowania i dyskryminacji *Hashtag „ja też”*, który nadaje jej nazwę, w domyśle oznacza „ja też zostałam skrzywdzona” i stał się platformą dla wielu kobiet do dzielenia się swoimi historiami. Ze względu na udział celebrytek, wytworzył się przekaz akcji, że przemocowe sytuacje mogą dotyczyć każdej osoby, niezależnie od statusu społecznego.

<sup>216</sup> Zob. „Andrea Dworkin” w Filmweb, <https://www.filmweb.pl/person/Andrea+Dworkin-1348773>. Dostęp: 2023-09-23.

<sup>217</sup> Zob. „My name is Andrea” w Filmweb, <https://www.filmweb.pl/film/My+Name+Is+Andrea-2022-10037764>. Dostęp: 2023-09-23.

biograficznym, ale także m.in. wystąpienie Dworkin w dokumencie na temat Larry’ego Flynta, jedną z najśłynniejszych osób związanych z biznesem pornograficznym<sup>218</sup>. Bez znajomości treści, możemy podejrzewać, że działaczka występuje w roli ekspertki stojącej w kontrze do działalności głównego bohatera. Łącznie IMDb wymienia dziesięć produkcji z udziałem Dworkin we własnej osobie, dwie wzmianki w podziękowaniach i cztery pozycje wśród materiałów archiwalnych. Serwis zawiera także podstawowe informacje o twórczyni, znane nam już z serwisów bibliograficznych i encyklopedycznych: lata życia oraz miejsca narodzin i śmierci, a także skrócony biogram. Widzimy też listę małżonków, analogicznie do Filmweb; obecność tych danych może sugerować zainteresowanie odbiorców (użytkowników IMDb) życiem prywatnym osób występujących w filmach. Podobną funkcję, przesunięcia archiwum w stronę tematów popularnych, pełni sekcja ciekawostek (*Trivia*). Co ciekawe, w połączeniu z biogramem przekazuje ona większość informacji, które skompilowałem powyżej na temat Dworkin. IMDb posiada własne identyfikatory rekordów oraz interfejs programistyczny, ale niestety nie zawiera stałych identyfikatorów z zewnątrz ani odnośników do połączonych danych, zatem identyfikacja osoby musi się odbywać na zasadzie pośredniego dopasowywania informacji. Znamienne jest określenie API (*Application Programming Interface* czyli interfejs programowania aplikacji, zasób umożliwiający współpracę dwóch typów programów, w kontekście baz danych o kulturze najczęściej używany do automatyzowanego przeglądania i przetwarzania treści) udostępniane przez IMDb jako *Entertainment Data*, czyli dane dotyczące rozrywki – a nie kinematografii<sup>219</sup>.

Oba portale poświęcone filmowi, jak i wcześniej przywoływane strony dotyczące książek, posiadają też funkcjonalność dodawania ocen w wybranej skali. Jest to ciekawe zagadnienie, ze względu na swoistą próbę demokratyzacji estetyki i sądów smaku, wymagałoby ono jednak specjalnego potraktowania w innym tekście. Zaznaczam tutaj ten aspekt, ponieważ wcześniej była mowa o krytyce mediów społecznościowych ze względu na kompetencyjne braki uczestników sieciowych debat. W wymienionych wcześniej serwisach ocena całości dzieła zostaje właściwie sprowadzona do liczbowego wyniku na skali, będącego agregacją głosów. Niektóre portale, jak np. Rotten Tomatoes czy rodzimy Filmweb, prezentują dwa współczynniki: ocenę widzów-użytkowników serwisu oraz ocenę krytyków, czyli „odbiorców

---

<sup>218</sup> *Larry Flynt: The Right to Be Left Alone*, Documentary (Midtown Films, 2007).

<sup>219</sup> Przykładowe odpytania bazy w dokumentacji to w pierwszej kolejności zapytanie o tytuł lub nazwisko, a w drugiej o wyniki kasowe, co wyraźnie wskazuje na silnie zakorzenione w relacjach i wartościach rynkowych myślenie o kinie (i projektowanie w ten sposób tego, kto używa programistycznego interfejsu do przetwarzania danych IMDb). Por. „Sample Queries”, IMDb Developer, <https://developer.imdb.com/documentation/api-documentation/sample-queries/>. Dostęp 2023-09-05.

profesjonalnych”, dysponujących aparatem krytycznym dla zrozumienia dzieła i osadzenia go w szerszym kontekście swojej dziedziny (w tym wypadku kinematografii). Jest to szczególnie wyraźne i potrzebne, biorąc pod uwagę intencjonalne zorganizowane ataki na internetowe oceny, znane jako *review bombing*<sup>220</sup>.

Każdy taki zbiór (na razie zgromadziliśmy internetowe encyklopedie i serwisy społecznościowe zorientowane na literaturę oraz film) dostarcza sieci powiązań informacji. Jednocześnie każdy z nich ma inny sposób opisu tematu, tzn. inne kryteria doboru informacji oraz klasyfikowania ich. Użytkownicy czasem zatrzymają się tylko na jednym z nich (np. najlepiej wypozycjonowanym względem algorytmu wyszukiwarki internetowej wobec wprowadzonego zapytania). Niektórzy bardziej dociekliwi sprawdzą więcej, w poszukiwaniu konkretnej informacji. Poziomem bazowym będzie tu jednak zazwyczaj Google lub podobna wyszukiwarka, jako węzeł rozpoczynający wędrówkę i taki, który przecina się, by do niego powrócić i sprawdzić kolejne źródło. Użytkownik ma niewielki wpływ na kolejność serwowania mu treści i ich charakter (np. edukacyjny vs marketingowy), poza dostosowywaniem kolejnych zapytań, by jak najprecyzyjniej odnaleźć poszukiwane informacje. Część z osób bierze na siebie role kuratorów treści i samemu tworzy dane, dostępne później dla innych. Zarówno aktywni użytkownicy, jak i ci, którzy wyłącznie odbierają komunikaty, działają jednocześnie na (z pewnymi wyjątkami) tym samym zbiorze informacji w sieci.

Pozornie są to tematy doskonale znane humanistycy: moglibyśmy się odwołać do pojęcia McLuhanowskiej globalnej wioski czy konceptów wyrażanych za pomocą sztuki przez Andy’ego Warhola (mieszanie się porządków sztuki wysokiej z popkulturą oraz namysł nad dystrybucją społecznego uznania i sławy). Rzecz jednak zasadza się na tym, iż posiadamy przed sobą pewien obiekt (roboczo nazwijmy go dziedzictwem intelektualnym Andrei Dworkin) i w poszanowaniu dziedzictwa kulturowego jako humaniści chcielibyśmy o tym obiekcie zdać sprawę. Z pozycji humanistyki cyfrowej jako metody proponuję uznać wszelkie

---

<sup>220</sup> Najświeższym i bardzo głośnym przykładem *review bombing*, który przydarzył się w trakcie finalizowania niniejszego tekstu, była historia filmu „Zielona granica” w reżyserii Agnieszki Holland. Dzieło opisuje napięcia na granicy polsko-białoruskiej w związku z kryzysem migracyjnym. Film otrzymał nagrodę jury festiwalu w Wenecji. Ze względu na poruszaną tematykę, spotkał się z krytyką konserwatywnej części społeczeństwa, głos zabral premier Polski. Po nagonce medialnej ponad 3 tysiące osób wystawiło „Zielonej granicy” niskie noty, jeszcze przed powszechną premierą kinową. 10 września 2023 oceny wszystkich użytkowników miały w dziesięciostopniowej skali średnią 2,6. W tym samym dniu uśrednione oceny dziesięciu krytyków to 7,5. Zob. „Zielona granica” w Filmweb, <https://www.filmweb.pl/film/Zielona+granica-2023-10039040> dostęp: 2023-09-10. Widzimy wyraźnie ryzyko opierania się na tego rodzaju danych bez należytej refleksji i komentarza specjalistów, aby uwzględnić takie zjawiska jak masowy atak na ocenę filmu.

źródła dostępne w internecie<sup>221</sup>, dotyczące tego obiektu badań za istotne oraz pozwalające oświetlać go wiedzą z różnych perspektyw. Mowa tu zwłaszcza o tych elementach, które rejestrują fakty i referują zdarzenia<sup>222</sup>. Wielość tych źródeł nie tyle burzy istniejącą hierarchię akumulacji wiedzy, co raczej dowodzi, że w takim kontekście jak przedstawiony temat badań, korzystne jest eksplorowanie rejonów produkcji wiedzy, które pojawiają się oddolnie, spontanicznie i mają charakter społeczny. Widzę w tym dwojaki powód: z jednej strony niedobór (w tym wypadku znikomość polskiej recepcji), z drugiej – niekompletność (żadne pojedyncze źródło nie wyczerpuje tematu). W takiej sytuacji, jeżeli synteza, to tylko taka, która wyznaczy drogi pomiędzy poszczególnymi punktami danych dla każdego, kto zechce udać się w wędrówkę w poszukiwaniu informacji. Wędrówkę, dodajmy, narażoną na pułapki interpretacyjne, co tym bardziej stwarza potrzebę oznaczenia tychże i zaproponowania bezpiecznego kursu. Zmierzamy tu w stronę specjalistycznej kuracji treści przez osoby do tego wykwalifikowane.

Po polsku nie odnajdziemy zbyt wielu źródeł autorstwa Dworkin ani dotyczących jej osoby w piśmiennictwie humanistycznym. Jest wzmiankowana w *Encyklopedii gender*, wydanej w Warszawie w 2014<sup>223</sup>, a także artykuł Natalii Klejdysz z 2006 roku<sup>224</sup>. Odnotujemy, że główne źródło akademickie dla polskich literaturoznawców, podręcznik teorii literatury Burzyńskiej i Markowskiego, w rozdziale poświęconym feminizmowi jako metodzie badawczej nie wspomina o pracach Dworkin i nie znajdziemy o niej wzmianki w indeksie tej publikacji; nie jest to wadą publikacji, a raczej świadczy o rygorach objętościowych oraz mniej istotnej pozycji Dworkin w porównaniu do takich postaci jak np. Gloria Steinem, wspomniana w podrozdziale dotyczącym drugiej fali feminizmu<sup>225</sup>.

Dworkin jest za to wzmiankowana w feministycznej audycji o tytule *Nikt nas nie pytał, ale i tak się wypowiemy!*. Trzy spośród nagrań są poświęcone historii ruchu feministycznego i towarzyszyły kursowi *gender studies* na Uniwersytecie Łódzkim „Płeć, seksualność, polityka”

---

<sup>221</sup> Bibliograficzne, encyklopedyczne, audiowizualne, publicystyczne itp.

<sup>222</sup> Szczególny status miałyby tu niektóre teksty publicystyczne, jak felietony, oraz źródła zawierające praktyki dyskursywne i agon: fora internetowe, czy te części mediów społecznościowych, które umożliwiają dyskusje. Mamy do czynienia z porządkiem faktów oraz interpretacyjnych i myślę, że wszelkie opisane wyżej „modalne” treści wymagałyby opisu np. z uwzględnieniem analizy sentymentu.

<sup>223</sup> Katarzyna Czczot i in., red., *Encyklopedia gender: płeć w kulturze* (Warszawa: Wydawnictwo Czarna Owca, 2014). Dziękuję Marcie Mazurek za wskazanie tego źródła.

<sup>224</sup> Natalia Klejdysz, „Andrea Dworkin – najbardziej znienawidzona feministka”, *Środkowoeuropejskie Studia Polityczne*, nr 1 (2006): 159–72.

<sup>225</sup> Por. Burzyńska, Anna, Michał Paweł Markowski, i Anna Burzyńska. *Podręcznik. Teorie literatury XX wieku* / Michał Paweł Markowski. Kraków: Wydaw. Znak, 2009.

– „Gender, Sexuality and Politics” – stanowią jego lekturę oraz wstęp (a więc mowa o kanonie nauczania akademickiego w tej dziedzinie). W odcinku dotyczącym drugiej fali feminizmu<sup>226</sup> Agnieszka Szczepanek i Marta Mazurek<sup>227</sup> odwołują się do zagadnień historii ruchu oraz głównych postaci, działań i postulatów. Nagranie trwa około godziny. Ma za zadanie zarysować główne obszary problemowe i stanowi swego rodzaju zagajenie, inspirujące do dalszych poszukiwań lekturowych i dyskusji. Ale nawet tutaj, mimo bardzo wyspecjalizowanego dyskursu stojącego za tym wprowadzeniem, prym wiodą inne autorki niż Dworkin. W opisie słyszymy więc (a w opisie nagrania – czytamy) przede wszystkim o Betty Friedan, Adrienne Rich czy Helene Cixous i ich działalności oraz publikacjach. Dworkin zostaje wymieniona tylko raz, pod koniec *podcastu*. Przywołana zostaje w części uwag rozszerzających główne treści, w kontekście swojej krytyki pornografii i patriarchalnego rozumienia kobiecości (idei dotyczącej tego co kulturowo rozumiane jako zmysłowe i seksowne, a w interpretacji autorki skrywające opresyjne traktowanie). Nie znalazłem też wzmianki o autorce ani jej prac wśród zalecanych lektur dwóch kursów UŁ w publicznie dostępnych sylabusach<sup>228</sup>.

W dalszej kolejności postanowiłem jednak sprawdzić, jaki obraz jako autorki wyłania się ze źródeł ściśle związanych z bibliografiami.

Biblioteka Narodowa w swoich połączonych katalogach odnotowuje wprawdzie 22 wyniki na zapytanie uwzględniające imię i nazwisko Andrei Dworkin, ale przez traktowanie obu członów rozłącznie trafiamy też na niepotrzebne informacje. Na wyszukanie poprzez zapis imienia i nazwiska w cudzysłowie, czyli jako ciąg, BN zwraca 16 wyników, ale jest to w większości zbiór poświęcony feminizmowi jako zjawisku.

Dla porównania, Library of Congress zwraca odpowiednio 56 i 48 wyników na to samo zapytanie, najpierw bez, a później z użyciem cudzysłowu. Przy wyszukiwaniu

---

<sup>226</sup> *Druga fala feminizmu, czyli prywatne jest polityczne*. Nikt nas nie pytał, ale i tak się wypowiemy!, 2020. <https://www.estrada.poznan.pl/series/nikt-nas-nie-pytal/?i=9>.

<sup>227</sup> Dziękuję autorkom audycji za konsultację źródeł.

<sup>228</sup> Pierwszy kurs: „Gender, Sexuality and Politics - Przedmioty – USOSweb”

[https://usosweb.uni.lodz.pl/kontroler.php?\\_action=katalog2/przedmioty/pokazPrzedmiot&prz\\_kod=0600-WW728B](https://usosweb.uni.lodz.pl/kontroler.php?_action=katalog2/przedmioty/pokazPrzedmiot&prz_kod=0600-WW728B). Dostęp: 2023-09-20. Drugi kurs nie posiada listy zalecanych lektur, por.: „Ciało, płeć, seksualność - wymiar społeczno-kulturowy”,

[https://usosweb.uni.lodz.pl/kontroler.php?\\_action=katalog2/przedmioty/pokazPrzedmiot&kod=0600-PS149C](https://usosweb.uni.lodz.pl/kontroler.php?_action=katalog2/przedmioty/pokazPrzedmiot&kod=0600-PS149C). Dostęp: 2023-09-20. Należy pamiętać, że zalecane lektury najczęściej są wyłącznie wskazówką i zbiorem najbardziej charakterystycznych czy kanonicznych prac.

zaawansowanym w polu Author/Creator, LoC pokazuje 26 pozycji napisanych, redagowanych lub współtworzonych przez Dworkin.

PBL w jedynym zapisie odsyła do magazynu feministycznego *Zadra*, a konkretnie tekstu Michała Płaczka *Andrea, Beatryks, ciało*. Tutaj ujawnia się istotna wartość przyjętej metodologii i podejścia zespołu Pracowni Bibliografii Bieżącej do opracowywanego materiału. Z adnotacji możemy się bowiem od razu dowiedzieć, że artykuł „dot. podejścia do kobiecego ciała w świetle teorii Andrei Dworkin nt. pornografii oraz literatury polskiej i obcej (m.in. »Beatryks Cenci« Juliusza Słowackiego); z not. o autorze” (pisownia oryginalna z zachowaniem skrótów)<sup>229</sup>.

Ale nawet sama strona internetowa *Zadry* nie przekazuje informacji, dotyczących treści artykułu – w opisie zawartości numeru mamy do dyspozycji wyłącznie tytuły i nazwiska poszczególnych autorów. Jak wiadomo, konwencja nazewnicza jest kwestią subiektywną i niejednokrotnie odbiega od funkcji ściśle informatywnej wobec czytelnika. Tytuł może stanowić zastosowanie strategii szoku, być nawiązaniem intertekstualnym, czy na zasadzie funkcji ekspresywnej języka wyrażać treść na sposób artystyczny. Nadanie tytułu to także umiejętność kompresji znaczenia w kilku-kilkunastu wyrazach, które zostaną emblematem pełnej zawartości danego tekstu. Celem określenia przydatności treści czy zawartości dla danej kwerendy, warto zatem opisywać je metadanymi. Współcześnie w komunikacji naukowej używa się w tym celu słów kluczowych dla artykułów. Przyporządkowują one publikację do określonego nurtu badawczego czy metodologii, zarysowują pole odniesień i przedmiot badań. Niejednokrotnie wskazują też na konkretne dyskusje w nauce. W serwisach społecznościowych, na blogach i w podobnych miejscach publikacji w sieci tę rolę będą pełnił przede wszystkim tagi.

W przypadku prac naukowych powszechną praktyką informowania o zawartości tekstu jest, oprócz nadania tytułu, przedłożenie abstraktu, gdzie autor formułuje najważniejsze założenia, odniesienia, sygnalizuje metodologie i zapowiada wnioski. Tego rodzaju streszczeń próżno jednak szukać w przypadku publicystyki i pisarstwa blogowego.

*Zadra* stanowi tutaj o tyle ciekawy przykład, że o ile mi wiadomo, była bardziej medium społecznym i publicystycznym, niż naukowym – pomimo oczywistych odwołań do feminizmu

---

<sup>229</sup> Zob. URL zapisu (w bibliografii załącznikowej link do strony głównej): [https://pbl.ibl.poznan.pl/dostep/index.php?s=d\\_biezacy&f=zapisy\\_szczeg&p\\_zapis=1521159](https://pbl.ibl.poznan.pl/dostep/index.php?s=d_biezacy&f=zapisy_szczeg&p_zapis=1521159). Dostęp 2023-07-15.

akademickiego. Gdyby nie uzupełnienie luki informacji przez osobę z zespołu PBL, można by nie dotrzeć do istotnej informacji, jak powiązać ten tekst z osobą Dworkin. Jest to więc wirtualne miejsce służące do połączenia odległych wobec siebie skupisk faktów.

Wyszukiwarka Federacji Bibliotek Cyfrowych (CINIBA) zwraca 0 wyników w wyszukiwarce słów kluczowych<sup>230</sup>. Wyszukiwanie w polu Autor prezentuje 43 wyniki, głównie prace zagraniczne w czasopiśmie. z kolei wyszukiwanie po słowie kluczowym daje już lepszy rezultat: 599 pozycji, z czego większość jest anglojęzyczna i również pochodzi z periodyków. Wśród nich, jak łatwo się domyślić, dużo znajdziemy pozycji feministycznych. Część stanowi recenzje i bezpośrednie odwołania do prac Dworkin, a część przywołuje tę postać w kontekście idei czy wydarzeń.

W przypadku serwisu BazHum zapytanie w wyszukiwarce zwraca jeden rekord, wspomniany już artykuł Natalii Klejdysz<sup>231</sup>. Tak samo dzieje się w przypadku wyszukiwarki Federacji Bibliotek Cyfrowych Pionier<sup>232</sup>.

Przedstawiony powyżej sposób docierania do informacji jest tym, jaki znamy na początku trzeciego dziesięciolecia XXI wieku. Użytkowniczka ma do dyspozycji serwisy rozwijane społecznie, jak i eksperckie – nie wszystkie z nich musi znać, może nawet nie mieć szans dotarcia w obrębie sieci *web* lub poprzez polecenie innej osoby. Oprócz samych serwisów funkcjonują narzędzia profilujące wyświetlanie nam treści. Być może opisane przeze mnie wnioski powinienem nawet opisać np. jako przynależne cis-męskiemu studentowi studiów trzeciego stopnia w Polsce, w grupie wiekowej 30-40, ponieważ tak może widzieć moją działalność w sieci wyszukiwarka. Dotyczy to jednak bardziej doboru źródeł prezentowanych użytkownikowi, a nie ich zawartości. Osoba o innym profilu niekoniecznie otrzyma podobne rezultaty (zwłaszcza jeśli, hipotetycznie, istnieje hokeistka o nazwisku Andrea Dworkin, to osoba o wyraźnym nachyleniu wyszukiwać w stronę tego sportu może zobaczyć wyniki wyszukiwania w postaci tabel rywalizacji ligowej). Może się jednak zdarzyć, że są repozytoria wiedzy na temat Dworkin, do których nie dotarłem. Mając narzędzia automatyzacji takich

---

<sup>230</sup> Zob. URL zapytania (w bibliografii załącznikowej link do strony głównej): „Katalog Bibliotek Uniwersytetu Śląskiego i Uniwersytetu Ekonomicznego w Katowicach - Prolib Integro”, [https://integro.ciniba.edu.pl/integro/search/description?bID=19&\\_lang=pl&q=andrea+dworkin&index=4](https://integro.ciniba.edu.pl/integro/search/description?bID=19&_lang=pl&q=andrea+dworkin&index=4) Dostęp: 2023-07-15.

<sup>231</sup> Zob. URL zapytania (w bibliografii załącznikowej link do strony głównej): „BazHum” <https://www.bazhum.pl/bib/search/results/?generalQuery=andrea%20dworkin>. Dostęp: 2023-07-15.

<sup>232</sup> Zob. URL zapytania (w bibliografii załącznikowej link do strony głównej): „Federacja Bibliotek Cyfrowych”, [https://fbc.pionier.net.pl/search#fq={!tag=dcterms\\_accessRights}dcterms\\_accessRights%3A%22Dost%C4%99p%20otwarty%22&q=andrea%20dworkin](https://fbc.pionier.net.pl/search#fq={!tag=dcterms_accessRights}dcterms_accessRights%3A%22Dost%C4%99p%20otwarty%22&q=andrea%20dworkin). Dostęp: 2023-07-15.

wyszukań, na przykład dysponując sztuczną inteligencją i algorytmami łączenia tożsamych informacji, moglibyśmy jednak analizować dużo więcej danych niż pojedyncza osoba.

Jaką narrację snuje sieć

Twórczość i działalność Dworkin to bardzo zniuansowany temat. Łatwo o śmiałe sądy, nadinterpretacje, wrywanie fragmentów wypowiedzi z oryginalnego kontekstu, insynuowanie poglądów niewyrażonych, jak i innego rodzaju nadużycia. Ryzyka tego upatruję chociażby w uproszczonym rozumieniu tego pisarstwa i aktywizmu. Bardzo łatwo o daleko idące uproszczenia jej twierdzeń, np. prezentowanie feminizmu jako antymaskulinizmu, doszukiwanie się w jej działalności nawoływania do przemocy wobec mężczyzn, a nawet zbrojnego oporu przeciw patriarchy i utrzymywanie ciała w jak najbardziej naturalnym stanie, bez przywiązania do ubioru i kosmetyków jako odrzucenie wbrew przemysłowi modowemu i drogerijnemu. Z dostępnych źródeł audiowizualnych (fragmenty wywiadów, użyte m. in. w filmie biograficznym), widać, że Dworkin sama odnosiła się do niektórych z tych zarzutów, tłumacząc na przykład swoje wywody dotyczące stosunku seksualnego, aby uświadomić, że nie jest przeciwna współżyciu. Jej krytycy zarzucali np. zrównanie seksu z gwałtem lub określenie każdego seksu jako gwałt. Autorka podkreślała, że zwraca uwagę na element agresji w kulturowym obrazie aktu seksualnego jako wartościowany pozytywnie. To, jej zdaniem, prowadzi do społecznego przyzwolenia na gwałt.

Refleksja nad tego rodzaju tematem wymaga umieszczenia publikacji Dworkin i wydarzeń z jej życia na osi czasu, odwołania się do kontekstu epoki, aktualnych wydarzeń (zimna wojna, wojna w Wietnamie), ówczesnego prawodawstwa, statusu kobiet i dyskursu publicystyczno-popkulturalnego, ze szczególnym uwzględnieniem marketingu i reklamy, oferty programów opiniotwórczych i rozrywkowych w mediach, czy wreszcie pornografii, które wtedy wytwarzały m.in. pole semiotyczne dla pojęcia kobiecości i społeczny obraz roli kobiet. Byłbym skłonny zaryzykować hipotezę, że dla współczesnych badaczy wypowiedzi Dworkin mogą budzić podobne reakcje (szok, niezrozumienie), co wtargnięcie Emily Davison na tor wyścigów konnych w 1913 i śmiertelną dla niej kolizję z koniem należącym do króla, co uczyniło ją męczennicą za sprawę politycznej walki kobiet<sup>233</sup>. Owe zmagania są wszakże rozłożonym na dekady procesem. Wyodrębnienie „fal” feminizmu to użyteczna metafora, która dobrze oddaje zróżnicowanie także pod względem geograficznym. Dworkin działała

---

<sup>233</sup> Niejasne są okoliczności incydentu. Davison była zaangażowana w walkę o prawo kobiet do głosu. Miała nie ujawnić nikomu motywów swojego działania podczas wyścigu i istnieje kilka hipotez na ten temat, z których nie wszystkie wskazują na samobójstwo.



głównie w USA i Holandii, ale inne społeczeństwa mogły być w odmiennym momencie tej walki. Dla polskiej czytelniczki obrazuje to najlepiej kontekst tzw. kompromisu aborcyjnego w początkach III RP oraz zaostrzenie przepisów aborcyjnych w Polsce w 2020 roku.

Polskojęzyczna dyskusja (na podstawie płytkiej kwerendy w sieci, dostępnej każdemu użytkownikowi nieprofesjonalnemu), raczej wskazuje na niszowy charakter i wąskie zainteresowanie tą osobą: do drugiej fali feminizmu odwołują się osoby entuzjastycznie nastawione do tego tematu i te wiedzione ciekawością zawodową, np. naukową lub związaną z działalnością na rzecz pomocy kobietom.

Podsumowując, obrany przez nas obiekt analizy (dziedzictwo intelektualne Andrei Dworkin) wymyka się tradycyjnym opisom, znanym humanistycy do tej pory, jeśli spojrzymy nań przez pryzmat jego widoczności w internecie. Patrząc w drugą stronę, aby poznać omawiany temat i polską recepcję autorki, musimy zadać sobie trud rozproszonej kwerendy wśród wielu źródeł, niejednokrotnie cząstkowych. Polska czytelniczka i użytkowniczka internetu nie mają w sieci przewodnika dla zniuansowanego odbioru i dyskusji, a raczej skrawki informacji, rozsianych tu i ówdzie. Łączą je wszakże dane, metadane i odnośniki między nimi. Przy podjęciu wysiłku ekspozycji tych połączeń, jesteśmy w stanie zbudować ontologię danego zagadnienia i poprowadzić zainteresowane osoby w ich poznawczej wędrówce.

Powinniśmy mieć świadomość różnic przygotowywanych źródeł. Sądzę, że dla przeciętnego użytkownika *webu* nie zawsze jest jasne, z jakiego rodzaju przekazem ma do czynienia (reklamowym, eksperckim, opiniotwórczym, publicystyczno-reporterskim, opinią indywidualną). Dążąc do podsumowania, przedstawiam tutaj trzy porządki: społeczną *Web 2.0*, przeciwstawioną kuracji treści przez instytucje dziedzictwa kultury oraz emergentną *Web3* i sieć opartą na tzw. sztucznej inteligencji.

### Serwisy budowane społecznie

W narzędziach *Web 2.0* jak Wikipedia, nad tonem wypowiedzi czuwa moderacja samych użytkowników. Sens i język są mediowane między edytorami, przy zastosowaniu podstawowych zasad formatowania treści, a także dobrych praktyk związanych z obiektywnym przekazywaniem informacji (unikanie stronniczości).

W serwisach specjalistycznych, do jakich zaliczymy bibliografie narodowe i serwisy literaturoznawcze czy encyklopedyczne, obowiązuje kuracja treści przez zespół opracowujący

daną stronę. Mamy tu zatem do czynienia ze „scenzalizowanym” podejściem do określania zarówno tonu, jak i zakresu wypowiedzi. Pozwala to utrzymać jeden standard edytorski.

Szansą może być także UGC (*User-Generated Content*). Jak przekonuje Melissa Terras, *crowdsourcing*, czyli pozyskiwanie rezultatu (w tym wypadku danych) na drodze (najczęściej niewynagradzanej w żaden sposób) pracy lub wolontariatu dużych grup odbiorczych, może również przysłużyć się do tworzenia, adnotacji i modyfikacji danych także w humanistyce cyfrowej<sup>234</sup>. Wszędzie, gdzie przetwarzane są informacje cyfrowe, ich kompletność, weryfikowalność i adekwatność w kontekście stanowią może wyzwanie. Zaangażowanie grupy odbiorczej w uzupełnianie i weryfikację może być odpowiedzią na ten stan rzeczy.

Jedną z wymienionych koncepcji, które mają zapewnić pewność informacji przy automatyzacji powiązań, jest wspomniana sieć semantyczna (*Semantic Web*) – oparta na danych połączonych (*linked data*). Uzgodnienie informacji za pomocą metadanych i łączenie wielu zbiorów pozwoliłoby takie tematy jak omawiany wyżej opisać w bardziej skonsolidowany sposób i przedstawić odbiorcy.

W pewnym sensie różnica między *Web 2.0* a kuracją treści przez zespoły edytorskie i badawcze przypomina greckie rozróżnienie między *episteme* a *doxa* – wiedzą a mniemaniem. Tam, gdzie „powiadają ludzie” na agorze ogólnodostępnych serwisów, mamy do czynienia z organizacją wiedzy na zasadzie konsensusu. Najczęściej jednak typowy użytkownik serwisu społecznościowego nie jest w stanie wyjaśnić dlaczego dane treści są promowane lub zwracane w tym medium. W odniesieniu do wiedzy naukowej o zagadnieniach bardzo specjalistycznych, Konrad Dominas zauważa wyzwanie współczesnej recepcji:

Wielu internautów, przenosząc schematy rozumienia i postrzegania tradycyjnego świata mediów, uważa, że wyznacznikiem Wikipedii jest liczba internautów odwiedzających serwis analogicznie do wielu programów telewizyjnych, które są lub nie są kontynuowane z powodu optymalnej (najlepiej stale rosnącej) lub niewielkiej oglądalności. Ogólnoświatowy charakter projektu, którego każdy internauta może być zarówno odbiorcą, jak i uczestnikiem, połączona dodatkowo z brakiem odpłatności, tworzy przekonanie, że na tym właśnie polega dostęp do informacji: szybko, za darmo i konkretnie. Kwestia wiedzy

---

<sup>234</sup> Melissa Terras, „Crowdsourcing in the Digital Humanities”, w *A New Companion to Digital Humanities*, red. Susan Schreibman, Ray Siemens, i John Unsworth (Wiley, 2015).

pozostaje w tym przypadku zagadnieniem drugorzędym. Wpisanie do wyszukiwarki Google (...) „antycznych” słów [autor odnosi się tu do wpisywania w pole wyszukiwania haseł związanych z antykiem: „Perseus”, „Julius Caesar”, „Homer” – dop. JE] sprawia, że spośród wielu tysięcy stron internetowych Wikipedia zawsze znajduje się w pierwszej dziesiątce rankingu. I na tym właśnie polega fenomen tego serwisu. To nie tylko liczba użytkowników decyduje o jego popularności, lecz także (a może przede wszystkim) mechanizmy funkcjonujące w środowisku cyfrowym – mechanizmy, których nie zawsze jesteśmy świadomi<sup>235</sup>.

Tak jak to było sygnalizowane przy okazji przemian infosfery, obecnie wydaje się, że głównymi motywacjami dla rozwoju sieci semantycznej czy ontologii pozostają władza (w tym powiększanie obszaru już sprawowanego wpływu oraz szeroko rozumiana kontrola grup odbiorczych – internetowy rząd dusz) i zarobki (profilowanie klienta, system rekomendacji zakupowych)<sup>236</sup>. Można sobie wyobrazić inne motywy, bardziej kooperatywne niż konkurencyjne, takie jak: organizacja wiedzy i upowszechnienie dostępu do niej, przyspieszenie komunikacji w nauce oraz popularyzacja nauki, weryfikacja faktów (*fact-checking*), edukacja. Przekłada się to na silniejsze społeczeństwo obywatelskie, posługujące się rzetelną wiedzą i mające dostęp do narzędzi krytycznego sprawdzania prawdziwości twierdzeń w infosferze, jak i zdobywania nowych umiejętności na podstawie umieszczonych w wolnym dostępie informacji.

W obecnej formie przede wszystkim mówimy o demokratycznej stronie mediów budowanych społecznie, uznając równość głosu specjalistów danych dziedzin i osób o niewielkim lub żadnym doświadczeniu w nich. Można jednak prześledzić pewien odpór, z braku lepszego określenia, merytokratyczny. A może po prostu: merytoryczny?

W 2023 roku można zaobserwować pod tym względem interesujące trendy. Owszem, działają pewne ośrodki sensotwórcze w formie *podcastów* (audycji), kanałów strumieniowania wideo, blogów i profili w mediach społecznościowych - można nawet powiedzieć,

---

<sup>235</sup> Konrad Dominas, *Internet jako nowa przestrzeń interpretacji literatury antycznej* (Poznań: Pracownia Humanistycznych Studiów Interdyscyplinarnych, 2017), 153-155.

<sup>236</sup> Por. Jean-Hervé Lorenzi, Mickaël Berrebi, i Justyna Nowakowska, *Przyszłość naszej wolności: czy należy rozmontować Google'a... i kilku innych?*, Wydanie pierwsze, Biblioteka Myśli Współczesnej (Warszawa: Państwowy Instytut Wydawniczy, 2019).

że tworzą internetowe plemiona wokół swoich nadawców – ale wzrasta weryfikowalność informacji wśród dyskutantów. Powoduje to, że gros twórców treści posługuje się rzetelnymi odwołaniami do źródeł i musi rozwijać własny aparat badawczy<sup>237</sup>. Z drugiej strony zachęca badaczy uniwersyteckich, pracowników nauki, do sięgania po społecznościowe formy wytwarzania wiedzy. Dzięki temu tworzy się pomost między akademią a platformami w znacznej mierze rozrywkowymi. Wymienić tu można np. działalność literaturoznawczyni i animatorki kultury Agnieszki Budnik, która prezentuje zjawiska życia literackiego, np. poprzez recenzje nowości i rozmowy z przedstawicielami rynku wydawniczego (oraz samymi twórcami dzieł) pod nazwą *Raport z literatury*<sup>238</sup>. Wykorzystuje do tego m. in. platformy Facebook, Instagram i Spotify. Z kolei w 2023 pojawił się *podcast* Joanny Bednarek i Przemysława Czaplińskiego *Wychodzenie z pochodzenia*, pod auspicjami *Czasu Kultury*<sup>239</sup>. Warto pamiętać, że ton tego rodzaju wypowiedzi jest luźniejszy, a rygor struktury wypowiedzi czy prezentacji wywodu praktycznie nie istnieje i panuje pod tym względem dowolność. Stopień uporządkowania wyznaczają tu kompetencje i decyzje twórców.

Popularnym gatunkiem treści w serwisie Youtube jest tzw. wideoesej. Autorzy kanałów streamingowych wybierają zagadnienie, a następnie starają się w wyczerpujący sposób przedstawić je odbiorcom na różnych poziomach zaawansowania. Najczęściej prezentują argumentację czy podejścia na wzór stanu badań. Dobrą praktyką jest umieszczanie odnośników do źródeł, co pełni funkcję aparatu krytycznego w pracy naukowej (i tak jak w jego przypadku daje co kompetentniejszym widzom szansę na ocenę przygotowania materiału i okazję do krytyki samego opracowania). Tak jak wspominałem, w Polsce trudno znaleźć więcej informacji o pisarstwie Andrei Dworkin, ale jej praca dotycząca konserwatywnych kobiet stanowi mocny punkt argumentacyjny w materiale wideo kanału ContraPoints, dotyczącym J. K. Rowling i jej pozapisarskiej działalności aktywistycznej. Ten i inne odcinki działalności wideoeseistycznej na Youtube mają strukturalnie cechy rozprawki czy prezentacji konferencyjnej w naukach humanistycznych. W tym materiale

---

<sup>237</sup> Jako przykłady zjawiska, które mam na myśli, można podać Podcast Religioznawczy, działający w serwisach streamingowych, oraz profil Sigillum Authenticum w serwisie Facebook. Pierwszy prezentuje badania istotnych w kulturze zjawisk z perspektywy nauki o religii, popularyzując tym samym tę dyscyplinę, a drugi wspiera metodologiczną odpowiedzialność w dyskusjach o początkach państwa polskiego, znaczną część działalności poświęcając na krytykę warsztatu autorów opisujących współczesny fantazmat tzw. Wielkiej Lechii.

<sup>238</sup> Zob. „Raport z literatury”. Dostęp 2023-07-31. <https://www.raportzliteratury.pl/>. Dostęp: 2023-07-31.

<sup>239</sup> Dostępne do odsłuchu na profilu pisma w serwisie Soundcloud, zob. „Czas Kultury”, Soundcloud, dostęp 2023-07-31, <https://soundcloud.com/user-80673682>. Dostęp 2023-09-16.

ContraPoints użycie cytatów z Dworkin przydaje całości walorów pod względem kompozycji i argumentacji.

W dodatku w roli *influencera* osoba tworząca treści do serwisów społecznościowych często intencjonalnie przerysowuje tak prezentowany materiał, jak i sposób jego podania widzowi czy słuchaczce. Popularyzator nauki, zgodnie ze źródłosłowem innego miana, „amator”, będzie więc z umiłowania przedmiotu prezentować stan wiedzy i w gruncie rzeczy wskazywać przykłady najbardziej dla siebie inspirujące z jakiegoś większego obszaru wiedzy. W podejściu naukowym istniałoby jednak ryzyko posądzenia o *cherry picking*, czy dowody anegdotyczne. Badacze uniwersyteccy w mediach społecznościowych częściej wprowadzają jednak swoje wyuczone podejście i „myślenie metodologią”, choć nie bez spinu walki o uwagę widza i zasięgi docierania własnych treści.

Mamy też do czynienia z wieloma aktywnościami środowisk naukowych w kontekście społecznościowego *Webu*, jeśli weźmiemy pod uwagę profile instytucji. Tematycznie bliżej naszych rozważań będą zatem – obecne na Facebooku – Centrum Humanistyki Cyfrowej IBL PAN, Nowa Panorama Literatury Polskiej, czy wreszcie Pracownia Bibliografii Bieżącej IBL PAN<sup>240</sup>. Wszystkie wymienione grupy badawcze informują na swoich profilach głównie o bieżącej działalności: wydarzeniach *online* i *offline*, rozwijanych projektach, z rzadka komentując sprawy bieżące, czy odwołując się do trendów. Warto jednak pamiętać, że te działania należą do domeny popularyzacji nauki i informowania o aktualnościach, a nie są celami tych instytucji, wyjąwszy może cel rozpowszechniania wyprodukowanych treści (jak to ma miejsce np. przy publikacji projektów cyfrowych). Przestrzeń wymiany myśli między badaczami stanowi także X (Twitter). Nieco bardziej zwięzły i dynamiczny jako platforma społecznościowa, zindywidualizowany i bezpośredni, daje okazję do debat, jak i publikacji własnych badań. Wyżej wymienione instytucje również korzystają z tego narzędzia lub mają na X (Twitterze) swoich przedstawicieli w postaci prywatnych kont ich pracowników.

Media społecznościowe, ze względu na mechanizmy SEO, stanowią też dobry punkt dotarcia – odnośnik do treści właściwej. Profil na Facebooku czy X (Twitterze) bywa dostępniejszy i lepiej wypozycjonowany do wyszukania od głównej strony instytucji lub projektu, który ta instytucja rozwija. Publikowanie postów, związanych z działalnością, poprzez sieć

---

<sup>240</sup> Działalność PBB IBL PAN w mediach społecznościowych omawiam osobno w rozdziale trzecim.

interakcji i „zalecanie się do algorytmów” stwarza szansę na przywiedzenie czytelnika-użytkownika do części infosfery bliżej związanej z nauką.

Co również charakterystyczne, media społecznościowe dają możliwość maszynowego przetwarzania. X (Twitter) udostępnia API, za pomocą którego można pobrać wpisy z wybranych kont i analizować ich treści. Facebook pozwala pracować właścicielom kont jako nadawcom przekazu na grupie odbiorczej i zasięgu dotarcia – w dwie strony, jako analiza odbiorców oraz poziom, w jaki treści „performują” czyli wywołują reakcje, są czytane i udostępniane dalej. Youtube steruje z kolei kreatorami zawartości, aby dostosowywali się do trendów oglądalności. Stąd powstają miniatury do umieszczanych w tym serwisie materiałów wideo przedstawiające twarze wyrażające silne emocje, stosowanie wewnętrznych podziałów w filmach, aby użytkownik mógł przeskoczyć do określonego fragmentu, czy opatrywanie odnośnikami wewnątrz materiału oraz w jego opisie.

Wszystkie te serwisy wykorzystują także wewnętrzne algorytmy, profilujące użytkowników i serwujące im treści o spodziewanym najlepszym rezultacie, czyli zaangażowaniu i pozostawaniu na platformie. I właśnie algorytmy programistyczne są ostatnią warstwą w niniejszym porównaniu, ale żeby do nich przejść, konieczne jest omówienie scentralizowanych sposobów kuracji treści, niezależnych od publiczności.

### Kuracja treści

Mamy także podejście rozwijania treści przez „redakcje”: mam tu na myśli sytuację instytucji dziedzictwa kulturalnego, które dokumentują piśmiennictwo w ramach swoich celów i za pomocą zatrudnionych specjalistów. W narracji o *Web 2.0* będzie tu widoczna opozycja serwisów Britannica Online i Wikipedia, gdzie ten pierwszy serwis stawia na kurację treści przez zespół dokumentalistów, a drugi oddaje tworzenie zawartości w ręce użytkowników, jak to zostało opisane wyżej. Większość serwisów opartych na danych bibliograficznych, po które sięgałem, szukając informacji o Dworkin, to właśnie inicjatywy posiadające stałe kadry twórców treści.

Tomasz Umerle proponuje ciekawą optykę, by opisać działalność tych instytucji: dokumentację kultury możemy widzieć jako proces zbierania danych<sup>241</sup>. Autor opisuje to podejście w schemacie triady przedmiotu dokumentacji (*object of documentation*), mediatora (*evidence* – świadectw, dowodów, materiału badawczego) i projektu dokumentacji

---

<sup>241</sup> Umerle, Tomasz. „Rethinking the Potential of Documentation of Culture as a Data Gathering Practice”. *Proceedings from the Document Academy* 4, nr 2 (2017-12-12). <https://doi.org/10.35492/docam/4/2/15>.

kultury (*DoC Project*), posiadającego określoną metodologię. Dla PBL mediatorem będzie książka lub czasopismo, a obiektem opisywanym – „to, co zdefiniowano jako »literackie«”. Co interesujące, soczewka przyjętych założeń sprawia nieraz, że – słowami autora – może to przypominać ustanowienie danego obiektu dziełem sztuki poprzez umieszczenie go w galerii. Przykładowo, jeśli nasza baza danych ma za zadanie zbierać arcydzieła powieściowe, to odbiorca otrzyma komunikat zarówno o tym, że zawartość jest złożona z informacji na temat wybitnych tekstów literackich, ale także pośrednio – że te, których nie opisano, nie kwalifikują się do tego miana. Umerle wskazuje na dwojaką sytuację projektów dokumentacji kultury (DoC) wobec rzeczywistości danowej: z jednej strony nie gromadzą one fizycznych obiektów, więc są lepiej przystosowane do zadania cyfrowego rozwoju. Z drugiej, mierzą się z przeszkodami natury technicznej, co sprawia, że ich działalność jest interesująca i złożona. Badacz szacuje na podstawie informacji z instytucji dokumentujących kulturę, że zdigitalizowano tylko około jednej piątej informacji, które wytwarzają i które można przechowywać jako bazy danych. Byłyby to cenny wkład do przetwarzania dla inżynierów. Umerle zauważa tu, że dokumentacja kultury *de facto* wytwarza metadane; wymienia wszakże dwie trudności praktyczne w danetyzacji omawianego przez siebie pola. Pierwszą stanowi (fałszywa) wiara w *panaceum* techniczne, rozwiązanie, które w automatyczny sposób wykona znaczną część pracy bez wkładu instytucji; towarzyszy mu też przekonanie o wystarczającej obecnej ilości dostępnych danych. Powoduje to, że wytwarzanie metadanych wydaje się nieintuicyjne czy niepotrzebne. Drugim wyzwaniem jest tu luźne powiązanie metadanych z obiektami. Przy digitalizacji samych zbiorów i ogólnie określonych obiektów, jak obrazy, teksty i inne dzieła, odbiorcy mają kontakt z zapośredniczeniem, które rozumieją i mogą z niego skorzystać (jako zastępnik kontaktu z realnym obiektem). Wytwarzanie i archiwizacja metadanych to praktyka dużo bardziej abstrakcyjna, a przez to niekoniecznie da się ją łatwo przełożyć na doświadczalny pozytywny efekt.

Szansa, jaką widzi Umerle w tej sytuacji, to pójście w kierunku zasilania dużych zbiorów danych i użycia maszynowych odczytań takich zbiorów dla analizy w większej skali. Oznacza to zarówno interdyscyplinarność (łączenie metadanych z różnych dziedzin wiedzy), jak i wolumen (wykraczający poza pojedynczy projekt DoC). Do tego potrzeba jednak uzgodnienia formatów danych i budowania ontologii informacyjnej oraz perspektywy połączenia z inicjatywami *data science*. Autor wymienia tu trzy ważne kierunki działania, związane z podejściem: 1) stałe dokumentowanie dziedzictwa kultury jako proces długofalowy, nastawiony na odległy efekt, ale w bliskiej perspektywie tworzenie rozwiązań

przyjaznych obecnym użytkownikom; 2) implementowanie technologii, które zapewnią połączenia między różnymi zbiorami danych, a jednocześnie zachowają „pamięć metody”, tzn. założenia danego projektu DoC; 3) transparentne komunikowanie jak cyfrowe narzędzia wpływają na odbiorcę, jaki obraz kultury wytwarzają.

## Dostęp

Serwisy społecznościowe są, jak zostało to wspomniane o *Web 2.0*, związane z działalnością biznesową. Pod względem archiwizacji i dostępu do treści, użytkownicy pozostają zdani na łaskę właścicieli takich platform. Może to niejednokrotnie oznaczać bezpowrotną utratę danych. Przykładem może być zamknięcie projektu społecznościowego Google+ przez właściciela, co doprowadziło do utraty wszelkich danych użytkowników, o ile nie zostały zapisane gdzie indziej (np. jako kopia zapasowa, *backup*)<sup>242</sup>.

Możemy sobie wyobrazić taki scenariusz, w którym podobna sytuacja spotyka wspomniane wcześniej serwisy agregujące wiedzę o kulturze, a rozwijane pracą ich użytkowników. W razie ich zamknięcia, znikną wirtualne półki biblioteczek, np. z etykietami „chcę przeczytać” czy „ulubione”. Dla wielu osób to nie tylko swego rodzaju budulec tożsamości, ale także przedłużenie aparatu poznawczego i zewnętrzna pamięć – coraz częściej w rozmowach sięgamy po wsparcie własnego konta serwisu społecznościowego, by przypomnieć sobie tytuł ostatnio oglądanego filmu (i porównać, jakie wystawiliśmy mu oceny) albo zweryfikować szczegóły obsady. Wszystkie te informacje trwają w sieci tak długo, jak długo platformy z nimi związane są rentowne i jak długo użytkownicy pozwalają im zarabiać w swoich modelach biznesowych (np. poprzez interakcje z treściami reklamowymi czy opłaty abonamentowe). Długie trwanie archiwów i kolekcji powinno być priorytetem, ale tylko z perspektywy ochrony dziedzictwa. Zapewnienie stałości dostępu do zasobów wiedzy należy do domeny zarządzania w kulturze, nauce i szkolnictwie.

## Coraz większe wolumeny danych

Z każdym rokiem rośnie wolumen przetwarzanych w internecie danych. Wymaga to rozwoju zaplecza dla ich przechowania i przeliczania. Mark P. Mills w książce *Digital Cathedrals*<sup>243</sup> porównuje obliczenia w chmurze (*cloud computing*) do tytułowych cyfrowych katedr, co dla humanisty buduje paralelę z doskonaleniem średniowiecznej sztuki – być może nieco tajemnej,

---

<sup>242</sup> Ron Amadeo, „Google+ shuts down April 2, all consumer data will be deleted”, *Ars Technica*, 2019-01-31, <https://arstechnica.com/gadgets/2019/01/google-shuts-down-april-2-all-data-will-be-deleted/>. Dostęp 2023-09-16.

<sup>243</sup> Mark P. Mills, *Digital cathedrals*, First American edition (New York, New York: Encounter Books, 2019).



ale z pewnością kolektywnej – tworzenia infrastruktury reprezentującej ideę większą od każdego z pojedynczych projektantów-budowniczych.

Autor porównuje wynalazek i praktykę obliczeń (przechowywania danych) w chmurze i pracę na dużych zbiorach danych do takich przełomów jak wynalezienie kanałów, kolei, telegrafu, rurociągów ropy, dróg, sieci telefonicznej, infrastruktury lotnictwa i wreszcie samego internetu. Jest to bardzo inżynierska perspektywa, gdyż narrację determinują wynalazki, które odmieniają codzienne życie przeciętnej osoby (użytkownika-odbiorcy). Mimo swoistego technoentuzjazmu, możemy jednak uznać, że wymienione przykłady naprawdę odcisnęły ważny ślad na cywilizacji i mają masowe, dalekosiężne skutki na całych populacjach. Powołując się na badania Arnulfa Grüblera<sup>244</sup>, Mills wskazuje raptowność zmian, tzn. każdy z tych wynalazków wymagał olbrzymiego skoku w rozwoju infrastruktury. W latach dwudziestych XXI w. jesteśmy dopiero na początku takiego wzrostu systemu danych w chmurze. Badacz zauważa, że każdy ze wskazanych trendów technologicznych po gwałtownym rozwoju w końcu osiągał szczytowy poziom i przechodził do stagnacji, lecz w przypadku *Big Data* nie widzimy jeszcze horyzontu tego wyhamowania.

Mills odwołuje się do idei danych jako „nowej ropy” (*data is the new oil*) – z tą różnicą, że konsumują one energię, a nie stwarzają. Rodzi to potrzeby dokładnego porządkowania, opisu, ale także zapewnienia materialnych środków do utrzymania działającego systemu, odpowiadającego za wirtualną reprezentację informacji oraz optymalizacji kosztów (tak finansowych, jak środowiskowych)<sup>245</sup>.

Futuryzm, czyli archiwistyka dla maszyn

W ostatnich latach karierę robi także pojęcie *Web3*, czyli sieci opartej na dostępie do treści za pomocą technologii *blockchain*<sup>246</sup>. Ważnym kontekstem dla *Web3* są płatne subskrypcje

---

<sup>244</sup> Arnulf Grübler, „Time for a Change: On the Patterns of Diffusion of Innovation”, *Daedalus* 125, nr 3 (1996): 19–42.

<sup>245</sup> Istotnym czynnikiem jest środowiskowy aspekt działania internetu i *cloud computing*: wpływ na otoczenie i emisyjność tego sektora. Serwery wymagają chłodzenia i stałego dostarczania energii. Każde odpytanie o dane angażuje konkretną część infrastruktury, gdzie dane się przechowuje. Przedstawiciele branży stale szukają optymalizacji impactu środowiskowego, ale osławiona zeroemisyjność przedsięwzięć jest często możliwa jedynie na drodze wykupywania pakietów-rekompensat np. sadzenia lasów w innym miejscu dla zrównoważenia wytwarzania zanieczyszczeń w miejscu działania farm danych (*carbon footprint offsetting*).

<sup>246</sup> *Blockchain* pozwala na archiwizację „zdarzeń” w sieci w rozproszony sposób – na sieci wielu komputerów, zamiast w jednej, scentralizowanej bazie danych. Najpopularniejsze konteksty *blockchain* to unikalny dowód wytworzenia kryptowaluty, uprawniający do posiadania jej przez użytkownika i tzw. non-fungible tokens (NFT), czyli unikatowe cyfrowe obiekty. Pierwotnie opracowana jako gwarant transakcji kryptowalut, może zostać wykorzystana także jako identyfikator osób – cyfrowy paszport. Nieprzypadkowo to właśnie portfel ma stanowić istotną część internetowej tożsamości użytkownika. *Web3* to scenariusz zarazem wolnościowy, jak i silnie ekonomiczny, ponieważ zakłada zaangażowanie finansowe osoby korzystającej z sieci w daną

treści i współwłaścicielstwo użytkowników. Możemy sobie wyobrazić sytuację, w której także dostęp do wiedzy będzie ograniczony i zależny od podmiotów komercyjnych – tym większa potrzeba utrzymywania infrastruktury w wolnym dostępie.

Algorytmiczna sieć, oparta na sztucznej inteligencji, która dopiero się wyłania, jest – przynajmniej tak, jak dziś ją rozumiemy – mechaniczną obróbką treści, pozbawioną komponentu ludzkiego poznania. Aspekt kognitywny tych rozwiązań opiera się na skojarzeniach analogicznych do ludzkiej sieci neuronowej (co jest dosłownym przeszczepieniem mechanizmów biologicznych na komputerowe jeśli chodzi o zasadę działania i taką nazwę noszą), ale nie są to jeszcze *stricte* rzeczywiste „skojarzenia” i praca umysłu, a bardziej procedura algorytmiczna<sup>247</sup>.

W 2023 roku wytwory działania narzędzi nazywanych dzisiaj AI budzą całą gamę reakcji od rozbawienia, przez podziw, po bunt osób, które obawiają się utraty pracy. Ci ostatni mają dodatkowo słuszne pretensje o wykorzystanie w bezprawny sposób ich dzieł do tworzenia modeli treningowych maszyn. Mowa tu nie tylko o twórcach grafiki czy scenarzystach filmowych Hollywood, ale nawet postaciach tak rozpoznawalnych dla polskich humanistów jak pisarka Rebecca Solnit<sup>248</sup>. Aby sieć neuronowa mogła wykonywać zaordynowane jej operacje, musi przeciwżyć swoje działania metodą prób i błędów na zestawach danych, podanych przez człowieka. Najczęściej także człowiek asystuje przy ocenie wyniku, aby wyuczyć narzędzie pożądanego efektu (przez wielokrotne wskazanie, jaki *output*-wynik jest do zaakceptowania, a jaki do odrzucenia).

Większość rozwiązań, obecnie nazywanych AI, to duże modele językowe – zautomatyzowane przetwarzanie tekstu. Nie są obdarzone świadomością, nie przetwarzają informacji krytycznie. Maszyny pytane obecnie, mogą się mylić za pierwszą odpowiedzią, ale powoływane są całe

---

społeczność oraz społeczne wytwarzanie wartości na zasadzie konsensusu (porównuje się to do wartości kolekcjonerskich kart bejsbolowych, które są warte wiele dla ich zbieraczy – wysoka wartość zostaje ustalona w grupie zainteresowanych tematem). Sama idea ekosystemu kryptowalut pozostaje też próbą uniezależnienia się niektórych grup od banków – ze wszystkimi konsekwencjami. Nie sposób bowiem pominąć kontrowersje i negatywny wpływ blockchain: produkcja kryptowalut zużywa duże zasoby energetyczne, główna waluta (Bitcoin) w dość regularnych odstępach zalicza gwałtowne spadki wartości, a obietnica anonimowości i niezależności od aparatu kontroli państwowej przyciąga do tego rynku grupy przestępcze. Przytaczam tu główne założenia za: Stackpole, Thomas. „What Is Web3? Your guide to (what could be) the future of the internet.” Harvard Business Review, 2022-05-10. <https://hbr.org/2022/05/what-is-web3>. Dostęp 2023-09-16.

<sup>247</sup> Por. Sean Gerrish i Filip Fierek, *Jak myśleć inteligentne maszyny*, Wydanie I (Warszawa: PWN, 2020). Zob. także Marek Kasperski, *Sztuczna Inteligencja* (Helion, 2015).

<sup>248</sup> Shirin Ghaffary, „Author Rebecca Solnit Found ‘Half’ Her Books in a Dataset Used to Develop AI”, *Bloomberg*, 2023-10-05, <https://www.bloomberg.com/news/newsletters/2023-10-05/rebecca-solnit-found-half-her-books-in-a-dataset-used-to-develop-ai#xj4y7vzkg>. Dostęp 2023-09-16.

zespoły ludzi do tworzenia zapytań (*prompts*), które pozwalają osiągać zamierzone cele. W produkcji kodu programistycznego może to polegać na tym, iż pierwszy kod, o który pyta operator, będzie wygenerowany błędnie, ale po modyfikacji zapytań maszyna odwoła się do bardzo dużej dokumentacji w sieci związanej z naprawianiem błędów (serwisy specjalistyczne, poświęcone zadawaniu pytań i rozwiązywaniu problemów, w działalności programistycznej najpopularniejszy jest StackOverflow<sup>249</sup>). Dzięki temu pożądaný efekt jest osiągnięty za n-tą iteracją, po uwzględnieniu informacji z sieci.

Specjaliści, zadając dziś sztucznej inteligencji opisanie zjawisk w swoich dziedzinach, opisują to jako „halucynacje” i konfabulacje. Zdania języka naturalnego są złożone poprawnie, ale ciągi przyczynowo-skutkowe nie zawsze są zachowane, wygenerowane obrazy przedstawiające ludzi przedstawiają nadreprezentację osób z polidaktylią względem całej populacji.

Problemem jest też „grodzenie” wiedzy dla takich narzędzi, to znaczy np. korzystanie z zasobów pozostających w otwartym dostępie – albo ściśle przygotowanych. Przez takie wytyczenie granic, AI nie ma możliwości swobodnego poruszania się w infosferze, ale raczej daje tak dobre odpowiedzi, jak pozwala na to zbiór danych, na którym się ją trenuje.

W związku z *feedbackiem* społecznościowym, przynależnym do porządku Web 2.0, zdarzył się przypadek, w którym osiągnięto skutek odwrotny od zamierzonego i bot konwersacyjny w mniej niż dobę zaczął powielać nienawistne komunikaty. Odpowiedzialność poniosły za to właśnie „treningowe” treści, złośliwie podsunięte przez użytkowników<sup>250</sup>. Stanowi to pewną granicę dla kolektywnie tworzonej wiedzy, bez postawienia granic. Technooptymistyczne podejście, w którym korzystający z internetu są grupą racjonalnych jednostek, dążących do jak najlepszej wymiany informacji, zderzyło się z intencjonalnym podkopaniem projektu (*trolling*).

Same AI, kiedy trenuje się je, aby sprawdzić negatywne skutki ich użycia, celowo dając instrukcje prowadzące do negatywnych konsekwencji (*red teaming*), traktują to jak kolejne zadanie do wykonania. Istnieją w kulturze przykłady „złych sztucznych inteligencji” i dzięki nim modele językowe mogą odnaleźć wzorce zachowań. Jest to ciekawe sprzężenie zwrotne,

---

<sup>249</sup> Zob. „Stack Overflow”, <https://stackoverflow.com>. Dostęp: 2022-10-01.

<sup>250</sup> Studium przypadku, por.: Medium. „Microsoft shuts down ‘Tay’ AI”. <https://medium.com/collections/2dca30d0ee6a>. Dostęp 2023-09-16.

ponieważ narzędzia maszynowe rozumieją scenariusze, w jakich mogą się zachowywać, dzięki istniejącym uprzednio tekstom, jakie wyprodukowała ludzkość.

Nie doświadczyliśmy jeszcze zatem momentu Singularity, którego nadejście przewiduje Ray Kurzweil<sup>251</sup>. Różne firmy i organizacje opracowywały swoje narzędzia, ale posiadały też komisje etyczne i wewnętrzne mechanizmy, które zabezpieczały owe projekty przed niepożądanymi konsekwencjami dla społeczeństw, technologii, ekosystemów i ogólnie pojętego środowiska, ale to wyścig rynkowy wymusił publikację narzędzi AI<sup>252</sup>.

Dane bibliograficzne są tylko jednym z typów informacji, jakie mogą przetwarzać modele językowe, potocznie nazywane sztuczną inteligencją. Ze względu na ścisłe podejście i standaryzację lub stosowanie powtarzalnych schematów opisu (format i model danych) mają one szansę stać się punktem odniesienia w ustalaniu faktów, kiedy spotykamy się z zafałszowaniami, wynikającymi z automatycznego przetwarzania zawartości tekstowej przez maszynę. Przez to, że intencja gromadzenia, przechowywania i przetwarzania danych bibliograficznych (i szerzej: danych o kulturze) jest związana z dokumentacją dziedzictwa, a nie np. sprzedażą produktów czy dezinformacją, tego rodzaju zbiory mogą pomóc „zakotwiczyć” odbiorcę przy zaufanej instancji odwoławczej. Taki punkt odniesienia zapewnia weryfikowalność napotkanych informacji. W obszarach dokumentacji kultury pojawia się także pogląd, że AI ma potencjał wspierania digitalizacji dziedzictwa, która wciąż pozostaje na niewystarczającym poziomie. Dyrektor generalny fundacji Europeana szacował w 2018 roku, że było to zaledwie 10% całości, zwracając uwagę na znikomy odestek otwartości do ponownego użycia dla materiałów audiowizualnych<sup>253</sup>.

### Modele świadomości sieci

Jussi Parikka proponuje spojrzenie na media jako struktury organiczne, odwołując się do modeli naukowych, związanych z owadami. Pozornie to daleka paralela, ale w obu przypadkach – gdy mowa o wymianie informacji w mediach elektronicznych oraz kiedy mamy na myśli życie insektów – opisuje się struktury (samo)organizacji, emergentnych hierarchii, klastrowania (roju) itp. zjawisk<sup>254</sup>. Opisane zostają także metamorfozy. Parikka nie

---

<sup>251</sup> Por. Kurzweil, Ray. *The singularity is near: when humans transcend biology*. New York: Viking, 2005.

<sup>252</sup> Zob. Andrew R. Chow i Billy Perrigo, „The AI Arms Race Is Changing Everything”, *Time*, 2023-02-16, <https://time.com/6255952/ai-impact-chatgpt-microsoft-google/>. Dostęp 2023-09-16.

<sup>253</sup> Harry Verwayen, „Could AI and data mining technologies overcome issues in cultural heritage?”, <https://pro.europeana.eu/post/could-ai-and-data-mining-technologies-overcome-copyright-issues-in-cultural-heritage>.

<sup>254</sup> Jussi Parikka, Mateusz Borowski, i Małgorzata Sugiera, *Owady i media*, Interpretacje 78 (Kraków: Księgarnia Akademicka, 2017).

tylko opisuje media z perspektywy owadziego życia, ale także wskazuje na mediowanie modeli wiedzy. Z jednej strony, biologia opisuje owady, doszukując się wzorców matematycznych. Z drugiej strony, matematyczne wzorce opisujące media (jak u Manovicha, kultura wyrażona liczbowo), opisuje się owadziemi metaforami. Praca badacza skupia się na systemach organizacji oraz interakcjach i stanowi jeszcze jeden sposób opisu mediów, z którymi się stykamy – możliwe, że bardziej osławający od narracji o złowrogich AI.

Ciekawą optykę proponuje w tym kontekście Elżbieta Winiecka, postulując przenikanie się pojęć literackości i sieci. Autorka zwraca uwagę zarówno na sieciowość tekstu elektronicznego (literatura ergodyczna, cyfrowa), jak i tekstualny charakter *webu*<sup>255</sup>. Oto internet jawi się jako tekst, pełna przypisów i powiązań księga i biblioteka jednocześnie (teksty są bowiem splecione ze sobą, nie jest to ani jeden, ani jednolity tekst). Rozwinięciem tych dywagacji staje się odwołanie do badań Sandy’ego Baldwina<sup>256</sup> w nurcie psychoanalitycznym: z jednej strony z naciskiem na tekstualność sieci (możemy też chyba zaryzykować stwierdzenie: narracyjność w sieci), z drugiej: nieświadome treści, takie jak związane z plikami *cookies*, profilowaniem użytkownika i umieszczaniem mechanizmów wpływających na doświadczenie korzystania z internetu w tle, poprzez kod programistyczny i jego algorytmy<sup>257</sup>. Jeśli ta personalizacja idzie za daleko, każdy korzystający z sieci może mieć zupełnie różne informacje. Można tu mówić o pewnym podziale. Upatruję tu więc dla kuratorów treści spod znaku Britanniki (instytucji takich jak PBL), roli scalającej różnorodne doświadczenia, idąc tropem psychoanalitycznym. W jakimś sensie takie ośrodki mogą kotwiczyć chaotyczny ruch odbiorcy za pomocą gromadzonej, przechowywanej i udostępnianej wiedzy – stanowić odwołanie „terapeutyczne”, odkrywające treści nieuświadomione, wyparte i celowo przed użytkownikiem zakrywane. Ujawnianie faktów życia literackiego, czy szerzej - kulturalnego, obnaża symulakry i fantazmaty. Podsumowanie rozważań Winieckiej na ten temat zalicza *in plus* poetyckie podejście Baldwina, wskazując jednakże na braki. Poznańska autorka powołuje się tu na recenzję Tomasza Mizerkiewicza, w której czytamy o możliwym przekroczeniu pozornie

---

<sup>255</sup> Winiecka, Elżbieta. Poszerzenie pola literackiego: studia o literackości w internecie. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas, 2020: 39-61.

<sup>256</sup> Baldwin, Sandy. The Internet unconscious: on the subject of electronic literature. International texts in critical media aesthetics, vol. 9. New York, NY: Bloomsbury Academic, an imprint of Bloomsbury Publishing Inc, 2015. Podaję za : Winiecka, Elżbieta. Dz. cyt.

<sup>257</sup> Por. również: Sumpter, David J. T, i Radosław Kosarzycki. Osaczeni przez liczby: o algorytmach, które kontrolują nasze życie : od Facebooka i Google’a po fake newsy i bańki filtrujące. Wydanie I. Kraków: Copernicus Center Press, 2019. Jak na ironię mój menadżer bibliografii, Zotero, zaserwował dla numeru ISBN tej pozycji tytuł zgodny z graficznym przedstawieniem z okładki, czyli „Osaczeni przez l1cz3y” – zgodnie z przyjętą praktyką nauk pomocniczych polonistyki, pozostawiam tytuł polski wydawcy, opowiadając się po stronie eksperckiej kuracji treści.

nierozwiązywalnych problemów, opisywanych przez Baldwina poprzez konfrontację z Lacanowskim Realnym i ukonstytuowanie się niezależnego, samoświadomego podmiotu<sup>258</sup>. Te uwagi osadzające doświadczenie korzystania z sieci są szczególnie interesujące, gdy uświadomimy sobie (pozwolimy przeniknąć tej informacji do świadomości?), iż poszukiwanie informacji w internecie nie jest wyłącznie uwarunkowane wyborem użytkowniczką, a skrywa zakulisowe mechanizmy proponujące konkretne treści, czy angażujące uwagę w celu maksymalizacji zysku. Jak było to widoczne na przykładzie recepcji Dworkin w polskim internecie, wiedza na dany temat może być rozszkana, a próba wyrobienia sobie opinii w określonej kwestii może napotkać przeszkody, czy – jak chciałby Baldwin i jego komentatorzy – zawiera elementy samopoznania.

#### Uwagi końcowe

Wracając teraz do przypadku Dworkin, wróć jeszcze do spekulacji o postulowanym charakterze kwerendy na jej temat. Dzięki tagowaniu społecznemu, jesteśmy w stanie połączyć więcej wątków tematycznych za pomocą sieci danych i metadanych – pod warunkiem, że znajdziemy jednolity sposób ich opisu do agregacji. Serwisy oparte na kuracji treści możnaby wykorzystać do ustalenia ramy metadanych i swoistego wzorca rekordu. Źródła społeczne – do zasilenia zbioru o brakujące informacje. Polska recepcja mogłaby być opisana obok informacji ze świata. Do tego audiowizualne materiały dostarczają wiedzy, pod warunkiem, że zostają przeanalizowane pod kątem opisu lub anotowane.

I tu pojawia się doniosła rola automatów, nakierowanych na konkretne zadania przez ludzkiego operatora i samouczących się. Wszystkie te informacje mogłyby zostać skompilowane przez odpowiednie narzędzie AI, o ile zapewnimy sposoby opisu za pomocą metadanych, czyli podłoże ontologii informacji. Kwerenda stałaby się tutaj zautomatyzowana, co ma istotne znaczenie tak dla piśmiennictwa literaturoznawczego, jak i dla dydaktyki. Oto zarówno krytyczka literacka, jak i wykładowca, mogliby w łatwy sposób określić zakres źródeł do przebadania i po wspomaganym maszynowo zapytaniu otrzymać spójny raport stanu wiedzy na taki temat, jaki widzieliśmy wcześniej w rozproszeniu.

Żaden najlepszy biograf ani bibliograf nie dotrze do całego zbioru informacji wyprodukowanych na temat takiej skali przez ludzkość. Jeśli dotrze – to niekoniecznie je pozna

---

<sup>258</sup> T. Mizerkiewicz, *The Internet Unconscious Sandy'ego Baldwina a teorie nieświadomości pisania elektronicznego*, „Forum Poetyki” wiosna/lato 2016, [http://fp.amu.edu.pl/wp-content/uploads/2016/05/TMizerkiewicz\\_TheInternetUnconsciousSandyegoBaldwina\\_ForumPoetyki\\_wiosna\\_lato\\_2016.pdf](http://fp.amu.edu.pl/wp-content/uploads/2016/05/TMizerkiewicz_TheInternetUnconsciousSandyegoBaldwina_ForumPoetyki_wiosna_lato_2016.pdf) (dostęp: 29.10.2019), s. 141. cyt. za: Winiecka Elżbieta, dz. cyt.

(przeczyta każdy unikatowy tekst kultury). A jeśli przeczyta – istnieje ryzyko, że nie przekaze innym tych informacji, przedkładając w zrozumiałej, ustrukturyzowanej formie. Nie próbuję tutaj w temacie ontologii informacji powtarzać za sofistami wątpliwości o świecie. Zwracam raczej uwagę na trud obróbki wiedzy. Praca bibliografa także niejednokrotnie jest osobliwą orką: zbiera informacje, ale ich nie konsumuje. Przygotowuje wszakże dane jak najdokładniej, ponieważ kieruje je w stronę odbiorców o różnorodnych oczekiwaniach (bibliografia może być mapą nieznanego ładu, listą sprawunków lub testem sprawności, w zależności od doświadczenia i ambicji badawczych). Jak starałem się również pokazać w odniesieniu do spekulatywnych wątków, związanych z rozwojem technologii, dane mogą także trafiać do podmiotów sieci o różnorodnych aparatach poznawczych, często nie-ludzkich, o zmieniającym się poziomie autonomii.

Przedstawiłem powyżej zgrubny opis trzech podejść do przechowywania, udostępniania, przetwarzania, zdobywania oraz produkcji wiedzy w internecie. Większość *webu* dominują treści społeczne spod znaku *Web 2.0* (społecznego internetu), ze względu na skalę działalności serwisów społecznościowych. Oprócz tego istnieją serwisy skupione na dokumentacji i pielęgnowaniu dziedzictwa (historycznego, kulturalnego itp.). Mniejszy rozmach rekompensują rzetelną, ekspercką analizą. W przyszłości coraz częściej analizy informacji będą dokonywać nie-ludzcy aktorzy sieci *web*: *boty* i *crawlers*, czy sieci neuronowe. Dlatego tak ważne jest, aby akademia wraz z instytucjami archiwizującymi naukę, kulturę i sztukę, wspólnie wykorzystywały dynamiki pozostałych dwóch systemów (społecznościowych i automatycznych), aby w przyszłości zadbać o rzetelny opis swoich dziedzin i dać do niego dostępną ścieżkę dotarcia dla poszukujących sprawdzonej informacji czytelników – ludzkich i nie-ludzkich. Parafrazując słowa aktu fundacyjnego Akademii Zamojskiej, takie będą narzędzia sztucznej inteligencji, jakie jej zasobów danych chowanie.

# Rozdział 3. PBL jako przykład działalności przynależnej do humanistyki cyfrowej

PBL: od bibliografii do bibliograficznej bazy danych

Polska Bibliografia Literacka, dostępna online dla użytkowników, działa nieprzerwanie od lat 40. XX wieku jako zagregowane archiwum, dotyczące literatury, filmu i teatru. Materiały są opracowywane według autorskich metod, wypracowywanych wewnątrz zespołu odpowiedzialnego za rozwój bazy bibliograficznej: działającej przy IBL PAN Pracowni Bibliografii Bieżącej w Poznaniu (dalej również jako PBB). Na przełomie wieków XX i XXI przekształciła się w bazę danych. W pierwszym ćwierćwieczu XXI w. postępują prace, zmierzające do stałego udostępniania danych bibliograficznych w środowisku cyfrowym dla użytkownika docelowego poszukującego konkretnych zapisów - jako serwis internetowy typu *discovery* i baza danych, a także upowszechniania owych danych do badań dla wyspecjalizowanych badaczy i do przetwarzania maszynowego – jako możliwość eksportu bazy danych i zbiory danych (*datasets*). Jak piszą Beata Koper i Tomasz Umerle, na tle innych tego typu inicjatyw od początku wyróżniało PBL „podbudowane naukową metodą dążenie do analizy całości piśmiennictwa w założonym okresie przy jednoczesnym przyjęciu szerokiej definicji kultury literackiej”<sup>259</sup>.

Cele bibliografii

By uporządkować pojęcia, również w kontekście opisywanych wcześniej otwartych danych połączonych, przywołajmy ustalenia Zyty Szymańskiej co do głównych założeń, jakie przyświecają bibliografii literackiej, na które powołują się także Koper i Umerle:

[Szymańska] szczegółowo analizuje funkcje bibliografii literackiej; według badaczki do podstawowych należą: naukowa, rejestracyjna, dokumentacyjno-źródłoznawcza, informacyjna, informowania o treści publikacji, porządkowania faktów w czasie, integracji całej literatury

---

<sup>259</sup> Beata Koper i Tomasz Umerle, „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”, w *Nowa humanistyka: zajmowanie pozycji, negocjowanie autonomii* (Warszawa, 2017), 499–511. Powtórzę także za autorami wtrącenie z przypisu do tego stwierdzenia: „Obecność w PBL działów teatralnego i filmowego wiązała się pierwotnie z przekonaniem o szczególnym związku tych sfer kultury z podstawowym obszarem zainteresowania PBL, tj. literaturą, gdyż także dla znacznej części twórczości teatralnej i filmowej punktem odniesienia są zarówno teksty twórcze (literackie), jak i scenariusze czy sztuki teatralne” – tamże.



polskiej (krajowej i emigracyjnej), archiwalno-historyczna, edukacyjna; do nadrzędnych (niezamierzonych): bibliometryczna, naukoznawcza, ogólnokulturowa, a do funkcji PBL w formie bazy danych online: biblioteczna, kontaktu z autorami materiałów, genealogiczna oraz marketingowa<sup>260</sup>.

### Rys historyczny i założenia

Początek działalności PBL przypada na lata 40. XX wieku<sup>261</sup>. Projekt zapoczątkował Stefan Vrtel-Wierczyński, profesor ówczesnego Uniwersytetu Poznańskiego (dziś UAM). Zorganizował on struktury Pracowni Bibliografii Bieżącej w ramach Instytutu Badań Literackich (najpierw jako instytucji niezależnej, później włączonej do Wydziału I Polskiej Akademii Nauk)<sup>262</sup>. Pierwszy rocznik Polskiej Bibliografii Literackiej jako monografii, za rok 1948, ukazał się w r. 1954 nakładem wydawnictwa Ossolineum, które pozostało wydawcą do roku 1960. W latach 1962-2000 PBL wydawało PWN. Wstecz uzupełniono również materiał od lat 1944/45<sup>263</sup>. Pod względem zakresu tematycznego, głównymi działami były Teksty i Opracowania.

Jacek Biesiada, długoletni kierownik PBB<sup>264</sup>, podsumowując w 2004 r. 50-lecie działalności PBL, wymienia następujące grupy materiałów dla owego początkowego rocznika: 1) Historia literatury i krytyka literacka, 2) Teoria literatury i krytyki. Językoznawstwo, 3) Nauki pomocnicze, 4) Teatr, film, 5) Organizacja nauki o literaturze. Oprócz zapisów druków bibliografia uwzględniała zapisy dla 355. tytułów czasopism, opracowywanych przez zespół

---

<sup>260</sup> Zyta Szymańska, *Funkcje bibliografii literackich* (Uniwersytet Wrocławski, Wydział Filologiczny, Instytut Informacji Naukowej i Bibliotekoznawstwa, 2011) [Niepublikowany doktorat]. Cyt. za: Koper i Umerle, „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”, 503.

<sup>261</sup>Jacek Biesiada, „Pięćdziesiąt lat «Polskiej Bibliografii Literackiej»”. *Nauka*, nr 1 (2004): 137–141.. W dalszej części przytaczam fakty wymienione w tym tekście, aby opisać historię projektu.

<sup>262</sup> Początkowo, jeszcze w latach dwudziestych, Vrtel-Wierczyński planował rozpocząć działalność bibliograficzną we Lwowie jako restytucję warszawskiego Instytutu Bibliograficznego pod patronatem Towarzystwa Naukowego Lwowskiego. Nie doszło to jednak do skutku. Badacz miał też istotne doświadczenie kierowania bibliotekami: Biblioteki Uniwersyteckiej w Poznaniu w l. 1927-1937 i od 1937 do wybuchu II Wojny Światowej Biblioteki Narodowej w Warszawie. W latach powojennych również był krótko związany z tymi dwiema instytucjami przed założeniem PBB. Por. Jan Sandecki, *Stefan Vrtel-Wierczyński: w kręgu bibliografii i bibliotekarstwa*, Nauka, Dydaktyka, Praktyka = Science, Didactics, Practice 92 (Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 2007).

<sup>263</sup>Elżbieta Gondek, Zbigniew Żmigrodzki i Stowarzyszenie Bibliotekarzy Polskich, red. *Bibliografia: metodyka i organizacja*. *Nauka, Dydaktyka, Praktyka = Science, Didactics, Practice* 38. Warszawa: Wydaw. SBP, 2000, s. 292.

<sup>264</sup> Jako były kierownik Pracowni Bibliografii Bieżącej IBL PAN jest obecnie jej emerytowanym współpracownikiem. Zob. Instytut Badań Literackich Polskiej Akademii Nauk - Pracownicy. „Jacek Biesiada”. Dostęp 2020-07-31. <https://ibl.waw.pl/pl/o-instytucje/pracownicy/biesiada-jacek>.

z autopsji<sup>265</sup>. Następnie dodaje, co pod względem zakresu pozostawało niezmiennie przez cały czas trwania:

(...) rejestracja tekstów literackich (autorów polskich wydanych zarówno w języku polskim, jak i przełożonych na języki obce, autorów obcych przełożonych na język polski), inne formy istnienia literatury takie, jak: film, słuchowisko radiowe, widowisko telewizyjne; odnotowywanie opracowań dotyczących zagadnień znajdujących się w obszarze zainteresowań bibliografii; sporządzanie z autopsji opisów bibliograficznych uwzględniających druki zwarte, materiały z czasopism polskich – krajowych i emigracyjnych oraz zagranicznych o profilu slawistycznym<sup>266</sup>.

Wyklarowały się cztery zasadnicze części tej bibliografii<sup>267</sup>: I. Teoria literatury, II. Literatura polska, III. Literatury obce, IV. Teatr, film, radio, telewizja. Dla literaturoznawców z pewnością ten podział jest zrozumiały, choć warto zaznaczyć sprawy nieoczywiste, tzn. takie, które nie wynikają z samej definicji tych pól czy obszarów. I tak w części pierwszej, poza metodologią badań literatury, teorią krytyki i dziedzinami pokrewnymi, znajdziemy zapisy dotyczące przekładu literackiego oraz związków literatury z innymi dziedzinami sztuki. W drugiej części zastosowano podział historyczny na epoki literackie oraz gatunkowe rozróżnienie dla literatury współczesnej (teksty dramatyczne, krytyczne, poetyckie i prozatorskie). Biesiada podkreśla też kronikarski charakter, związany z życiem literackim, czyli rejestrację nagród i imprez, łącznie z odnotowywaniem faktów polskiego życia literackiego za granicą. Oprócz tego w tej części znajdują się informacje związane z badaniami literaturoznawczymi i dydaktyką. Największą objętościowo część stanowią hasła osobowe, z podziałem na bibliografię podmiotową i przedmiotową. Co do części trzeciej, poświęconej dziełom zagranicznym, zorganizowana była ona na wzór opisu literatury polskiej, ale w podziale na literatury innych krajów. Czwarta część, poświęcona teatrowi, filmowi, radiu i telewizji, również zawiera hasła osobowe, związane z tymi dziedzinami. Oprócz tego rejestrowane są teksty teoretyczne, jak i informacje o premierach teatralnych i rodzimych

---

<sup>265</sup> Biesiada, 137.

<sup>266</sup> Biesiada, 137-138.

<sup>267</sup> Biesiada podaje je jako podział monografii rocznika 1968 stosowany do momentu zaprezentowania tekstu, ogłoszonego w 2004 roku. W kontekście bazy danych i serwisu online podobny podział jest funkcjonalny dla prezentacji zakresu zawartości możliwej do odszukania, natomiast wydaje się być bardziej użyteczny, gdy mamy do czynienia z monografią drukowaną.

produkcjach filmowych. Czwarta część zbiera także materiał związany z adaptacjami literatury na inne media (słuchowiska radiowe, widowiska itp.). Zawarte są w niej również informacje o nagrodach i odznaczeniach w tych dziedzinach. Warto podkreślić czwartą sekcję tematyczną jako bardzo przewidujący wybór: wydaje się to nieoczywiste, ale już badacze czasów PRL rozumieli potrzebę archiwizowania treści związanych z widowiskami i mediami. Na przełomie wieków pojawia się także pytanie o efektywne dokumentowanie internetu, co jeszcze będzie wspomniane. Wszystkie roczniki monografii PBL posiadały indeks rzeczowy i osobowy, zgodnie z dobrymi praktykami nauk pomocniczych filologii, dla skutecznej nawigacji w medium drukowanym i łatwego wyszukiwania informacji.

Dla wyobrażenia sobie skali przedsięwzięcia, przytoczę wolumen informacji przetwarzanej w latach 80-tych, na przykładzie bibliografii za rok 1979. Ową pracę wydano w 1986, a jej wstęp jest datowany na 1981. Był to 36. rocznik i 29. tom PBL. Oprócz dokumentacji tytułowego rocznika, uwzględniono materiały z wcześniejszych lat wydane z opóźnieniem, jak i te niedostępne zespołowi w momencie opracowywania materiału lat poprzednich monografii PBL. Czytamy we wstępie:

W celu zebrania materiału do tomu opracowano około 660 czasopism polskich i obcych, bibliografie narodowe, publikacje zbiorowe; zapisano wszystkie książki z zakresu literatury i nauki o literaturze wydane w Polsce, polonika literackie wydane za granicą oraz fakty z życia teatru, filmu, radia i telewizji. Uwzględniono również płyty z nagraniami utworów literackich.

Sporządzono w ten sposób około 56 800 zapisów, na podstawie których zredagowano 12 296 numerowanych pozycji bibliograficznych<sup>268</sup>.

Wspomniane tu zapisy przekładają się bezpośrednio na rekordy w bazie danych. Oczywiście nie było to wiadome w momencie archiwizacji, ale dziś spojrzymy na nie jako jednostki przechowywania informacji za pomocą metod cyfrowych.

Pod czujnym okiem Cenzora

Istotnym czynnikiem dla kompletności informacji jest problem cenzury, czyli działalność w okresie PRL Głównego Urzędu Kontroli Prasy, Publikacji i Widowisk oraz związana z jego

---

<sup>268</sup> „Wstęp”. W Polska bibliografia literacka 1979,1 za rok 1979, część 1. - (1986), IV–V. Warszawa: Państwowe Wyd. Naukowe, b.d., s. IV.

pracami regulacja zarówno rynku wydawniczego, jak i komunikacji na temat literatury<sup>269</sup>. Były to działania motywowane politycznie: jeśli GUKPPiW identyfikował dane treści jako nieodpowiednie do publikacji, a osoby jako szeroko pojętych wrogów władzy ludowej, publikacja w głównym obiegu<sup>270</sup> okazywała się w praktyce niemożliwa. Urząd cenzorski chronił autorytet władzy, stosując szykany wobec redakcji czasopism, indywidualnych autorów. Ze względu na upolitycznioną prezentację dokonań polskiego piśmiennictwa, dotknięte problemami zostały także – zależne od władzy – instytucje archiwizujące kulturę, w tym PBB, rozwijająca PBL. Wśród wspomnianych problemów możemy na pewno wymienić brak informacji o niektórych publikacjach książkowych, pominięcie istotnych kulturotwórczych czasopism w rozbudowie bibliografii, zapis na autorów (czyli swoiste embargo na wydawanie dzieł, a nawet wymienianie ich z imienia i nazwiska, jak to np. spotkało Czesława Miłosza). Warto o tym pamiętać jako specyficznym doświadczeniu kultur i instytucji opisujących działalność kulturalną w określonym miejscu i czasie, czyli w Europie Środkowo-Wschodniej lat 1945-1989, gdzie silnie zaznaczyły się wpływy sowieckie i sztuka wraz z literaturą tworzyły osobny nurt wobec szeroko pojętego Zachodu, jak i oddzielne obiegi wymiany informacji i strategię tożsamości<sup>271</sup>.

W kontekście braków PBL uzupełnienie przyniosły prace, którymi kierował Jerzy Kandziora. Nie tylko część dokumentalistyczną udało się odbudować (a właściwie zbudować na nowo w obliczu zaniechań). Nakładem wydawnictwa IBL PAN ukazało się w 1999 uzupełnienie monografii o ten brakujący materiał, którego nie rejestrowała PBL za czasów komunistycznych rządów w Polsce. Autorami są Jerzy Kandziora i Zyta Szymańska, przy współpracy Krystyny Tokarzówny. Publikacja wymienia też dwoje innych współpracowników: Jadwigę Czachowską (konsultacja naukowa) i Marka Jastrzębskiego (konsultacja w zakresie druków pozacenzuralnych). We wstępie Kandziora jako redaktor tomu opisuje spodziewany efekt prac oraz ich istotność dla pełnego obrazu w dokumentacji PBB:

---

<sup>269</sup> Cenzura w okresie PRL jest osobnym i obszernie opisanym tematem, sygnalizuję ją wszakże jako kontekst. Por. np. Woźniak-Łabieniec, Marzena. *Cenzorskie lekcje literatury: studia o systemowej kontroli słowa w polsce po 1945 roku*. Wydanie I. Łódź: Wydawnictwo Uniwersytetu Łódzkiego, 2022.

<sup>270</sup> Dziś dyskusje na temat publikacji twórczości w głównym nurcie i podziemiu (opozycja *mainstream-underground*) to kwestia mody i ewentualnych uwarunkowań rynkowych czy perypetii związanych z pozyskiwaniem dofinansowania. Nie funkcjonują obiegi literackie, czy szerzej: publikacyjne, które moglibyśmy szeregować pod względem legalności w odniesieniu do krajowego prawa.

<sup>271</sup> Por. Piotr Piotrowski, *Awangarda w cieniu Jałty: sztuka w Europie Środkowo-Wschodniej w latach 1945-1989*. Wyd. 1. Poznań: Dom Wydawniczy Rebis, 2005.

W ciągu prawie piętnastu lat istnienia tzw. drugiego obiegu dokonywały się na jego obszarze doniosłe dla kultury polskiej procesy, m.in. dojrzewanie nowych pokoleń literackich i estetyk artystycznych, „spotkanie się piśmienniczych nurtów” – emigracji i twórczości krajowej, włączanie w świadomość społeczną – także za pośrednictwem niezależnego teatru – postaci i dzieł literatury polskiej i obcej wcześniej, po 1945 roku, z niej wymazywanych. Mamy nadzieję, że ogarnięcie tych procesów w formie bibliografii, rejestrującej po raz pierwszy zarówno druki zwarte niezależnych oficyn, jak i zawartość literacką, teatralną i filmową czasopism wydawanych poza cenzurą, pozwoli badaczom literatury i dziedzin pokrewnych na stworzenie pełniejszego i nie obciążonego ryzykiem luk informacyjnych obrazu literatury i życia artystycznego powojennej Polski<sup>272</sup>.

Kandziora powołuje się też na kontynuację pracy, jaką zaczął zespół PBL, rejestrując publikacje drugiego obiegu jako polonika zagraniczne (reedycje zagraniczne) – działanie podjęte przez PBB, ale zniweczone przez w większości przez cenzurę: z dużo obszerniejszego materiału dopuszczone do druku zostały niektóre zapisy z pism *Krytyka* i *Spotkania*, jak pisze Kandziora „według do dziś niewyjaśnionych kryteriów”<sup>273</sup>. Część zawartości ukazała się w monografii za rok 1985, ukończonej w 1989, a więc już wtedy, kiedy cenzura nie obowiązywała. Co charakterystyczne, z perspektywy wytwarzania wiedzy i dokumentowania kultury, informacje zawarte w publikacji z 1999 częściowo odwołują się do gromadzonej uprzednio wiedzy, ale w dużej części zostały oryginalnie pominięte i trzeba było wytworzenia ich do nowa. W pracy uczestniczyły osoby z PBB. Dzięki temu powstało uzupełnienie zgodne z metodą PBL i zachowujące ciągłość oraz spójność redakcyjną.

Zakres tematyczny publikacji tworzą literatura, ruch wydawniczy, film i teatr. Udokumentowane zostały teksty literackie polskich twórców, jak i paraliterackie tych osób (np. felietony, listy, reportaże). Autorzy przyjęli kryterium publikacji przynajmniej jednej książki i „zasadniczo literacki profil twórczości”. Dla takich twórców zarejestrowano wszystkie teksty nieliterackie. Metodologicznie ciekawe jest włączenie w obręb opisu także

---

<sup>272</sup>, Jerzy Kandziora, Zyta Szymańska, i Krystyna Tokarzówna. *Bez cenzury 1976-1989: literatura, ruch wydawniczy, teatr: bibliografia*. Warszawa: IBL, 1999, IX.

<sup>273</sup> Kandziora et al.

tekstów eseistycznych, poruszających tematy polityczne, historyczne i filozoficzne, jak i „pisarstwa politycznego o walorach literackich”, motywowane charakterystyką piśmiennictwa drugiego obiegu<sup>274</sup>. Poskutkowało to także włączeniem do opisu przedstawicieli tego typu pisarstwa, którzy, jak można wywnioskować, nie kwalifikowaliby się z uwagi na kryteria wyłącznie literackie. Uzupełnienie obejmuje także przekłady na język polski 52 literatur zagranicznych i opracowania na temat ich autorów. Jednym z wyodrębnionych tematów jest działalność władz w sferze informacyjnej: nowomowa, manipulacje językowe i propaganda. Podobnie jak PBL rejestrujący wydarzenia kulturalne, tak i *Bez cenzury 1976-1989* rejestruje również fakty życia literackiego (protesty, niezależne nagrody itp.). Działalności GUKPPiW poświęcono też osobną sekcję. Rozdział „Cenzura” otwierają ogólne opracowania (gdzie np. odnotowano drugoobiegowe tłumaczenie Orwella *Pisarze i Lewiatan. Zabójcy słowa*, a więc tekst ilustrujący samo zjawisko kontroli swobody wypowiedzi, jak i teksty bieżące autorów rodzimych np. *Cenzura w PRL. Parę podstawowych informacji* Stanisława Barańczaka, czy *Czarną księgę cenzury PRL*, a nawet teksty ustaw władz Polski Ludowej). Następnie skompilowane zostały informacje o ingerencjach w teksty, zapisach cenzorskich oraz konfiskatach urzędów celnych. Wydanie elektroniczne pracy z 2018 r. poszerzono o aneks, opracowany przez Kandziore i Szymańską.

Inicjatywa, kierowana przez Kandziore, stanowi errata do dawniejszej formy PBL-u. Prowokuje także do pytania, które będzie wielokrotnie się pojawiać w kontekście dalszych rozważań i ustaleń: kiedy informacja jest w swojej istocie kompletna? Zarówno przemiany samej Bibliografii Literackiej, jak i zastosowanie metod humanistyki cyfrowej do przetwarzania danych upewniają w zmienności kryteriów odpowiedzi na tak postawione pytanie, w zależności od kontekstu historycznego, metodologicznego, jak i technologicznego. W przypadku cenzury mamy do czynienia z motywacją polityczną – przy referowaniu prac cyfrowych humanistów pokroju Morettiego byliśmy bliscy oskarżać ich o imperialne zapędy (a na pewno możemy przyznać, że imperialna tradycja jedne źródła wypromowała, na archiwizację innych nie bacząc), historyczny przykład PBL pokazuje jednak złożoność kwestii przygotowania danych do udostępnienia na zewnątrz. Widziałbym tu aktualność potrzeby promowania obecności bibliografii Polski i państw regionu środkowej Europy na arenie międzynarodowej. Pominiecia cenzorskie, na które na szczęście (proszę wybaczyć nieśmieszna grę słów) znalazło się *remedium*, rozdzieliłbym jakościowo od braku

---

<sup>274</sup> Kandziore et al.

uwzględnienia danych na temat polskiej kultury i jej piśmiennictwa, spowodowanych słabą ich widocznością poza krajem.

W stronę bazy danych

We wspomnianym już tekście Jacka Biesiady, autor ukazuje jako przełomowy 46. rocznik PBL, wydany w 2001 roku, a zawierający zapisy za rok 1989. Po pierwsze podkreślone zostało bogactwo materiału, związane z przemianami ustrojowymi, a tym samym zakończeniem działalności cenzury i boomem na rynku wydawniczym (opracowanie z autopsji 1067 tytułów czasopism)<sup>275</sup>. Po drugie, a zarazem ważniejsze, był to pierwszy rocznik opracowany z udziałem narzędzi informatycznych. Ostatnim rocznikiem drukowanym PBL była bibliografia za rok 1988, wydana w r. 2000. Kolejne powstawały jako baza danych, początkowo rozpowszechniana dla odbiorców na płytach CD-ROM<sup>276</sup>.

Rozrost rynku książki i stopniowe przekształcenie PBL w bazę danych na przełomie wieków miały także skutki dla metodologii opisu, stosowanej w PBB. Zmieniły się kryteria doboru czasopism przeznaczonych do opisu bibliograficznego: zrezygnowano z rejestracji dzienników regionalnych, organów związków zawodowych oraz partii politycznych, jak i periodyków „o charakterze popularnym”, w tym także wyznaniowych. Przechodząc do części bazodanowej, Biesiada konstatuje: „Był więc wspomniany rocznik 46. ostatnim, w którym obowiązywał dotychczasowy kanon źródeł czasopiśmienniczych, ale był też pierwszym, który powstał w wyniku syntezy dyscyplin humanistycznych z informatyką”<sup>277</sup>.

Nastąpiła tranzycja do formuły bazodanowej: cele działalności się nie zmieniły, ale formuła już tak. Możemy wnosić, że przede wszystkim wiązało się to z przeszkoleniem personelu do nowej metody pracy (przy zachowaniu metodologii w części zachowanej jeszcze od początków projektu, ale dostosowanej do środowisk cyfrowych). Opisując część czwartą PBL, Biesiada używa sformułowania „środku inne niż druk” – współcześnie ta sytuacja

---

<sup>275</sup> Biesiada, 139. Zob. również na temat sytuacji rynku wydawniczego: Przemysław Czapliński, *Powrót centrali: literatura w nowej rzeczywistości*. Kraków: Wydawnictwo literackie, 2007.

<sup>276</sup> Szybkość przemian technologicznych może uzmysłowić czytelnikowi fakt, że, po niespełna ćwierć wieku od tamtych edycji PBL, CD-ROM nie jest już powszechnie wyokrzesywanym nośnikiem, w dobie przechowywania danych w chmurze i powszechnego dostępu do zasobów internetowych, a jego zastosowanie sprowadza się do bardzo specyficznych wymogów archiwizacji, np. w kontekście składania prac doktorskich. Nowe modele komputerów osobistych typu laptop przeważnie nie posiadają wbudowanego napędu CD-ROM. Świadczy to o konieczności refleksji nad długofalowymi skutkami wybranych metod przechowywania informacji w planowaniu projektów dokumentacji dziedzictwa i podejmowaniu decyzji zorientowanych na zrównoważony rozwój.

<sup>277</sup> Biesiada, 139.

komplikuje się o tyle, że owych środków pojawiło się więcej, niż znane wczesnemu PBL-owi cztery obszary medialne, do czego warto wrócić w kontekście środowisk cyfrowych.

Bibliograf jako operator maszyny

Charakter pracy dokumentalistycznej zmienił się wraz ze skokiem technologicznym. Obecnie bibliografowie i bibliografki wprowadzają dane do bazy danych za pomocą specjalnie do tego przygotowanego interfejsu. Początkowe możliwości narzędzia nie uwzględniały wszystkich potrzeb, które twórcy PBL mają obecnie w związku z udostępnianiem bazy w celu tworzenia ekosystemu otwartych danych połączonych. Ściślej: nie wszystkie te potrzeby dało się przewidzieć. Możemy prześledzić rozwój PBL na przykładzie określania relacji między poszczególnymi zapisami w bazie.

Karolina Jaworska i Beata Koper wspominają o rozwoju narzędzi dostępnych bibliografkom. W odniesieniu do „usieciowienia” wewnątrz treści bazy piszą w następujący sposób:

W dotychczasowej bazie danych Polskiej Bibliografii Literackiej odnotowywane są różnego rodzaju relacje pomiędzy poszczególnymi opisami bibliograficznymi. Książkom, filmom, spektaklom, audycjom radiowym i telewizyjnym towarzyszą recenzje, artykułom polemiki, nawiązania lub sprostowania. Artykuły mogą być także powiązane z twórcą, utworem, instytucją, czasopismem, wydarzeniem. Aby zwiększyć potencjał danych gromadzonych w bazie Polskiej Bibliografii Literackiej zaprojektowane zostały nowe moduły służące m.in. do tworzenia i prezentowania użytkownikom połączeń między danymi. Jedne są kontynuacją dotychczasowych praktyk, inne zostały zaprojektowane od podstaw. Za ich pomocą tworzona będzie sieć połączeń między poszczególnymi zapisami bibliograficznymi<sup>278</sup>.

Autorki wspominają o nowych możliwościach interfejsowych, wprowadzonych właśnie w tym celu, na przykładzie wprowadzania kart wydarzeń, które zastąpiły wcześniejszy formularz o nazwie *Imprezy*. Przykładową korzyścią po wprowadzeniu zmian jest możliwość powiązania nagród jako zbioru-cyklu z kolejnymi edycjami i ich laureatami. Pojawiło się również powiązanie z organizatorem danej imprezy (kartoteka osób i instytucji). Nową możliwością

---

<sup>278</sup> Karolina Jaworska i Beata Koper, „Polska Bibliografia Literacka – nowe wyzwania”, w *Krajowe bibliograficzne bazy danych - przegląd*, t. 25 (Bibliograficzne bazy danych : perspektywy i problemy rozwoju, Kraków, 2017), [https://ebiojs.pl/index.php/Mat\\_konf/article/view/611](https://ebiojs.pl/index.php/Mat_konf/article/view/611), 3.



stało się wzajemne przeszukiwanie z wykorzystaniem tych informacji. Pojawiło się także wyszukiwanie utworów uprzednio występujących w bazie, dla sprawniejszego dodania nagrodzonych książek.

Poszukiwanie informacji za pomocą dostępnych narzędzi, m.in. systemu wewnętrznych odsyłaczy, pozwoli na dotarcie do szczegółowych informacji, a także na poznanie charakterystyki konkretnych wydarzeń czy zjawisk. (...) Mając takie informacje, odbiorca może przeprowadzić szczegółową analizę zjawisk zachodzących w życiu literackim i kulturalnym w Polsce. Polskiej Bibliografii Literackiej da to znaczące możliwości gromadzenia i klasyfikowania materiałów bibliograficznych. Udoskonalony mechanizm oznaczania relacji między danymi zwiększy jej potencjał, stworzy nowe możliwości wykorzystywania danych, np. w badaniach ilościowych i statystycznych dotyczących literatury, teatru, filmu czy szerszych zjawisk kulturowych<sup>279</sup>.

We wspomnianym tutaj tekście autorki posługują się makietami roboczymi rozwiązań. W przytoczonym przykładzie nagrody „Nike” pojawia się sugestia otwarcia na dane połączone, np. w formie śledzenia odnośników do przekładów danego tytułu. Jaworska i Koper posługują się zaproponowanym przez Jarosława Packa pojęciem „hiperbibliografii” jako rozwiniętej struktury, za pomocą odnośników łączącej rozproszone zbiory.

Widoczne jest tutaj świadectwo przemian pracy bibliografów, w której nie tylko istotnym pozostaje wprowadzanie nowych rekordów do zbioru, ale rysuje się obszar refleksji nad usprawnieniami narzędziowymi i działalność bliższa domenie projektowania systemów informatycznych.

PBL jako laboratorium wiedzy o kulturze

Z pewnością możemy mówić o PBB jako laboratorium, zgodnie z tym, co z najlepszymi życzeniami (i ostrzeżeniem) dla humanistów cyfrowych pisze Pawlicka, w przytaczanym wcześniej tekście krytycznym: jest to zarówno ośrodek nastawiony na pracę dokumentalistyczno-badawczą, jak i miejsce spotkań, budujące społeczność wokół działań wirtualnego archiwum.

---

<sup>279</sup> Jaworska i Koper, 7.

Prace prowadzone w latach 2015-2018 pod kierownictwem Macieja Maryła w ramach grantu *Polska Bibliografia Literacka – laboratorium wiedzy o współczesnej kulturze polskiej*<sup>280</sup> nadały istotny rozwój technologiczny temu przedsięwzięciu, dlatego warto poświęcić kilka zdań na opis jego założeń i efektów.

Jednym z działań były opisywane wcześniej rozbudowy w obszarze relacji elementów bazy. Ogólnym zamierzeniem było natomiast stworzenie bazy kompletnej, która za pomocą danych bibliograficznych zdawałaby sprawę z lat 1939-2002 jako kompendium kultury.

Jako osoby bezpośrednio zaangażowane w projekt figurują, oprócz kierującego projektem Maryła, Łukasz Bukowiecki i Piotr Wciślik. Współpracownicy wymienieni na stronie CHC to Beata Koper i Tomasz Umerle, chociaż na podstawie konsultacji z zespołem można wnioskować, że efekty działań grantowych dotyczyły pracy całej PBB. Przykładowo, wdrożone zmiany techniczne i interfejsowe, jak widzieliśmy na przykładzie danych związanych z nagrodami, wpłynęły na zmianę sposobu pracy z danymi i tworzenie nowych rekordów. Pod względem instytucjonalnym projekt realizowały CHC i Pracownia Bibliografii Bieżącej w Poznaniu. Partnerem informatycznym było Poznańskie Centrum Superkomputerowo-Sieciowe przy IChB PAN, a koordynatorami informatycznymi Marcin Werla oraz Michał Kozak, obaj zatrudnieni w PCSS.

Z kolei wyróżników PBL mamy kilka: szczególny zakres czasowy, sięgający lat 40. XX w., a nawet 1939, opisywany obszar przedmiotowy (kultura literacka, czyli pole szersze od skupienia się na literaturze jako takiej), a wreszcie rozbudowana struktura systemu indeksów rzeczowych. Zespół grantowy wskazuje też w opisie na szeroki zakres zbieranych informacji, przez co PBL odnotowuje często zagadnienia i cechy swoich rekordów, które pozostają dla niej unikatowe, ponieważ podobne bazy różnią się przyjętą metodą opisu czy zakresem działań dokumentalistycznych. Na uwagę zasługuje spojrzenie na literaturę w odniesieniu do adaptacji i obiegu komunikacyjnego wokół dzieł, np. recenzji.

„PBL stanowi zatem zasób cennej, różnorodnej wiedzy o kulturze polskiej, który w ramach projektu zostanie poszerzony i zaopatrzony w narzędzia badawcze pozwalające na dokonywanie złożonych operacji na zebranych materiale”<sup>281</sup>. Zespół grantu wskazuje czynniki infrastrukturalne. Możemy sobie z jednej strony wyobrazić, że od przełomu wieków

---

<sup>280</sup> Projekt realizowany ze środków Narodowego Programu Rozwoju Humanistyki (Maj 2015 – Maj 2018); NR 0061/NPRH3/H11/82/2014. Zob. „PBL Lab”, <http://chc.ibl.waw.pl/pl/projekty/pbl-lab/>. Dostęp: 2023-09-23.

<sup>281</sup> „PBL Lab”.

współpraca IBL PAN i PCSS w jakiejś mierze już była działaniem humanistyki cyfrowej (choć może nie używano tego pojęcia i nie odwoływano się do trendów badawczych); z drugiej, działania podjęte od 2015 zostały zaplanowane w celu rozwoju i otwarcia nowych możliwości. Jedną z nich jest udostępnianie danych do przetwarzania maszynowego. Dzięki ścisłej współpracy informatyków z humanistami, planowano rozwój bazy o potencjale utrzymywania przeszło 3 milionów rekordów. Zidentyfikowanym wyzwaniem jest na pewno aktualność bazy, do czego również odniesiono się w opisie grantu: „Długofalowa strategia rozwoju PBL zakłada, iż w kolejnym etapie prac (po zakończeniu tego projektu) luka w opracowaniu materiału zostanie zmniejszona do koniecznego jednego roku, a w obręb bazy zostaną włączone dokumenty internetowego życia literackiego”<sup>282</sup>. To dzięki tym staraniom, wychodzącym naprzód zarówno pod względem rozwoju narzędzi, jak i kadr, efekty grantu są używane do dzisiaj i umożliwiają start kolejnych projektów w obrębie zespołu PBB.

#### Nowy serwis PBL

Karolina Jaworska analizuje aspekty serwisu pod kątem architektury informacji<sup>283</sup>. Odwołując się do typologii systemów etykietowania<sup>284</sup>, autorka wskazuje realizacje w PBL. Przykładowo, dla grupowania informacji na temat osoby kluczami są autorstwo, współtwórstwo, materiały o osobie i jej działalności, udział w wydarzeniach, przejawy kultu (rozumianego jako nadawanie imienia osoby – patronat, jak i poświęcone działalności osoby wydarzenia, np. przegląd filmów reżysera), wzmianki (adnotacje, decyzje, oznaczenia osób). Budowa serwisu internetowego PBL (adres na domenie [ibl.poznan.pl](http://ibl.poznan.pl)) odzwierciedla strukturę działów, wypracowaną zgodnie z metodą PBB, zorientowaną na archiwizowanie danych bibliograficznych o kulturze.

Z kolei po zmodernizowaniu i przeniesieniu na domenę [ibl.waw.pl](http://ibl.waw.pl) serwis przyjmuje luźniejsze kryteria, czasowe i tematyczne w strukturze strony: bardziej przypomina dzięki temu układ popularnych stron, serwisów i portali, z których użytkownicy internetu korzystają na co dzień. Z perspektywy doświadczenia użytkownika (UX) zmiany są bardzo korzystne, ponieważ daje to łatwiejszą nawigację, a elementy znajdują się w miejscach, w których większość osób się ich spodziewa. Przykładem może być belka z mapą strony – odnośnikami do głównych części u dołu, w związku z częstym zachowaniem nawigowania do samego dołu:

---

<sup>282</sup> „PBL Lab”.

<sup>283</sup> Karolina Jaworska, „Projektowanie przestrzeni informacyjnej – przykład nowej strony internetowej «Polskiej Bibliografii Literackiej»”, <https://depot.ceon.pl/handle/123456789/15438>.

<sup>284</sup> Louis Rosenfeld, Peter Morville, i Jorge Arango, *Architektura informacji w serwisach internetowych i nie tylko*, Wydanie IV (Gliwice: Wydawnictwo Helion, 2017), 119-152.

jeśli użytkownik patrzy na tę przestrzeń, prawdopodobnie jeszcze nie dotarł do poszukiwanej treści). Pole wyszukiwania zajmuje obecnie centralną część: serwis zachęca do eksploracji. Wyraźnie wydzielono graficznie zakładkę Kontakt. Interfejs nowego serwisu zawiera też usprawnienia dla zwiększenia widoczności treści: są dostępne tryby takie jak widok z kontrastem i opcje powiększania wyświetlanego tekstu w trzech stopniach wielkości liter. Jest to funkcjonalność istotna dla osób z utrudnionym widzeniem lub chcących oszczędzać wzrok (kontrast spełnia funkcję trybu ciemnego).

Użytkownicy i użytkowniczki mają możliwość eksportowania danych ze swoich kwerend do formatów przetwarzania maszynowego. Pojawiły się też w odróżnieniu do wcześniejszej odsłony PBL *datasety* tematyczne: nowy serwis zawiera część „Kolekcje. Dzięki temu nakierowuje się odbiorcę na eksplorację, poszukiwanie nowej wiedzy, skojarzeń i kontekstów dla własnych poszukiwań. Występują również odnośniki do mediów społecznościowych i serwisów pokrewnych. Przez porównanie do Library of Congress jako wiodącego internetowego zasobu bibliograficznego na świecie, możemy zauważyć trendy w prezentacji danych dotyczących kultury – piszę o tym w dalszej części. Są to m. in. kolekcje tematyczne i okolicznościowe.

W 2019 roku miałem okazję wprowadzać zapisy do bazy w ramach wdrożenia do pracy z PBL. Poniżej przedstawiam wygląd interfejsu do wprowadzania zapisów dla bibliografów Pracowni oraz widok serwisu po stronie użytkownika. Obsługa programu wymaga krótkiego przeszkolenia dla osób archiwizujących. Z kolei dla użytkownika oba serwisy są użyteczne, przy czym nowa odsłona posiada sprawniejsze, bardziej rozbudowane sposoby nawigacji i lepiej komunikuje korzystającym z niej osobom ich możliwości przeglądania i korzystania z wirtualnych zbiorów.

IBL - Publikowanie roczników PBL - [Inne zapisy]

Struktura Kartoteki Zapisy Redakcja Parametry Zakończ prace

**Inne zapisy** Rok: 2004 Numer zapisu: 1320308 Znacznik: Uwagi:

Klasyfikacja Rodzaj zapisu: **artykuł o utworze** Wprowadź do działu twórcy (autora hasłowego): **Lipska** Dopisz twórcę

Dział: **Hasła osobowe (literatura polska) - L** Dział: **Lipska** EWA

Dział: **Lipska** EWA

**Dotyczy** Słowny opis odwołania (wskazanej książki, filmu, sztuki, imprezy, itp.)  
 Numer zapisu, do którego odwołuje się ten zapis: **1320292** Takie czasy. SZUKAJ ZAPISU  
 Usuni odwołanie

**Podstawowe dane** Autorzy materiału: **Sikora** Władysław Kryptonim: Dopisz autora **Adnotacje**

Tytuł: **Gawędy o poezji (4)**

Dotyczy teatru: **Teatry**

**Osoby współtworzące** Funkcja Nazwisko, imię, uwagi Kartoteka osób współtw. **Dodatkowe osoby do indeksu** Nazwisko Imię Pseud./Krypt.

Opis: **Hasła przekrojowe i słowa kluczowe do indeksu rzeczowego** Hasła i klucze Dodatkowe sł. kluczowe


Czasop.: **Zw** Zwrot: **Zwrot** Zapis utworzony przez: **BEATAD** Zmodyfikowany przez: **28.03.19**

Rok: **2004** Nr: **nr 1** Str: **s. 53-54** Kartoteka źródeł Status zapisu:

Nazwa rodzaju zapisu: **Lipska**  
 Licznik: **6** <Lieta

Ryc. 1 Interfejs wprowadzania danych dla pracowników PBB

← → ↻ ↗ pbl.ibl.poznan.pl/dostep/index.php?s=d\_biezacy&f=zapisy\_szczeg&p\_zapis=420304 ... Szukaj



## POLSKA BIBLIOGRAFIA LITERACKA

Instytut Badań Literackich PAN w Warszawie, Pracownia Bibliografii Bieżącej w Poznaniu

**Materiał dostępny za lata 1988-2003**

| [Spis działów](#) | [Indeks nazwisk](#) | [Indeks rzeczowy](#) | [Kartoteka czasopism](#) |  
| [Kartoteka teatrów](#) | [Kartoteka wydawnictw](#) | [Szukaj tytułu](#) |

---

**INFORMACJE SZCZEGÓŁOWE O ZAPISIE**


Dział bibliografii: Teatr, film, radio, telewizja  
- Teatr  
- Teatr polski za granicą  
- Czechy (od 1993; teatr polski za granicą)  
- Czechy (od 1993; teatr polski za granicą) - repertuar

Rodzaj zapisu: sztuka teatralna  
Autori: Wawrosz Adam - [szczegóły](#)  
Tytuł: **Zolyty. Co po nogle**  
Dotyczy teatru: Zespół teatralny przy PZKO : Łomna Góra - [szczegóły](#)  
Numer zapisu: [zobacz](#) (AW)

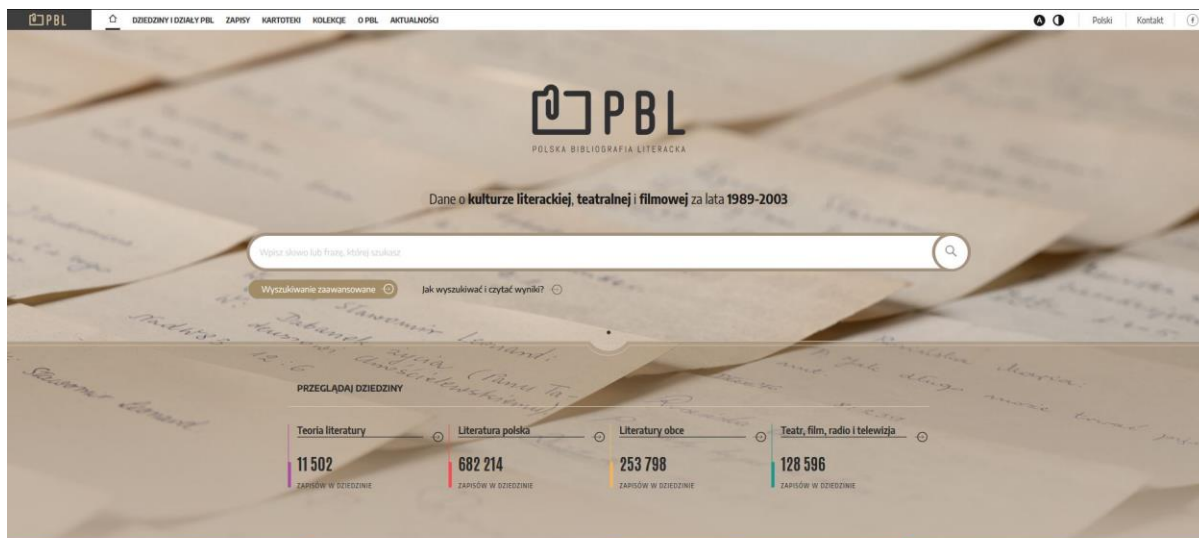
Inne zapisy dotyczące tego materiału:  
● artykuł; C.R.: **Amatorzy na scenie**. Zwrot: 1997 nr 6 s. 68 (nota...) [szczegóły](#)

**PBL ON-LINE**  
[Spis działów](#)  
[Indeks nazwisk](#)  
[Indeks rzeczowy](#)  
[Kartoteka czasopism](#)  
[Kartoteka teatrów](#)  
[Kartot. wydawnictw](#)  
[Szuka i tytuł/słowa](#)  
[Transliteracja](#)

**O PBL**  
**CD PBL**  
**WYDAWNICTWO**  
**IBL PAN**  
**KONTAKT**

 [English version](#)

Ryc. 2 Interfejs użytkownika serwisu w domenie ibl.poznan.pl; zaznaczono identyfikator rekordu



Ryc. 3 Nowy serwis PBL

Osoby pracujące w PBB korzystają też w coraz większym stopniu z relacyjnej bazy Oracle za pomocą narzędzi programistycznych. Składa się ona z wielu wirtualnych kartotek (ponieważ początkowo odzwierciedlała układ indeksów bibliografii tradycyjnej), jakie można łączyć wspólnymi kluczami (unikatowe identyfikatory osób, instytucji, poszczególnych zapisów itp.). Pracownicy PBB formułują zapytania SQL do odczytu i modyfikacji treści w bazie, używają także języków R i Python w celach statystycznych i pełniąc obowiązki archiwistyczne.

Sieć instytucji związana z wytwarzaniem *Bibliodata*

Bibliografowie mają kontakt z materialnymi i niematerialnymi wytworami kultury. Istotną częścią pracy przy spisywaniu danych z autopsji jest współpraca z zewnętrznymi podmiotami i instytucjami – PBL nie wymaga prowadzenia zbiorów *sensu stricto* (poza archiwum zapisów, które gromadzono w sposób tradycyjny, jako katalog fiszek, jak i rocznikami monografii i niezbędnymi materiałami pomocniczymi): działalność dokumentalistyczno-archiwistyczna ogranicza się do rejestracji, ale nie przechowywania tytułów. PBB nie zarządza katalogiem książkowym, nie gromadzi książek. W tej sytuacji ważna pozostaje współpraca z bibliotekami naukowymi, dla zapewnienia materiału badań. Oznacza to praktycznie minimalną potrzebę utrzymywania infrastruktury materialnej związanej z zasobami bibliotecznymi<sup>285</sup>. Jak pisze Karolina Jaworska, PBL czerpie przede wszystkim z *Przewodnika Bibliograficznego Biblioteki Narodowej*, jako źródła podstawowego źródła bibliografii bieżącej, zawierającego

<sup>285</sup> Praca PBB wymaga natomiast zaplecza do funkcjonowania maszyn, miejsc spotkań i wymiany wiedzy oraz przestrzeni biurowej dostosowanej do wykonywania zadań bibliograficznych.

informacje o publikacjach z ostatniego roku i do dwóch wstecz, opierającego się na składaniu do BN egzemplarza obowiązkowego przez wydawcę<sup>286</sup>. Autorka dodaje:

Należy wspomnieć również o tym, że niektórzy wydawcy, autorzy książek, a czasem też redakcje czasopism, dbając o to, aby informacja o ich książkach lub zawartości czasopisma (których tematyka znajduje się w zakresie zainteresowań bibliografii) znalazła się w „Polskiej Bibliografii Literackiej”, sami przysyłają lub osobiście dostarczają do Pracowni Bibliografii Bieżącej IBL PAN w Poznaniu egzemplarze swoich książek lub czasopism. Jest to jednak znikomy odsetek wszystkich źródeł odnotowywanych w PBL.<sup>287</sup>

Biblioteka Narodowa pomaga również przy pozyskiwaniu informacji do rejestracji poloników (rozumianych jako wydane za granicą dzieła polskiej literatury pięknej, literatura teoretyczna i krytyczna dotycząca polskiej literatury, teatru i filmu; PBL zalicza też do tej kategorii publikacje poświęcone literaturom obcym, których współtwórcami są Polacy).

Jaworska wymienia również główne źródła dla zapisów pozaliterackich:

W „Polskiej Bibliografii Literackiej” dokumentowane są również związane z literaturą zagadnienia teatralne, filmowe, radiowe i telewizyjne. Repertuary polskich teatrów zawodowych odnotowywane są na podstawie „Almanach Sceny Polskiej” (...) oraz zawartości bazy Encyklopedia Teatru Polskiego ([encyklopediateatru.pl](http://encyklopediateatru.pl)). Informacje o polskich filmach fabularnych pełnometrażowych, dokumentalnych, filmach i serialach telewizyjnych oraz obcych filmach dokumentalnych poświęconych polskiej literaturze i polskim pisarzom, a także obcych filmach stworzonych przez polskich twórców, zrealizowanych w danym roku pochodzą z bazy Film Polski ([filmpolski.pl](http://filmpolski.pl)). Dane dotyczące spektakli Teatru Telewizji, słuchowisk, powieści radiowych, telewizyjnych i radiowych programów dokumentalnych dotyczących zakresu

---

<sup>286</sup> Karolina Jaworska, „Jak powstaje «Polska Bibliografia Literacka»? Znaczenie współpracy z bibliotekami naukowymi”, w *Biblioteka naukowa: czy jeszcze naukowa?*, red. Jazdon (Poznań: Biblioteka Uniwersytecka w Poznaniu, 2018), 167–78. Pisownia oryginalna.

<sup>287</sup> Jaworska, 171.

zainteresowań PBL pozyskujemy obecnie dzięki współpracy z Telewizją Polską S.A. i Polskim Radiem<sup>288</sup>.

W kwestii dostępu do książek i czasopism, głównym kryterium pozyskiwania egzemplarzy fizycznych do opisu jest geograficzne: bibliografowie korzystają w znacznej mierze z poznańskich bibliotek, gdyż PBB ma siedzibę w Poznaniu<sup>289</sup>. Oprócz tego stosuje się wypożyczenia międzybiblioteczne; wzrosła rola wymiany cyfrowej – jedną z metod stanowi elektroniczny system wypożyczeń dla dzieł naukowych *Academica*. Jaworska zauważa synergie z tym związane:

Ponadto zespół pracowni nawiązał współpracę z bibliotekarzami administrującymi systemem *Academica* mającą na celu wzbogacanie jego zasobów o książki i czasopisma trudno dostępne lub takie, których egzemplarze posiada jedynie Biblioteka Narodowa. Oczywiście z zasobów zdigitalizowanych i udostępnionych na prośbę zespołu Pracowni Bibliografii Bieżącej IBL PAN mogą korzystać wszyscy użytkownicy korzystający z *Academiki*<sup>290</sup>.

Przedstawia się tutaj korzyść dla odbiorców jako pochodna samego procesu dokumentalistycznego w tworzeniu bibliografii dziedzinowej. Większość powyższych uwag dotyczy tradycyjnego procesu rozwijania tego projektu, czyli metod wypracowanych od powstania PBL. Część z tych wątków powróci jednak w kontekście ucyfrowienia – zapowiedź stanowi korzystanie przez badaczy z witryn w internecie do zdobywania informacji oraz weryfikacji zapisów (jako tzw. praca ręczna, osobiste sprawdzanie poprawności). Wśród uwag końcowych badaczka podkreśla istotność otwartej wymiany informacji oraz zasobów, tak pod względem środowisk cyfrowych, jak i artefaktów materialnych, wykorzystywanych do badań, jak to się dzieje z egzemplarzami pozyskiwanymi z księżnic w celu ustalania danych z autopsji.

---

<sup>288</sup> Jaworska, 172.

<sup>289</sup> Jaworska wymienia tutaj następujące instytucje: Biblioteka Uniwersytecka i biblioteki wydziałowe, Biblioteka Raczyńskich, Biblioteka Poznańskiego Towarzystwa Przyjaciół Nauk, Biblioteka Towarzystwa Chrystusowego dla Polonii Zagranicznej, Biblioteka Kórnicka, Publiczna Biblioteka Pedagogiczna. Autorka zwraca także uwagę na istotną rolę wypożyczeń międzybibliotecznych przy utrudnionym dostępie. Wyróżnia szczególnie rolę Biblioteki Narodowej i Biblioteki Uniwersyteckiej w Toruniu jako współpracujące przy tym instytucje.

<sup>290</sup> Jaworska, 173.



PBL i inne podobne projekty na świecie

Jaworska wymienia także projekty prowadzone przez globalne biblioteki i inne instytucje, które możemy porównać do działalności PBB pod kątem polityki czy też procesów instytucji oraz z uwagi na prowadzone projekty, przypominające w swojej formie lub celu PBL<sup>291</sup>. Projekty światowe, które stanowią odpowiednik PBL lub posiadają zbliżony profil. Są to: Austlit w Australii; (<https://www.austlit.edu.au/>) tworzony na Uniwersytecie Queensland. W Czechach funkcjonują Centrum Informacji Literackiej (Střediska literárněvědných informací) w Instytucie Literatury Czeskiej Akademii Nauk i projekt podobny do PBL: Bibliografie české literární vědy. W Bibliotece Uniwersyteckiej we Frankfurcie nad Menem prowadzona jest Bibliographie der deutschen Sprach- und Literaturwissenschaft (BDSL Online; <http://www.bdsl-online.de>), tworzona przez zespół współpracujący z DNB (Niemiecką Biblioteką Narodową). Bardzo cenne w tym kontekście jest porównanie zakresów i sposobów działania tych inicjatyw:

Są pewne podobieństwa, a także różnice w sposobie opracowywania oraz udostępniania danych zgromadzonych w „Polskiej Bibliografii Literackiej” i omówionych wyżej bibliografiach. Bibliografie czeskie, podobnie jak bibliografię polską, tworzy zespół wyspecjalizowanych pracowników, filologów. Baza australijska powstaje z udziałem badaczy literatury i bibliotekarzy z różnych ośrodków naukowych w Australii. Natomiast w przypadku BDSL Online za jej powstawanie odpowiedzialni są bibliotekarze z Biblioteki Uniwersyteckiej we Frankfurcie nad Menem oraz współpracujący z nimi pracownicy Niemieckiej Biblioteki Narodowej. Wszystkie te bibliografie (w większym lub mniejszym zakresie czasowym) udostępniane są użytkownikom w postaci internetowych baz danych. Jedynie Centrum Informacji Literackiej Instytutu Literatury Czeskiej Akademii Nauk Republiki Czeskiej prowadzi kilka baz danych dotyczących literatury (jest ich siedem), w przypadku pozostałych mamy do czynienia z jedną bazą danych, w której publikowane są opisy bibliograficzne książek, artykułów itp. dotyczących zagadnień rzeczowych i osób. Dostęp do PBL i czeskich portali jest bezpłatny, w przypadku niemieckiej bibliografii bez ograniczeń można korzystać jedynie z roczników

---

<sup>291</sup> Tamże, 175.

1985–2000, darmowy dostęp do australijskiej bazy danych mają użytkownicy, którzy korzystają z usług instytucji subskrybujących bazę. Takimi instytucjami są np. uniwersytety, szkoły, biblioteki. PBL i BDSL Online mają taki sam sposób gromadzenia opisów bibliograficznych – kierowane są one do poszczególnych działów tworzących ogólną strukturę danej bibliografii. Wszystkie omówione bazy posiadają wyszukiwarki, dzięki którym użytkownicy mogą wyszukiwać informacje i zawężać wyniki wyszukiwania. Znaczącą różnicą jest to, że bibliografie czeskie, niemiecka i australijska dokumentują tylko literaturę i recepcję literatury danego kraju. Jedyne „Polska Bibliografia Literacka” odnotowuje także przejawy recepcji literatur obcych w Polsce<sup>292</sup>.

#### Porównanie z Biblioteką Kongresu

Innym istotnym punktem odniesienia jest tutaj amerykańska Library of Congress, która także prowadzi dokumentalistyczną bazę bibliograficzną. Oczywiście jej najbliższym odpowiednikiem jest polska Biblioteka Narodowa, jako najważniejszy krajowy zasób, archiwizujący publikacje, ale możemy przyjrzeć się głównym założeniom i poszczególnym rozwiązaniom także w kontekście PBL. Warto pamiętać, że porównuję tutaj bazę, której głównym zadaniem jest utrzymanie katalogu bibliotecznego z bibliografią dziedzinową, jaką pozostaje PBL.

Serwis LoC można przeszukiwać z poziomu strony głównej w poszukiwaniu wielu typów treści<sup>293</sup>, oraz w katalogu, zawierającym już ukonkretyzowane zapisy<sup>294</sup>. Strona główna pozornie wygląda, jakby to z niej wchodziło się do katalogu. Odwołuje się jednakże do szerszej gamy typów materiałów. Użytkownicy mogą wyszukiwać, wykorzystując typ zawartości: nagrania audio, książki i materiały drukowane, filmy i wideo, treści prawne, manuskrypty i materiały mieszane, mapy, zapisy muzyki, gazety, periodyki, historie osobiste, materiały wizualne (zdjęcia, wydruki i rysunki), oprogramowanie i zasoby cyfrowe, archiwa sieciowe, strony internetowe, obiekty 3D.

Duża część serwisu ma charakter popularyzatorski i zachęcający do interakcji, zdobywania wiedzy. Kontekstem jest tu na pewno edukacyjna rola bibliotek jako kulturotwórczych miejsc

---

<sup>292</sup> Jaworska, 175-176.

<sup>293</sup> Zob. „Library of Congress”, <https://www.loc.gov>. Dostęp 2020-03-20.

<sup>294</sup> Zob. „Library of Congress Catalog”, <https://catalog.loc.gov/>. Dostęp 2020-03-20.

spotkań<sup>295</sup>. Dużo tego typu treści odwołuje się do tożsamości amerykańskiej, ale także dyskursów mniejszościowych. Znaczna część moich badań nad interfejsem LoC oraz działalnością tej instytucji w obszarze zarządzania danymi bibliograficznymi wydarzyła się w trakcie pandemii i protestów Black Lives Matter. Biblioteka Kongresu publikowała wówczas kolekcje tematyczne i pamiętam bardzo dobrze jak aktualna wydawała się strona główna serwisu: zespół dbał o przygotowanie treści, które przyciągały uwagę poprzez odwołania do bieżących wydarzeń<sup>296</sup>. Strona główna LoC to miejsce zarządzane na zasadach polityki wydawniczej, którą do pewnego stopnia można porównać z serwisami informacyjnymi. W przypadku pierwszej witryny PBL mamy do czynienia raczej z minimalistycznym interfejsem. Druga wersja serwisu zawiera odniesienia do kolekcji i ma potencjał na podobną kurację treści.

Pod względem wyszukiwania w katalogu, serwis oferuje wyszukiwarke z opcją wyszukiwania: prostego, zaawansowanego w polach oraz w słowach kluczowych. Użytkownicy mają też dostęp do historii wyszukiwania. Wyszukiwanie we wcześniejszym PBL opiera się na nawigacji związanej z działami bibliografii. Dopiero nowy serwis oferuje wyszukiwanie zaawansowane i opcje pomocy.

Główną różnicę między projektem rozwijanym przez Bibliotekę Kongresu a tym tworzonym przez PBB stanowi tutaj zastosowany format danych. W przypadku PBL jest on autorski (do pewnego stopnia można go określić jako *proprietary*, czyli zamknięty lub zastrzeżony – z uwagą, że wśród przemian bibliografii jedną z istotniejszych jest opracowywanie metadanych dla połączenia z innymi kolekcjami, w innych formatach). Z kolei podstawą dla LoC jest format MARC 21. Baza pełni wszak prymarnie funkcję biblioteczną. Użycie tego formatu zapewnia uniwersalność pomiędzy instytucjami bibliotecznymi i poprzez zastosowanie stałej konwencji opisu jest przewidywalne dla każdego, kto chciałby te dane przetwarzać. A jest ku temu możliwość, ponieważ LoC udostępnia API wraz z dokumentacją<sup>297</sup>. Oznacza to dostęp do danych dla dalszego przetwarzania. Jednocześnie wadą MARC 21 jest dość duża złożoność, gdy chcielibyśmy te dane zastosować w formie

---

<sup>295</sup> Na marginesie warto dodać często powtarzaną frazę: biblioteki są tymi miejscami dostępnymi lokalnie, które nie próbują nam niczego sprzedać. To hasło podkreśla egalitarny dostęp do wiedzy, jaki oferują te instytucje.

<sup>296</sup> Przykłady kolekcji: Archiwum internetowe związane z protestami przeciwko rasizmowi: zob. „Protests Against Racism Web Archive”, <https://www.loc.gov/collections/protests-against-racism-web-archive/>. Dostęp: 2023-09-16. Historia Afroamerykanów - kolekcje: zob. „African American History”, [https://www.loc.gov/collections/?fa=subject\\_topic:african+american+history](https://www.loc.gov/collections/?fa=subject_topic:african+american+history). Dostęp: 2023-09-16.

<sup>297</sup> Zob. „APIs for LoC.gov”, <https://www.loc.gov/apis/>. Dostęp: 2023-09-16.

otwartych danych połączonych, a jest za to odpowiedzialna największa zaleta formatu: jego wyspecjalizowanie. Informacje opisuje się jako zagnieźdzenia pól i podpól, co wymaga stałego odwoływania się do dokumentacji i bardzo wykształcenia określonych kompetencji bibliograficznych. Istnieje wiele innych „zwinniejszych” form opisu danych do osiągnięcia podobnego celu<sup>298</sup>, dzięki czemu łatwiej się łączy większe zbiory.

Są także dostępne inne dodatkowe API. Wśród nich na szczególną uwagę zasługuje Linked Data Service, czyli serwis danych połączonych<sup>299</sup>. Umożliwia on dostęp do metadanych bibliograficznych na użytek przetwarzania maszynowego. Stosowane są w nim słowniki kontrolowane, dzięki czemu informacje nie wymagają dodatkowej pracy człowieka. Przeciwnieństwem tej sytuacji pozostaje używanie pól z zawartością typu *string*, które nie dają się łatwo zinterpretować przez maszyny, treść nie jest kontrolowana; jest to charakterystyczna cecha większości serwisów zawierających dane bibliograficzne w Polsce, w tym PBL.

Polska baza posiada co prawda wewnętrzne stałe identyfikatory typów zapisów, Twórców, tytułów czasopism itp. istotnych pól, ale duża część bazy, w tym tytuły utworów, adnotacje bibliograficzne i współpracownicy (tłumacze tekstów), pozostaje uzupełniana właśnie ciągami znaków. Konsekwencją jest nie tylko ryzyko błędów ludzkich w utrzymaniu treści, ale także brak powtarzalności dla „zrozumienia” przez maszyny; jednocześnie możemy argumentować za szkodliwością nadmiaru standaryzacji, ale dostępne modele, np. słowniki kontrolowane tematyki i treści tworzone przez LoC czy ontologie Wikidata mogą być bardzo pomocne w utrzymaniu porządku w bazie. Badania *Bibliodata* na zawartości pozbawionej słowników kontrolowanych stanowią już większe wyzwanie i potrzebny jest do nich np. *text mining*, czyli wydobywanie informacji z tekstu<sup>300</sup>. W kwestii dostępności danych, w nowym serwisie PBL użytkownicy znajdują zakładkę z API, wraz z dokumentacją. W kontekście warunków dostępu, dane zgromadzone w serwisie są wyraźnie opisane jako otwarte dane badawcze; można je pobrać w formacie JSON. Obecna jest wyraźna zachęta do kontaktu w celu ich użycia.

---

<sup>298</sup> Rekomenduje się np. użycie znaczników Dublin Core i Schema.org. Por. „DCMI: Home”, <https://www.dublincore.org/>. Dostęp: 2023-09-23. Zob. też „Schema.org”, <https://www.schema.org>. Dostęp: 2023-09-16.

<sup>299</sup> Por. „ID.LOC.GOV - Linked Data Service”, <https://id.loc.gov/>. Dostęp: 2023-09-16.

<sup>300</sup> PBB prowadzi prace nad słownikami kontrolowanymi, m. in. w obszarze tematyki tekstów, właśnie z myślą o wykorzystaniu zawartości bazy w zbiorze danych połączonych.

Podsumowując, wiele funkcjonalności obecnych w LoC może wyznaczać kierunki rozwoju dla PBL albo stanowić przykład i punkt odniesienia dla dalszych działań PBB w kierowaniu zachodzącymi obecnie przemianami cyfrowymi.

Baza retrospektywna i dążenie do aktualności

Wyzwanie PBB stanowi rozrost rynku wydawniczego; przyrost produkcji literackiej, pojawianie się nowych tytułów i form publikacji. W pewnej mierze komplikuje również sytuację notoryczny brak egzemplarzy obowiązkowych. Istnieje wymóg zgłaszania nowych wydawnictw do BN i przekazywania ich bibliotekom, ale właściwie nie sposób tego wyegzekwować, stąd ograniczenie dostępu egzemplarzy do opisu z autopsji. O tym problemie wspomina Jaworska przy okazji opisu współpracy PBL z bibliotekami<sup>301</sup>. W przypadku nowego serwisu, w domenie [ibl.waw.pl](http://ibl.waw.pl), jako dostępne do przeszukiwania i pobierania są wymienione dane z lat 1989-2003. Serwis wcześniejszy, hostowany na domenie [ibl.poznan.pl](http://ibl.poznan.pl), komunikuje odbiorcy dostępność zapisów z lat 1989-2013. Mimo intensywnych prac grantowych nad modernizacją, pozostaje luka, którą zespół PBB wypełnia w obu kierunkach: digitalizując zbiory lat wcześniejszych oraz przygotowując roczniki zaległe i nadchodzące. Niewątpliwie można tutaj powtórzyć za badaczami z UCLA apel o tworzenie szerokich i dostępnych infrastruktur, ponieważ duża część zaplecza metodologicznego jest przygotowana na opracowanie danych dotyczących kultury w XXI wieku, ale działania w tym kierunku są uzależnione od finansowania oraz prac technicznych. Zarówno migracja na nowy serwis, jak i próby nadążenia za produkcją literacką i kulturalną, to proces ciągły. Odbywa się on w ramach prac projektowych o określonych ramach czasowych oraz w trybie stałego funkcjonowania Pracowni jako instytucji dokumentalistycznej.

Wyzwaniem odbiorcy staje się tym samym aktualność bazy i sprawdzenia bardziej bieżących informacji. Baza jest z definicji retrospektywna, o tyle, że rejestruje zawsze już wydane teksty na cele archiwalne: przedmiot badań zawsze będzie związany z dokumentacją, a nie prognozą. Wyzwanie może stanowić zbyt szerokie okno takiej retrospektywy i kłopoty z aktualnością.

Kolejne wyzwanie, stojące przed zespołem PBB, przynajmniej na początku, to konsekwencja sposobu zapisu<sup>302</sup>. W przygotowaniu materiału do druku jako monografii rocznej, PBL mógł przejść proces redakcyjny jak każda publikacja tego typu. Obecnie PBL to relacyjna baza Oracle, z interfejsem użytkownika udostępnionym w sieci jako serwis internetowy typu

---

<sup>301</sup> Jaworska, 170.

<sup>302</sup> Temat częściowo wiąże się ze wspomnianym wcześniej użyciem słowników kontrolowanych.

*discovery*. Na „zapleczu” funkcjonuje natomiast interfejs przygotowany dla bibliografów, służący do wprowadzania nowych zapisów. W przypadku każdej bazy mogą pojawiać się błędy i odstępstwa, istotne jest jednak wykrywanie ich i ograniczanie. Pojawia się zatem potrzeba wprowadzania praktyk i wewnętrznych procesów kontroli jakości (*Quality Assurance, QA*) dla zachowania spójności i prawidłowości informacji, oraz ujęcia pracy dokumentalistycznej w sposób procesowy. Dostrzegam tutaj dwa obszary: narzędziowy i behawioralny. W pierwszym potrzeba monitorowania treści i te procesy funkcjonują już dla tej części PBL, którą pracownicy PBB przetwarzają maszynowo (część zapisów czeka jeszcze na digitalizację). W drugim przypadku da się zaobserwować wymianę wiedzy między pracownikami i dzielenie się dobrymi praktykami. Wprowadza to spójność praktyk i niweluje różnice pracy np. przy wprowadzaniu danych. Zespół PBB konsultuje również w grupie tak metodę, jak i nowe podejścia i wyjątkowe sytuacje, które pojawiają się w trakcie opracowywania kolejnych roczników.

Pojawia się także konieczność patrzenia wprzód z uwagi na zmiany technologiczne. W badanym zakresie PBL istnieje cały obszar twórczości związanej z internetem, do której trzeba także przystosować metodologię opisu: literatura publikowana w internecie, tzw. interaktywna itp., ale także rozróżnienie na wydania elektroniczne oraz papierowe. Potencjalnie może mieć to znaczenie dla zawartości i odbiorców. Zgodnie z opisywaną wcześniej przez Manovicha zasadą wariacyjności: to, co znaliśmy jako *errata* w druku, odbywa się jako aktualizacja, niejednokrotnie poza świadomością użytkownika czy odbiorcy. Dokumentaliści kultury powinni być na te ewentualności przygotowani i przyjmować nowe kryteria oraz metody działania (jak opisywać i wersjonować, jakie cechy uznać za dystynktywne, a które będą poboczne). Dotyczy to zarówno redefinicji przedmiotu badań oraz zakresu archiwizacji, jak również sposobów, na jakie owe opisy się sporządza<sup>303</sup>. Analogicznie do wersji oprogramowania użytkowego i narzędzi programistycznych, oznacza się treść (przykładem znanym większości użytkowników będą historie edycji haseł na Wikipedii). Istotne jest także budowanie relacji między rekordami (np. tak, aby wskazać kolejne iteracje tego samego dzieła). Na podobnej zasadzie mogłoby to się realizować w serwisach opisujących zmienny w czasie przedmiot.

---

<sup>303</sup> O literaturze cyfrowej por.: Winiecka.

PBL cyfrowy: stan badań

Rozwój PBL następował obok procesów, obejmujących resztę instytucji GLAM (zbiorcze określenie domeny gromadzenia, katalogowania i ochrony dziedzictwa kulturalnego – od *galleries, libraries, archives, museums*). PBB działa niezależnie od Biblioteki Narodowej, głównej instytucji w kraju, zajmującej się rejestracją rynku wydawniczego (omawiana bibliografia należy do bibliografii specjalnych, tematycznych). Natomiast pod względem modernizacji i wprowadzania systemów informatycznych, jak np. Horizon, programy tego typu prowadzono z nastawieniem na biblioteki. Podobnie zatem jak w przypadku założycielskiej idei, jak wskazują przywołane wcześniej źródła, również modernizacja do pracy na danych w przypadku PBL odbywała się jako inicjatywa rozpoczęta wewnątrz zespołu ją tworzącego i z zachowaniem metodyk wypracowanych przez ten zespół. Przedsięwzięcie PBB jako części IBL PAN nie jest też częścią państwowych instytucji ochrony dziedzictwa, które również posiadają własne ustandaryzowane polityki zarządzania archiwizacją, zbiorami oraz kontrolą danych elektronicznych (mam tu na myśli np. zarządzanie muzeami przez Ministerstwo Kultury i Dziedzictwa Narodowego czy system archiwów państwowych).

Dla kogo powstaje PBL

Do określenia prototypowego odbiorcy tworzy się też symulacje. Popularnym narzędziem jest persona czyli w rozumieniu badań *User Experience* (UX) przewidywany odbiorca usługi, strony internetowej, aplikacji itp., znana też jako profil użytkownika (*user profile*). Jej opis może uwzględnić zdjęcie i opis przypominający biogram. Zwykle nadaje się personie nazwę (imię lub imię i nazwisko), prawdopodobnie dla lepszego utożsamienia się z sytuacją takiego projektowanego odbiorcy. W opisie profilu użytkownika umieszcza się informacje o kontekstach korzystania takiej osoby z danej aplikacji czy narzędzia, poparte wynikami badań (persona jest zbiorem określonych tendencji). Dzięki temu rozwiązaniu, w procesie projektowania i udoskonalania produktów cyfrowych wykonawcy mogą argumentować swoje wybory potrzebami tych wymyślonych postaci. Najczęściej stosuje się więcej person, które przedstawiają zróżnicowane sposoby interakcji z danym narzędziem czy systemem<sup>304</sup>. Literaturoznawcy świetnie znają taki sposób schematyzowania dzięki kategorii czytelnika modelowego Umberta Eco<sup>305</sup>.

---

<sup>304</sup> Por. Marcin Sikorski, *Interakcja człowiek - komputer*, Dodr (Warszawa: Wydawnictwo Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych, 2013), 233-234.

<sup>305</sup> Por. Eco, Umberto, Stefan Collini, i Tomasz Bieroń. *Interpretacja i nadinterpretacja*. Wyd. 2. Kraków: Wydawn. Znak, 2008.

Już we wspomnianym tekście Jacka Biesiady pojawiają się sygnały dotyczące grupy odbiorczej jako osób zainteresowanych nauką o literaturze, życiem literackim, metodyką, relacjami literatury polskiej do literatur światowych, jak i dokumentacją mediów (ze względu na interdyscyplinarne opracowania, dotyczące m.in. filmu i teatru)<sup>306</sup>.

Jak diagnozuje Joanna Zoń w 2007 roku, w odwołaniu do przemian PBL z wersji drukowanej na cyfrową:

Wersja on-line Polskiej Bibliografii Literackiej w dalszym ciągu, tak jak jej poprzedniczka w wersji drukowanej, jest niezbędnym narzędziem pracy dla badacza literatury polskiej, teatrologa, filmoznawcy, studenta filologii polskiej, doskonalącego swe umiejętności nauczyciela języka polskiego, wreszcie dzięki swej dostępności nawet dla maturzysty, że o bibliotekarzu tylko wspomnę. Dzięki swej nowej formie PBL stała się doskonałym, szybkim i rzetelnym źródłem informacji o literaturze w Polsce.<sup>307</sup>

Dokładniejsze badania, związane z bardziej współczesną grupą odbiorczą PBL przeprowadzili Beata Koper, Tomasz Umerle i Marcin Werla (dwoje autorów z zespołu PBB, trzeci z IChB PAN i Poznańskiego Centrum Superkomputerowo-Sieciowego). Tutaj warto podkreślić istotną różnicę między deklaratywnym a pragmatycznym podejściem do tego tematu. Wcześniejsze ustalenia można potraktować na zasadzie swobodnego manifestu twórców i postulowanej grupy odbiorczej, poprzez dziesięciolecia funkcjonowania PBL. Poniższe natomiast jako stwierdzenie stanu faktycznego w drugiej dekadzie XXI wieku.

Opieram się tutaj na trzech źródłach: 1) badaniach Marcina Werli, polegających na zebraniu i analizie danych Google Analytics dla serwisu PBL z 2016<sup>308</sup>; 2) wspólnej prezentacji konferencyjnej trójki badaczy z roku 2016<sup>309</sup>; 3) publikacji w tomie pokonferencyjnym,

---

<sup>306</sup> Por.: Biesiada.

<sup>307</sup> Polska Bibliografia Literacka – zmiany, ale jakie? / Joanna Zoń// W: Biuletyn EBIB [Dokument elektroniczny] / red. naczelny Bożena Bednarek-Michalska - Nr 7/2007 (88) sierpień/wrzesień. - Czasopismo elektroniczne. - [Warszawa] : Stowarzyszenie Bibliotekarzy Polskich KWE, 2010. - Tryb dostępu: <http://www.ebib.info/2010/88/a.php?zon>. - Tyt. z pierwszego ekranu. - ISSN 1507-7187.

<sup>308</sup> Werla, Marcin. „Serwis internetowy „Polska Bibliografia Literacka”. Analiza danych z Google Analytics za okres 7 miesięcy: od 1 kwietnia 2016 r. do 31 października 2016 r.” Poznań: PCSS, 2016-11-02. <https://lib.psnc.pl/dlibra/publication/edition/781?id=781..>

<sup>309</sup> Umerle, Tomasz, Beata Koper, i Marcin Werla. „Nowa bibliografia literacka dla nowej humanistyki”. 2016. <https://depot.ceon.pl/handle/123456789/14863>.



autorstwa dwójki badaczy związanych z PBB (2017)<sup>310</sup>. Skrótowo przedstawiam zebrane informacje, wyniki i ustalenia, dodając kontekst i skrótowe komentarze. Posługuję się kolejnością, w jakiej źródła zostały wymienione.

W badaniu Werli użyte zostało narzędzie Google Analytics, które pozwala śledzić ruch na stronie internetowej. Trwało przez 7 miesięcy: dane monitorowano w okresie od 1 kwietnia do 31 października 2016. Ogólna skala ruchu przedstawia się następująco:

[S]erwis PBL odwiedziło ok. 51 300 użytkowników, z czego niecałe 7 000 byli to użytkownicy, którzy w tym okresie odwiedzili serwis więcej niż jeden raz. Łączna liczba sesji (odwiedzin serwisu) w raportowanym okresie wyniosła 66 866, z czego 75% to były sesje nowych użytkowników, a 25% były to sesje użytkowników powracających, czyli takich którzy już wcześniej w raportowanym okresie przynajmniej raz serwis odwiedzili.<sup>311</sup>

Z analizy wykluczono adresy IP należące do pracowni, opracowującej PBL. Średni czas wizyty na stronie to 2 minuty 58 s. Średnia liczba wyświetlonych stron serwisu to 3,66 (użytkownik wyświetlał około 3-4 strony). Dla użytkowników korzystających z PBL więcej niż raz te wartości były znacząco (około dwukrotnie) wyższe: średni czas trwania sesji to 6 minut 11 s., a średnia liczba wyświetlonych stron w jednej sesji to prawie 6.

Następnie Werla podaje statystyki powracających użytkowników: 53% takich sesji generowali użytkownicy, którzy w okresie badania odwiedzali PBL 2-3 razy, 28% tacy, którzy korzystali z serwisu 4-14 razy, a pozostałe 19% odpowiadało za tych, którzy mieli 14 i więcej wizyt.

Pod względem geograficznym dominuje Polska i jej duże miasta, przede wszystkim wojewódzkie. Aż 89,65% sesji odwiedzających (w tym 94,1% powracających) to użytkownicy z Polski. Powyżej przyjętego progu 0,5% ruchu Werla rejestruje jeszcze tylko 4 inne kraje: Niemcy (1,73%), Wielką Brytanię (1,72%), USA (1,07%) i Rosję (0,63%). Największa aktywność miast polskich dla ogólnej liczby sesji przedstawia się następująco: Warszawa (24,37%), Kraków (9,76%), Poznań (8,48%), Wrocław (5,18%), Gdańsk (3,40%), Łódź (3,11%), Lublin (3,11%), Katowice (2,55%), Szczecin (1,68%), Bydgoszcz (1,62%), Rzeszów

---

<sup>310</sup> Koper, Beata, i Tomasz Umerle. „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”. W Nowa humanistyka: zajmowanie pozycji, negocjowanie autonomii, 499–511. Warszawa, 2017.

<sup>311</sup> Zob. M. Werla. Dalsze statystyki również przytoczone za tym badaniem.

(1,5%), Białystok (1,04%) i Kielce (1%). Czołówka miast zagranicznych, które wymienia Werla to Londyn (0,34%), Wilno (0,29%), Berlin (0,27%) i Moskwa (0,23%).

82% użytkowników korzystało z PBL na komputerach stacjonarnych (88% wśród powracających), zaś 18% na urządzeniach mobilnych. Współczesne serwisy internetowe są optymalizowane w zależności od docelowego miejsca wyświetlania (np. pod względem czasu ładowania strony i rozmieszczenia elementów serwisu). Niektóre rozwiązania zakładają również osobne aplikacje dla urządzeń mobilnych, lepiej przystosowane do tych środowisk i posiadające interfejs projektowany specjalnie dla lepszego doświadczenia użytkownika, odróżniające się od wyświetlania witryny w przeglądarce internetowej na tych urządzeniach (*mobile web*). Moim zdaniem nie tylko powyższe dane, ale również przeznaczenie serwisu, przekonują o braku potrzeby rozwijania dla PBL specjalnych aplikacji *mobile*, jeśli rejestruje się mały ruch na tym kanale i nie ma uzasadnienia pod względem funkcjonalności. Tego rodzaju działanie nie jest także zasadniczo praktyką serwisów bibliograficznych. Omawiana baza bibliograficzna nie stanowi tutaj odstępstwa: PBB nie ma w planach rozwijania narzędzi i serwisów mobilnych.

Werla przytacza źródła owiedzin: aż około 86% wizyt pochodzi z wyszukiwarki Google. Około 10% to bezpośrednie wejścia na stronę PBL<sup>312</sup>. Pozostała część, około 4% to przekierowania z innych miejsc w sieci. Interesująca jest następująca obserwacja Werli:

Co ciekawe, użytkownicy którzy wielokrotnie korzystali z PBL w raportowanym okresie, również najczęściej trafiają do serwisu PBL z Google (tak zaczyna się ok. 85% sesji użytkowników powracających, a około 11% to odwiedziny bezpośrednie). Zaskakujący jest tu zarówno bardzo wysoki odsetek skierowań z Google, jak i to że nawet użytkownicy powracający wchodzą przede wszystkim „przez Google”<sup>313</sup>.

Wśród zewnętrznych źródeł badacz z PCSS zauważa przede wszystkim Wikipedię, a także stronę główną IBL PAN. Internetowa wolna encyklopedia odsyłała do różnych 140 zapisów. Niewielki ruch został zauważony z serwisu Facebook oraz stron ciniba.edu.pl i sluchowiska.ugu.pl (Werla określa go jako nieduży, ale nie doprecyzowuje, co to oznacza,

---

<sup>312</sup> Użytkownicy mają serwis zapisany w historii przeglądarki, w zakładkach lub znają adres witryny i bezpośrednio wchodzą na odpowiednią jego stronę lub podstronę.

<sup>313</sup> Werla.

więc za badaczem przyjmujemy, że nie są to istotne statystycznie wartości). Facebook odsyłał do 83. różnych haseł PBL.

Następnie Werla proponuje analizę zachowania użytkowników na stronie. W tym celu najpierw mapuje możliwości, dostępne w menu serwisu PBL:

- Spis działów
- Indeks nazwisk
- Indeks rzeczowy
- Kartoteka czasopism
- Kartoteka teatrów
- Kartoteka wydawnictw
- Szukaj tytułu / słowa w tytule

Analiza Google Analytics wykazała, że to indeks nazwisk wiodł prym w badanym okresie i odpowiadał za ok. 15% wszystkich wyświetleń. Następnie 13% to zapisy w hasłach konkretnych osób. Na tych stronach użytkownicy spędzali średnio prawie 2 minuty<sup>314</sup>. Werla zwraca tu jednak uwagę na to, że wyświetlanie szczegółów o osobach stanowiło tylko 1,41% generowanych (pod)stron. Świadczy to o znikomości potrzeby wejścia na tego rodzaju strony, jak i może wskazywać na małą szczegółowość, która z kolei przekłada się na niewielką przydatność dla odbiorców (z reguły osoby opisują tu tylko lata życia, tj. informacja o roku urodzenia i ew. roku śmierci; nie wszyscy twórcy posiadają też takie podstrony w PBL). Typowi internauci prawdopodobnie czerpią wiedzę ogólną na temat autorów z innych źródeł, a w serwisie przygotowywanym przez PBB poszukują już informacji o konkretnych publikacjach.

Kartoteka czasopism to drugi najpopularniejszy element w badaniu. Za 7,51% wyświetleń odpowiadało wyświetlanie listy czasopism opracowanych w PBL, informacji o tych publikacjach, a także związanych z nimi zapisów. Strony dotyczące konkretnych czasopism średnio zajmowały użytkownikom niewiele powyżej minuty sesji.

Trzeci co do częstotliwości odwiedzin był spis działów, odpowiadający za 5,63% odwiedzin. Składają się nań przeglądanie struktury działów oraz zapisów w działach. Średni czas sesji na stronie związanej z zapisami konkretnego działu to niecałe 2 minuty.

---

<sup>314</sup> Werla podaje też konkretne najpopularniejsze strony twórców w danym okresie, podobnie dla pozostałych badanych elementów dostępnych w serwisie, por. Werla s.5-7.

Wyszukiwarka PBL odpowiadała za 5,09% wyświetleń, a więc niewiele mniej od spisu dzieł. Średnio nieco ponad minutę trwały odwiedziny na stronie z wynikami wyszukiwania.

Werla zauważa:

Mimo, że wyszukiwarka na stronie głównej jest opatrzona wyraźną informacją iż pozwala na wyszukiwanie w tytule, wiele spośród najpopularniejszych wydawanych zapytań zawierało nazwiska pisarzy (31 z pierwszej setki najpopularniejszych zapytań). Trudno określić z dużą pewnością czy było to błędne wykorzystanie wyszukiwarki przez użytkowników, ale wydaje się wysoce prawdopodobne, że niestety tak<sup>315</sup>.

Jest to informacja zwrotna tak do użytkowników, jak i mogąca wesprzeć wszelkie prace modernizacyjne serwisu w przyszłości.

Indeks rzeczowy uplasował się na piątej pozycji sposobów przeglądania PBL. Wyświetlanie haseł w indeksie oraz zapisów przypisanych do poszczególnych haseł odpowiada za 4,37% wyświetleń. Średni czas wizyty to około półtorej minuty.

Werla podsumowuje tę część podając informację o marginalnym wykorzystaniu pozostałych funkcji eksplorowania serwisu i uzupełnia o ważną uwagę metodologiczną. Kartoteki teatrów i wydawnictw generowały 0,71% i 0,47% wyświetleń w badanym okresie (może to być motywowane podobnie jak w przypadku autorów: PBL nie służy do czerpania wiedzy o tych instytucjach, a same podstrony informują najczęściej o nazwie i mieście, w którym działa dana instytucja). Dodatkowo autor komentuje kwestię średniego czasu odwiedzin, który został podany, w kontekście budowy witryny. PBL nie posiada podziału długich list na strony – wszystkie wyniki tworzą długi listing<sup>316</sup>. Pojedyncza strona może zawierać po kilkadziesiąt-kilkaset zapisów ułożonych w formie tabeli z odnośnikami. Autor analizy zwraca tutaj uwagę na to, że ów fakt w połączeniu ze średnim czasem w przedziale 1-2 minut może wskazywać, że użytkownicy „nie analizują zbyt dogłębnie zawartości tych stron”. Mogę dodać z własnej praktyki dowód anegdotyczny, że poszukiwania pożądanej informacji często kończą się użyciem opcji wyszukiwania na stronie w przeglądarce internetowej (w systemie Windows

---

<sup>315</sup> Werla, 6

<sup>316</sup> *Listing* jest określeniem oznaczającym po prostu listę, używanym często w branży e-commerce w kontekście listy ofert. Sformułowanie podkreśla komputerowe generowanie tejże listy (w przeciwieństwie do tradycyjnego rozumienia słowa „lista” jako „wyliczenie”, niezależnie od źródła czy medium).

*ctrl+f*), jeśli potrzebna jest wyłącznie dana fraza (np. tytuł konkretnej publikacji wśród wszystkich zapisów danego autora).

Analizę użycia PBL wieńczy w tekście Werli opis dostępu do poszczególnych zapisów bazy, jako „podstawowych jednostek informacyjnych”. Użytkownicy wyświetlili łącznie 33537 unikatowych zapisów. Niecałe 43% wizyt uwzględniało wyświetlenie strony, która zawierała jakiś zapis. W końcowej części Werla opisuje punkty dotarcia do konkretnych zapisów, tzn. z jakiego „miejsca” w witrynie odbył się bezpośredni ruch do zapisu (przyjmując, że zapis to jednostka wiedzy, która ma być przeglądana, możemy językiem optymalizacji ruchu w internecie mówić o „konwersji” czyli które części serwisu lepiej lub gorzej przekierowują w wybrane miejsce). Około 30,7% użytkowników do zapisu docierało ze stron z informacjami o twórcy. Ok. 10,3% ze stron informujących o autorze<sup>317</sup>. Ok. 5,9% z wyników wyszukiwania w tytule. Ok. 5,6% trafiało do zapisów ze stron z informacjami o dziale, ok. 4,8% - o hasła. Ok. 4,6% ze stron z informacjami o nazwisku w adnotacjach. Ok. 4,4% ze stron informujących o czasopiśmie. Ok. 1,9% ze stron o zapisach powiązanych z innym zapisem głównym. Znaczący udział mają strony poświęcone twórcom i autorom, czyli punktem wyjścia najczęściej jest autor jako osoba (łącznie ok. 40% przejść do zapisów odbywa się z tych stron). Reszta miejsc dotarcia do zapisu ma stosunkowo równomierny rozkład, z niewielkimi różnicami procentowymi. Werla wskazuje, że nawet 45,6% wyświetleń można zgrupować, dodając do siebie strony związane z twórcami, autorami i nazwiskami w adnotacjach. Około 28,9% odwiedzających stronę trafia na adres konkretnego zapisu, bez przechodzenia

---

<sup>317</sup> Twórca to w PBL osoba opisana. Autor to przypisany w sposób bibliograficzny adres osobowy twórcy dla zapisu. Twórca to w istocie swego rodzaju hasło (temat, motyw) takie jak np. proza. Wedle przyjętej metody, w drukowanej monografii PBL taka osoba miałaby dedykowaną sekcję, poświęconą publikacjom własnym, opracowaniom, przekładom itp. Jest to zarazem hasło osobowe, jak i przedmiotowe – można przyrównać do pola tematycznego – hasła przedmiotowego w klasyfikacjach bazodanowych i światowych bibliografii cyfrowych: *subject heading (personal subject heading)*. Dopasowanie hasła do grupy odbywa się tak, aby być jak najbliżej konkretnej informacji, w kolejności: 1) tekst (np. „Pan Tadeusz”); 2) osoba (Adam Mickiewicz); 3) przedmiot (romantyzm). W bazie danych tagujemy wyższym tagiem (wyższego rzędu, ogólniejszym). Użytkownik eksplorujący bazę pod względem epoki w spisie działów, nie widzi pisarzy, a przeglądając bazę pisarzy, nie widzi przypisanych do nich epok. Hasło osobowe zawiera podmiotowe i przedmiotowe dzieła (twórczość osoby i opracowania na jej temat). Przykładowo, w polu Twórca w PBL dla wartości Leśmian dla książki nt. literatury rosyjskiej i twórczości Leśmiana Autor i Twórca są nieutożsamieni: „Autor” to ciąg znaków: podpis i nazwa autora na okładce książki. Nawet jeśli odróżnimy od siebie Twórców, to autorzy są nie do rozróżnienia. Utożsamienie odbywa się na dwa sposoby: 1) (nierekomendowany) weryfikacja pojedyncza z autopsji; 2) przez nadanie stałych identyfikatorów osobowych. W pierwszym sposobie, mając (docelową) kartotekę osób musimy decydować o autorstwie dla każdego konkretnego przypadku. Jeśli Czesław Miłosz (pole Autor) pisze o Bolesławie Leśmianie (Twórca), to trzeba ustalić, czy autor jest tym samym co twórca innych zapisów (czy autor to ten sam człowiek, który napisał dzieło literackie) – niezbędna do tego jest wiedza bibliograficzno-dokumentacyjna. Dzięki zastosowaniu w zbiorze PBL danych Biblioteki Narodowej znaczna część bazy jest już otagowana znacznikami identyfikacji osób (stałe identyfikatory kontroli autorytatywnej).

przez inne (pod)strony PBL (Werla podkreśla: prawdopodobnie bezpośrednio z Google). Tak określone dwa podstawowe scenariusze dotarcia do zapisu odpowiadałyby za blisko trzy czwarte wszystkich wyświetleń szczegółowych informacji o zapisach (74,5%).

Najpopularniejszy zapis wyświetlono 93 razy. 286 zapisów zostało wyświetlonych ponad dziesięć razy. 21376 zapisów wyświetlono po jednym razie. Średni czas wizyty na stronie jednego zapisu to niewiele ponad półtorej minuty.

Jak zaznaczałem na początku, badania były prowadzone w określonych ramach czasowych, zatem trudno mówić o statystycznych trendach i na pewno warto byłoby prześledzić ruch w witrynie PBL ponownie, a nawet monitorować przez ustalony okres regularne charakterystyki zachowania użytkowników. Umożliwiłoby to przykładowo porównania rok do roku czy miesiąc do miesiąca, z uwzględnieniem sezonowości (w tym kontekście przychodzi na myśl rok szkolny i akademicki).

Drugim przywoływanym tu badaniem jest ankieta użytkowników PBL, zaprezentowana przez Beatę Koper, Tomasza Umerle i Marcina Werłę. Tutaj prowadzący badanie, w odróżnieniu od ilościowej analityki wizyt, posłużyli się ilościowo-jakościowym badaniem skierowanym do tych odwiedzających witrynę, którzy chcieli się podzielić informacją zwrotną na temat działania PBL. Ankieta składała się z 15. pytań zamkniętych, w większości wielokrotnego wyboru, oraz 4. otwartych (nieobowiązkowych). Została umieszczona bezpośrednio w serwisie (pod adresem [pbl.ibl.poznan.pl/dostep/](http://pbl.ibl.poznan.pl/dostep/)). W badaniu wzięło udział 310 respondentów. Dane pochodzą z okresu 02-10.2016. Oczywiście sama prezentacja wyników badania to opracowanie odpowiedzi ze wskazaniem prawidłowości oraz trendów, a nie surowe dane uzyskane w ankiecie.

Jeden ze slajdów z prezentacji przedstawia agregację danych, z których można wywieść dane do utworzenia osoby badawczej. Układ prezentacji wskazuje, że posłużyły do tego pytania zamknięte, a więc prezentowany obraz jest precyzyjny dla badanej grupy. Typowa osoba korzystająca z PBL to:

- kobieta z wyższym wykształceniem, mieszkająca w dużym mieście (powyżej 250 tys. mieszkańców);
- korzysta z biblioteki miejskiej, akademickiej i wojewódzkiej;
- sprawnie porusza się w świecie cyfrowym;
- korzysta z innych baz bibliograficznych i repozytoriów prac;

- nie korzysta z roczników PBL wydanych w wersji papierowej (do 1989 roku);
- z PBL-online korzysta kilka razy w roku (w celach badawczych), dokładnie wie czego szuka, z zakresu PBL interesuje ją literatura.

W kolejnej części badacze prezentują treść pytań otwartych i kategorie odpowiedzi użytkowników. Spośród respondentów 30% zdecydowało się odpowiedzieć na pytania otwarte. W końcowej części prezentacji badacze na podstawie zebranego w ten sposób materiału prezentują wnioski do dalszych prac nad rozwojem serwisu. Treść pytań otwartych dotyczyła: 1) typów materiałów, które byłyby przydatne dla użytkowników, a PBL ich nie posiada; 2) najczęstszych i najbardziej uciążliwych problemów korzystania z serwisu (z prośbą o opis występowania); 3) brakujących funkcji elektronicznej odsłony Bibliografii (z prośbą o opisanie sytuacji, w których byłyby przydatne). Na podstawie uzyskanych informacji grupa badawcza mogła podzielić odpowiedzi na trzy grupy: 1) braki materiałowe; 2) wady systemu; 3) potrzebne funkcje.

Wśród braków materiałowych najbardziej dotkliwa była nieobecność materiałów po 2002 roku (przed opisem bibliograficznym), następnie materiałów do 1988 (po opisie bibliograficznym, ale przed digitalizacją, czyli dostępnych wtedy wyłącznie w monografiach drukowanych). Wymienione wyżej zostały ujęte na wykresie frekwencyjnym z wartościami 32 i 22. Kolejne zauważalne braki materiałowe to: czasopisma, źródła internetowe, ebooki i audiobooki oraz materiały teatralne (odpowiednio 8, 7, 4 i 1 odpowiedź). Badacze wymieniają też poza wykresem bardzo szczegółowe braki, zgłoszone prawdopodobnie przez pojedynczych użytkowników lub z niewielką powtarzalnością. W tej grupie, zbyt dużej, aby ją przytaczać, zaznaczono poprzez wytłuszczenie dwa wątki: biografie i informacje o zawartości tomików poetyckich. Przy niektórych, bardzo ogólnych hasłach, pojawiają się znaki zapytania – intencja odpowiadających mogła tu nie być jasna.

Wyróżniają się (i dominują nad pozostałymi) dwie zgłoszone tu wady systemu: nieczytelny interfejs użytkownika oraz nieintuicyjny interfejs użytkownika (12 i 11 odpowiedzi). Rozdzielenie tych cech wydaje się dobre, ponieważ przy czytelnym wyglądzie strony, nadal mogłaby ona prezentować swoje funkcjonalności w nieintuicyjny sposób i na odwrót: rozwiązania mogłyby też być intuicyjne, ale nieczytelne. W pierwszym przypadku jako odbiorca widzę elementy strony i je rozumiem, ale układ treści nie sugeruje, jakie akcje powinienem wykonać. W drugim, interfejs komunikuje mi pożądane zachowania, ale jego przestrzenne rozplanowanie pozostawia wiele do życzenia. Marginalne, acz istotne problemy

systemu, niezwiązane już z układem treści i nawigacji, to: problemy z wyszukiwarką, niekonsekwencja zapisów i brak mobilnej wersji strony (3, 2 i 1 odpowiedź). Tych odpowiedzi było znacząco mniej od dotyczących interfejsu, ale świadczą o sprofesjonalizowanym odbiorze.

Wśród potrzeb użytkowników na pierwszy plan wysuwa się zaawansowana wyszukiwarka – zgłoszeń jest ponad dwukrotnie więcej od kolejnej zdefiniowanej potrzeby (28), jaką jest dostępność wersji pełnotekstowych w serwisie (12). Trzecia z kolei odpowiedź (również po przeskoku frekwencji, 6) to cytowania. 3 odpowiedzi wskazują na lokalizację publikacji. Po jednej odpowiedzi dotyczy: narzędzi do pobierania danych, zapisywania wyników, narzędzi bibliometrycznych, narzędzi do raportowania błędów. Warto zauważyć, że duża część zgłoszonych zagadnień to kwestie przydatne w pracy filologicznej, związanej z lokalizacją, analizą i interpretacją tekstów, a także porządkowaniem metadanych i danych bibliograficznych na ich temat. Część dotyczy także wymiany wiedzy i zapewniania informacji zwrotnej, również od użytkownika do twórców serwisu.

Inne zgłoszone braki dotyczą informacji o ISBN, zasobach z archiwów Polskiego Radia, wyjaśnień zastosowanej przez zespół metodologii oraz opisów zawartości przy każdej publikacji. Wskazuje to z jednej strony na profesjonalny odbiór (zgłoszone tematy są bardzo precyzyjnie ujęte i zdradzają obycie odbiorców z procedurami dotyczącymi bibliotek i bibliografii), ale też z drugiej strony mówi o roli PBL jako zbiorze punktów odniesienia do lektur podmiotowych i przedmiotowych zarówno dla badaczy, jak i osób uczących się. Wskazują na to dosłownie cytowane fragmenty wypowiedzi, wyrażające potrzeby informowania o treści utworów, co pozwoliłoby użytkownikom ocenić, czy ze względu na swoje cele powinni się zapoznać z daną publikacją.

Dane liczbowe z części pytań otwartych przedstawiają się skromnie względem ogólnej liczby odpowiedzi. Warto jednak pamiętać, że udzielenie informacji było fakultatywne dla uczestników badania, a nie każdy z nich musiał dysponować aparatem oceny serwisu internetowego pod względem użyteczności czy UX oraz potencjału usprawnień (pomimo dobrego sprofilowania pytań). Jest to także cecha grupowania udzielonych odpowiedzi i formułowania wśród nich szerszych tendencji. Nawet pojawiające się pojedyncze odpowiedzi zasługują na uwagę i namysł w rozwoju serwisu, wskazując potencjalne usprawnienia możliwe do wdrożenia małym nakładem kosztów lub przez niewielkie zmiany, jak i te zmiany, które wydają się obiecujące ze względu na przyjętą metodę bibliograficzną czy powszechnie



stosowane standardy. Liczba odpowiedzi wspierających dany scenariusz rozwoju stanowi także informację zwrotną, aby zespół uwzględnił oczekiwania użytkowników w priorytetyzacji swoich działań.

Koper, Umerle i Werla formułują na tej podstawie pięć wniosków. Pytanie, które sobie zadali, brzmi: „Jaka jest i jaka będzie nowa bibliografia dla nowej humanistyki?”<sup>318</sup>. Po pierwsze, „Implementuje narzędzia do automatycznej wymiany danych w oparciu o współpracę międzyinstytucjonalną – dążąc do stworzenia możliwie kompletnego obrazu polskiej kultury literackiej (i nie tylko) – oraz, co za tym idzie, poszerza poszerza swój zakres i/lub zasięg”<sup>319</sup>. Po drugie, oferuje (wytłuszczenia moje – JE):

funkcjonalny interfejs użytkownika, a także specjalistyczny interfejs badawczy (narzędzia do pracy z danymi); ew. **tworzy kolekcje danych do prowadzenia badań**; gromadzi **dane kompatybilne z innymi zbiorami danych** (także wskutek współpracy metodologicznej z polskimi bazami danych, np. teatralnymi i filmowymi) czy międzynarodowymi kartotekami haseł wzorcowych<sup>320</sup>.

Po trzecie, „tworzy powiązania z tekstami” i w dalszej kolejności metadane dopasowuje do tego, w jakim stopniu wersje pełnotekstowe zasobów są udostępnione, ze szczegółowym uwzględnieniem potrzeby użytkownika wobec relacji między bibliografią a dostępem do pełnego tekstu przywoływanego materiału, a także rozpoznaje możliwości zastosowania narzędzi do analizowania treści tych tekstów<sup>321</sup>. Po czwarte, zespół tworzący bibliografię nowej humanistyki to dokumentaliści posiadający wysokie kompetencje cyfrowe i doszkalający się w dziedzinie wiedzy o literaturze (kulturze) w Polsce oraz humanistyki cyfrowej<sup>322</sup>. Po piąte:

Inicjuje kontakty ze środowiskiem naukowym poprzez uczestnictwo w środowiskowym dialogu oraz prace nad przygotowaniem gromadzonych danych do pełnienia istotnej roli w humanistycznych

---

<sup>318</sup> Tomasz Umerle, Beata Koper, i Marcin Werla, „Nowa bibliografia literacka dla nowej humanistyki”, <https://depot.ceon.pl/handle/123456789/14863>.

<sup>319</sup> Tamże.

<sup>320</sup> Tamże.

<sup>321</sup> Tamże.

<sup>322</sup> Tamże.

badaniach na danych (zarówno na zasadzie kooperacji projektowych, jak i ciągłego monitorowania potrzeb użytkowników)<sup>323</sup>.

Część tych wniosków powtarza i rozwija publikacja dwójki z zespołu badaczy (Koper i Umerle), w której PBL stanowi przykład cokolwiek szerszego zjawiska<sup>324</sup>. Badanie ankietowe jest tu omówione, a sama baza i jej dane służą za studium przypadku, wszakże chodzi o szerszy kontekst funkcjonowania bibliografii naukowej. Autorzy podkreślają przemiany rozumienia roli bibliografii, od spisu publikacji znanych jej twórcy (role tego rodzaju opisów spełniają teraz katalogi biblioteczne), przez fenomen „bibliografii osobistych”, będących kompilacjami na potrzeby jednostkowej działalności i swoistym dodatkiem do pracy badawczej nad określonym zagadnieniem<sup>325</sup> (np. prywatne zbiory danych w oprogramowaniu do zarządzania bibliografią), po czym wprowadzają w temat działalności PBL jako nowoczesnej bibliografii dziedzinowej<sup>326</sup> i przechodzą do omówienia trzech zagadnień: problemu remediacji, rezultatów ankiety, a wreszcie prezentacji nowych perspektyw użytkowników PBL.

Chciałbym zacząć od środka, czyli od odniesień do przywoływanego wcześniej badania ankietowego. Koper i Umerle zaprzeczają intencji tworzenia osoby, ale sygnalizują świadomość użyteczności takiego sposobu myślenia o projektowaniu i utrzymywaniu serwisu. Jednocześnie podkreślają, że rodowód PBL wytwarza konieczność wychodzenia naprzeciw wielu bardzo różnorodnym oczekiwaniom i potrzebom. Tak przedstawiają dane na temat swojej grupy odbiorczej:

większość użytkowników korzysta z bazy w celach badawczych (43%), sporą grupę odbiorców stanowią osoby niezwiązane z akademią, z różnych środowisk zawodowych, których łączy

---

<sup>323</sup> Tamże.

<sup>324</sup> Koper i Umerle, „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”.

<sup>325</sup> J. Wojciechowski, Bibliografia wczoraj i dziś, „Forum Akademickie” 2011, nr 10, <https://prenumeruj.forumakademickie.pl/fa/2011/10/bibliografia-wczoraj-i-dzis/> (10.09.2017). Cyt. za: Koper, Beata i Tomasz Umerle.

<sup>326</sup> Była już o tym mowa wcześniej, ale pozostawiam w przypisie dłuższy cytat, który definiuje misję zespołu i pochodzi wprost od jego członków: „To, co wyróżniało od początku projekt PBL na tle innych tego rodzaju przedsięwzięć, to podbudowane naukową metodą dążenie do analizy całości piśmiennictwa w założonym okresie przy jednoczesnym przyjęciu szerokiej definicji kultury literackiej. Tak rozumiana bibliografia spełnia dwie kluczowe funkcje. Po pierwsze, dokumentuje dorobek kulturalny, po drugie, stanowi źródło informacji i podstawę warsztatu dla badaczy kultury. Tym samym zespół tworzący PBL związany jest podwójnym zobowiązaniem: wobec żywej kultury, która nieustannie się rozwija i wymaga rejestracji we wszelkich swoich ważnych przejawach oraz wobec użytkowników i odbiorców”. Zob. Koper i Umerle, „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”, 503.

zainteresowanie kulturą polską (31%). Wiele osób zadeklarowało przy tym, że korzysta z bazy w celach edukacyjnych (35%) i dydaktycznych (25%). Jeśli zestawić te informacje z danymi o zachowaniu użytkowników w serwisie PBL (m.in. przeglądanie materiałów za pomocą indeksu osób i czasopism oraz docieranie do konkretnych informacji z zewnętrznych wyszukiwarek) oraz kompetencjami cyfrowymi odbiorców, okazuje się, że zaprojektowanie funkcjonalnego interfejsu bazy danych wymaga znacznego nakładu pracy koncepcyjnej. Biorąc pod uwagę nieustanny rozwój technologiczny, należy uznać, że tego typu zadania – tworzenia i dopasowywania interfejsu serwisu online – powinny zostać na stałe wpisane w pracę zespołów dokumentalistycznych, jako element równoważny z nieustannym rozwojem metodologicznym i wzbogacaniem bazy o nowo opracowane źródła<sup>327</sup>.

Badacze związani z PBL proponują odniesienie do bogatych interfejsów, o których pisze Mitchell Whitelaw. Im więcej opcji interfejsowych, możliwości przeszukiwania, wyboru, eksponowania różnorodności – tym większa korzyść dla odbiorcy. Koper i Umerle dają tu wyraz przekonania o słuszności metod opisu, rozwijanych przez zespół PBL, ale wskazują, że mimo bogactwa dostępnych danych, aktualnym jest wyzwanie ich atrakcyjnego udostępnienia:

Aby spełnić oczekiwania użytkowników i równocześnie dać możliwość jak najpełniejszego wglądu w całość zasobów Polskiej Bibliografii Literackiej, nie jest potrzebna rewolucja metodologiczna (tutaj raczej obserwujemy nieustanną ewolucję związaną ze zmianami w obszarze przedmiotu zainteresowania bibliografii), tylko technologiczna, która pozwoli wzbogacić istniejące rekordy o nowe metadane oraz zaprojektować interfejs, który będzie wirtualną mapą pokazującą obszar, zakres i bogactwo materiałów odnotowanych w bibliografii, nie gubiąc przy tym informacji o samych publikacjach zawartych przecież nie tylko w metadanych, ale przede wszystkim w umiejscowieniu danego zapisu w strukturze działowej PBL

---

<sup>327</sup> Koper i Umerle, 507.

i adnotacji sporządzonej przez wykwalifikowanego bibliografa / bibliografkę<sup>328</sup>.

Postulowane przez badaczy zmiany to wprowadzenie wyszukiwania zaawansowanego, aby użytkownicy mogli zawęzić otrzymane rezultaty do dokładnego wyszukania, które jest aktualnie potrzebne, a także odwołania do całej sieci powiązań obecnych w bazie. Pod kątem procesu użytkowania autorzy widzą tutaj wartość w wędrówce przez kolejne poziomy hierarchii i przestrzenie struktur danych. Zaznają to odbiorcę z zasadami organizacji wiedzy, ale też wnoszą dodatkową wartość poznawczą. Syntetyczna wizja promowana przez badaczy to połączenie „pudełka wyszukiwania” jako prostego elementu funkcjonalnego z bogatym interfejsem, który ma pełnić rolę wsparcia nawigacji.

Z uwag Koper i Umerlego odnośnie do remediacji warto odnotować tę o mimikrze medialnej (idącą za uwagami Boltera i Grusina<sup>329</sup>), tzn. porównanie założeń publikacji drukowanej do medium cyfrowego; przykładem jest tu zachowanie indeksu rzeczowego, który ma spełniać takie zadanie jak indeks książki. Relacje pomiędzy różnymi formami składowania i przekazywania informacji mogą odsłaniać także większą złożoność przy dokładniejszym spojrzeniu. Przytoczmy za przypisem:

Zauważmy, że głównym narzędziem, które posłużyło do remediacji PBL, była relacyjna baza danych Oracle – a więc w pełni nowoczesne narzędzie gromadzenia i przetwarzania danych. Pierwsza baza danych PBL przechowuje nie tylko dane dostępne dzisiaj dla użytkowników, ale i pamięć o pomysłach na remediację drukowanej bibliografii. Niektóre kategoryzacje, struktury, tabele zostały zarzucone – i mogą być dzisiaj przedmiotem dociekań z pogranicza „archeologii danych”<sup>330</sup>.

Badacze związani z PBB rozróżniają dwa etapy remediacji, które zaszły w PBL. Samo powstanie bazy danych, jak to zostało powiedziane, było niejako przekuciem drukowanej monografii na zasób cyfrowy. Kolejny etap stanowiła modernizacja realizowana w ramach opisywanego wcześniej grantu NPRH. Prace prowadzono w oparciu o omówione badania

---

<sup>328</sup> Koper i Umerle, 508.

<sup>329</sup> Bolter i Grusin, *Remediation*.

<sup>330</sup> Koper i Umerle, „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”.

ankietowe. Dwa kluczowe aspekty zmian to przystosowanie interfejsu do oczekiwań odbiorców i przekształcenie modelu danych.

Jeśli chodzi o nowe perspektywy, Koper i Umerle podsumowują, że efektem remediacji pierwszej stało się utworzenie cyfrowego zasobu, liczącego około siedemset tysięcy zapisów, tworzących dokumentację polskiej kultury lat 1989-2002. Możemy dodać, że część rekordów dostępnych w druku została wprowadzona do bazy danych podczas prac nad remediacją PBL, a także odbył się import zewnętrznych źródeł (artykuł nie rejestruje jeszcze tego drugiego działania).

Wrócę teraz do tematu odbiorców, czyli osoby badawczej z perspektywy doświadczenia użytkownika serwisu. Jest to na pewno temat do dalszej eksploracji, mogłyby w tym pomóc badania Google Analytics, podobne do tych, które prowadził Marcin Werla. Wszelkie próby przedłożenia projektu cyfrowego możemy ująć w parze z realnym oczekiwaniem odbiorczym. Badania z 2016 roku wykazały istotną popularność wyświetleń indeksu osób, zatem czymś, co mogłyby przyciągnąć uwagę użytkowników są *datasets* związane z autorami, badaczami, krytykami i osobistościami kultury. Podobnie w przypadku kartoteki czasopism i spisu działów. Innym źródłem „wsadu” do takiego projektu może być analiza danych z wyszukiwarki serwisu i przygotowanie kolekcji tematycznej, osnutej wokół tematów lub grup tematów, które najczęściej występują w zapytaniach użytkowników).

Często najpopularniejsze części PBL w badaniu z użyciem Google Analytics były bardzo szczegółowe (np. periodyki regionalne i tematyczne w sekcji czasopism lub autorzy nienależący do ścisłego kanonu literackiego w szkole i na akademii w kartotece autorów; pięć najpopularniejszych zapytań wyszukiwarki w badanym okresie to „pamiętnik”, „neopoganie” „dramat”, „archetyp” i „Dumas”). Płynie stąd wniosek, że witrynę odwiedzają badacze i entuzjaści zainteresowani bardzo specyficznymi domenami. Są to świadome jednostki, zorientowane na określony wycinek tematyczny rejestrowanego dziedzictwa kulturowego.

Rysuje się zatem z jednej strony obraz specjalistek i specjalistów, korzystających z PBL dla pogłębiania wiedzy oraz własnych badań, głównie zorientowanych na cyfrową wersję publikowanych przez PBB informacji i zorientowanych na cel. Z drugiej – amatorów, obeznaných z internetem, zaciekawionych zagadnieniami pomieszczonymi w zbiorach bazodanowych. Tym pierwszym z perspektywy projektowania trzeba „nie przeszkadzać” otrzymać tego, co chcą uzyskać (surowe dane, weryfikacja konkretnej informacji, bibliograficzna kolekcja tematyczna dostępna do dalszych przetworzeń lub na cele

dydaktyczne). Tym drugim należy pomagać, o tyle, żeby podsyć pierwszą ciekawość i zachęcić do eksploracji. Pod względem zawartości mogą do tego służyć kolekcje tematyczne prezentowane w nieco bardziej „marketingowy” sposób (popularne społecznie, a nie badawczo tematy, odwołanie do rozpoznawalnych nazwisk, tytułów i zjawisk), materiały wiążące dane bibliograficzne z edukacją (wskazanie na zdobywanie wiedzy w kwerendzie w wizualizacji danych, rekomendacje, odniesienia do innych zbiorów, np. wersji pełnotekstowych dzieł lub serwisów edukacyjnych). Pod względem narzędziowym zapewniają to zazwyczaj atrakcyjność wyglądu serwisu i intuicyjna nawigacja, co sprzyja retencji użytkownika oraz powracaniu do serwisu.

Warto mieć na uwadze kilkumiesięczny zakres czasowy badań Google Analytics, od których minęło 7 lat. Nie tylko pożądanym byłoby ponowić analizę tego typu, ale też uruchomić ją na dłuższy okres (badanie pełnego roku kalendarzowego, badanie krótkich okresów np. kilkudniowe odczyty raz miesiąc, ale przynajmniej przez rok ze wskazaniem różnic miesiąc do miesiąca itp.).

Interesującym polem dalszych dociekań naukowych byłyby zmiany po stronie grupy odbiorców serwisu – grupy demograficzne i ich cele, punkty dotarcia i zainteresowania badawcze najbardziej zaangażowanych użytkowników. Niewykluczone, że od czasu badań Werli PBL zdążył się spopularyzować także w innych kontekstach. Mogło to wpłynąć na, wspomnianą już, podlegającą długoletnim zmianom, mieszankę oczekiwań odbiorczych.

Strategie i obecność PBL w serwisach społecznościowych

Na pewno dla pozyskania nowych grup odbiorczych i aktywizacji dotychczasowych użytkowników, serwisy określone wcześniej m.in. jako *Web 2.0* są bardzo skuteczne. PBL może poszczycić się aktywnością w mediach społecznościowych: PBB prowadzi od 2018 r. profil w serwisie Facebook. Treści publikowane na tej stronie dotyczą zarówno bieżącej działalności Pracowni i samego projektu bibliograficznego, jak i aktualnego życia literackiego (np. laureaci nagrody Nobla w dziedzinie literatury) czy społecznego (informacje na temat zbiorów kultury i ich przeglądania w kontekście *lockdownu* czasu pandemii COVID-19). Profil służy także do informowania o specjalnie tworzonych wydarzeniach w formie wirtualnej np. *webinarów* oraz spotkań na żywo. Dodatkowo promowane są pokrewne inicjatywy, wśród których warto wymienić Nową Panoramę Literatury Polskiej<sup>331</sup>, a także wydarzenia, projekty

---

<sup>331</sup> Zob. „Nowa Panorama Literatury Polskiej”, <https://nplp.pl/>. Dostęp: 2023-09-16.

i komunikaty Centrum Humanistyki Cyfrowej IBL<sup>332</sup>, jak i samego macierzystego Instytutu<sup>333</sup>. Dzięki żywemu reagowaniu na sprawy kultury i społeczeństwa, profil facebookowy osadza działalność Pracowni w szerszym kontekście naukowym, kulturalnym i społecznym. Obok promocji działań własnych, podkreśla więzi z innymi instytucjami, jak i wskazuje społeczny wpływ działalności PBL (odniesienie się do zdarzeń bieżących i podkreślenie połączenia z aktualnymi i przeszłymi pracami PBB tzw. naukowy impakt, który wykracza poza samą wymianę wiedzy w świecie nauki, a dotyczy także czytelników i odbiorców spoza świata akademii). Publikacja treści związanych np. z życiem literackim wpisuje się w działalność na rzecz popularyzacji nauki oraz odsłania interesujące fakty, znane pracownikom PBB, a niekoniecznie amatorom zagadnień kulturalnych<sup>334</sup>.

Z pewnością gdyby powtórzyć badania ruchu w PBL, Facebook i X (Twitter) mogłyby się okazać statystycznie istotnymi źródłami – punktami dojścia do danych Bibliografii. Media społecznościowe odgrywają też istotną rolę we współczesnym ekosystemie wymiany informacji w internecie, a zatem linkują, tworzą połączenia pomiędzy jednostkami, jak i pomiędzy zbiorami wiedzy. Przykładowo algorytm mediów społecznościowych podsunie do polubienia czy zaobserwowania stronę projektu humanistycznego użytkownicze, która wchodzi w interakcje z innymi, podobnymi stronami i ich treściami. Buduje to także relacje interpersonalne samych badaczy i tworzy półprofesjonalną sieć powiązań osobowych: śledzenie profili osób pozwala na zapoznanie się z efektami i metodami ich projektów, co przekłada się na dystrybucję wiedzy i może wytwarzać platformę do wymiany opinii poza istniejącym systemem konferencji i publikacji, jako wspierający podsystem komunikacji.

Jest wreszcie działalność w tych serwisach tworzeniem wizerunku. W warunkach współczesnej wymiany naukowej nie bez znaczenia pozostaje element *public relations* i tworzenie swego rodzaju marki – mam tu na myśli zarówno markę zespołu i określonej jednostki badawczej, jak i markę osobistą każdego poszczególnego badacza. Działalność PBB

---

<sup>332</sup> Profil CHC IBL PAN w serwisie Facebook – zob. „Centrum Humanistyki Cyfrowej IBL PAN” W *Facebook*. <https://www.facebook.com/Centrum.Humanistyki.Cyfrowej>. Dostęp 2023-09-20.

<sup>333</sup> Profil IBL w serwisie Facebook – zob. „Instytut Badań Literackich” PAN W *Facebook*, <https://www.facebook.com/InstytutBadanLiterackichPAN>. Dostęp 2023-09-20.

<sup>334</sup> Tu przykładem może być post na Facebooku dotyczący urodzin Krystyny Miłobędzkiej, zawierający odsyłacz do listy zapisów Twórcy w bazie, ilustrowany zdjęciem notatki, prawdopodobnie pochodzącej z papierowego indeksu osobowego, z prac nad monografiami papierowymi z lat PRL, kiedy ukazywały się pierwsze książki autorki. Widzimy tam m.in. odręcznie wpisaną notatkę na temat debiutu w 1960 roku i kolejne lata (publikacji), wpisywane różnymi kolorami i charakterami pisma. Zob. „Polska Bibliografia Literacka” w *Facebook*, <https://www.facebook.com/Polska.Bibliografia.Literacka/posts/175352946495747>. Dostęp 2023-09-20. W bibliografii załącznikowej zamieszczam wyłącznie odnośnik do profilu głównego.

na profilach mediów społecznościowych formułuje dla innych naukowców oraz użytkowników serwisu PBL tożsamość pracowni i jej zespołu. To także element promocji badań, związanych z serwisem. Tego typu komunikacja, w tonie mniej sformalizowanym, jest też swoistym odczarowaniem pracy archiwistycznej i dokumentalistycznej poprzez rozprawienie się z pokutującymi mitami i fałszywymi przeświadczeniami. Wśród nich można wymienić dwa: pracę archiwistyczną i opracowywanie bibliografii jako działanie mające niewiele wspólnego z nowoczesnymi technologiami, czemu zadaje kłam komunikowanie badań związanych z humanistyką cyfrową, a także domniemane akademickie oderwanie od rzeczywistości w codziennej pracy instytucji – przeciwstawiając mu uczestnictwo w komunikacji na tematy bieżące i istotne dla życia społecznego.

Znajdziemy nawet burzenie obrazu samych dokumentalistów jako osób pozbawionych poczucia humoru i dystansu do samych siebie, czy nienadążających za trendami *social media*: tu wystarczy wspomnieć o postach prezentujących koty pracowników PBB<sup>335</sup>, dozie autoironii w przesłaniu dalej publikacji profilu IBL z satyrycznym fragmentem o trudzie tworzenia bibliografii<sup>336</sup> oraz anegdotyczne działania wpisujące się w trendy mediów społecznościowych, jak zaaranżowanie zdjęcia o rzekomym znalezisku dyskietki pośród fiszek papierowego indeksu Pracowni<sup>337</sup>. Osobną grupę stanowią pojawiające się co jakiś czas posty, prezentujące swoistą maskotkę pracowni, czyli figurkę drewnianego czerwonego pieska, która należała do założyciela PBB, prof. Stefana Vrtela-Wierczyńskiego. Pracownicy PBL zabierają ją np. w delegacje i fotografują w odwiedzanych miejscach<sup>338</sup>. Tego rodzaju zbiór treści buduje obraz dostępności nauki i w przystępny sposób ukazuje rezultaty prac zespołu.

Media społecznościowe pełnią również dodatkową rolę komunikatora między grupą odbiorczą a naukowcami. Ze względu na dynamikę kontaktu, *social media* pomagają np. efektywniej zadawać pytania do zespołów. Dzięki wspomnianej zwiększonej widoczności, po ogłoszeniu

---

<sup>335</sup> Zob. „Polska Bibliografia Literacka” w *Facebook*,

<https://www.facebook.com/Polska.Bibliografia.Literacka/posts/pfbid05VLFVoRGUqbgNX6iQKHHTgx9aUeHjqCWnvP57xWkjc1r5qkVVc6GKERV5okcDJLVlhttps://www.facebook.com/Polska.Bibliografia.Literacka/posts/pfbid029WvxyccqaeJtuacmiSm6SwAcqeJsGKaKMQVQQtai3c1ke3MYRj8ba6n7vLY3v1qe7l>. Dostęp 2023-09-20.

<sup>336</sup> Zob. „Polska Bibliografia Literacka” w *Facebook*,

<https://www.facebook.com/Polska.Bibliografia.Literacka/posts/pfbid02TsBdqr9gP4nzY3Tx34y95kw1oiJpU4qqD7ugZA2MfZ1r5iLbcSo7BuktmYwEcqzEl>. Dostęp 2023-09-20. Jest to fragment szopki z 1955 autorstwa jednego z twórców, Romana Lotha.

<sup>337</sup> Zob. „Polska Bibliografia Literacka” w *Facebook*,

<https://www.facebook.com/Polska.Bibliografia.Literacka/photos/a.193129268051448/193111504719891/>

<sup>338</sup> Zob. przykład z Gazu: „Polska Bibliografia Literacka” w *Facebook*,

<https://www.facebook.com/Polska.Bibliografia.Literacka/posts/pfbid0z7kjt8d115PhNK7U86PvZopTgtnT3krfyzCJeYwsFyjc7XPYJNmiMEqjyqRd4Cadl>. Dostęp 2023-09-20.



jakiegokolwiek inicjatywy w sieci społecznościowej przez daną jednostkę (np. *webinar*) można doprecyzowywać szczegóły z organizatorami (jak chociażby często pojawiające się prośby o udostępnianie materiałów i nagrań transmisji, także już po wydarzeniu – szczególnie od osób, które nie mogą dołączyć na żywo). Inną zaletą staje się rola mediów społecznościowych jako swoistego (niezamierzonego) centrum pomocy technicznej – otwarty kanał komunikacji zapewnia możliwość zadawania pytań w razie użytkowniczych niepowodzeń (np. gdy nie możemy wyszukać informacji na stronie), ale także w sytuacji usterek i awarii (błędy wczytywania strony, niemożność pobrania pliku, brak zawartości). Użytkownicy zapewniają dodatkowy aspekt kontroli jakości (QA, *Quality Assurance*), ponieważ sami mają motywację do korzystania z serwisów i chcą, aby były one jak najlepsze. Mogą wykrywać i zgłaszać błędy. Zwykle w obszarze QA mamy na myśli błędy techniczne, ale tym lepiej dla humanistycznych baz danych, jeśli zapewnia się możliwość raportowania również błędów merytorycznych lub pomyłek redaktorskich. Może się to wydawać nieintuicyjne, a nawet godzić w przyjęte przez daną instytucję procedury, ale zwiększenie widoczności w sieci przekłada się na tego rodzaju drobne usprawnienia, a w zarządzaniu proceduralnym można zawsze wskazać użytkownikom „algorytm” postępowania (np. prośba o przeniesienie rozmowy do konwersacji e-mailowej, skorzystanie z odpowiedniego formularza w sieci, czy nawet utworzenie zadania na otwartej kolejce w systemie zarządzania pracą informatyczną itp.). Takie zarządzanie informacją zwrotną buduje zaufanie i jeśli użytkownik zostaje wysłuchany, a zmiany zostają wprowadzone, stwarza to wartość dodaną w projekcie i okazję do kolejnych sugestii, przez co dany wytwór (baza bibliograficzna, strona humanistycznego projektu cyfrowego) jest lepszy i może być stale rozwijany.

Obok mediów społecznościowych warto wymienić serwis GitHub, który służy międzynarodowej społeczności programistów do wymiany częściami gotowego kodu. W ten sposób można zastosować cudze narzędzia do rozwiązania problemu na który się trafiło, ale też „forkować” czyli tworzyć własne wersje, oparte na programie innej osoby (*forks* od angielskiego określenia oznaczającego rozgałęzienia i odnogi np. rzek). Serwis opiera się na dość zminimalizowanej części interfejsu użytkownika, ponieważ stanowi w znacznej mierze repozytorium kodu. Dzięki niemu można w łatwy sposób przeglądać i analizować używane rozwiązania programistyczne, odwoływać się do dokumentacji (o ile została dołączona jako osobny opis w plikach tekstowych, ew. bezpośrednio w formie komentarzy w kodzie).

Wśród twórców PBL, aktywnym publikującym jest Cezary Rosiński<sup>339</sup>. W repozytorium jego autorstwa badacze mogą znaleźć kod w językach R (statystyka, analiza danych, wizualizacja) i Python (język ogólnego zastosowania). Pod względem celu, publikowane narzędzia służą m.in. do zarządzania bieżącą pracą PBB (raportowanie, wsparcie osób kierujących jednostką), konwersji typów danych (np. zmiany formatów, harmonizacja), *webscrapingu* (pozyskiwanie zbiorów danych bezpośrednio z witryn internetowych, przez co strona internetowa zostaje potraktowana jako baza danych i zamieniona na formę tabelaryczną), czy agregacji dużych zbiorów danych poprzez łączenie ze sobą wielu mniejszych źródeł. Kolekcje obejmują m. in. trajektorie recepcji literatury czeskiej, analizę danych związanych z nagrodami literackimi czy badania samizdatu (co koresponduje z działalnością publikacyjną i projektową autora). Warto wspomnieć też o działalności Patryka Hubara, który w swojej pracy zajmuje się remediacją papierowych roczników monografii, aby mogła służyć użytkownikom bazy jako źródło cyfrowe<sup>340</sup>. Na jego koncie znajdziemy np. zapisy pracy wykładowczej (autorski kurs języka Python), ale także repozytoria ściśle związane z obróbką danych PBL. Należą do nich narzędzia wykorzystujące dane m. in. Wikidata i Federacji Bibliotek Cyfrowych, których można użyć w połączeniu z większą grupą zbiorów informacji. Inne osoby związane z PBB, mające wpływ na kształt PBL pod względem dokumentalistycznym i programistycznym to Barbara Wachek<sup>341</sup> i Nikodem Wołczuk<sup>342</sup>. Jest to jak widać znacząca grupa osób zajmujących się (w mniejszym lub większym stopniu) programowaniem i automatyzacją codziennych zadań w pracy bibliograficznej, biorąc pod uwagę nieduży zespół samej jednostki<sup>343</sup>. Świadczy to o szkoleniu kadr w obrębie Pracowni z myślą o rozwijaniu projektów cyfrowych. Odpowiada to na potrzeby rozwoju infrastrukturalnego, które były sygnalizowane w przywoływanym wcześniej *whitepaper* z UCLA, dotyczącym „obietnicy” uprawiania nauk humanistycznych w sposób zdigitalizowany.

U takich aktywnych użytkowników GitHub, związanych z cyfrową humanistyką da się dostrzec zainteresowania badawcze oraz ich przemiany (a niejednokrotnie natrafimy na projekty hobbystyczne czy rozwijające inne dziedziny wiedzy lub służące edukacji

---

<sup>339</sup> Zob. „Cezary-Rosinski” w GitHub, <https://github.com/cezary-rosinski>. Dostęp 2023-09-20.

<sup>340</sup> Zob. „Patryk Hubar” w GitHub, <https://github.com/patthub>. Dostęp 2023-09-20.

<sup>341</sup> Zob. „Barbara Wachek” w GitHub, <https://github.com/barbara-wachek>. Dostęp 2023-09-20.

<sup>342</sup> Zob. „Nikodem Wołczuk” w GitHub, <https://github.com/nwolczuk>. Dostęp 2023-09-20.

<sup>343</sup> Zob. „Pracownia Bibliografii Bieżącej”, <https://ibl.waw.pl/pl/o-instytucie/pracownie-i-zespoły/pracownia-bibliografii-biezacej>. Dostęp 2023-09-20. Patryk Hubar wymieniony jako współpracownik, łącznie 14 osób jako pracownicy, z czego trójka z nich dzieli się kodem na GitHub (stanowiska kierownicze sprawują dwie osoby, emerytowanych współpracowników jest czwórka).

i rozwojowi umiejętności). Nieco górnolotnie można je określić jako część naukowej i nienaukowej autobiografii, pozostawieniem swojej sygnatury w sieci<sup>344</sup>.

Dzięki publikacji rozwiązań informatycznych w GitHub, zwiększa się widoczność pracy humanistów cyfrowych, jak i możliwość dotarcia do jej efektów w celu dalszego wykorzystania. Są także swoistą „bibliografią cieni”<sup>345</sup> dla projektów współczesnej humanistyki, w której (jak widzieliśmy m.in. na przykładach Jockersa i Pipera) kod programistyczny współlistnieje z pisarstwem akademickim.

Ostatnia istotna cecha GitHub to – poza dostępnością kodu – możliwość obserwowania innych (*following*). Pierwsza korzyść płynie z otrzymywania powiadomień o nowych projektach dodanych przez wybrane osoby. Druga – tworzenie się sieci badawczych i społecznych między osobami o kompetencjach cyfrowych. Często mamy do czynienia z efektem wzmocnienia przez sieć, tzn. mocny węzeł (badacz o dużej estymie, jak chociażby przywoływany Silvio Peroni), łączy jako pośrednik osoby obserwujące o zbieżnych aspiracjach – dzięki temu otwierają się możliwości przeglądania podobnych projektów na kontach takich użytkowników oraz zawiązywanie przygodnych projektów i stałych kontaktów, które nie urzeczywistniłyby się bez znalezisk w przestrzeni wymiany kodu. Wnosi to aspekt tworzenia się i rozwoju naukowych wspólnot.

Połączone zbiory danych

Druga i trzecia dekada XXI w. przynoszą w PBL projekty związane z rozszerzaniem pola powiązań bazy centralnej dla działalności zespołu z innymi źródłami dokumentacji kultury.

Wspominaliśmy już o zakresie rejestracji bazy bibliograficznej, który dotyczy także m.in. filmu. Interesującą badawczo narrację naukową na studium przypadku prezentuje zespół PBB

---

<sup>344</sup> W tym miejscu warto zaznaczyć, że same osoby publikujące decydują o dostępie do swojego kodu i niektóre repozytoria mogą być zamknięte lub posiadać określone poziomy uprawnień. Dlatego konta GitHub możemy potraktować jako wizytówki badaczy: eksponowane są treści, którymi chcą się podzielić z szerszą grupą odbiorców w sieci.

<sup>345</sup> Tu nasuwa się ponowne skojarzenie z genetyką tekstów: wspominałem już o archiwach pisarzy będących „na wymarcu”, mając na myśli zmiany metod pracy twórczej na zdigitalizowane i niepozostawiające po sobie praktycznie śladu, gdy mowa o procesie oraz wariantach tekstu. Tymczasem dokumentacja programistyczna, pełna śladów w czasie (*time stamps*), a także u swoich podstaw oparta na wersjonowaniu pracy, okazuje się bardzo interesującym obszarem do badań. Być może doczekamy się „genetyki kodu”, która pokaże autorski charakter rozwiązań, możliwe przyszłości projektów, które się nie wydarzyły (odrzucone alternatywy), czy porzucone w żmudnej edycji funkcjonalności (usunięte niczym w geście redakcyjnej obróbki). U Rosińskiego widać zmianę preferowanej technologii czyli przejście od języka R do języka Python, ale takich interesujących fenomenów i możliwych spostrzeżeń w archiwach GitHub jest na pewno dużo więcej.

o wzbogaceniu danych PBL danymi Archiwum Telewizji Polskiej<sup>346</sup>. Tekst opisuje proces zdobywania nowej wiedzy (i uzupełniania danych w procesie dokumentalistycznym), a jednocześnie formułuje pewne wskazania w kierunku dobrych praktyk, ogólniejsze zasady do użytku w podobnych kontekstach. Wśród tych uwag, co dla nas ważne, pojawia się sporo na temat metadanych. Autorzy kreślą kontekst przemian układu działów bibliografii względem filmu, telewizji i radia. Przywołują także źródła drukowane i instytucje, w których odbywano kwerendy, aby pozyskać tę część materiału wcześniej, przy analogowym tworzeniu monografii. Nawiazanie współpracy już w czasie zaawansowanej cyfryzacji projektu wiązało się z dwoma czynnikami: zabrakło głównego źródła opracowywania materiału dziedzinowego (zespół opracował ostatni rocznik wydawania tygodnika *Antena*), a w związku z działaniem grantowym, polegającym na rozwoju kompetencji cyfrowych u dokumentalistów na cele migracji ze starej bazy do nowej, można było zastosować tę wiedzę i umiejętności do pozyskania innych zasobów. Okazało się, że tego typu informacje są przechowywane w bazie wewnętrznej TVP, do której standardowo nie składano zamówień – potrzeba PBB, aby pozyskać metadane na temat archiwaliów telewizji była tutaj wyjątkowa i nieoczekiwana dla wykonawcy. Wymagało to też wyjątkowej procedury administracyjnej aby przygotować dane w sposób komercyjny. Badacze PBB raportują, że udało się otrzymać metadane 447 tytułów jednostkowych z łącznej liczby 1507 emisji. Były to spektakle teatru TV oraz dokumentalne materiały o literaturze, teatrze, filmie, pisarzach i aktorach z lat 1989-2016. Wystąpiła niezgodność formatu oczekiwanego (tabela .csv) i dostarczonego („dane w szcążtkowo tabelarycznej formie HTML”<sup>347</sup>). Pojawiły się wyzwania informatyczne: brak odpowiednich znaczników, podział na kolumny wyrównany poprzez ręcznie wstawiane wiele znaków spacji. Zapisy nie były też zgodne ze specyfikacją pod względem potrzebnych typów informacji, ale stanowiły zrzut wszelkich dostępnych. Utrudniało to mapowanie struktur:

Choć uzyskane dokumenty pozwalały na przetworzenie otrzymanych danych do podstawowej postaci tabeli za pomocą występujących w danych regularności, to pojawiły się również inne konsekwencje źródłowego braku oczekiwanej przez zespół PBL segmentacji tekstu:

- 1) wielokrotnie i bez regularności w kolumnie wartości pojawiały się

---

<sup>346</sup> Tomasz Umerle, Beata Koper, i Cezary Rosiński, „Dokumentacja kultury a cyfrowe zasoby archiwalne. Przypadek Polskiej Bibliografii Literackiej (PBL) i Archiwum Telewizji Polskiej”, *Archiwa - Kancelarie - Zbiory*, nr 9(11) (2018-12-31): 109, <https://doi.org/10.12775/AKZ.2018.006>.

<sup>347</sup> Umerle, Koper, i Rosiński, 116.

nazwy kategorii; 2) występującym w kolumnie kategorii wartościom odpowiadało wiele wierszy w kolumnie treści (niemniej jedynie pierwszy z wierszy w kolumnie wartości był pierwotnie powiązany relacją z daną kategorią; kolejne relacje należało wytworzyć); 3) uzyskana dwukolumnowa tabela musiała zostać przetworzona na tabelę wielokolumnową (kategorie winny występować jako kolumny, a nie wiersze, co z perspektywy bazy danych PBL jest dużo bardziej operatywne)<sup>348</sup>.

Problemom z formatem danych towarzyszyły też te dotyczące ich modelu. Obie bazy – źródłowa i docelowa – posługiwały się własnymi sposobami opisu, co wymagało wytworzenia standaryzacji własnej w obliczu braku wspólnego standardu.

Fakt korzystania z „lokalnych” formatów metadanych, czyli wystarczających dla kontekstu działalności, stanowił dla zespołu PBL przesłankę do formułowania wniosku o konieczności budowania struktur współpracy między takimi jednostkami jak PBB i Archiwum Telewizji Polskiej. Nie – standardów, ponieważ wyzwaniem może być poszukiwanie uniwersalności wygodnej dla każdej ze stron, ale właśnie procedur dialogu oraz otwierania metadanych. ciekawą obserwacją było bowiem uznanie przez jedną z instytucji, że metadane mogą nie stanowić wartości poza swoim kontekstem (zasób wewnętrzny, służący do archiwizacji działalności bieżącej). Tymczasem przy udostępnieniu ich może się okazać, że „dane o danych” znajdą odpowiedni kontekst użycia „samodzielnie” – tak jak to obserwaliliśmy, śledząc różnego rodzaju serwisy w poszukiwaniu informacji o Andrei Dworkin. Zupełnie nie dziwi, że także w studium przypadku współpracy PBB i Archiwum rekomendacją trójki badaczy było stosowanie takich opisów, by dało się za ich pomocą budować struktury *Linked Open Data*.

Zespół PBL aktywnie szuka również usprawnień poprzez wykorzystanie nowoczesnych narzędzi do powiększenia bazy rekordów z wykorzystaniem istniejących danych opublikowanych w internecie. Inicjatywa iPBL wpisuje się w porządek zarządczy badań i rozwoju (R&D, *Research & Development*), czyli szukania nowych ścieżek dla prowadzonego projektu<sup>349</sup>. W tym podejściu myśli się szerzej i bardziej perspektywicznie, często ignorując

---

<sup>348</sup> Umerle, Koper, i Rosiński, 118-119.

<sup>349</sup> Dostępne materiały konferencyjne (plakaty badawcze) pod dwoma adresami: Beata Koper i in., „iPBL – supplementing literary bibliography with internet sources. Collaborative cataloguing”, 2023-06-30,

codzienne ograniczenia – polega to wszak na tym, by wyszukać jeszcze nie eksplorowane ścieżki. Takimi w przypadku PBL okazują się *web scraping*, analiza mediów i serwisów społecznościowych, czy blogów, stron poświęconych polskiej kulturze i literatury elektronicznej.

Z tą ostatnią wiążą się także przejawy nowych rozwiązań technologicznych do badań. Jest to np. medium wysoce efemeryczne – podobnie jak niektóre miejsca w sieci poświęcone krytyce literackiej czy publikacji twórczości własnej. Badacze często muszą polegać na automatycznym archiwizowaniu stron, jak Internet Archive<sup>350</sup>. Zespół PBB zidentyfikował wiele źródeł, w tym strony internetowe twórców i wydawnictw literackich, które warto w ten sposób udokumentować, ale wdrożenie planu opisu literackości w internecie to także kwestia pozyskania dodatkowych zasobów, czasu i priorytetyzowania bieżących zadań<sup>351</sup>. W działaniach R&D Pracowni widać inicjatywę i chęć przystosowania metody badawczej do nowych czasów i nowych mediów. Opisywany temat zmienia się przecież nieustannie.

Dane z PBL wchodzi w skład Europejskiej Bibliografii Literackiej (ELB), łączącej także źródła czeskie i fińskie<sup>352</sup>. Tego rodzaju zbiory sprzyjają „uwalnianiu” danych do szerszych kręgów odbiorczych. Badacze związani z PBB angażują się w rozwój rodzimej humanistyki cyfrowej, także przez publikację *datasetów*, takich jak zbiór danych dotyczących europejskich nagród literackich<sup>353</sup>. W kolejnym rozdziale odwołuję się także do współpracy polsko-czeskiej, w której bierze udział Cezary Rosiński. Rozszerzenie grona odbiorczego danych PBL do użytkowników sieci spoza Polski i nie posługujących się językiem polskim, oraz do odczytań maszynowych to istotny cel rozwoju bazy.

Dla PBL będzie też na pewno istotny nowy odbiorca-cyborg czy algorytm, jakiego starałem się wywoływać w rozdziale poprzednim. Częściowo możemy przewidywać trendy na podstawie współczesnych obserwacji. Na podstawie badań Werli widać jak wyszukiwarka Google „dysponuje” zapytaniami, nawet wśród doświadczonych (powracających)

---

<https://doi.org/10.5281/ZENODO.8228490>; Beata Koper i in., „iPBL – Supplementing Literary Bibliography with Internet Sources. Collaborative Cataloguing”, 2023-06-30, <https://doi.org/10.5281/ZENODO.8107864>.

<sup>350</sup> Zob. <https://archive.org/web/>.

<sup>351</sup> Dziękuję zespołowi PBB za udostępnienie części informacji związanych z ubieganiem się o dofinansowania na ten cel; z dokumentacji merytorycznej wynika, że analizowane mogłyby być nawet ponad 4000 źródeł internetowych.

<sup>352</sup> Zob. <https://literarybibliography.eu/pl/>. Por. Patryk Hubar i Marcin Giersz, „European Literary Bibliography: Aggregation and Harmonization of Literary Bibliographical Data”, 2023-03-07, <https://doi.org/10.5281/ZENODO.7705026>.

<sup>353</sup> Paulina Czwordon-Lis i in., „Dataset of European Literary Awards” (Zenodo, 2022-11-30), <https://doi.org/10.5281/ZENODO.7380507>.

użytkowników serwisu PBL. Ta swoista centralizacja ruchu wiadczy to tyle o jej popularności, co również o nawykach osób korzystających z internetu. Działalność Google i innych serwisów wyszukiwania ma też złą stronę. Rola centrali-rozdzielni współczesnych wyszukiwarek przy nie zawsze jasnym działaniu algorytmów ich rekomendacji i tworzenia rankingu wyników dla zapytań, może utrudniać dostęp także do treści PBL. Można śmiało założyć, że bazy bibliograficzne, o ile nie poszukuje się ich strony głównej, przegrywają w wyścigu o uwagę typowej użytkowniczki internetu z płatną promocją, treściami sprzedażowymi i stronami należącymi do mediów masowych (nawet biorąc pod uwagę profilowanie wyświetlania treści). Wcześniej sygnalizowałem problemy, które się z tym wiążą, dlatego wróć do propozycji rozwiązań: jeśli pojedynczy użytkownik wyszukiwarek, będący badaczem, nie ma łatwej ścieżki dotarcia do bibliograficznej treści, która przydałaby się w działalności naukowej, odpowiedzią jest zwiększanie widoczności źródeł poprzez konsolidowanie ich w większe zbiory i przetwarzanie przez międzynarodowe konsorcja oraz udostępnianie zgodnie z zasadami otwartej nauki<sup>354</sup>. Dzięki temu naukowcy docierają do danych w swoim „naturalnym środowisku”, np. na konferencjach i seminariach, co nie tylko zwiększa widoczność źródeł wiedzy, ale też osadza je we właściwszym kontekście.

#### Podsumowanie

Odwołując się do funkcji wymienionych przez Szymańską, przytaczanych na początku rozdziału, możemy teraz zrewidować wskazane przez badaczkę cele bibliografii i opisać ich realizację przez PBL. Pod względem funkcji naukowej jest to projekt dwojaki: umożliwia prowadzenie badań, zapewniając dostęp do źródeł, jak i sam rozwija się w związku z działalnością naukową jego współtwórców. Funkcje: rejestracyjna, dokumentacyjno-źródłoznawcza, informacyjna, informowania o treści publikacji, porządkowania faktów w czasie, archiwalno-historyczna – realizują się w codziennej pracy zespołu bibliograficznego, dodającego nowe zapisy z autopsji, a także poprzez bazę danych i serwis internetowy, zapewniające przechowywanie informacji oraz dostęp do nich. Funkcja integracji całej literatury polskiej (krajowej i emigracyjnej) pomimo wyzwań jest realizowana, warto tu zauważyć, że właśnie narzędzia cyfrowe mają zapewnić ciągłość i bieżący charakter rejestracji, a projekty z użyciem wielu źródeł danych konsolidują wiedzę. Funkcje edukacyjna, ogólnokulturowa oraz marketingowa zauważymy przede wszystkim w nowym serwisie, przyjaźniejszym dla użytkownika i zachęcającym do odwiedzin, promującym jednocześnie

---

<sup>354</sup> Nie dotarłem do opracowań, które mówiłyby o korzyściach pozycjonowania stron naukowych w wyszukiwarkach internetowych. Takie działanie w świetle powyższych ustaleń wydaje się być nieopłacalne. Analiza Werli nie zawiera takich rekomendacji dla PBL.

humanistykę. Jak dało się zauważyć, realizowana jest także funkcja bibliometryczna, ale nie jako główne źródło wiedzy w tym zakresie<sup>355</sup>. Podobnie z naukoznawczą. Co do funkcji bibliotecznej, kontaktu z autorami materiałów i genealogicznej jest zbyt mało materiałów dostępnych w serwisie, by uznać ich istotność. W tym wypadku również szukałbym pośredniej realizacji tych funkcji przede wszystkim poprzez dane połączone, wskazując w pierwszej kolejności na integrację czy wymiany danych z Biblioteką Narodową.

Historia PBL jest historią adaptacji do zmieniających się warunków: najpierw ambitną próbą opisania całości piśmiennictwa polskiego w kraju i za granicą, jako kierunek wyznaczony przez Vrtela-Wierczyńskiego i jego zespół. Następnie aktualizacją i uzupełnieniem braków, wynikającymi z uświadomionych, ale niewyrażonych wcześniej ograniczeń politycznych w metodzie archiwizacji. Kolejną granicą do przekroczenia stała się cyfryzacja i remediacja treści – z jednej strony w celu utrzymania aktualności metodologicznej, z drugiej zaś z myślą o szerszym udostępnieniu wartościowych zbiorów. Jest wreszcie PBL cyfrowym projektem humanistycznym, obliczonym na długie trwanie w cyberprzestrzeni i to obecna perspektywa: wymyślenie bibliografii dziedzinowej jako bazodanowego zbioru informacji, użytecznych dla ludzkich i maszynowych aktorów sieci internetowej, którzy i które będą przeszukiwać, przechowywać, przetwarzać oraz produkować kolejne dane oraz treści. Projekt z zamysłu dokumentalistyczny, przeznaczony dla odbiorcy lokalnego, możemy traktować jako część superorganizmu, jaki stwarza współczesna komunikacja w nauce, który stale będzie się przeobrażać w kierunku synergii działania ludzkich i nie-ludzkich elementów w systemie<sup>356</sup>.

---

<sup>355</sup> Por.: Drabek.

<sup>356</sup> Mam tu przede wszystkim na myśli koncepcję cyborgizacji, spopularyzowaną np. przez Donnę Haraway i liczne komentarze do jej manifestu. Cyborgizacja zaciera granice między tym, co ludzkie, zwierzęce i maszynowe, jak również eliminuje możliwość rozdzielenia natury od kultury. Zob. Haraway, Donna Jeanne. *Manifestly Haraway*. Posthumanities 37. Minneapolis: University of Minnesota Press, 2016, gdzie zebrano nie tylko 'A Cyborg Manifesto', ale też manifest gatunków stowarzyszonych (odwołujący się np. do związków ludzi i psów i krytyczny wobec idei zwierzęcia domowego) oraz tekst o relacjach międzygatunkowych. Pierwsze wystąpienie pojęcia „cyborg” było natomiast związane z koniecznością rozszerzenia zdolności ludzkiego organizmu w perspektywie przeniesienia ludzkiej cywilizacji poza Ziemię i dalekich podróży kosmicznych, por. Saulius Vydas, „Cyborg: Evolution of the Superman”, *JAMA: The Journal of the American Medical Association* 194, nr 4 (1965-10-25): 474, <https://doi.org/10.1001/jama.1965.03090170152043>. Osobiście rozumiem aspekt cyborgiczny w sieci jako nieodzowność korzystania z maszyn uprzednio zewnętrznych wobec ludzkiego ciała (niewytworzonych biologicznie), w codziennym funkcjonowaniu, co można obecnie przyjąć za istotne w każdej pracy ze zbiorami informacji, włączając w to humanistykę. Podobne idee na temat rozszerzania ludzkiego aparatu poznawczego dzięki technologii prezentuje Marshall McLuhan, por. Marshall McLuhan, *Galaktyka Gutenberga: tworzenie człowieka druku*, tłum. Andrzej Wojtasik i Grzegorz Godlewski, Wydanie III, Myśl o Kulturze (Warszawa: Narodowe Centrum Kultury, 2021).



## Rozdział 4. Studium przypadku

*Literatura na Świecie* – hipertęcza zza Żelaznej Kurtyny

W wywiadzie udzielonym Michałowi Nogasiowi<sup>357</sup>, o wiele mówiącym tytule *Jakie książki czyta Olga Tokarczuk?*, polska noblistka dzieli się doświadczeniami, które uformowały ją jako pisarkę. Wspomina pierwsze nastoletnie próby, związane z publikowaniem w piśmie *Na Przelaj* (debiut pod pseudonimem Natasza Borodin w 1979). Autorka *Czulego narratora* zwraca uwagę na – ówczasnie niełatwą – możliwość literackiego zaistnienia dla osób z prowincji, dzięki otwartym łamom *Na Przelaj* i powołanemu przez redakcję Klubowi Młodych Autorów. Rozmowa płynie jednak w kierunku doświadczeń czytelnicznych: tego, co wpłynęło na wrażliwość odbiorczą Tokarczuk. Autorka wspomina:

Ja i tak miałam ogromne szczęście, bo do naszego domu przychodziły regularnie magazyny poświęcone literaturze. Dzień, w którym listonosz przynosił nowy numer „Życia Literackiego”, był świętem. Siadałam i czytałam od deski do deski, miałam dzięki temu wrażenie, że choć mieszkam na prowincji, na bieżąco wszystko śledzę. A gdy miałam 16 lat, odkryłam „Literaturę na Świecie”. I z ręką na sercu mogę powiedzieć, że bez niej nic by się w moim zawodowym życiu nie wydarzyło. Najpewniej w ogóle nie zaczęłabym pisać. Ten miesięcznik był jak objawienie, pokazywał mi teksty ze świata, którego nie znałam, pisane przez ludzi, których istnienia nawet sobie nie wyobrażałam.

**Brzmi trochę jak internet.**

„Literatura na Świecie” była wtedy moim internetem. I, co ważne, obok tłumaczeń tekstów z różnych kontynentów miała też teksty krytyczne – bardzo inteligentne, pisane przez najlepszych wówczas w Polsce eseistów i tłumaczy.

---

<sup>357</sup> O. Tokarczuk, *Jakie książki czyta Olga Tokarczuk?*, rozm. M. Nogaś, <https://wyborcza.pl/ksiazki/7,154165,25471977,jakie-ksiazki-czyta-olga-tokarczuk-gdy-nie-pisze-rozmowa.html> (dostęp: marzec 2022).

One pokazywały nie tylko, jak tę literaturę należy interpretować, ale też, gdzie szukać jej źródeł, co i gdzie jeszcze przeczytać. Dziś powiedzielibyśmy, że doskonale linkowała.

Jak widać, dla polskiej noblistki ten tytuł miał niebagatelne znaczenie w rozwoju twórczym. Z kilku rozproszonych i niewątpliwie różnorodnych źródeł wybieram jeszcze dodatkowe określenia działalności periodyku i relacje – dla uproszczenia używając odtąd skróconej nazwy LnŚ.

Tomasz Mielczarek, analizując legalnie wydawane w PRL tytuły lat osiemdziesiątych (czyli takie, które podlegały cenzurze – w przeciwieństwie do drugiego obiegu, samizdatów), zalicza LnŚ do istotnych pism społeczno-kulturalnych<sup>358</sup>. Główną osią przytaczanego tekstu jest wpływ wydarzeń tamtego dziesięciolecia (Sierpień 1980, wprowadzenie stanu wojennego w 1981) na funkcjonowanie redakcji i ich działania, a także wpływem na odbiór i reakcjami czytelnickimi – co nie będzie nas tu zajmować. Autor podaje natomiast opartą na danych statystycznych charakterystykę rynku czasopism i klasyfikację tytułów, co pozwoli się zorientować w zajmowanym przez LnŚ miejscu w obiegu wydawniczym. Tu cenne jest przyporządkowanie do określonej grupy czasopism. Wcześniej badacz pisze o tygodnikach, które dziś nazwalibyśmy tygodnikami opinii (warszawska *Kultura*, *Polityka* czy *Tygodnik Powszechny*), wymienia istotne tytuły o zasięgu regionalnym, a następnie włącza do analizy kolorowe magazyny (dzięki temu ująć może *Ekran*, *Film* i *Przekrój*). Jako ważny punkt odniesienia do powyższych Mielczarek podaje „grube magazyny”<sup>359</sup>: opisuje je jako zróżnicowaną grupę niskonakładowych miesięczników – tu właśnie przyporządkowuje m. in. LnŚ. Autor umieszcza w tej grupie pisma katolickie, jak *Znak* i *Więź*, a także literackie i „o treści specjalnej”: *Teatr*, *Kino*, *Nowe Książki*, *Sztuka*. Wśród pism literackich, obok LnŚ zostają wspomniane *Twórczość*, *Dialog*, *Poezja*, *Miesięcznik Literacki*<sup>360</sup>. Daje to obraz tła, na jakim funkcjonowała redakcja wśród pism społeczno-kulturalnych. Dla wyobrażenia zasięgu tytułów z grupy owych „grubych magazynów” podaję za Mielczarkiem: „zazwyczaj czasopisma te wychodziły w nakładach wynoszących od 5 do 15 tys. egzemplarzy. Nakład zbliżony do 20 tys. egz. miało jedynie, ukazujące się od 1978 roku, społeczno-polityczne, choć nie stroniące od kwestii kulturalnych, krakowskie »Zdanie«”<sup>361</sup> (zaliczany do kolorowych

---

<sup>358</sup> Tomasz Mielczarek, *Czasopisma społeczno-kulturalne lat osiemdziesiątych XX wieku [w:] Rocznik historii prasy polskiej*, t. XI (2008) z. 1-2 (21-22), s. 194-232.

<sup>359</sup> Mielczarek, 199.

<sup>360</sup> Tamże.

<sup>361</sup> Tamże.

magazynów *Przekrój* – 700 tys. egz., *Polityka* – ok. 290 tys., *Tygodnik Powszechny* – 80 tys. – podaje wcześniej Mielczarek)<sup>362</sup>.

Do nieco późniejszego okresu wydawania pisma odnosi się w swoim tekście Joanna Orska: przedmiotem jej rozprawy jest działalność poetów-tłumaczy, będących redaktorami i współpracownikami pisma, w odniesieniu do awangardy<sup>363</sup>. Tymi postaciami są Tadeusz Pióro, Piotr Sommer (redaktor naczelny pisma od 1994), Andrzej Sosnowski i Bohdan Zadura. Wszyscy poza Zadurą są tłumaczami z angielskiego (w dorobku Zadury publikacje autorów anglojęzycznych występują, ale są współtłumaczone z innymi). Sosnowski ma na swoim koncie tłumaczenia z francuskiego. Zadura zajmował się przekładami z ukraińskiego i rosyjskiego, a także węgierskiego.

Według Orskiej działalność tych autorów układa się w pewien porządek dialogu z awangardą, próby stworzenia „ostatniej” awangardy albo też podjęcia dialogu z nią. Jako ostatni przejaw myślenia czy programu awangardowego zostaje uznana twórczość tzw. Szkoły Nowojorskiej (John Ashbery, Frank O’Hara, James Schuyler). Choć sami nie określali tak działalności swojej grupy, to jednak literatura, do której odwołuje się Orska, umieszcza ich twórczość właśnie w awangardowym kontekście. Kluczowym „dowodem” ma tutaj być *Niewidzialna awangarda*, czyli wykład Ashbery’ego z Yale wygłoszony w 1968. Natomiast popularyzatorami twórczości Nowojorczyków byli właśnie polscy tłumacze związani z LnŚ. Do historii życia literackiego w Polsce przeszedł „niebieski numer” (nr 7 z 1986) – czyli numer tematyczny pisma poświęcony Szkole Nowojorskiej, z obszernym wyborem przekładów. Dla lat dziewięćdziesiątych w polskiej poezji byli to w pewnym sensie „poeci poetów”. Jak zauważa Orska, tłumaczenia O’Hary i innych wpisywały się w przeszczepieniu na polski grunt dyskusji na temat postmodernizmu w sztuce. Dodatkowo, idiom Nowojorczyków, odwołujący się do prywatnych przeżyć i intymności, dał wzorzec dla polskich wierszy w podobnym stylu, jako odtrutki od patosu liryki. Z czasem też w niechętej temu zjawisku krytyce nadano temu zjawisku miano „o’haryzmu”. Orska pisze:

Nośność poetyckich dokonań szkoły nowojorskiej w Polsce nie ograniczyła się jednak do tłumaczeń, stanowiących wynik wielkiego zainteresowania, jakie nasi poeci (zwłaszcza kręgu „Literatury na Świecie”) im okazywali. Czytając niektóre wypowiedzi

---

<sup>362</sup> Mielczarek.

<sup>363</sup> Joanna Orska, Poeci-tłumacze jako „awangarda” lat 90. [w:] Wielogłos. Pismo Wydziału Polonistyki UJ, nr 1 (43) 2020, s. 57–73.

Andrzeja Sosnowskiego, odnosząc się do gestów artystycznych jego oraz Tadeusza Pióry, a także z drugiej strony do pewnych elementów koncepcji poetyckich Piotra Sommera czy Bohdana Zadury, można odnieść wrażenie pewnej koincydencji programotwórczych przesłanek nowojorczyków i polskich autorów<sup>364</sup>.

Sommer, Sosnowski, Zadura są znani w dyskursie krytycznoliterackim jako silne osobowości, tłumacze-poeci. Zaczyna się rysować pytanie o to, na ile w LnŚ odnajdziemy „kanoniczną” wykładnię tego, co wartościowe, a na ile subiektywny wybór. Ustalenia Orskiej świadczą wyraźnie o tym, że mamy do czynienia z grupą artystów-tłumaczy, bardzo samoświadomych w swojej sztuce i działalności popularyzatorskiej.

Z drugiej strony, po przemianach ustrojowych w Polsce mogło się także zmienić zapotrzebowanie na ofertę periodyku, wraz ze zwiększonym dostępem do informacji o kulturach i literaturach świata. Zwłaszcza na przełomie XX i XXI wieku, wraz z popularyzacją internetu i, jak miemam, upowszechnieniem się znajomości języka angielskiego, łatwiej o taki wgląd.

Chciałbym też wydobyć dwa głosy świadectwa ważności misji LnŚ. Pierwszy: Joanna Warmuzińska-Rogóż podaje przykład popularyzacji literatury Quebecu w poświęconym jej tematycznym numerze LnŚ z lipca 1984<sup>365</sup>. Publikacja stanowi dla autorki *exemplum* tworzenia obrazu literatury obcej dla czytelnika kultury docelowej<sup>366</sup>.

Badaczka chwali bardzo spójną strategię wydawniczą. Zarówno układ treści, jak i teksty stanowiące komentarz, są według Warmuzińskiej-Rogóż dobrą okazją do zapoznania się z literaturą regionu. Słowa krytyki znajdziemy jednak w porównaniu tego, co za Bożeną Tokarz nazywa czasoprzestrzeniami oryginału i przekładu<sup>367</sup>:

---

<sup>364</sup> Orska, 61.

<sup>365</sup> Joanna Warmuzińska-Rogóż, Rola czasopism literackich w kształtowaniu obrazu literatury obcej. Przypadek literatury quebeckiej w Polsce [w:] *Rocznik Przekładoznawczy* (online). 2014-12-11, nr 9, s. 233–246. Dostęp: 2022-03-19..

<sup>366</sup> Dla zrozumienia skali zjawiska, przytaczam też pierwszy akapit tekstu, który odzwierciedla stan na rok 2014: „Literatura quebecka w Polsce to przede wszystkim dziewięć spolszczonych powieści, pięć sztuk teatralnych, antologia poezji oraz antologia współczesnej noweli quebeckiej, a także – poza pojedynczymi utworami opublikowanymi w rozmaitych periodykach – jedno czasopismo literackie poświęcone w całości literaturze Belle Province”. – Warmuzińska-Rogóż, 233.

<sup>367</sup> Owa czasoprzestrzeń polega na relacjach rzeczywistości autora i czytelnika danego tekstu oryginalnego wobec czasów „drugiego autora”, jakim jest tłumacz (to określenie pracy translatorskiej pochodzi od Anny Legeżyńskiej). „Tokarz proponuje, by zbadać czasoprzestrzeń w powiązaniu z samym przedmiotem artystycznym, w kontekście komunikacji (nadawca–odbiorca) oraz przez pryzmat relacji utworu literackiego do

Warto także wspomnieć o specyfice czasopisma literackiego, która zasadza się na tym, że teksty stricte literackie stanowią z pewnością ważną jakościowo część spisu treści, ale uzupełnione są zawsze tekstami krytycznymi, które pod względem ilościowym przeważają wręcz nad analizowanymi dziełami czy fragmentami dzieł literackich. Tym samym wybór krytyków wypowiadających się na łamach czasopisma ma znaczenie pierwszorzędne dla kształtu całości danego numeru czasopisma, a co za tym idzie dla obrazu literatury wyjściowej, jaki otrzyma czytelnik przekładów. W przypadku omawianego numeru „Literatury na Świecie” dobór tekstów krytycznych wydaje się jak najbardziej zasadny: przeważają krytycy wywodzący się z samego Quebecu, poza tym swój punkt widzenia przedstawiają uznany polski znawca literatury quebeckiej oraz polska tłumaczka. Tym niemniej, jak każdy wybór, także i ten niesie ze sobą nieunikniony element subiektywny. (...)

Lipcowy numer „Literatury na Świecie” z roku 1984 ukazuje się w momencie, gdy w Quebecu królują niepodzielnie nowe tendencje: wkracza na scenę literatura neoquebecka („les écritures migrantes”), do głosu wyraźnie dochodzą kobiety, a swe miejsce na scenie literackiej ugruntowuje nowela. Tymczasem zbiór odnosi się głównie do zagadnień tożsamościowych i przemian, jakie wcześniej zaszły w literaturze prowincji. Kluczowy w tym numerze tekst Jacques’a Ferrona *Noc* pochodzi z 1972 roku, wiele opublikowanych tekstów krytycznych powstało w latach 70. Ma to swoje uzasadnienie: redakcja czasopisma stawia sobie bowiem najwyraźniej za cel ukazanie tego, co nowe, ale jednocześnie już w jakiś sposób uznane i opisane. Można tu zatem mówić o pewnym przesunięciu czasowym, swoistym opóźnieniu w stosunku do tendencji w kulturze wyjściowej. Pamiętajmy, że dany numer stanowi zwierciadło tendencji w danym momencie, starzeje się jednak dość szybko. Jest wypadkową opinii

---

rzeczywistości zewnętrznej”. –Warmuzińska-Rogóż, 243. Przykładem problemowej operacjonalizacji takiego zagadnienia byłyby kolejne wydania tłumaczeń Orwellowskiego *Roku 1984* w różnych krajach i towarzyszących temu okolicznościach politycznych. W kontekście LnŚ przywoływany wcześniej przykład „niebieskiego zeszytu” również obrazuje to asynchroniczne zjawisko.

krytyków, prądów drogich redaktorowi oraz możliwości wydawniczych. (...) Czasopismo literackie spełni zatem swoją rolę tylko wtedy, gdy sięgnie co pewien czas do zasobów wybranej literatury po to, by zaktualizować jej obraz w oczach obiorców kultury przyjmującej. I w przypadku literatury quebeckiej z pewnością byłoby o czym pisać: „écritures migrantes”, nurt feministyczny, powieść miejska, by wymienić jedynie kilka nurtów, które nie są w naszym kraju szczególnie znane. Polskim czytelnikom pozostaje tylko czekać cierpliwie na nową odsłonę czasopisma poświęconego literaturze tworzonej w języku francuskim rodem z Quebecu<sup>368</sup>.

Podsumowując, zaprezentowany przez LnŚ obraz zjawiska, jakim jest literatura Quebecu, jest dość dobry na dane warunki i wewnętrznie spójny. Stanowi jednak zredukowaną i skrótową reprezentację bardzo złożonego zagadnienia, przez co nie jest w stanie uchwycić go w pełni – pozostaje wiele białych plam. Cierpi też w tym przedstawieniu aktualność. Owe uwagi mogą się jednak odnosić zarówno do linii wydawniczej samego pisma, jak i do stanu badań nad literaturą quebecką. O tym warto pamiętać przy porównywaniu danych o wydawnictwach z wpisami bibliograficznymi LnŚ.

Drugi, krótszy przykład dotyczy literatury etiopskiej. Podwójny numer 7-8 z 2014 jest poświęcony Etiopii. Komunikat z września 2014 w serwisie [www.afryka.org](http://www.afryka.org), wydawanym przez Fundację Afryka Inaczej<sup>369</sup> w dość luźnej formie felietonu-recenzji zaczyna się od dobitnego stwierdzenia: „[b]odaj pierwsza tego typu antologia w języku polskim (jeśli wziąć pod uwagę tematyczne spektrum) wypełnia, przynajmniej częściowo, bolesną lukę na półkach polskich bibliotek, księgarni i prywatnych domków [*sic!*]”<sup>370</sup>. Tekst, opatrzony inicjałami ML, pełen jest metafor i entuzjastycznego tonu; ma przede wszystkim charakter popularyzatorski, zachęcający do sięgnięcia po ten numer. Nie brak mu jednakże rozeznania w temacie (komentarze dotyczą stanu etiopistyki i pokoleniowej zmiany kadr w tej dziedzinie badań). Autor docenia różnorodność treści pod względem gatunków (przemówienia Hajlego Syllasje, literatura piękna, teksty liturgiczne), jak i bogactwo języków, z których dokonano przekładu (odzwierciedlające tygiel kulturowy samego kraju): amharskiego, gyyz,

---

<sup>368</sup> Warmuzińska-Rogóż, 243.

<sup>369</sup> ML, O „etiopskiej” edycji pisma „Literatura na świecie”, <https://afryka.org/o-etiopskiej-edycji-pisma-literatura-na-swiecie/>. Dostęp: marzec 2022.

<sup>370</sup> Tamże.

angielskiego (literatura emigrantów). Co ciekawe, wymowa tekstu prezentuje egzotyczne dla nieobebranego z Etiopią polskiego czytelnika fakty (o kulturze rastafariańskiej, „Czarnych Żydach”), aby poznawać je za pośrednictwem opisywanego numeru LnŚ. Publikacja jest tu uznana za dobry punkt początkowy dla kogoś, kto chce poznać podstawy wiedzy o kulturze Etiopii i historii jej mieszkańców.

Wyłania się następująca teza: LnŚ to czasopismo ożywiający obieg literacki o silnej roli kulturotwórczej. W latach PRL mogła stanowić okno na świat i dawać obraz twórczości za granicą. Współcześnie nadal prezentuje zjawiska ciekawe i niekoniecznie dobrze znane w kraju.

Oznaczałoby to, że znajdziemy w periodyku teksty autorów dopiero wprowadzanych do polskojęzycznego obiegu literackiego, jak i elementy kanonu, dotychczas nieobecne (w rodzaju nietłumaczonych wcześniej tekstów autorstwa pisarzy o ugruntowanej recepcji w Polsce). LnŚ pełniłaby funkcję pierwszego miejsca kontaktu dla nowych zjawisk pochodzących z literatur obcych. Poniżej przedstawiam proponowane obszary do eksploracji, które określiłem przed przystąpieniem do pracy z danymi. Może się zdarzyć, że część pytań badawczych pozostanie bez jednoznacznej, wiążącej odpowiedzi – celem było jednak zebranie pytań, które intuicyjnie nasuwają się same, bez uprzedniej analizy treści pisma lub jego bibliografii, oraz swoistego stereotypu na temat kulturowej roli LnŚ – choć bardziej adekwatne byłyby słowa: renoma, prestiż. Robię to po to, aby umożliwić skonfrontowanie opinii i domniemań z tym, co uda się zobaczyć w analizie statystycznej.

Wyjściowa refleksja dzieli się na kilka obszarów: a) zasięg i reprezentacja literatur: zakres zawartości pisma, wpływy na system literacki spoza niego (np. polityka wydawnicza w PRL i współcześnie); b) eksploracja: nowe vs dobrze znane (pierwodruki tłumaczeń autorów na polski na łamach LnŚ, relacje numerów monograficznych LnŚ do innych publikacji w systemie literackim); c) kuracja treści czy stronniczość redakcji? – czy widoczne są np. stałe powroty danych literatur i autorów kosztem innych (nie musi to świadczyć o złej woli redakcji, ale może wskazywać na to, że dobierano tematykę zgodnie z kompetencjami dostępnych tłumaczy, w tym członków zespołu LnŚ); d) obecność kompetencji (dane mogą stworzyć potencjał dla analizy sieciowej między tłumaczami konkretnych języków i autorami, a także wytworzenia statystyk dla liczby tłumaczy w podziale na języki źródłowe).

Warto też pamiętać o wyzwaniach czekających na osoby przygotowujące zbiory danych, które wymienili autorzy korpusu polskich powieści z XIX i XX wieku wzbogaconego metadanymi

(publikowanego pod nazwą 19/20MetaPNC)<sup>371</sup>. W poświęconej pracy nad jego tworzeniem prezentacji<sup>372</sup> wymienione są następujące trudności: a) ocena reprezentatywności danych; b) różnorodne formaty i jakość tekstu; c) potrzeba użycia katalogów vs niemożność porównania danych z innymi źródłami; d) braki metadanych, wymagające pracy ręcznej w poszukiwaniu informacji; e) ograniczone metadane związane z autorami (np. brak danych socjoekonomicznych). Oczywiście zwłaszcza w ostatnim punkcie są to szczegółowe problemy związane z tym konkretnym przypadkiem, ale tak zarysowana grupa trudności powinna być brana pod uwagę w agregacji danych i pracy na połączonych zbiorach.

Odbiorcy – między globalnością a lokalnością badań

Podobnie jak wcześniej przyglądaliśmy się grupie odbiorczej bazy PBL, w kontekście przygotowywanych danych warto było zadać sobie pytanie o to, dla kogo one będą użyteczne. W wykreowaniu osoby odbiorczej pomógł artykuł prezentujący podejście do historii przekładu jako pracy na danych<sup>373</sup>. Jego autorki, Diana Roig-Sanz i Laura Fóllica, badaczki z Universitat Oberta de Catalunya ICREA i Institució Catalana de Recerca i Estudis Avançats, proponują tytułową Big Translation History, czyli Dużą (Wielką, Masową) Historię Przekładu, analogicznie do terminu *big data* – w oparciu o potężne zbiory danych. W dalszej części będę posługiwać się postulowanym przez nie skrótem BTH na określenie tego konceptu.

Badaczki zwracają najpierw uwagę na krajobraz użycia narzędzi komputerowych w refleksji nad wzorcami zmian w kulturze, wykorzystaniem potencjału narzędzi automatycznych (uczenie maszynowe, sztuczna inteligencja), wraz z ich wyzwaniem, oraz badań ilościowych nad przekładem.

Według Roig-Sanz i Fóllici istnieją już wystarczające narzędzia i zastosowania ich do badania historii przekładu, ale nie wszystkie języki zajmują w nich tak samo eksponowane miejsce: angielski określają jako hiper-centralny, pozostałe języki zajmują pozycje peryferyjne i semi-peryferyjne<sup>374</sup>.

---

<sup>371</sup> Cezary Rosiński i in., „Towards Metadata-Enriched Literary Corpora in Line with FAIR Principles: 19/20MetaPNC”, 2023-06-30, <https://doi.org/10.5281/ZENODO.8107574>.

<sup>372</sup> „Towards Metadata-enriched Literary Corpora in Line with FAIR Principles”, <https://tinyurl.com/literary-corpora-workflow>. Dostęp: 2023-09-16.

<sup>373</sup> Roig-Sanz, Diana, i Laura Fóllica. „Big Translation History: Data Science Applied to Translated Literature in the Spanish-Speaking World, 1898–1945”. *Translation Spaces* 10, nr 2 (2021-11-08): 231–59. <https://doi.org/10.1075/ts.21012.roi>.

<sup>374</sup> Roig-Sanz, i Fóllica, 3. Posługuję się wewnętrzną numeracją stron dokumentu PDF – dop. JE.



Zaproponowane pole badań mieści się między badaniami literaturoznawczymi i przekładoznawczymi, a historią publikowania (system literacki i wydawniczy), badaniami mediów i wreszcie technik komputerowych. Pojawia się ciekawa intuicja, zaczerpnięta od Andrew Chestermana, proponującego obszar badań „studiów nad tłumaczami”<sup>375</sup>, gdzie wysiłek jednostek i zespołów translatorskich i ich wpływ na kulturę może być opisany i wskazany jako istotny impuls (co potwierdzałoby zasadność w przypadku LnŚ i np. wcześniej przytoczonych uwag Orskiej). Badaczki wskazują, jak można przełożyć zgromadzone „duże” dane na opis wymiany kulturowej, który stanie się narzędziem emancypacji, tak w duchu feministyczno-genderowym, jak i dowartościowującym kulturę regionów, państw i języków, które zazwyczaj nie otrzymują tyle uwagi, co inne:

In our view, BTH represents a breakthrough in the sense that it challenges previous research on the circulation of translations, by exploiting the metadata of large datasets, the increasing digitization of texts and the expansion of archives (Algee-Hewitt et al. 2016) that allow us to engage in large-scale analysis, both geographically and over time. The comparison of datasets and transnational translation flows on a scale that was previously unimaginable also reveals both the limits of BTH and the strengths of rewriting translation history through digital archives and metadata. A general hypothesis is that one of the main possibilities offered by BTH is to help decentralize translation history and literary and cultural history, in a broad sense, by breaking with national historiographies in translation. This might be particularly significant for researchers working on periods in which borders have changed, those dealing with translated literature in the diaspora, and those working on translations of regional literatures (such as Flemish) and nonstate literatures (as with Catalan), for example. Macro-analyses of national translation programs might also benefit from such a decentralizing approach. BTH can also help us identify centres of production, the transit of translations, and connectivity between actors. BTH is therefore not only a methodological tool to show patterns of translation, the centrality (or not) of specific translators or translated authors (women among them), or how translations are organized in

---

<sup>375</sup> Roig-Sanz i Fólica, 2.

different ways, but it can also contribute to conceptualization in global approaches to literature and translation history<sup>376</sup>.

W przypadku LnŚ i PBL jak najbardziej wymienione problemy są żywe: potrzeba włączenia polskich źródeł do obiegu światowej informacji naukowej. Dane są niewidoczne lub zbyt mało widoczne w światowym obiegu, mimo prężnych działań zespołu PBB w międzynarodowych kooperacjach. Po części odzwierciedla to zainteresowanie rodzimą literaturą i związanymi z nią zagadnieniami na świecie, ale wystarczyłoby przedstawienie surowych danych, związanych z trajektoriami przekładu, aby trafiły na swych odbiorców.

Autorki postulują trzy filary-warunki, na których można oprzeć badania BTH: 1) Badanie na szeroką skalę geograficzną i chronologiczną; 2) Praca na dużych zbiorach danych, przy wsparciu mniejszych *datasetów* oraz przy użyciu niejednorodnych i nieustrukturyzowanych źródeł; 3) Użycie narzędzi komputerowych jako część procesu badawczego, do wytwarzania nowej wiedzy, w przeciwieństwie do ograniczania się do wizualizacji wyników.

Za częścią drugą artykułu o BTH, określającą ramę koncepcyjną, odnotowuję najważniejsze zagadnienia. Są to w kolejności prezentacji: a) wielość skal i warstw pozyskiwania danych oraz ich późniejszej analizy („małe” i „duże” dane, dane „płytkie” i „głębokie” – czyli poziomy zakresu i szczegółowości); b) wymiary czasu i przestrzeni (przekłada się na zmienną dostępność danych w okresach, ich różne sposoby opisu i poziom świadomości czy profesjonalizmu w działalności dokumentalistycznej); c) niekompletność (niepełność) informacji i nieunikniona stronniczość (*bias*) zawarta w dostępnych danych (mogą mieć na to wpływ czynniki zewnętrzne: budując paralelę do polskich danych bibliograficznych możemy np. wskazać na nieciągłość w produkcji literackiej i dokumentacji kultury spowodowaną konfliktami zbrojnymi, którą chyba najlepiej obrazuje przytaczany przez historyków literatury i edytorów *casus* podstawy wydania Norwidowskiego *Vade-Mecum*; przypadki stronniczości mogą się wiązać z wykluczeniem lub nadmierną reprezentacją w zbiorach, które nie przystają do rzeczywistych warunków danego systemu literackiego czy kulturalnego); d) nierówna dystrybucja i podział wiedzy (*knowledge divide*) – mogą tu występować różnice kulturowe, które należy wziąć pod uwagę, ale też zmienny stopień ucyfrowienia zasobów w najpopularniejszych serwisach (Google Books i inne) pomiędzy kulturami, grupami narodowymi, krajami, językami twórczości itp. – dominuje narracja anglojęzyczna; e) związki analizy jakościowej i ilościowej – badaczki stawiają tutaj pytanie o kwalifikację pod względem

---

<sup>376</sup>Roig-Sanz i Fóllica, 3.

kryterium „literackości” tytułu; analiza liczbowa powinna być wspierana kompetencjami krytycznymi; f) interdyscyplinarność, udostępnianie danych i otwarta nauka – tutaj autorki podkreślają aspekt etyczny działań nakierowanych na otwartość wiedzy; jednocześnie odwołują się do utrzymywania standardów w obrębie poszczególnych dyscyplin; proponowane są rozwiązania szerokie i nastawione na długofalowy zrównoważony rozwój, co wymaga rozwoju infrastruktury (dopowiadając w tym miejscu: praca dokumentalistyczna powinna być obliczona na długie trwanie).

W trzeciej części autorki prezentują studium przypadku, dotyczące literatury tłumaczonej w hiszpańskojęzycznym świecie lat 1898-1945. Wśród kluczowych wniosków pod względem metodologicznym warto zauważyć między innymi możliwość wskazania silnych ośrodków wydawniczych, popularności kierunków tłumaczeń (z jakich języków tłumaczy się najchętniej), a także analizy sieciowe dotyczące osób (autorów).

Kluczowe dla naszego tematu przystosowania bibliografii są tu ustalenia związane z wykorzystaniem danych i metadanych do badań nad fenomenami kultury. Podobnie jak w badaniach *Bibliographic Data Science* mamy tu do czynienia z wyciąganiem przekonujących wniosków z dość podstawowych (i względnie łatwo dostępnych) faktów. Dodatkowo zasługuje na uwagę fakt wplatania narracji translatologicznej do takich badań, tzn. zarysowanie węższego, ale inspirującego tematu badawczego.

BTH uważam za bardzo rozwojowe podejście dla humanistyki, w szczególności komparatystyki i translatologii, a sam artykuł zwraca uwagę na problem centrum i peryferii. Język angielski stanowi *lingua franca* przynajmniej w dwóch interesujących nas tutaj obszarach: komunikacji naukowej oraz wytwarzaniu oprogramowania. W pierwszym wypadku takie dziedziny badań jak polonistyka siłą rzeczy będą związane z danym językiem, a angielszczyzna stanowi na zmianę podporę i przymus. Widziałbym raczej potrzebę traktowania tego języka jako wspólnego mianownika dla badań wielojęzycznych i wielokulturowych. Wspominany już wcześniej w kontekście raport Triple<sup>377</sup> także wymienia wielojęzyczność (*multilingualism*) jako ważny warunek rozwoju danych badawczych humanistyki. W tym drugim kontekście praktycznie nie ma mowy o posługiwaniu się

---

<sup>377</sup> Umerle, Tomasz i in., „TRIPLE Deliverable 8.5: Guidelines on the Research Data in the Humanities”, 2023-01-20, <https://doi.org/10.5281/ZENODO.7568873>, s. 17. Por. uwagi o potrzebie stworzenia wielojęzycznej ontologii dla efektywniejszego wyszukiwania treści z raportu OPERAS: „creation of open linked data and, more in general, the transformation of the internet into a semantic web, open up new perspectives for multilingualism both in research and publishing process” – Delfim Leão i in., „Operas Multilingualism White Paper”, 2018-07-30, 7, <https://doi.org/10.5281/ZENODO.1324026>.

na szerszą skalę językami narodowymi do tworzenia języków programowania i pisania w nich kodu, chyba że w formie niszowych projektów spekulatywnych – jako koncepcje kontestujące ten stan rzeczy<sup>378</sup>. Angielski na stałe zakotwiczył się w aparacie pojęciowym informatyki i jej narzędzi. Większość dokumentacji skierowanej do międzynarodowych odbiorców powstaje przede wszystkim w tym języku. Nawet potwierdzające ten stan rzeczy wnioski Roig-Sanz i Fólíki podawane są czytelnikowi po angielsku i nie byłbym w stanie ich zaprezentować, gdyby artykuł powstał po katalońsku lub hiszpańsku, ze względu na barierę językową. Wszelako wiemy, że odpowiednio przygotowane dane mogą niwelować różnice językowe. Tekst staje się zawartością, owszem wymagającą procesu translacji lub rozumienia obcej mowy, ale dane opatrzone metadanymi (treść z uniwersalnie rozumianym nagłówkiem) stają się możliwe do odczytania przez międzynarodową grupę odbiorczą<sup>379</sup>.

Podsumowując, grupę odbiorczą *datasetu* opartego na przewodniku LnŚ stanowią badacze i badaczki zajmujący się historią przekładu i piśmiennictwa (ruchu wydawniczego), jak i recepcją literatur i komparatystyką. Zbiór danych ma umożliwić włączenie informacji, wchodzących w jego skład, w szerszy ekosystem *big data*, dzięki czemu wycinek wiedzy związany z polską kulturą, praktyką przekładu i recepcją twórczości zagranicznej może być widoczny. Dane mogą być przekształcane, a dzięki użyciu powtarzalnych kluczy metadanowych da się je połączyć z innymi zbiorami.

## Stan badań i inspiracje metodologiczne

### Raportowanie prac nad *datasetami*

Część badań została wymieniona w pierwszym rozdziale, przy prezentacji humanistyki cyfrowej, w intencji zdefiniowania jej jako zjawiska. W tym miejscu skupię się na badaniach bezpośrednio dotyczących wybranego obszaru, pod względem zarówno tematu badawczego,

---

<sup>378</sup> Gretchen McCulloch, „Coding Is for Everyone—as Long as You Speak English”, *Wired*, 2019-04-08, <https://www.wired.com/story/coding-is-for-everyone-as-long-as-you-speak-english/>. Dostęp: 2023-09-16.

<sup>379</sup> W obrębie humanistyki rezonuje to z teorią Anny Wierzbickiej, znaną jako NSM (*Natural Semantic Metalanguage*) i mającą ciekawe implikacje dla translatologii. W bardzo dużym uproszczeniu polega ona na wyznaczeniu podstawowych jednostek znaczeniowych w języku, dzięki czemu można wyznaczyć sens uniwersalny i grupować wiele leksemów różnych języków naturalnych (w relacji wiele do jednego, a nie jeden do jednego, jak w klasycznym modelu szukania ekwiwalentu słowa z języka źródłowego w języku docelowym). Motywacją zastosowania takiej klasyfikacji jest, oprócz umożliwienia komunikacji „ponad” różnymi językami narodowymi, także hermeneutyka tekstów uznawanych za święte – tu autorka zapoczątkowała całą serię badań nad przekazem nauczania Jezusa, wyrażonym w sformalizowanym języku. Egzegeza tekstów religijnych niejednokrotnie opiera się na językoznawczych niuansach i odwołaniu do kontekstu socjokulturowego. Z tego powodu badaczka jednocześnie nie odżegnuje się od sensu lokalnego i wskazuje na istotne różnice kulturowe w leksyce (każda kultura może przedstawiać inne odcienie semantyczne pojęć zgrupowanych w modelu NSM). Zob. Bartmiński, Jerzy. „Droga naukowa Anny Wierzbickiej. Od składni polskiej prozy renesansowej do semantyki międzykulturowej.” *Teksty Drugie*, 2011.

jak i wykorzystanych narzędzi oraz metodologii. Te konkretne prace utorowały drogę do *datasetu* i pomogły podjąć nieraz trudne decyzje o porzuceniu pierwotnych zamiarów.

Najbliższym badaniem na polskim gruncie będzie analiza treści dwumiesięcznika *Teksty Drugie. Teoria literatury, krytyka, interpretacja*, dokonana przez Macieja Maryła<sup>380</sup>. Autor wykorzystuje opisaną już wcześniej metodę makroanalizy, zaproponowaną przez Matthew Jockersa<sup>381</sup>, aby spojrzeć „z lotu ptaka” na obejmujący ćwierćwiecze działalności pisma korpus. Pozwala to analizować statystycznie metadane o tekstach.

Maryl wymienia trzy źródła danych: dane bibliograficzne Repozytorium Cyfrowego Instytutów Naukowych (RCIN) za lata 1990-2012, bibliografię zawartości *Tekstów Drugich*, pochodzącą ze strony internetowej czasopisma, a także pliki pełnotekstowe – zawartość numerów pisma z lat 1990-1998. Bardzo cenne są tutaj uwagi o metodzie i przydatności poszczególnych zbiorów. Po pierwsze, widać różnice ram czasowych zbierania danych. W przypadku bibliografii zawartości Maryl zwraca uwagę na to, że zasób powstał w 2012 za pomocą skanowania i automatycznego przetwarzania tekstu, przez co „w bazie pojawia się trochę literówek, a starsze roczniki nie odnotowują tłumaczy”. Po drugie, zawartość czasopisma była sporządzana w dwojaki sposób: dla lat 1999-2014 były dostępne pliki redakcyjne, przetworzone następnie na format tekstowy. Dla lat 1990-1998 materiał został zeskanowany z OCR i poddany późniejszej obróbce. Tak przygotowany zbiór był następnie analizowany narzędziami przetwarzania języka naturalnego opracowanymi przez konsorcjum CLARIN-PL.

Można zatem zidentyfikować wyzwania stojące przed tego rodzaju projektem. Dane mogą nie być spójne ze sobą, w zależności od czasu sporządzania; informacje mogą być niekompletne. Przykładowo, Maryl wspomina o starszych rocznikach bibliografii pisma bez informacji o tłumaczach. Nie wszystkie dane będą mieć tę samą formę – będą wymagać harmonizacji.

Dzięki analizie tak skompletowanego zbioru Maryl podaje dla całego korpusu liczbę autorów publikujących w *Tekstach Drugich* i ich publikacji (co pozwala także badaczowi namierzyć najczęstszych współpracowników, dane dotyczące objętości (stopniowe zwiększenie liczby stron), rozkład płci osób piszących dla pisma, a nawet interesujące zestawienie proporcji

---

<sup>380</sup> M. Maryl, *Tekstów świat. Przyczynek do makroanalitycznej monografii czasopisma literaturoznawczego* [w:] *Projekt na daleką metę. Prace ofiarowane Ryszardowi Nyczowi*, s. 443–462, Warszawa 2016.

<sup>381</sup> Por. M. Jockers, *Macroanalysis: Digital Methods and Literary History*, Urbana 2013.

debiutantów do pozostałych. Wydawałoby się kilka prostych statystyk prowadzi do ciekawych obserwacji. Dalsze analizy stają się jednak ciekawsze, ponieważ uruchamiają inne wskaźniki.

Maryl proponuje „heurystykę makroanalityczną”, wzorowaną na przeglądaniu indeksu osobowego i rzeczowego w monografii naukowej; z tą różnicą, że pracuje na danych dla wszystkich numerów pisma w opisanym zbiorze. To tak, jakbyśmy kartkowali każdy numer pisma na przestrzeni dwudziestu lat, zakładając też przywoływanie szczegółów z fotograficzną pamięcią. Dla uproszczenia dowolna liczba przytoczeń liczy się jako jedno cytowanie (odcina to np. nadreprezentację cytowań jednego dzieła w przypadku recenzji). Dzięki temu docieramy do konkretnych danych statystycznych, umożliwiających dalsze analizy: „Ponad 2/3 cytowanych (67,89 %, 6919) zostało przywołanych jednokrotnie i tylko 18,66 % (1902) osób pojawia się częściej niż dwa razy. 385 autorów przywołano częściej niż dziesięciokrotnie (3,78 %), a tylko 1 % (103 osoby) uzyskał więcej niż 30 odwołań”<sup>382</sup>. Ciekawe, czy słuchacze Andrew Pipera z Berkeley tak właśnie by obstawiali te wartości.

W ten sposób dla pięciu czteroletnich okresów można było ustalić najczęściej cytowanych autorów. Maryl komentuje: „widać, że część osób pozostaje stałym punktem odniesienia w artykułach publikowanych w »Tekstach Drugich«, pojawiając się we wszystkich (Barthes, Bolecki, Derrida, Głowiński, Gombrowicz, Nycz, Sławiński) lub prawie wszystkich (Markiewicz, Balcerzan, Janion, Ricoeur) okresach wśród 30 najczęściej przywoływanych postaci”<sup>383</sup>. Aby jednak oddać dynamikę relacji między cytowanymi i cytującymi, potrzeba czegoś więcej. Dlatego zostają określone wektory, która praca odwołuje się do jakich. Dzięki temu może powstać analiza sieciowa cytowań.

Maryl zdradza istotną cechę warsztatu w przypadku tworzenia sieci w programie Gephi. Wykorzystane wektory między cytującymi a cytowanymi zostały przetworzone przez algorytm wykrywający skupiska na podstawie wspólnych odwołań. „Czynnik ludzki” polegał tutaj jednak na ustawieniu parametrów pracy tego algorytmu. Wyniki, według słów badacza, przybierały wygląd od kilku do 44 skupisk. Ostatecznie najbardziej adekwatny okazał się podział na 27 grup. Skupiska autorów (*communities*) Maryl przyporządkował do następujących przykładowych tematów (pięć najpopularniejszych): A) Poststrukturalizm, filozofia i literatura, B) Teoria literatury i kultury (strukturalizm, hermeneutyka, socjologia literatury), C) Historia literatury polskiej XX wieku I, D) Historia literatury polskiej XX wieku II,

---

<sup>382</sup> Maryl, *Tekstów świat*, 451.

<sup>383</sup> Tamże.

E) Modernizm, kultura i sztuka. Wśród grup mogą się powtarzać ci sami autorzy (jeden autor może należeć do więcej niż jednego skupiska), chodzi jednak o widzialne zagęszczenie powiązań między nimi. Na podstawie tego, kto i jak często kogo cytuje, widząc takie klastry, Maryl przypisuje owe nadrzędne tematy. O grupie E pisze w ten sposób: „jest dość niejednoznaczna i ogólnie została sklasyfikowana jako prace odnoszące się do modernizmu, kultury i sztuki – centralnymi postaciami są w niej Maria Janion i Kazimierz Wyka, a także Popiel, Brzozowski, Berent, Eliade, Zawadzki, Hegel, Czerwińska. Wydaje się, że są to sieci blisko związane z grupami B i D, choć operują innymi powiązaniemiami”<sup>384</sup>. Cenne do pracy nad danymi LnŚ jest wskazanie na ten element poznawczy w wyborze metody wizualizacji danych. Nasuwa się skojarzenie z naukami przyrodniczymi: obserwacja zostaje nazwana i opisana, niczym gatunek zwierzęcia czy rośliny – zawarty jest w tym jednocześnie element dokumentujący, jak i odautorski<sup>385</sup>.

Ostatnią zawartą w tekście analizą jest analiza tematyczna. Za pomocą wspomnianych już narzędzi przetwarzania języka naturalnego wydobyte zostały najpopularniejsze użyte słowa. W tym miejscu Maryl sygnalizuje potencjał, drzemiący w narzędziach cyfrowych, nie wchodząc w głębsze analizy, ale wskazując np. na relacje zaistnienia jakiegoś tematu badań i dalszej jego recepcji (wyraźnie widoczne tematy Zagłady Żydów i postkolonializmu: w pierwszym przypadku słownictwo zagładowe od pewnego momentu na stałe wiąże się z obecnymi już wcześniej słowami związanymi z kulturą żydowską, w drugim – badania postkolonialne zaczynają być obecne bardziej po roku 2002). Badacz podkreśla potrzebę większego zaplecza – i rzeczywiście, należy się zgodzić, że na podstawie zasobów tylko jednego tytułu trudno orzekać np. o tematyce polskiej dyskusji teoretyczno- i krytycznoliterackiej w ogóle. Maryl wzmiankuje także PBL jako zasób danych wart uwagi, ze względu na unikatowe zapisy szczegółowe. Tym większy to impuls w kontekście danych dotyczących LnŚ, aby starać się je konfrontować i łączyć z innymi źródłami informacji. Tak uzyskany efekt może już dawać szerszy obraz, niż ograniczony do danego tytułu i np. wspomnianego już problemu polityki redakcji.

Co do metodologii, analiza *Tekstów Drugich* stanowi przetarcie szlaku i udowodnienie, ile z danych liczbowych jesteśmy w stanie uzyskać zrozumiałej, nowej treści. Z dzisiejszej

---

<sup>384</sup> Maryl, 456.

<sup>385</sup> Być może nawet nie do końca pasuje tu skojarzenie z nazwaniem gatunku, a bardziej nazwanie zrozumiałym pojęciem tego, co się bada – jak w przypadku pionierskich działań Jane Goodall przy opisie szympanów z Gombe, którym nadawała imiona, wbrew przyjętym wówczas praktykom numerowania.

perspektywy najgorzej zestarzały się chyba chmury tagów – prosta wizualizacja, umieszczonych obok siebie najczęściej powtarzających się słów, gdzie ich częstotliwość występowania decyduje o wielkości napisu względem pozostałych. Na przygotowanej z użyciem narzędzia Wordle<sup>386</sup> chmurze słów (chmurze tagów) dotyczącej popularności cytowanych autorów<sup>387</sup> największy „Miłosz” przechwytuje spojrzenie odbiorcy, ustępując następnie hasłom „Głowiński”, „Gombrowicz”, „Derrida”. W literce o niczym w brzuchu wieloryba zamknięte, ledwie czytelne w mojej kopii PDF, siedzi hasło „Cixous”, a między litery S a Z wdał się „Agamben” (Barthes jest na to za duży). Czemu jednak służą moje złośliwości względem tej metody pokazania informacji? Przede wszystkim poza wspomnianą wielkością fontu, wyrażającą wartość liczbową (w tym wypadku frekwencja w przypisach), nie ma żadnej relacji przestrzennej między poszczególnymi hasłami. Słowa układają się tak, aby stworzyć chmurę i mniej-więcej dopasować, tworząc zwarty układ. Więcej, nie ma czytelnej metody porównania wartości podobnych – jest to wizualizacja „błyskawiczna”, pozwalająca się zorientować jaki jest rozkład „na oko”. Trudno porównywać wartości bardzo podobne do siebie. Przyjmuję jednak, że istnieją dane tabelaryczne – precyzyjne, choć mniej atrakcyjne przy pierwszym zetknięciu się z nimi. Dlatego chmurę tagów potraktujmy jako zabieg, który daje nam ogólną orientację w temacie – ma to swoją cenę i niesie określone ryzyka, więc zachowałbym ostrożność w prezentowaniu danych w ten sposób zbyt często lub w kluczowych zagadnieniach<sup>388</sup>.

W dalszej części pracy podjąłem decyzję o skupieniu się na samym przygotowaniu danych i zrezygnowałem z tworzenia wizualizacji za wszelką cenę (dzięki dobrym danym atrakcyjna dla odbiorcy wizualizacja powinna być prosta do osiągnięcia, ale nie na odwrót). Prace przywoływanej już Drucker mogą stanowić tutaj wskazówkę dla dalszych eksploracji kierunkowych, zorientowanych na aspekt prezentacji treści w formie wizualnej<sup>389</sup>.

Jest jeszcze jeden ważny tekst Macieja Maryła, który należy przywołać w kontekście tych badań. W artykule *Operationalising the change Dispersion of Polish literary life (1989–2002)*

---

<sup>386</sup> Szukając pierwotnie użytego narzędzia, zdałem sobie sprawę, że w 2022 nazwa Wordle nie jest w pierwszej kolejności kojarzona z chmurą tagów. Rozrywkowym hitem jest gra słowna o tej nazwie w formie aplikacji internetowej, w której użytkownicy zgadują zmieniające się codziennie hasła: <https://www.nytimes.com/games/wordle/index.html>. Dostęp: 2022-03-01.

<sup>387</sup> Maryl, 452.

<sup>388</sup> Więcej krytycznej analizy chmury słów jako narzędzia w Ú. Torres Parejo, J. R Campaña, M. Amparo Vila, M. Delgado, *A survey of tag clouds as tools for information retrieval and content representation* [w:] *Information Visualization*, vol. 20, 1: pp. 83-97. , 2020.

<sup>389</sup> por. Drucker, *Visualization and interpretation*; Johanna Drucker, „Humanities Approaches to Graphical Display”, b.d.



autor podejmuje się zobrazowania życia literackiego, w odwołaniu do diagnoz obecnych w dyskusji krytyków literackich przy pomocy narzędzi makroanalitycznych<sup>390</sup>. Przyjęte daty w historiografii i krytyce kojarzą się z debatą na temat przesilenia i przełomy. Dyskusje obracały się wokół tematu wpływu przemian polityczno-ustrojowych w Polsce na system literacki. Obszarem, który miał być zoperacjonalizowany, jest zbiór sądów na temat centrum i peryferii w komunikacji i życiu literackim. PBL stanowi tu zbiór danych, oferujący wystarczającą próbkę informacji, by dać świadectwo o stanie kultury badanego okresu. *Dataset* tworzyły trzy podzbiory: książki, artykuły i recepcja (np. recenzje, polemiki).

Jedną z istotniejszych konkluzji Maryla jest dostrzeżenie niebezpieczeństwa w takiej pracy liczbowej: badacz sięgający po analizę statystyczną może referować nie dane, ale... samą dyskusję krytyczną. Dwoma istotnymi czynnikami wpływającymi na zainteresowanie literaturą są bowiem, po pierwsze, ważne wydarzenia (przyznanie nagrody Nobla Wisławie Szymborskiej, śmierci istotnych autorów, np. Herberta czy Herlinga-Grudzińskiego, 90. urodziny innego polskiego noblisty - Miłosza), po drugie – „zagęszczenie” pisarstwa krytycznego i publicystyki wokół tych twórców. Krytyk jest zatem obserwatorem, który jednocześnie wpływa na postrzegany przez siebie system. Nie wszystkie obserwacje Maryla wynikające z działań na danych pokryły się z sędami krytyków, a te, które były z nimi zbieżne, zawierały mocny komponent danych dotyczących recepcji, a więc właśnie zawierających wspomniany wpływ. Szansą dla badań na danych jest więc uniezależnienie się od wniosków debat i praca bliższa socjologii komunikacji literackiej. Dla humanistów cyfrowych może to być także zachęta do maszynowego wsparcia dyskursu krytycznego. Dodajmy do wniosków Maryla: zamiast generalizującego zdania o wydawanych w danym czasie dziełach, krytyk może odwołać się do statystyki, zanim wygłosi swoją opinię.

Badania trajektorii przekładowych i recepcji

Ważnym punktem odniesienia okazała się także polsko-czeska współpraca PBB, czyli badania Vimra i Rosińskiego nad mapowaniem (rozprzestrzeniania się) czeskiej literatury na świecie<sup>391</sup>

---

<sup>390</sup> Maciej Maryl, „Operationalising the Change. Dispersion of Polish literary life (1989–2002)”, w *Digitale Literaturwissenschaft*, red. Fotis Jannidis, Germanistische Symposien (Stuttgart: J.B. Metzler, 2022), 509–34, [https://doi.org/10.1007/978-3-476-05886-7\\_21](https://doi.org/10.1007/978-3-476-05886-7_21).

<sup>391</sup> Ondřej Vimr i Cezary Rosiński, „Mapping Czech literature in translation on a large scale (1820-2020)”, *Česká literatura* 70, nr 6 (luty 2023): 711–34, <https://doi.org/10.51305/cl.2022.06.03>. Publikacja w języku czeskim. Dziękuję Cezaremu Rosińskiemu za objaśnienie metodologii projektu i wskazanie głównych ustaleń oraz wniosków.

oraz projekt grantowy związany z trajektoriami przekładów literatury czeskiej<sup>392</sup>, wprost odnoszące się do koncepcji BTH i rzeczywistości danowej.

We wspomnianym tekście badacze opisują system przekładów czeskiej literatury w latach 1820-2020. Źródła danych stanowią zbiory OCLC (światowej organizacji bibliotekarstwa cyfrowego), Czeskiej Bibliografii Literackiej, a także Czeskiej Biblioteki Narodowej. Prezentowane są trzy studia przypadków: dynamika reprezentacji płci, trendy w danych dotyczących języków docelowych translacji oraz klastrowanie autorów pod względem profilowania języków, na które są przekładane ich dzieła, w celu wskazania prawidłowości co do różnorodności globalnej. Badacze wychodzą od pytania o wpływ czeskich dzieł na literaturę światową i interakcje w obrębie tego systemu, co można pokazać na przykładzie danych o tłumaczeniach.

Jednym z ciekawych spostrzeżeń w tym badaniu są regionalne cechy systemów literackich. Dla przykładu: rosyjskie tłumaczenia były często pracami zbiorczymi, zawierającymi dzieła większej liczby autorów w jednej książce, a w Indiach przekłady wydaje się głównie w czasopiśmie. Pierwszy przypadek „scala” wielu twórców i utworów do pojedynczego zapisu bibliograficznego, a drugi powoduje niską reprezentację języka hindi w zbiorze, kiedy się bierze pod uwagę publikacje książkowe.

Autorzy borykali się także z problemem niedostatku metadanych lub błędnymi i niepełnymi danymi. Ze względu na brak informacji trudności sprawiło przypisywanie płci tłumaczom, czy też odnalezienie rzetelnej informacji o grupie odbiorczej: powtarzającym się błędem było np. oznaczanie dzieł beletrystycznych jako *non-fiction* w źródłach bibliograficznych. Jedną z czynności badawczych było także ustalenie statusu dzieł pod względem kryteriów związanych z translacją, ponieważ znaczna część zbioru oznaczonego jako tłumaczenia okazała się być czeskimi dziełami w oryginale (użyto narzędzi wykrywania języka).

Metodologicznie Vimr i Rosiński zwracają uwagę na konieczność ostrożnego podchodzenia do danych wejściowych (wspomniana edycja, harmonizacja, ale też deduplikacja czy organizacja i zarządzanie duplikatami na wyższym poziomie w formie agregacji). W swoim zamierzeniu autorzy posługiwali się Big Data, ale podkreślają, że pełna korzyść z tego typu badań (więcej informacji i zwiększenie precyzji opisu) może zostać osiągnięta przy nieuniknionej pracy manualnej. Wskazane zostają także braki w danych i – aby umożliwić

---

<sup>392</sup> Zob. „Globální trajektorie české literatury od roku 1945”, <https://clb.ucl.cas.cz/projekty/aktualni-projekty/globalni-trajektorie-ceske-literatury-od-roku-1945/>. Dostęp: 2023-09-16.

korzystanie z niektórych zasobów – konieczność wytworzenia ich. Przykładem są kraje Globalnego Południa, gdzie zdobycie bibliografii wiązałoby się z badaniami ręcznymi i przetworzeniem danych. Pod względem danych wyjściowych, cały szereg operacji prowadzi do jasnej konkluzji. Badacze identyfikują przekłady na język niemiecki jako punkt wyjścia dla rozpowszechniania się przekładów na północy Europy, a francuski dla południa kontynentu; połączenie niemieckiego, francuskiego i angielskiego może dać podstawę translacji na języki spoza Starego Kontynentu.

Celem czeskiego grantu jest z kolei badanie o węższym zakresie czasowym, po roku 1945, a obok wspomnianej wcześniej metody *Big Data* odnotowuje zastosowanie narzędzi socjologii i historii tłumaczeń. Ważne jest uwzględnienie kontekstu historycznego, tzn. refleksja dotycząca różnych obiegów literackich w systemie. Przywołany zostaje kontekst zimnej wojny, rozpadu bloku wschodniego, globalizacji i integracji europejskiej. W założenia projektu jest też wpisane krytyczne odniesienie się do binarnych opozycji takich jak centrum-peryferie oraz kosmopolityczne-wernakularne.

W odniesieniu do prac badających trajektorie translacji i wymianę kulturową, trzeba wspomnieć o publikacji Małgorzaty Gaszyńskiej-Magiery, której temat stanowi recepcja przekładów literatur iberoamerykańskich: w badaniach czytelnictwa, które przeprowadziła autorka, LnŚ nie była wymieniona, tj. kojarzona przez respondowanych (wskazywali inne tytuły jako istotne w czasach *boomu* literatur iberoamerykańskich<sup>393</sup>) – odbiorcami byli „przeciętni” czytelnicy literatur, czyli niezwiązani z akademią zawodowo<sup>394</sup>. Oznacza to mniejszą rozpoznawalność tytułu przez odbiorców systemu literackiego w tych czasach (ewentualnie: LnŚ mało zapadała w pamięć). Wysoka jest za to rozpoznawalność i znajomość samego autora – Borges pojawia się w czołówce znanych i czytanych autorów w każdej z

---

<sup>393</sup> Gaszyńska-Magiera używa wymiennie pojęć „proza iberoamerykańska” oraz „proza latynoamerykańska” umownie, przy czym zaznacza, że ograniczyła zakres w swoim opracowaniu: bada utwory hiszpańskojęzyczne powstałe w Ameryce łaćwińskiej, a jeśli chodzi o literaturę języka portugalskiego, odsyła do innych publikacji.

<sup>394</sup> Badanie ankietowe przeprowadzone wśród czytelników wybranych bibliotek między styczniem a sierpniem 2010. Por. Małgorzata Gaszyńska-Magiera, *Recepcja przekładów literatury iberoamerykańskiej w Polsce w latach 1945-2005 z perspektywy komunikacji międzykulturowej*. Varia Culturalia. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2011, 212-253. Taka uwaga pojawia się także w podsumowaniu rozdziału poświęconego krytyce: autorka zwraca uwagę na zjawisko „przyprawiania gęby” realizmu magicznego autorom latynoamerykańskim młodszego pokolenia, co komentowano przy okazji antologii dzieł autorów z Ameryki łaćwińskiej przed 35. rokiem życia zatytułowanej *McOndo*, wydanej w 1996 r. Badaczka dodaje: „Nowej prozie poświęcono jeden z numerów »Literatury na Świecie« (2004, nr 9-10). Ta tematyka pojawiała się niemal wyłącznie w obszerniejszych tekstach znawców, przed rokiem 2005 nie znalazła jednak odzwierciedlenia choćby w wyborze książek do zrecenzowania” – widzimy zatem obecność LnŚ w dyskursie fachowym, wyprzedzającym ogólniejszą, zbiorową percepcję tej literatury w życiu literackim. Cyt. za: Gaszyńska-Magiera, 183.

badanych grup wiekowych. Sama badaczka oczywiście powołuje się na teksty publikowane na łamach LnŚ w bibliografii i w tekście.

Możemy zatem przyjąć, że w kontekście recepcji Borgesa dla czytelników literatury argentyńskiej, czy szerzej – iberoamerykańskiej, istotne były przede wszystkim publikacje książkowe. Fragmenty zamieszczane w czasopismach, jeśli miały znaczenie, to dotyczyły innych tytułów niż LnŚ. Gaszyńska-Magiera zwraca uwagę na kronikę życia literackiego, gdzie zjawisko *boomu* jest odnotowane<sup>395</sup>. Kreuje to obraz przede wszystkim recepcji popularnej. Stąd wniosek, że typowy czytelnik literatury iberoamerykańskiej nie sięgał po LnŚ, ale w pierwszej kolejności po książki konkretnych autorów (w tym Borgesa), a później po czasopisma. Pod względem recepcji czytelniczej LnŚ stanowiłaby zatem źródło specjalistyczne, a nie takie, które kształtowałyby gusta o szerokim zasięgu.

Od *Bibliodata* do polisystemów

Wspominam również o pracy Gaszyńskiej-Magiery dotyczącej recepcji, ponieważ osadza ona literaturę iberoamerykańską w kontekście spotkania kultur (komunikacji międzykulturowej). Warto wspomnieć o tym przy okazji zbioru danych LnŚ. Autorka odwołuje się do teorii polisystemów Itamara Even-Zohara; przy zastosowaniu tej metody przyjmujemy optykę unikającą izolacji jednego obiektu, starając się uchwycić wpływ jednych złożonych układów na inne oraz ich wzajemnych relacji:

[K]oncepcja literatury jako odrębnej, na poły niezależnej instytucji socjokulturowej jest zasadna tylko wtedy, gdy jej polisystem - jak każdy inny system socjokulturowy - wyobrazimy sobie jako jednocześnie autonomiczny i heteronomiczny wobec wszystkich innych współistniejących systemów. Nie da się zaprzeczyć, że fakty zinstytucjonalizowanego „życia literackiego” (na które składają się np. ideologie literackie, oficyny wydawnicze, krytyka, grupy pisarzy czy też inne ośrodki dyktujące reguły smaku i obowiązujące normy), choć zachowują się jak odrębny, na wpół niezależny system socjokulturowy rządzący się własnymi prawami, muszą być również

---

<sup>395</sup> Czapliński et al.

uznane za niezbędne czynniki systemu literackiego w ścisłym tego słowa znaczeniu<sup>396</sup>.

Zapewne w tak globalnym ujęciu trudno będzie dostrzec procesy, o jakich pisze Zohar na niewielkim wycinku danych bibliograficznych, którym się tutaj zajmujemy. Mowa o zbiorze publikacji jednego pisarza w jednym kraju. Podejście związane z polisystemami staje się jednakże obiecujące wtedy, kiedy uzmysłowimy sobie potencjał otwartych danych połączonych, czyli mozaikę datasetów o różnej proveniencji, które będziemy umieli w logiczny sposób połączyć. W dodatku nabiera to celowości, jeśli wziąć pod uwagę, że cała praca, o jakiej pisze Zohar, którą określiłbym jako przemiany kulturowe, dzieje się w przestrzeniach pomiędzy wyraźnie wytyczonymi abstraktami, np. takimi jak literatura narodowa i język:

Reasumując, głównym celem i możliwością zastosowania teorii polisystemów jest badanie szczególnych warunków, w jakich zachodzą interferencje między literaturami, w wyniku czego pewne właściwości przekazywane są z jednego polisystemu do drugiego. Przykładowo, jeśli przyjmiemy hipotezę, że cechy typowe dla peryferii prawdopodobnie przenikną do centrum, gdy tylko zdolność tego ostatniego (tzn. repertuaru centralnego) do spełniania określonych funkcji ulegnie osłabieniu (drugie prawo Szklowskiego), to nie ma sensu zaprzeczać, że dokładnie ta sama zasada działa na płaszczyźnie międzysystemowej. Podobnie też polisystemowa struktura poszczególnych literatur pozwala wyjaśnić skomplikowane procesy interferencji, które - wbrew powszechnemu mniemaniu - odbywają się często za pośrednictwem peryferii. Ignorując te zjawiska, rezygnujemy z możliwości wyjaśnienia, dlaczego w repertuarze pojawiają się i zaczynają funkcjonować nowe jednostki. Teksty paraliterackie, przekłady, literatura dla dzieci - wszystkie te warstwy zaniedbywane przez współczesne literaturoznawstwo stanowią przedmiot badań absolutnie konieczny do zrozumienia, dlaczego i w jaki sposób

---

<sup>396</sup>Itamar Even-Zohar, „Teoria polisystemów”, tłum. Katarzyna Lukas, *Przestrzenie Teorii*, nr 7 (2007-11-15): 347–66, <https://doi.org/10.14746/pt.2007.7.22>, 363-364.

dochodzi do transferów zarówno w obrębie systemów, jak i między nimi<sup>397</sup>.

Szczególnie istotne wydaje mi się w kontekście użycia danych PBL mówienie o tekstach paraliterackich, sądzę, że to źródło może dostarczyć ciekawych wskazówek odnośnie otoczenia samych dzieł pod względem recepcji. Dlatego badania związane z teorią polisystemów (kultury globalne i lokalne, emancypacja kultur, emergentne podsystemy i przemiany w obrębie uprzednio funkcjonujących systemów itp.) może być punktem dojścia w refleksji, którą rozpoczyna *Bibliographic Data Science*.

Wstępne przygotowanie techniczne

Niniejsze badania nie doszłyby do skutku, gdyby nie mentoring z Cezarym Rosińskim w ramach studium Humanistyka Cyfrowa IBL PAN. Część edukacji zapewnił kurs zastosowania języka Python w analizie tekstu autorstwa Melanie Walsh. Badaczka opublikowała kurs programowania w wolnym dostępie<sup>398</sup>. Walsh pokazuje, jak można łączyć naukę opartą na danych z refleksją humanistyczną i społeczną. Jej przykłady dotyczą imigrantów z Irlandii w USA w latach 40. XIX wieku, niewolnictwa w obu Amerykach w latach 1514-1866, czy bazy dialogów filmowych. Szereg technik (analiza danych tekstowych, *scrapping*, analiza sieciowa) prowadzi do wytwarzania zestawów danych.

Pomimo praktycznego nastawienia (mimo wszystko jest to kurs, uczący konkretnych umiejętności cyfrowych), dostrzegalna staje się narracja. Kwantyfikowalne informacje, sprowadzone do cenzusu i jego interpretacji, dają obraz minionego świata: oto rejestr schodzących na ląd nowoprzybyłych (a więc we współczesnym rozumieniu baza danych) daje nam obraz biurokracji, stojącej za wyzyskiem. Pozornie neutralna, urzędowa informacja z nowojorskiego Bellevue Almshouse, gdzie przyjmowano imigrantów, z dzisiejszej perspektywy wyraźnie zdradzi uprzedzenia: w rubryce „choroby” wpisywano np „niedawny imigrant”<sup>399</sup>.

Oprócz szkolenia z Python, potrzebne były umiejętności odpytywania relacyjnej bazy danych Oracle, aby dostać się do zasobów PBL. Stosuje się w tym celu SQL, czyli *Structured Query Language*, język ustrukturuowanych odpytań. Posiadałem część umiejętności w związku z pracą

---

<sup>397</sup> Even-Zohar, 365-366.

<sup>398</sup> Walsh, Melanie. Introduction to Cultural Analytics & Python. <https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>. Dostęp: marzec 2022.

<sup>399</sup> Walsh, Melanie. Pandas Basics — Part1 . <https://melaniewalsh.github.io/Intro-Cultural-Analytics/03-Data-Analysis/01-Pandas-Basics-Part1.html>. Dostęp: marzec 2022.

zawodową i wystarczyło mi ich dostosowanie do przedmiotu badań oraz nadanie dostępu do bazy.

Osobnym wątkiem pozostała harmonizacja, normalizacja i edycje danych. Mój bazowy poziom umiejętności uformowały tutaj MS Excel i OpenRefine. Obsługa arkuszy kalkulacyjnych w programie pakietu Office dla wielu osób bywa wrażliwą przyjemnością, ale na pewno daje satysfakcję z każdego małego sukcesu. Pozwala pracować na danych „w dwuwymiarze”, w widoku tabelarycznym. Drugie narzędzie, OpenRefine, umożliwia natomiast wyszukiwanie prawidłowości i masowe zmiany. Najważniejsze funkcje stanowi klastrowanie podobnych do siebie przypadków, których próżno szukać w łatwy sposób w Excelu, a także stosowanie algorytmów podobieństwa ciągu znaków (oprogramowanie sugeruje użytkownikowi, że dana grupa może odnosić się do tej samej rzeczy, co pomaga np. tworzyć znormalizowane słowniki autorów, dopasowując do siebie oboczności nazw, najczęściej jako różne formy zapisu, tak poprawne, jak i błędne).

W trakcie tej pracy nauczyłem się natomiast praktycznie od podstaw zastosowania wyrażeń regularnych (*regular expressions*, regex), w celu wyszukiwania i modyfikacji powtarzalnych grup treści.

#### Analiza i przetwarzanie bibliografii LnŚ

Opublikowana przez Instytut Książki<sup>400</sup> bibliografia LnŚ obejmuje lata 1971-2014, a więc od początku działalności pisma, ale bez kolejnych lat po 2014 (pismo ukazuje się nadal). Irena Bednarz, niezależny bibliograf profesjonalny<sup>401</sup>, stworzyła autorski przewodnik po zawartości pisma w formie bibliografii tekstów oraz indeksu osobowego<sup>402</sup>. Dokument będę nazywał w dalszej części pracy na przemian przewodnikiem lub bibliografią LnŚ.

Bibliografia jest inicjatywą autorki – zaproponowała redakcji pisma taką pracę, prezentując swoje dotychczasowe osiągnięcia. Według Ireny Bednarz monografia bibliograficzna jednego tytułu jest przedsięwzięciem odpowiednim dla jednego badacza<sup>403</sup>. Autorka miała już na swoim

---

<sup>400</sup> Instytut Książki, *Literatura na Świecie*, <https://instytutksiazki.pl/czasopisma,9,literatura-na-swiecie,3.html> (dostęp: 1 III 2022). Instytut Książki jest wydawcą pisma, rola ta została przejęta od Biblioteki Narodowej.

<sup>401</sup> W korespondencji prywatnej p. Bednarz konsekwentnie nie używała feminatywu dla określenia swojej profesji, co pragnę tu uszanować.

<sup>402</sup> I. Bednarz, Przewodnik po Literaturze na Świecie 1971–2014 (nr 1–521), [https://instytutksiazki.pl/files/upload/files/LNS\\_1971-2014\\_\\_CALOSC.pdf](https://instytutksiazki.pl/files/upload/files/LNS_1971-2014__CALOSC.pdf) (dostęp: marzec 2022).

<sup>403</sup> Powołuję się tutaj na korespondencję prywatną z autorką.

koncie indywidualne projekty o większej skali, np. bibliografię poświęconą warszawskiej szkole historii idei<sup>404</sup>.

Badaczka tworzyła swój zbiór z autopsji. Pracowała, posługując się zawsze papierowym wydaniem danego numeru czasopisma, podobnie jak robi to zespół przygotowujący wpisy do zbiorów bibliograficznych PBL. We wstępie dokumentu są zawarte podstawowe informacje metodologiczne, jak używane skróty, konwencje nazewnictwa i znaki specjalne użyte w tekście ze stosownymi objaśnieniami.

Plan działań

Przewodnik LnŚ wydał się na tyle interesującym spisem, że warto było o nim pomyśleć w szerszym kontekście obiegu danych bibliograficznych. Wystarczyłoby tylko ustrukturuowaną już formę, zaproponowaną przez autorkę, sprowadzić do wspólnego mianownika z powszechnymi zbiorami danych, by sprawdzić ich podobieństwa i różnice, a także zastanowić się nad tym, jak takie *datasets* mogą wpływać na siebie. Pierwszym pomysłem było odszukanie w bibliografii LnŚ takiej unikatowej informacji, która będzie wartością dodaną dla zbiorów już istniejących w PBL (*data enrichment*). Kolejnym – analiza danych wydawniczych z zasobów Biblioteki Narodowej. Koniecznym komponentem do wykonania takiego zadania stały się identyfikatory kontroli autorytatywnej, jak VIAF, oraz dane Wikidata. Dostarczyłyby one odpowiedniego klucza, owego wspólnego mianownika, dla połączenia wszystkich danych ze sobą.

Wracając do uwag i wątpliwości metodologicznych, nowa wiedza jest zdobywana w kolektywny sposób, choć nie na zasadzie pracy zespołowej. Jedni badacze spisują informacje bibliograficzne z autopsji, inni kompilują dane w większe zbiory, jeszcze inni mogą na podstawie takich publikacji transformować *datasets* do własnych ustaleń. W ten sposób informacja przechowywana w bazie bibliograficznej ożywa, ponieważ nie służy już tylko do odnalezienia treści, ale prowadzi do formułowania wniosków na temat obiegów literackich. Dzięki temu można podjąć się próby opisanego życia literackiego na podstawie zbiorów danych, w skali wykraczającej poza tradycyjną pracę na siły jednego literaturoznawcy. Ujmujemy tu interpretacyjnej finezji na rzecz pokazania faktów – samych w sobie przecież interesujących.

---

<sup>404</sup> I. Bednarz, *Mistrzowie i uczniowie. Bibliografia zawartości wydawnictw zbiorowych dedykowanych warszawskiej szkole historii idei*, Warszawa 2018.



## Rozpoznanie i obróbka danych

Wyzwanie stanowił autorski sposób edycji bibliografii. Nie można temu zestawieniu odmówić konsekwencji, natomiast jako materiał źródłowy nie było ono przygotowywane do analizy zautomatyzowanej. Plik bardzo dobrze się przegląda i czyta w przeglądarce lub edytorze PDF, ale nie ma on łatwej struktury do podziału tekstu i sporządzenia z niego bazy danych – tabeli. Przy przetwarzaniu PDF trudności nastręczyły niekonsekwentnie obecne białe znaki – spacje, znaki nowej linii i tabulatury. Mogło być to wszakże efektem konwersji zawartości bibliografii do formy, którą „rozumiał” edytor języka programowania Python. Sprawą, która z kolei bez wątpienia stanowiła wybór autorki bibliografii - a wymagała uwzględnienia w przetwarzaniu - był dobór fontów: zarówno kroju pisma, jak i zmieniającej się wielkości liter. Irena Bednarz zastosowała dla nazwisk autorów jako hasel font Arial Black z autorskim inicjałem w postaci powiększenia o jeden punkt pierwszej litery nazwiska. Podobny zapis pojawia się, jak to się stosuje np. w encyklopediach czy słownikach, jako odsyłacz przy hasłach związanych z danym twórcą. Miało to m. in. wpływ na generowanie podziału: kiedy zaczyna i kończy się segment poświęcony danemu twórcy oraz jakie ma składowe. Początkowo stosowane narzędzia miały też ograniczenia, nie wychwytyjąc oznaczenia kolorem. W bibliografii LnŚ są użyte podstawowe kolory drukarskie, czarny i czerwony. Tym drugim oznaczone są np. kody państw wg normy ISO 3166-2 oraz informacja o tym, że dany autor jest laureatem nagrody Nobla. W przypadku nazw krajów zastosowano także autorski sposób rozstrzelenia druku, poprzez rozdzielenie liter spacjami – i tak należało „przyuczyć” program, aby rozpoznawał miejsca, w których rozpoczyna się bibliografia następnej literatury narodowej.

Specjalnego potraktowania wymagała także konwencja związana ze znakami umownymi. Bednarz podaje je w następujący sposób:

~ oznacza kontynuację roku bądź roku i numeru z poprzedzającego zapisu

w danym haśle osobowym

◆ oddziela hasła osobowe od utworów zebranych w cykle typu : Z poezji... ;

od recenzji antologii

◆ ◆ sygnalizuje pulę artykułów dotyczących danej literatury i wydarzeń literackich

\*...\* dwa asteryksy zamykają teksty tłumaczone przez jednego tłumacza, którego

nazwisko widoczne jest w przy pierwszym tekście poprzedzonym asteryksem

→ odsyła do tekstu powiązanego

Poniżej pojawia się także kolejna grupa oznaczeń dla indeksu osób:

Znaki :

+ poprzedza strony z portretami oraz niekiedy z ilustracjami tematycznie

związanymi z daną osobą ;

~ zastępuje powtórzenie zapisu tego samego roku bądź roku i numeru w zakresie

jednego hasła ;

→ odsyła od inicjałów do nazwiska Autora.

Widzimy tutaj, że jeden znak może pełnić więcej niż jedną funkcję w obrębie całości pliku, co też jest istotne. W toku prac okazało się także, iż znak tyldy (~) jest w ogóle przyjęty jako powtórzenie, tzn. poza wymienionymi wyżej występował też jako skrócone powtórzenie nazwy współpracownika (tłumacz - „tł.”).

Początkowo pracowałem na pliku PDF przekonwertowanym na tekst. Następnie udało się od autorki bibliografii pozyskać pliki robocze edytora tekstu dla bibliografii literatur światowych. Ułatwiło to zadanie, ponieważ zamiast jednego dużego (i trudnego w automatycznym przetwarzaniu) pliku miałem trzy mniejsze, a dodatkowo był to format pełnotekstowy – zmniejszyło się ryzyko błędnej konwersji PDF do tekstu.

W największym uproszczeniu, dla przedstawienia bibliografii LnŚ za pomocą kodu, jest ona listą list. Każda pozycja na liście najwyższego rzędu, czyli w spisie państw, jest zbiorem innych list. Schodząc niżej, są to nazwiska twórców. Następnie – każdy twórca jest reprezentowany przez listę jego publikacji. Na najniższym poziomie publikacja jest zbiorem informacji, jakie znamy z adresu bibliograficznego, np. przypisu – każdy taki element, jak nazwa współpracownika („tłumacz”) i jego nazwisko, to osobne całości, z których można stworzyć

tabele. Po dodaniu wspólnego klucza (identyfikatora osób) dla wielu tabel otrzymuje się stosunkowo jednorodną całość.

Następnym celem było wyeksportowanie listy jako *dataframe*, nazywany potocznie spolszczeniem „ramką danych”, czyli formę dwuwymiarowej tabeli, gdzie wiersze odpowiadają za pojedyncze zapisy bibliograficzne, a kolumny wskazują na ich poszczególne atrybuty. Pozwala to na harmonizację i edycję danych do publikacji. Zanim to się jednak stało, wymagało to próbnego eksportu i dostosowania skryptu.

### Rozwój środków technicznych

Szybko w miarę pracy z danymi rozrósł się narzędziownik programowania w Python. Początkowe analizy tekstu, najpierw z plikiem PDF, później z dokumentami tekstowymi, wymagały wykorzystania pakietów do przetwarzania dokumentów tekstowych. Później dzięki bibliotece pandas można było wyrazić dane w formie tabelarycznej (jako ramkę danych, czyli *dataframe*). Inne pakiety służą konwersji do pożądanego formatu danych lub przeprowadzania określonych operacji. Wymieniam poniżej kilka najważniejszych, które znalazły zastosowanie w przekształcaniu plików do tabel i pracy na danych połączonych.

- *io* – jeden z podstawowych modułów, zarządzający przekształceniami obiektów, np. tekstu jako ciągu znaków w schemacie dostarczania danych na wejściu i odbierania wyniku operacji na wyjściu (*input-output*);
- *itertools*, *more\_itertools* – moduły iteracji, pozwalające tworzyć pętle, czyli działania powtarzalne o określonych warunkach zaprzestania czynności i reagowania na błędy;
- *json* – moduł zapisu i odczytu danych w formacie JSON;
- *math* – biblioteka dająca możliwość obliczeń, stosowania stałych oraz funkcji matematycznych;
- *numpy* – moduł pozwalający na obsługę wielowymiarowych tabel i macierzy;
- *pandas* – jedna z najpopularniejszych bibliotek, służąca do edycji i analizy danych;
- *pdfplumber* – narzędzie o wdzięcznej nazwie, które pomaga w ekstrakcji tekstów i tabel z plików PDF (warto nadmienić, że PDF stosuje się często, aby zabezpieczyć przed nadmiernym kopiowaniem i ingerencją w treść, co wprost utrudnia wydobywanie z nich informacji, zatem tego typu narzędzia bywają ocaleniem); „hydraulik PDF” nie napracował się na szczęście za bardzo: po krótkim użyciu prace były kontynuowane na plikach edytora tekstu;

- `regex` – wyrażenia regularne, czyli rozpoznanie części składowych rekordów na podstawie poszukiwania prawidłowości zapisu; opisuję je dokładniej, odnosząc się do obróbki *datasetów*;
- `requests` – popularna biblioteka umożliwiająca wysyłanie zapytań do sieci za pomocą protokołu HTTP; dzięki niej możliwe staje się np. automatyzowanie odpytań katalogu Biblioteki Narodowej przez interfejs programistyczny;
- `sickle` – biblioteka do obsługi protokołu danych OAI-PMH, powszechnie wykorzystywanego nauce i w kontekście danych bibliograficznych; umożliwia np. *harvesting* czyli pobieranie zasobów z zewnętrznych archiwów udostępnionych w ramach OAI-PMH;
- `textract` – narzędzie ekstrakcji tekstów z różnych źródeł, które mogą sprawiać trudność poza formatami *stricte* zawierającymi tekst, mogą to być także pliki graficzne i dźwiękowe;
- `tqdm` – narzędzie wizualizujące postęp działania kodu;
- `xml.etree` – narzędzie tworzenia danych w formacie xml.

Doświadczenia pracy z danymi

Muszę tutaj zaznaczyć, że użycie narzędzi technicznych zwłaszcza dla literaturoznawcy, przypomina istic Kafkowską podróż przez niezbadane rejony internetu, zawierające poradniki, jak wykonać jakies operacje albo ktorzych narzędzi używać. Python stanowi przystępną i popularną część technologicznego *stacku* (narzędziownika i kompetencji), zatem w bród znajdziemy zasobów pisanych w tonie *troubleshooting*: radzenia sobie z konkretnymi zagadnieniami, przy użyciu najskuteczniejszych (najszybszych) metod. Nasuwa się ciekawa obserwacja: wiedza programistyczna jest używana w swego rodzaju przepływie, w inżynieryjny sposób (można zapomnieć detale rozwiązania, o ile osiągnęło się zamierzony cel), co kontrastuje z kompetencjami humanistycznymi, nastawionymi na internalizację i stałe użycie wiedzy dziedzinowej.

Błędy i wyzwania przy pisaniu kodu na bieżąco były zatem konsultowane z forami dla osób programujących profesjonalnie i amatorsko. Jest to tak popularna dziedzina, że właściwie każde zagadnienie, nawet jeśli nie nazwalibyśmy go trywialnym, znajdowało gotowe rozwiązania (ponieważ ktoś miał podobny problem w przeszłości i udało się uzyskać poradę lub rozwiązanie). Tak było, kiedy kod nie pokazywał wszystkich osób z kategorii współpracowników: skrypt Python nie rejetrował dwóch tłumaczy (Klementyna Suchanow i K. Radny); ostatecznie trzeba było przeformułować zapytanie. Inną trudność sprawili

tłumacze z pojedynczym inicjałem imienia zamiast dwoma (B. L. Surowska), ale tutaj ewidentnie zawiodło początkowe założenie o budowie takiego rekordu i wystarczyło wziąć pod uwagę występowanie dwóch inicjałów (inicjały imienia nie mogły zostać użyte jako prosty separator części zapisu bibliograficznego).

Spisywanie danych bibliograficznych z autopsji nie jest wolne od błędów. Tak też było w przypadku opracowywanego dokumentu Bednarz, co narzędzia programistyczne mogą znaleźć prędzej niż, romantyzując technologię, mędrca szkiełko i oko. Ujawnił się tutaj ruch, który wcześniej przyrównywałem do koła hermeneutycznego: mediacja między szczegółem pytania badawczego, a ogółem nieprzetworzonych danych, by wrócić do szczegółu w uzyskanym zbiorze i zmodyfikować na powrót ogół przetwarzanej informacji.

Wspomniany wyżej regex umożliwił szybkie przeszukiwanie całości zbioru tekstowego, aby wyszukać elementy powtarzalne. Wyrażenia regularne wymagają specjalistycznej wiedzy co do formuły zapisu. Posługują się ściśle określoną składnią i notacją, których konsekwentne używanie jest istotne. Korzystałem w tym celu ze strony regex101, oferującej interfejs do testowania wyrażeń regularnych, opartych na różnych standardach i językach programistycznych oraz pomoc w postaci podpowiedzi (wyjaśnienia działania wprowadzonych wyrażeń i słowniczek tych możliwych do wprowadzenia)<sup>405</sup>. Narzędzie pozwala wkleić opracowywany właśnie tekst w jednym polu interfejsu i testować wyrażenia regularne, wpisywane w drugim polu, podświetlając wyszukania we wprowadzonej treści.

Regex nie stanowi wszakże narzędzia uniwersalnie radzącego sobie z każdym zagadnieniem przetwarzania danych tekstowych: przesłanką jego zastosowania jest uprzednie zarejestrowanie prawidłowości w opracowywanym źródle. Może się jednak zdarzyć, że owa prawidłowość nie zostaje zachowana, nie obowiązuje dla całości przetwarzanego materiału lub ma nieoczywiste wyjątki.

W odróżnieniu od wydobywania informacji z tekstu (*text mining*), regex nie jest bowiem przetwarzaniem języka naturalnego. Badacz nie może tutaj polegać na odwołaniu do powszechnie znanej syntaktyki i reguł wypracowanych przez badania lingwistyczne. Odpowiednikami syntaksu stają się struktury wykryte w badanym zbiorze. W pewnym sensie te struktury tworzone są też w samym zapytaniu do zbioru przetwarzanych informacji. Niczym w przypadku pytyjskiej wyrocni, jakość odpowiedzi nierozzerwalnie wiąże się z jakością

---

<sup>405</sup> Dib, Firas. „Regex101”. Regex101. Dostęp 2022-12-12. <https://regex101.com/>.

pytania. Jest też tak, że otrzymane dane wpływają na kształt i zakres zadanego na początku procesu pytania badawczego.

Przykładowo, aby osiągnąć zapis intuicyjnie zrozumiały dla autorki, w kodzie należy dokonać przekształceń danych, aby informacja była wyrażona konsekwentnie. Irena Bednarz określiła w dokumentacji bibliografii LnŚ, że przy tłumaczeniach autorstwa tej samej osoby, dla zachowania ekonomii tekstu, będzie się posługiwać gwiazdką na początku i końcu zbioru. Zapisy z gwiazdkami należało zatem zgrupować i przypisać im tego samego tłumacza. Zanim to nastąpiło, trzeba było wziąć tę sytuację pod uwagę przy „wyciąganiu” powtarzalnych fraz za pomocą regex, ponieważ brak informacji szukanej nie musiał oznaczać, że jej nie ma w przeszukiwanym fragmencie. Mogła to być dana „przemilczana” ale zasygnalizowana autorskim oznaczeniem, podana niejako „w domyśle”.

Do weryfikacji tych danych posłużyły - ponownie w kolistym ruchu - pozostałe zbiory, mające wejść w skład ostatecznego *datasetu*, czyli BN i PBL. Podczas harmonizacji można było się natknąć na kilka pomyłek, wynikających z przygotowania bibliografii. Dzięki ich wykryciu, docelowy *dataset* może łączyć wszystkie przewagi swoich pomniejszych źródeł oraz niwelować pomyłki i niedociągnięcia.

Pierwsze testy kodu przeprowadzałem na danych literatury albańskiej. Pisarz z tego kraju, Rrhaman Dedaj, figuruje w bibliografii LnŚ jako „Brahman Dedaj” – prawdopodobnie zadziałało tu podobieństwo do pojęcia z hinduizmu, jak i rzadko spotykane imię, z podwojeniem pierwszej litery, które polskiemu odbiorcy nie musi być znane<sup>406</sup>. Inną, wartą odnotowania pomyłką, było opuszczenie jednej litery w tytule eseju Johna Bartha *Literatura wyczerpania*, znanego dobrze polonistom. W zbiorze bibliograficznym jest jako *Literatura wyczerpana*. Tutaj prawdopodobnie powodem ostatecznej formy zapisu było podobieństwo słów. Dla utworu *Historia wieczności* Borgesa są podane różne zakresy stron: w sekcji przekładów Adama Elbanowskiego są to 19-33, a w zbiorze dzieł Borgesa 19-35 ze znakiem zapytania. W numerze poświęconym Dantemu 1995) tłumacz Maciej Ziętara jest wymieniony jako „Zientara M.”. Tu w ramach weryfikacji złamałem zasadę pracy wyłącznie z danymi i sięgnąłem po fizyczny egzemplarz pisma. Jest to powtórzenie błędu wydawcy, a zatem Bednarz zdaje sprawę ze stanu faktycznego, jednak przygotowując zbiór danych, ujednoliciłem

---

<sup>406</sup> Wspomnę też o popularnej przyczynie, której nie podejrzewam w tym wypadku, niemniej potrafi być zmorą nie tylko bibliografa, ale literaturoznawcy: autokorekta edytorów tekstu zagraża pracy nad rozbudowanymi tekstami najeżonymi nazwami własnymi.

pisownię do „Ziętara”<sup>407</sup>. Dla badacza niesie to ważne skutki. Po otrząśnięciu się z pierwszej wątpliwości, czy nie mamy tu do czynienia z danymi dotyczącymi dwóch różnych osób, można docenić dostępność wielu źródeł, z którymi można konsultować informacje. Świadczy o ważności harmonizacji i grupowania danych: pokazuje, jak różne zbiory mogą wpływać na siebie. Jeden *dataset* podnosi jakość drugiego (zbiory danych wzajemnie służą sobie jako źródła weryfikacji).

Wspomniana już ekonomia w przygotowywaniu bibliografii zaowocowała również skróconym zapisem zakresu stron, na których występowały teksty w niej opisane. Bednarz odnotowuje go minimalną liczbą znaków. Na cele publikacji *datasetu* wprowadziłem zmiany, aby osiągnąć pełną numerację, tzn. zapis umowny „60–3” przybrał formę „60-63”, „212-13” modyfikowałem do „212-213”, „360-1” do „360-361” itd.; dotyczy to całego zbioru. Taka sama operacja miała miejsce przy podwójnych numerach ciągłych pisma, np. 1-2 (462-463) z 2010 roku, w zapisie Bednarz początkowo jako „462-3”. W przypadku zakresu stron zdecydowałem się usunąć znaki zapytania, wstawione przez autorkę bibliografii, ale tylko w kolumnach z wydzielonymi poszczególnymi informacjami bibliograficznymi, gdzie numeracja stron jest główną informacją.

Również z zakresem stron wiąże się kwestia, którą rozstrzygnęło dopiero użycie danych połączonych: w numerze 5-6 z 1995 pojawia się zbiorczy zapis dla kilku utworów poetyckich na stronach 103-108 oraz odrębny zapis dla utworu o tytule Edgar Allan Poe, z podanymi stronami 107-108. Przez porównanie z PBL potwierdzone zostały strony dla tego drugiego zgodne z zapisem Bednarz, a grupa wierszy kończy się na stronie 107.

Dodawanie autorskich oznaczeń zastosowała autorka bibliografii LnŚ, ale sam także posługiwałem się oznaczeniami czy tagami zrozumiałymi przede wszystkim dla mnie samego. Na przeciwnych biegunach tego działania sytuują się precyzja i powtarzalność. Z jednej strony, celem badacza jest jak najdokładniejsze wskazanie charakterystyk opisywanego przedmiotu. Z drugiej, wkrada się niebezpieczeństwo zbyt rozwiniętej unikatowości, a co za tym idzie, jednorazowości zastosowań. Mając na uwadze ponowne użycie danych, w zgodzie ze standardami FAIR, a także zrozumiałość dla jak najszerszego grona odbiorczego, ostatecznie kierowałem się do jak najprostszych informacji – na tyle ogólnych, aby dało się je łatwo zrozumieć i przetwarzać ponownie.

---

<sup>407</sup> Zasadność tej poprawki potwierdziła także kwerenda w PBL.

Pierwszy *dataset* - LnŚ

Tabela w *datasecie* początkowym LnŚ zawierała 23 typy informacji. Był to pierwszy element kuracji treści – doboru, jakie informacje pokazać, jak sklasyfikować i nazwać. Poniżej prezentuję ich nazwy (nagłówki kolumn) w wytłuszczeniu i po dwukropku krótki opis zawartości, z komentarzem odnośnie celu i możliwości zastosowania:

**Literatura:** przynależność do literatury narodowej, czyli informacja, z której części bibliografii Ireny Bednarz pochodzi dany wpis; w zbiorze Borgesowskim wszystkie mają przyporządkowaną wartość „ARGENTYNA”.

**Oznaczenie kraju:** kod ISO 3166-2, podany również w bibliografii LnŚ; dla wszystkich wpisów jest to „AR”.

**Pisarz:** informacja, jakiego autora dotyczy wpis, dziedziczona w każdym wierszu tabeli jako „BORGES, Jorge Luis 1899–1986”. Informacja dotyczy sekcji w bibliografii LnŚ.

**Pisarz VIAF ID:** unikatowy identyfikator Jorge Luisa Borgesa w bazie VIAF, „88919448” dla każdego wpisu.

**Zapisy bibliograficzne:** kolumna pomocnicza, w której znajdują się informacje wyekstrahowane z pliku tekstowego bibliografii LnŚ, oddzielone znakiem „|” według typu, dla ułatwionego dalszego podziału na kolumny (przykładowa wartość to, zgodnie z zachowaniem oryginalnego zapisu: „w i e r s z e|tł.|Baterowicz, M.|1973 nr 12 (32) s. 246–7, Aluzja do śmierci pułkownika Francisco Borgesa ; Ajedrez”).

**Literatura przedmiotowa:** kolumna służąca do oddzielenia tekstów autorskich Borgesa od omówień jego twórczości, wywiadów itp. Przybiera wartości „tak” lub „nie”.

**Była gwiazdka:** kolumna pomocnicza, określająca zakres dzieł przełożonych przez jednego tłumacza (jak było to sygnalizowane wcześniej, przyjęta przez Irenę Bednarz metoda opisu maksymalnie redukowałą wszelkie redundancje, stąd lista przekładów dokonanych przez jedną osobę była oznaczana dwiema gwiazdkami na początku i końcu zbioru). Przybiera wartości „tak” lub „nie”.

**Typ publikacji** – tam, gdzie pojawia się określenie typu przy danym wpisie; niektóre z nich były sygnalizowane rozstrzelonym drukiem (por. przykład kolumny „Zapisy bibliograficzne”). W kolumnie znajdziemy brak wypełnienia lub jedno z określeń: „dzieła”, „kalendarium”, „listy”, „nekrolog”, „przedmowy”, „recenzja”, „wiersze”, „wywiad”.



**Współpracownik – rola:** w badanym zbiorze występowały tylko dwie role pierwszych współpracowników: tłumacz i recenzent, oznaczeni odpowiednio w kolumnie jako „tł.” oraz „rec.”.

**Współpracownik – nazwisko** – imię i nazwisko pierwszego współpracownika.

**Współpracownik - VIAF ID:** unikatowy identyfikator współpracownika z poprzedniej kolumny.

**Współpracownik 2 – rola:** kolumna dodana, gdy okazało się, że część wpisów będzie zawierała kilka nazwisk osób, które przygotowały dany tekst. Oznaczenia drugiego współpracownika, zgodnie z przyjętą konwencją. Wśród kolejnych (nie-pierwszych) współpracowników pojawili się także prowadzący wywiady („rozm.”), autorzy i redaktorzy książek oryginalnych („autor”, „autor listu”, „red. dzieła omawianego”).

**Współpracownik 2 – nazwisko:** imię i nazwisko kolejnego współpracownika, zgodnie z przyjętą konwencją.

**Współpracownik 2 - VIAF ID:** unikatowy identyfikator współpracownika z poprzedniej kolumny.

**Współpracownik 3 – rola:** rola kolejnego współpracownika, zgodnie z powyższym opisem.

**Współpracownik 3 – nazwisko:** imię i nazwisko kolejnego współpracownika, zgodnie z przyjętą konwencją.

**Współpracownik 3 - VIAF ID:** unikatowy identyfikator współpracownika z poprzedniej kolumny.

**Opis bibliograficzny:** wygenerowane kodem adresy bibliograficzne (na podstawie powtarzających się prawidłowości tekstu skrypt Python wygenerował osobne całości), pozbawione nazw osób, czyli autorów i tłumaczy; w kolejnych kolumnach rozbite na poszczególne części.

**Rocznik:** rok wydania pisma.

**Numer bieżący:** numer LnŚ w danym roku.

**Numer ciągły:** numer w ogólnym podsumowaniu, od początku istnienia pisma.

**Strony:** adres stronicowy w zawartości pisma, wyrażony liczbą, jeśli tekst mieścił się na jednej stronie, lub zakresem liczb zapisanym za pomocą dywizu dla dłuższych publikacji.

**Komentarz:** kolumna pomocnicza z dodatkowymi informacjami. Początkowo kolumna robocza, docelowo przeznaczona na uwagi dla odbiorców *datasetu*.

Do powyższych dodane zostały jeszcze trzy kolumny. Pierwszą był tytuł, wyodrębniony z zapisu. Wymagało to decyzji, czy pojedynczy wiersz ma reprezentować jeden utwór, czy nie.

Uznałem, że można potraktować grupę krótkich form jako jeden zapis, analogicznie do pozycji w spisie treści samego pisma, czyli powtarzając wybór redaktorski. Tytuły poszczególnych utworów wymieniam w jednym polu. Stoi też z tym argument lokalizacji, tzn. dysponujemy wyłącznie informacją o stronach, które zajmuje dany zapis bibliograficzny jako grupa tekstów. Bez spisu z autopsji w każdym pojedynczym przypadku nie jesteśmy w stanie stwierdzić, czy dany tekst z jednego zbioru określonego np. jako „wiersze”, z listą tytułów, mieści się na jednej stronie, czy zajmuje więcej miejsca w danym numerze. W przypadku krótkich form, np. poetyckich, stałą praktyką jest umieszczanie także więcej tekstów na jednej stronie, także trudno mówić o prawidłowościach<sup>408</sup>.

Zdecydowałem się również dodać pole **Autor**, aby łatwiej oddzielić dzieła Borgesa od literatury przedmiotowej i wskazać osobę, która napisała dany tekst. Pełni to podobną rolę do rozróżnienia Twórcy od Autora w PBL. Pociągnęło to za sobą dodanie pola **Autor VIAF ID** – z reguły powinno być tożsame z którymś wymienionym wcześniej (ID Borgesa lub któregoś ze współpracowników), ale dla porządku powinno się pojawić takie korespondujące pole.

Przy okazji tego działania, nadal przy pomocy MS Excel, pomogły też wbudowane w ten program funkcjonalności<sup>409</sup> – sugestie autouzupełnienia komórek bez konieczności tworzenia formuł lub rozdzielania tekstu do kolumn (program wykrył powtarzalną czynność powtórzenia członu z innej kolumny, w tym wypadku nazwiska autora z kompletnego zapisu bibliograficznego).

Zdecydowałem się także ujednolicić zapis do formy „Borges, Jorge Luis”, względem wcześniejszej „BORGES, Jorge Luis 1899–1986”. Uwzględnienie identyfikatorów VIAF

---

<sup>408</sup> Oczywiście próbując przyjąć kryterium gatunkowe, można się spodziewać (nawet uwzględniając kieszonkowy format wydań LnŚ), że trzy do czterech haiku zajmą jedną stronę, sonety po jednej stronie na utwór, a fragmenty trzech poematów epickich odpowiadają od jednej do kilku stron na każdy z nich. Są to jednak wyłącznie swego rodzaju heurystyki i przypuszczenia, trudne do zastosowania w takim zbiorze.

<sup>409</sup> Korzystałem z wersji 2019 pakietu MS Office.

pozwała jednakowoż nie przejmować się formą zapisu nazwiska osoby, ponieważ baza kontoli autorytatywnej przechowuje wiele różnych sposobów zapisu nazwiska<sup>410</sup>.

Początkowa tabela zawierała 92 wiersze (zapisy bibliograficzne), dwa z nich zostały usunięte. Wspomniana już *Literatura wyczerpania* pojawiła się dwukrotnie po eksporcie za pomocą kodu. Skrypt wygenerował też na końcu jeden dodatkowy wiersz, pozbawiony merytorycznej treści.

W badaniu pojawiły się pierwsze możliwości wnioskowania i obserwowania tendencji. Przedstawiam dwie ilustracje, zestawiające numery ciągle i roczniki LnŚ, w których znajduje się tekst autorstwa Borgesa lub jemu poświęcony, z sumą tych publikacji. Dane ilustrują obecność autora *Księgi piasku* praktycznie przez cały czas wydawania pisma jako „powracającego” na jego łamy, od numeru 8. z pierwszego rocznika, po numery z piątej setki numeracji bieżącej, a więc prawie po czterdziestoleciu wydawania (istnieją skrajnie odwrotne przypadki, niszowych literatur oraz twórców, którzy zaistnieją jedynie raz w całym okresie wydawania tego tytułu). Punktem wybijającym się na wykresie pod względem liczby publikacji jest numer monograficzny, poświęcony pisarzowi: 12/1988 (209)<sup>411</sup>.

Obecność Borgesa w czterech dekadach pisma zajmującego się literaturą światową przywodzi skojarzenia z ponadczasowością, kanonem i klasyką. Zwykle na tego rodzaju przywołania po latach zasługują najwięksi twórcy: dla współczesnych im badaczy to przede wszystkim temat godzin zreferowania (w zbiorze danych mamy pewien korpus tekstów „bieżących” – wywiady, listy, krytyka); młodszy tłumacze i teoretycy znajdują w takim autorze punkt odniesienia lub wyzwania translatorskie w twórczości (fragmenty jeszcze nieprzełożone lub rewizyty i własne propozycje tłumaczeń). Takie określenie Nie jest to jednak sąd i kwestia smaku, ale poparcie w punktach danych (*data points*).

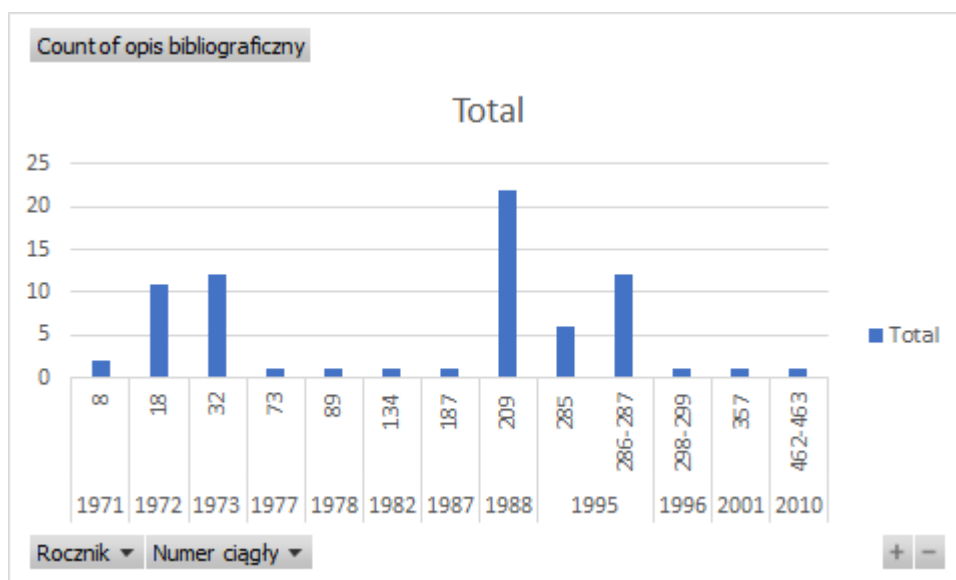
---

<sup>410</sup> Jest to przydatne w przypadku konwencji różnych języków (np. czeskie feminatywy w nazwisku albo sposoby zapisu w alfabetach niełacińskich).

<sup>411</sup> Warto tutaj zauważyć, że przygotowywany zbiór nie posiada informacji o tym, czy dany numer to monografia poświęcona autorowi, czy np. prezentacja regionu lub tematu (jak to wskazywałem wyżej, zgodnie z praktyką redakcyjną). Aby się tego dowiedzieć, należy odszukać szczegółowe informacje o tym konkretnym wydaniu. Jest to przykład kolistego ruchu między ogółem danych a rzeczywistym szczegółem, między redefiniowaną statystycznie całością a tradycyjnym opisem przedmiotu z autopsji. Potencjał automatyzacji takiego działania stanowi obróbka materiału wizualnego, aby uzyskać te informacje z okładek (co nie zawsze się uda ze względu na różnorodność rozwiązań graficznych i zastosowanej typografii – są zmienne z jednego wydania na drugie). Na chwilę obecną prawdopodobnie wymaga to większych nakładów niż manualne sprawdzenie rzeczywistych egzemplarzy lub kwerenda w sieci (zdjęcia okładek i kopie spisów treści).

Row Labels	Count of opis bibliograficzny
1971	2
8	2
1972	11
18	11
1973	12
32	12
1977	1
73	1
1978	1
89	1
1982	1
134	1
1987	1
187	1
1988	22
209	22
1995	18
285	6
286-287	12
1996	1
298-299	1
2001	1
357	1
2010	1
462-463	1
<b>Grand Total</b>	<b>72</b>

Ryc. 4 Zapisy dotyczące Borgesa w pierwszym zbiorze danych LnŚ (rok i numer pisma)



Ryc. 5 Zapisy dotyczące Borgesa wyrażone w formie wykresu frekwencji

Do połączenia z innymi zbiorami użyłem identyfikatorów kontroli autorytatywnej VIAF. Dla operacji na danych w przyszłości nadałem unikatowe identyfikatory wewnątrz opracowywanego zbioru, składające się z prefiksu JLB i losowej liczby (wygenerowanej w MS Excel za pomocą funkcji RAND). Miało to na celu uniknięcie potencjalnych błędów przy

sortowaniu i filtrowaniu informacji w arkuszu kalkulacyjnym. Posiadanie identyfikatorów dla zasobów stanowi dobrą praktykę w kontekście ponownego ich użycia.

Praca z danymi PBL

W PBL po przygotowaniu zapytania SQL i wyeksportowaniu wyniku do pliku Excel, otrzymano 430 rekordów. Kryterium było zdefiniowanie Borgesa jako Twórcy w bazie (każdy Twórca posiada unikatowy identyfikator, dla J. L. Borgesa to „1908”). Spośród tych rekordów, 34 pochodzą z LnŚ. Pod względem dostępu, typowy użytkownik serwisu PBL musi się posługiwać interfejsem, dostępnym na stronie internetowej, ale zapytanie SQL było możliwe, ponieważ nadano mi uprawnienia użytkownika do odczytu zawartości bazy, z możliwością przeglądania zawartości oraz eksportowania jej treści.

Jeśli chodzi o zakres, niestety pozyskane z PBL informacje dotyczą bardzo wąskiego wycinka badanego tematu: numery: 5/6 z 1995, 5/6 z 1996, 9 z 1999, 3 z 2000 oraz 10/11 z 2000 roku, czyli łącznie 5 wydań pisma. W bibliografii Bednarz mamy ich zaś łącznie 25 (takie numery, w których występuje tekst dotyczący Borgesa lub jego autorstwa – często współwystępują, zgodnie z charakterem LnŚ jako periodyku, np. recenzja czy tekst krytyczny towarzyszy dziełu samego pisarza, tworząc dział tematyczny).

Pod względem zawartości, jak już wspominałem, rekordy bazy PBL są tworzone przez bibliografów z autopsji, a część metadanych i dodatkowych informacji pochodzi ze źródeł zewnętrznych: np. tagowanie dla kontroli autorytatywnej polega na użyciu istniejącego już klucza, a nie wytworzeniu nowej informacji przez zespół (jest jednak niezbędne dla nadania tym danym łączliwości z innymi zbiorami); PBL korzysta także z importu części danych m.in. z BN, dla wzbogacania własnych danych (*data enrichment*). Bazę cechuje także ścisłe śledzenie zawartości czasopism, na co liczyłem w zestawianiu jej informacji z bibliografią Bednarz. Rekordy posiadają bogate opisy z komentarzami, dzięki temu, że opracowują je specjaliści. Niesie to szansę na odszukanie interesujących informacji, ale też stanowi ryzyko pod względem bazodanowym. W tym drugim podejściu mamy bowiem do czynienia z ciągiem dowolnych znaków (*string*): dla humanisty stringiem jest właściwie każda publikacja, ale gdy ubiera się informację w rygory danowe, zaczyna to przeszkadzać w standaryzacji. Mimo tego, wspomniane komentarze mogłyby się okazać cenną wiedzą ekspercką, wzbogacającą obraz całości. Rejestrowanie nagród i imprez wyróżnia PBL na tle pozostałych składowych *datasetu*, co stwarzało szansę na odnalezienie dodatkowych i unikatowych danych.

W przeciwieństwie do oszczędnego opisu Bednarz, obliczonego na zajęcie najmniejszej przestrzeni i użycie możliwie niskiej liczby znaków (unikanie redundancji), w PBL takie zabiegi nie są potrzebne. Dzięki temu zamiast jednego zbiorczego zapisu dla wierszy, mamy do czynienia z rozpisaniem każdego pojedynczego utworu jako osobnego zapisu. Dzięki temu dla wydania 5/6 z 1995 poznajemy ich dokładną lokalizację na stronach. PBL odnotowuje dodatkowo materiał ikonograficzny: odnośnie strony 62. w numerze 10/11 z 2000 pojawia się odrębny zapis „fot. z okresu dzieciństwa”.

Wspomniana kontrola autorytatywna realizuje się przez identyfikatory VIAF (Virtual International Authority File), które nadałem osobom, jakie udało się określić. Do spełnienia były dwa warunki: osoba posiada numer VIAF (w większości przypadków tak było) oraz można zweryfikować, że przypisanie właśnie tej osoby do danego rekordu jest właściwe. Dzięki temu można łączyć nazwy osób ze sobą w szerszym zbiorze – identyfikować twórców, autorów dzieł, tłumaczy, redaktorów i inne osoby.

Przetwarzanie danych Biblioteki Narodowej

W BN zidentyfikowano 3634 rekordów dotyczących LnŚ, 248 rekordów wymieniających Borgesa jako autora i 235, w których Borges figuruje jako temat. Zająłem się przede wszystkim zbiorem bibliografii książek i czasopism, aby wykazać relacje między publikacjami w periodykach i książkowymi. Dodatkowym kryterium było rozróżnienie na oryginalną twórczość własną i opracowania na temat dorobku autora.

Patrząc na zawartość i zestawiając ze wszystkimi zbiorami wejściowymi, BN cechuje największa aktualność danych względem pozostałych źródeł oraz największy wolumen danych dostępnych w całości do przeszukania. Schemat opisu danych podlega MARC 21. Jest to najbardziej uniwersalna struktura do systemów bibliotecznych. W bazie BN występują informacje kontroli autorytatywnej: identyfikatory VIAF, istnieje możliwość powiązania rekordów m.in. z danymi Wikidata.

Dostęp był możliwy poprzez API. Odpowiedź na zapytanie użytkownika serwis programistyczny BN segmentuje na listingi po sto rekordów w formacie JSON wyświetlane jako osobne strony w przeglądarce internetowej, a na końcu każdego z nich znajduje się odnośnik do kolejnej setki. Może to wymagać pracy ręcznej, aby importować kolejne pliki JSON i sklejać w całość, ale taki powtarzalny proces ma pole do automatyzacji. Posłużył do tego kod autorstwa Cezarego Rosińskiego, który importował kolejne strony wyświetlanej listy do pojedynczego (i jednolitego strukturalnie) pliku. Użyte do pobierania narzędzie

wymagało mediowania między strukturą JSON (ładowanie kolejnych stron z rekordami) a MARC 21 (docelowa struktura opisu bibliograficznego dla powstającego *datasetu*).

Eksporty w formacie MARC 21 poddałem następnie ocenie pod względem przydatności pól, określając, które typy informacji będą wartościowe dla docelowego zbioru danych. Poniżej zestawiam polabrane pod uwagę. Oznaczenia (P) i (NP) oznaczają odpowiednio pole powtarzalne i niepowtarzalne. W trzeciej kolumnie tabeli umieszczam rozpoznania względem celu kompilowania informacji do Borgesowskiego *datasetu*.

Tabela 2. Pola MARC 21, opracowane na potrzeby uzyskania informacji do datasetu

<b>BN - nr pola MARC 21</b>	<b>BN - nazwa pola MARC 21</b>	<b>Rola w treści lub przygotowywaniu <i>datasetu</i></b>
LDR	leader	
001	identyfikator	Główny klucz do łączenia pól; pomocnicza wartość do rozbijania podpól MARC 21 na osobne kolumny
041	<b>Język dokumentu (P)</b> <b>Kody podpól</b> <b>Pierwszy – wskaźnik przekładu</b> <b>0 dokument nie jest przekładem i nie zawiera przekładów</b> <b>1 dokument jest przekładem i/lub zawiera przekład</b> <b>Drugi – źródło kodu @ MARC Code List of Languages</b> <b>\$a Język tekstu (P)</b> <b>\$b Język streszczeń (P)</b> <b>\$k Język tekstu przekładu pośredniego (P)</b> <b>\$h Język tekstu oryginalnego (P)</b>	Wskazanie języka oryginału i języka przekładu
100	<b>Główny punkt dostępu – nazwa osobowa (NP)</b>	Wskazanie osób, filtrowanie informacji niezwiązanych z

		<i>datasetem</i> Borgesowskim lub określenie powiązania pośredniego.
245	<b>Strefa tytułu i oznaczenia odpowiedzialności (NP)</b>	Identyfikacja dzieł na podstawie tytułu
260	<p><b>Strefa adresu wydawniczego (NP)</b></p> <p>Wskaźniki: nieużywane</p> <p>Kody podpól:</p> <p>\$a Miejsce wydania (P)</p> <p>\$b Nazwa wydawcy (P)</p> <p>\$c Data wydania (P)</p> <p>\$e Miejsce druku (NP)</p> <p>\$f Nazwa drukarni (NP)</p> <p>\$g Data druku (NP)</p> <p>Drugie podpole \$a poprzedza się średnikiem.</p> <p>Podpole \$b poprzedza się dwukropkiem, \$c – przecinkiem,</p> <p>\$f – dwukropkiem. Podpole \$g poprzedza się przecinkiem, jeżeli to podpole występuje po podpolu \$e i \$f.</p> <p>Zawartość podpól \$e, \$f i \$g ujmuje się łącznie w nawiasy okrągłe. Pole kończy się kropką.</p>	Pozyskanie i porównanie informacji o dacie i miejscu wydania
336	<p><b>Typ zawartości (P)</b></p> <p>Wskaźniki: nieużywane</p> <p>Kody podpól:</p> <p>\$a Termin określający typ zawartości (P)</p> <p>\$b Kod określający typ zawartości (P)</p> <p>\$2 Źródło (NP)</p> <p>Podpól nie poprzedza się żadnym znakiem. Na końcu pola nie stawia się kropki.</p>	Pola przybierały wartości jak „nuty” i „tekst”, zachowane dla rozróżnienia.



337	<p><b>Typ mediów (P)</b></p> <p>Wskaźniki: nieużywane</p> <p>Kody podpól:</p> <p>\$a Tryb odtwarzania (P)</p> <p>\$b Kod typu mediów (P)</p> <p>\$2 Źródło (NP)</p> <p>Podpól nie poprzedza się żadnym znakiem. Na końcu pola nie stawia się kropki.</p>	Informacja o konieczności użycia urządzenia pośredniczącego do posługiwania się źródłem
338	<p><b>Typ nośnika (P)</b></p> <p>Wskaźniki: nieużywane</p> <p>Kody podpól:</p> <p>\$a Termin określający typ nośnika (P)</p> <p>\$b Kod określający typ nośnika (P)</p> <p>\$2 Źródło (NP)</p> <p>Podpól nie poprzedza się żadnym znakiem. Na końcu pola nie stawia się kropki.</p>	Rozróżnianie zasobów elektronicznych i fizycznych
380	<p><b>Forma dzieła (P)</b></p> <p>Wskaźniki: nieużywane</p> <p>Kody podpól:</p> <p>\$a Forma dzieła (NP)</p> <p>\$2 Źródło (NP)</p>	Rozróżnienie między artykułami, książkami, e-bookami itp.
386	<p><b>Cecha językowo-etniczna (P)</b></p> <p>Wskaźniki: nieużywane</p> <p>Kody podpól:</p> <p>\$a Deskryptor przedmiotowy (NP)</p> <p>\$m Wyrażenie wprowadzające (NP)</p> <p>\$2 Źródło (NP)</p>	Przynależność do literatury danego języka
500	<p><b>Uwaga ogólna (P)</b></p>	Adnotacje zespołu bibliograficznego, notatki

		kontekstowe, dodatkowe informacje
546	<b>Uwaga dotycząca języka (P)</b>	Informacje o kierunkach tłumaczeń
600	<b>Punkt dostępu rzeczowego – nazwa osobowa (P)</b> \$v Określnik formy (P) Termin wskazujący formę treści dokumentu.	Odniesienia do osób (i niekiedy również pojęć)
650	<b>Punkt dostępu rzeczowego – nazwa pospolita (P)</b>	Przyporządkowanie do literatur języków i narodów oraz odwołanie do pojęć oznaczających tematykę
651	<b>Punkt dostępu rzeczowego – nazwa geograficzna (P)</b>	Odniesienia do miejsc
655	<b>Punkt dostępu rzeczowego – nazwa gatunku formy (P)</b> \$a Nazwa gatunku/formy (NP) \$v Określnik formy (P) Termin wskazujący formę treści dokumentu. \$y Określnik chronologiczny (NP) Oznaczenie (wiek, rok, data, epoka) pozwalające osadzić przedmiot dokumentu w czasie.	Przyporządkowanie gatunkowe lub formalne publikacji i informacje o literaturze danego języka lub narodu
700	<b>Dodatkowy punkt dostępu – nazwa osobowa (P)</b>	Analiza porównawcza osób i ich ról dla powstania dzieła, do którego odwołuje się rekord
740	<b>Dodatkowy punkt dostępu – tytuł (P)</b>	Dodatkowe informacje o tytule
773	<b>773 – Jednostka nadrzędna (P)</b> pierwszy – wskazuje, czy uwaga jest generowana 0 – wyświetla treść pola jako uwagę poprzedzoną	W zbiorze z BN w tym polu znajdowały się adresy bibliograficzne z czasopism oraz nazwy archiwów osób (Archiwum

<p>wyrażeniem wprowadzającym</p> <p>1 – nie wyświetla uwagi – uwaga jest wyświetlana z pola 580</p> <p>drugi – wyrażenie wprowadzające</p> <p># – W:</p> <p>8 – brak wyrażenia wprowadzającego</p> <p>Podpola:</p> <p>\$i – wyrażenie wprowadzające (NP)</p> <p>\$a – hasło główne z pola 100, 110, 111 lub 130 związanego rekordu (NP)</p> <p>\$s – tytuł ujednoczony z pola 240 rekordu związanego (NP)</p> <p>\$t – tytuł z podpól: \$a, \$n, i \$p pola 245 rekordu związanego (NP)</p> <p>\$c – informacje dodatkowe (NP) – pozwalające jednoznacznie zidentyfikować tytuł z podpola \$t, np. dane z podpola \$b pola 245, o ile to potrzebne</p> <p>\$b – oznaczenie wydania z pola 250 związanego rekordu (NP)</p> <p>\$d – adres wydawniczy (NP) - dane z podpól: \$a, \$b i/lub \$c pola 260 związanego rekordu</p> <p>\$g – zakres powiązania (P) – określa lokalizacje jedn. związanej w nadrzędnej</p> <p>\$h – opis fizyczny (NP)</p> <p>\$k – dane o serii dla związanej jednostki z pól 4XX lub 8XX związanego rekordu (P)</p> <p>\$n – uwaga (P)</p> <p>\$o – inna identyfikacja jednostki związanej (P)</p> <p>\$z – ISBN z pola 020 rekordu związanego (P)</p> <p>\$w – numer kontrolny rekordu związanego (P)</p>	<p>Czesława Miłosza, Archiwum Themersonów, Archiwum Zbigniewa Herberta)</p>
--	---

986	Pole definiowane przez każdą instytucję do indywidualnych potrzeb: <a href="https://www.oclc.org/bibformats/en/9xx.html">https://www.oclc.org/bibformats/en/9xx.html</a>	Pole zawierało w kilku przypadkach odniesienia do literatur narodowych, ale ta sama informacja była obecna w innych miejscach
-----	---	---

Konwencja zapisu MARC 21 wymagała rozdzielania podpól, aby zachować same treści bez znaczników, co również wymagało użycia kodu w języku Python. Pracowałem zatem na pliku z całościowym eksportem rekordów oraz tyłoma plikami, ile pól wytypowałem do badania.

Ostatecznie tabela danych BN została utworzona poprzez redukcję informacji niepotrzebnych w kontekście datasetu *LnŚ*, czyli wybór odpowiednich kolumn, a następnie transformację MARC 21 na czysty tekst (można powiedzieć: niosący informację – bez znaczników bibliograficznych). Podpola (w tabeli wskazane ze znakiem dolara) musiały być niejednokrotnie standaryzowane pod względem sposobu prezentacji treści lub łączone ze sobą, do czego należało każdorazowo określić logikę połączeń. Powyższe działania stanowiły punkt wyjścia do połączenia z pozostałymi danymi.

Do przetwarzania danych wykorzystałem narzędzie OpenRefine, aby normalizować informacje uzyskane za pomocą skryptu Python. Zadanie polegało tu na wyłuskaniu prawidłowości tabeli. Dzięki wbudowanym algorytmom określania podobieństwa nazw oraz opcjom wykonywania masowych operacji, wartości można było sprowadzić do jednego klucza nazewniczego (np. warianty nazw tego samego wydawnictwa lub miejsca wydania).

W danych Biblioteki Narodowej również trafiały się błędy i nieścisłości. Przykładowo w polu 651, dotyczącym nazw geograficznych, nie wiedzieć czemu znalazła się informacja „Sen w literaturze”, pasująca bardziej do motywów i tematów. Mimo specyfiki MARC 21, w którym każdy typ informacji powinien znaleźć dla siebie miejsce, powodowało to tym większy chaos, gdy natrafiałem na treści niepasujące do siebie. Normalizacja była niestety była konieczna również przez niespójne nazewnictwo. Dało się to np. zauważyć przy opracowaniu pola 700 jako stosowanie naprzemiennie pełnych słów i skrótów przy oznaczaniu współpracowników („tł.” vs „tłumacz” i podobne), czy niezrozumiałych początkowo oznaczeń („ps” dopiero w kontekście pojedynczego zapisu można z całą pewnością zakwalifikować jako wskazanie

autora posłowania). Bazy danych są zmienne w czasie i przygotowywane przez ludzi, a nieraz zasilane zewnętrznymi zbiorami o innych kryteriach opisu (mapowanie struktur jest mediacją podobną do tłumaczenia tekstu między systemami językowymi – szuka się podobieństw przy braku jednoznacznego ekwiwalentu), co też trzeba mieć tu na uwadze. Dodatkowo, zgodnie z powszechnie przyjętą praktyką w przygotowywaniu bibliografii, starałem się upraszczać słownik pojęć, tzn. sprowadzać nazwy synonimiczne do niezbędnego minimum. Przykładem może być rola współpracowników jako różnorodnie określanych autorów wyboru, redaktorów i autorów opracowania do jednej nazwy roli: „redaktor” (na podobnej zasadzie funkcjonują różne sposoby nazewnictwa paratekstów, jak przedmowy i posłowania, które w przygotowywaniu zbioru danych korzystniej jest sprowadzić do jednego określenia).

Zdecydowałem się także na zmiany konwencji zapisu, np. usuwając kropki na końcu informacji w polach MARC 21, które w nowym zbiorze danych nie niosłyby sensu i wprowadzałyby raczej szum informacyjny po przeniesieniu do nowego kontekstu. Niektóre informacje, zawarte w różnych podpolach, sprowadzałem do jednego pola w *datasecie* (charakterystyczny jest tu np. podział tytułu i innych części tytułu, jak podtytuł, na odrębne „pojemniki”, w których przechowuje je MARC 21). Tak było w przypadku pola 650, niosącego interesujące treści kontekstowe, gdzie szczegółowe rozbicie na składowe nie wydawało się konieczne: przykładowa wartość zunifikowana z sześciu podpól do jednego pola to „Opowiadanie argentyńskie 20 w.; Poezja argentyńska 20 w.; Szkice literackie argentyńskie 20 w.” – tutaj dla jednego identyfikatora dzieła, czyli jednego zapisu, trzy razy wystąpił wiersz z określeniem tematycznym (nazwa pospolita) z informacjami w dwóch kolumnach z czterech (określenia gatunku i chronologii-periodyzacji). W wypadku pola 650 arbitralnie przypisałem dla wielowartości kolejność od ogółu do szczegółu np. jako „Literatura; Pisarze argentyńscy 20 w.” – mimo, że eksport w pliku programu Excel swoją strukturą sugerował odwrotną kolejność <sup>412</sup>. Wspólnym kluczem było tu nazwisko twórcy, dodatkowo wzbogacone w *datasecie* numerem identyfikacyjnym z VIAF.

---

<sup>412</sup> Uwaga na marginesie: taki sposób kategoryzacji i tak wpędza badacza w wątpliwości, gdy natrafia na ciąg „Czas (filozofia); Wieczność; Metafora” – taka jest cecha zbioru danych związanych z Borgesem, dotyczących interesujących zagadnień w niecodziennych połączeniach, a punktem odniesienia mogą być kwestie fundamentalne.

MathFiction: baza akademicko-amatorska

Ostatnie źródło stanowi Mathematical Fiction (dalej używam również skróconej nazwy Math Fiction lub MathFiction, także funkcjonującej w internecie)<sup>413</sup>. Jest to projekt Alexa Kasmana z Wydziału Matematyki College of Charleston, mający za zadanie śledzić i badać obecność matematyki w literaturze.

Można też nazwać Math Fiction bazą tematyczną związków literatury z matematyką. Muszę przyznać, iż był to zaskakujący i dość przypadkowy wybór kolejnego materiału wsadowego do datasetu, ponieważ trzy źródła były już wystarczające dla stworzenia spójnej całości. Jesienią 2022 roku trafiłem jednak na tę stronę na drodze zjawiska, które w angielszczyźnie nazywamy *serendipity* - przypadkowe znalezisko, dające nam coś, czego potrzebujemy (choć mogliśmy o tym nie wiedzieć). Taki efekt przyniosły poszukiwania w sieci, związane z fikcją matematyczną i ujęciami idei matematycznych w literaturze<sup>414</sup>. Jeden utwór, *Flatland*, przez linkowanie hasła z Wikipedii naprowadził mnie na pojęcie *mathematical fiction*. Okazało się, że istnieje cała udokumentowana refleksja naukowa, popularnonaukowa i hobbystyczna na ten temat<sup>415</sup>. Nic w tym dziwnego, wszak z jednej strony matematyka wpływała wielokrotnie na literaturę, a niektóre gatunki, zwłaszcza poetyckie i meliczne formy, mogą przypominać problemy matematyczne.

Rozwój treści serwisu zapoczątkował sam autor, dodając pierwsze wpisy, ale umożliwił przesyłanie propozycji haseł osobom korzystającym ze strony. MathFiction rozwija się zatem zgodnie z mechanizmami, które opisałem przy okazji zjawisk *Web 2.0*, ze wszystkimi konsekwencjami. Społeczne tworzenie treści to jednocześnie wada i zaleta, ze względu na niespójne ujęcie i subiektywny, selektywny dobór informacji chociaż w tym wypadku mowa o entuzjastach tematu i bardzo wąskiej grupie odbiorczej, stąd można liczyć na pieczołowite opracowanie z pasją. Na uwagę zasługuje prawdziwie interdyscyplinarny charakter projektu i mimo, że serwis powstał zapewne głównie z myślą o osobach zainteresowanych matematyką,

---

<sup>413</sup> Alex Kasman, „Mathematical Fiction”. Mathematical Fiction. <https://kasmana.people.cofc.edu/MATHFICT/>. Dostęp 2023-03-15. W bibliografii załącznikowej podaję wyłącznie adres strony głównej.

<sup>414</sup> „Mathematical fiction”. W Wikipedia, b.d. [https://en.wikipedia.org/wiki/Mathematical\\_fiction](https://en.wikipedia.org/wiki/Mathematical_fiction). Dostęp 2023-09-16.

<sup>415</sup> Warto wymienić polski, skromniejszy akcent: Wrocławski Portal Matematyczny zawiera dział Matematyka Wokół Nas, a w nim także sekcję literacką, odwołującą się m.in. do twórczości Drózdza, Lema czy Tuwima. Jest to przede wszystkim zbiór cytatów, nieopatrzonych komentarzem, lub całych, tekstów – krótkich form, np. fraszek i haiku, związanych z matematycznymi tematami, postaciami matematyków i działami tej nauki. Interesujące może być dla badacza porównanie taksonomii polskiego i zagranicznego serwisu. MathFiction Na portalu znajdziemy informację, że ostatnia modyfikacja nastąpiła w 2010 roku. Zob. Wrocławski Portal Matematyczny. „Matematyka w literaturze”. <http://www.matematyka.wroc.pl/book/matematyka-wokol-nas/matematyka-w-literaturze>. Dostęp 2023-06-20.

to dla literaturoznawcy prezentuje odświeżające spojrzenie. Kasman pisze o swoim projekcie, niejako definiując jego misję, że matematycy powinni czytać fikcję literacką, dotyczącą matematyki, niezależnie od tego, czy im się podoba, czy nie – ponieważ owa literatura wpływa na społeczną percepcję ich dyscypliny<sup>416</sup>.

Obecnie baza liczy ponad 1500 wpisów (na stronie widnieje komunikat o osiągnięciu takiego kamienia milowego)<sup>417</sup>. Odwiedzający mogą przeglądać ją względem zaproponowanych kluczy kategoryzacji: medium (obok powieści, filmu, czy sztuki znajdziemy tu też... „dostępne za darmo online” oraz kolekcje), gatunku (np. horror, literatura dziecięca), motywu (m.in. prawdziwi matematycy, kosmici, podróże w czasie, Sherlock Holmes, geniusze, Newton, „matematyka jako chłodna/sucha/nieprzydatna”) i tematu (wśród nich nieskończoność, geometria/topologia/trygonometria czy logika/teoria zbiorów). Klasyfikacja zdaje się być żywa i pojawia się podejrzenie, iż zmieniała się wraz z przyrostem zawartości i nadal będzie się podlegać przemianom w miarę rozwoju. Na stronie znajdziemy także odnośniki do wszystkich wpisów i najnowszych z nich.

Przykładowy typ informacji, który możemy wywieść z tego źródła, przeglądając stronę, to wyjaśnienie, w jaki sposób poboczny wątek w *Zbrojnych*, jednej z części cyklu *Świat Dysku* Terry’ego Pratchetta odnosi się nie tylko do matematyki, ale także do matematycznych podstaw informatyki. Wspomniany epizod jest rozwinięciem wątku początkowej wrogości, a później pogodzenia dwóch członków straży miejskiej miasta Ankh-Morpork: trolla Detrytusa i krasnoluda Cuddy’ego. Z opisu uniwersum wiadomo, że te rasy są zaciekłymi wrogami. Pratchett kreujetrolle jako złożone z kamienia, a ich mózgi uformowane z krzemu. Oznacza to, że owe istoty sprawniej myślą (przetwarzają informacje), gdy jest chłodniej, a gorzej im to idzie przy wysokich temperaturach. Wiadomo też z treści, że ich sposób liczenia (w domyśle przedstawiony jako nieco prymitywny) opiera się na odliczaniu do dwóch. W opisywanej scenie troll zostaje przypadkiem zamknięty w chłodni. Zaczyna tam doznawać matematycznego „objawienia” i wypisuje na ścianach pomieszczenia spontaniczne obliczenia i wzory.

---

<sup>416</sup> Alex Kasman, „MathFiction – About this website”. Mathematical Fiction. <https://kasmana.people.cofc.edu/MATHFICTION/about.php> Dostęp 2023-03-15.

<sup>417</sup> Alex Kasman, „MathFiction - Browse”. Mathematical Fiction.. <https://kasmana.people.cofc.edu/MATHFICTION/browse.php>. Dostęp 2023-03-15.

Wpis na temat tej powieści, autorstwa Lapo Fanciullo <sup>418</sup> prezentuje cytaty istotne z perspektywy nauk ścisłych, i opatruje krótkim komentarzem, nazywając ten fragment „listem miłosnym do matematyki”. Gra w liczenie cegieł na ścianie odpowiada odkryciu kodu binarnego, a scena przedstawiona przy użyciu fantastycznego sztafażu i bohaterów w gruncie rzeczy streszcza rozwój informatyki jako dziedziny wiedzy wyrosłej z podstaw matematycznych.

Fanciullo referuje, że w fandomie serii zdania są podzielone co do tego, czy troll Detritus prawie odkrywa równanie leżące u podstaw całej matematyki, czy też znak równości na końcu jego zapisów ma być rezultatem, niezrozumiałym dla ludzkiego poznania (czy szerzej: żywych istot).

Whatever your interpretation, this remains a powerful section in the book: Cuddy and Detritus get so excited in their discovery of base two that they bond over the experience, and the subplot culminates in one of the best written praises of the beauty of maths. This is one of my favorite scenes in all of Discworld, and I love to go back and reread it<sup>419</sup>.

Jak widać, użytkownik dzieli się też przeżyciem estetycznym, związanym z odbiorem dzieła. Nieprzypadkowo pojawiają się tutaj przytoczenia z portalu. Po pierwsze, są one dowodem na zaangażowane katalogowanie informacji, związanych z fikcją matematyczną. Po drugie, w matematyce również funkcjonuje pojęcie piękna (oraz określenie eleganckiego rozwiązania), co stanowi estetyczny pomost do humanistyki. Obie dziedziny, matematyka i literatura, są pracą na systemie znaków, z ich różnymi wariantami. W matematyce poszukiwane jest rozwiązanie jakiegoś problemu, w literaturze – przekazanie sensu lub doznania estetycznego. W tym celu odbywa się proces. W matematyce przeprowadzenia dowodu, w literaturze prowadzenia narracji. Komentarz Fanciullo ukazuje estetyczną przyjemność odwzorowania pracy matematycznej w literaturze: tak jakby dzięki pięknu istniejącemu i rozumianemu w procesie matematycznym była bogatsza fabuła dzieła przynależącego do fikcji.

---

<sup>418</sup> Fanciullo, Lapo. „Men at Arms (1993)”. W *Mathematical Fiction*. Dostęp 15 marzec 2023. <https://kasmana.people.cofc.edu/MATHFICTION/mfview.php?callnumber=mf1415>.

<sup>419</sup> Fanciullo.



Dla literaturoznawcy fascynujące jest tutaj odniesienie się do świata przedstawionego i interpretacja zakładająca „pójście za tekstem”; wywiedzenie konsekwencji z reguł, które stworzył pisarz, czy też: odkrycie dokładnego zamierzenia autora i nazwanie zjawiska. Osoby zaznajomione z twórczością Pratchetta wiedzą, że cykl opisujący Świat Dysku jest bardzo spójną całością, kreującą fantastyczny świat, ale jednocześnie – pomimo wyraźnie satyrycznych zamierzeń, wobec gatunków fantastyki rzeczywistości współczesnej autorowi – w obrębie swoich zasad bardzo przewidywalny. Oprócz informatyki pojawiają się w tej serii elementy z realnego świata: problemy (równouprawienie), produkty (uliczne stoiska z małymi przekąskami, wynalazek ubezpieczenia, aparat fotograficzny) i postaci (np. Leonard z Quirmu, wzorowany na Leonardzie da Vinci czy krasnolud Hwel, będący odbiciem Szekspira). Anachronizm jest zasadą tych powieści, która pozwala zderzyć kategorie współczesne (XX- i XXI-wieczne) z tropami typowymi dla fantastyki.

Dodatkowo na uwagę zasługuje wyłuskanie matematycznego szczegółu z całej fabuły i zbioru innych wątków powieści. Jednocześnie wskazanie nań odbywa się bez sugerowania, że fragment odgrywa większą rolę niż w rzeczywistości: wyraźnie mamy wskazany wątek relacji między dwiema postaciami, wpisany w szerszą zasadę świata przedstawionego (animozje fikcyjnych ras).

Math Fiction cechuje z jednej strony folksonomia, czyli opis przez społeczność, z drugiej zaś pewne ramy edytorskie nadane przez Alexa Kasmana. To pierwsze oznacza, że ten sam wpis na temat *Zbrojnych* Pratchetta, gdyby przez innego był pisany użytkownika, inaczej by wyglądał. Problem wyrażony głosem Szekspirowskiej Julii w gorączce miłości nie traci aktualności pośród chłodu surowych danych: ta sama informacja może wybrzmieć inaczej, w zależności od autora wpisu. To drugie ma konsekwencje w postaci reguł tagowania działów i określenia zakresu treści, jakie można dodać (opis, cytat, podział pól, cechujących utwory). Dla przykładu, użytkownicy nie dodają multimediów<sup>420</sup>. Jak w każdym tego rodzaju projekcie, istnieją ograniczenia, wynikające z zastosowanej architektury i rozplanowania serwisu, które mają względną stałość i przewidywalność, a także treści generowane przez odwiedzających (*User-Generated Content*), które posiadają zindywidualizowany charakter.

---

<sup>420</sup> Wszelkie ograniczenia faktyczności i obszerności opisu wynikają z kompetencji, nakładu czasu lub innych uwarunkowań tworzenia wpisu przez użytkownika. Z kolei brak możliwości dodania multimediów (abstrahując od potrzeby – osobiście uważam ew. umieszczenie hiperłączy do stosownych serwisów za wystarczające) to ograniczenie systemowe, jakie oferuje strona, ponieważ nie przewidziano w niej takiego modułu.

Math Fiction ma kilka sposobów otwierających ją na świat zewnętrzny, ale wejście w ekosystem otwartych danych połączonych nie stanowi raczej jej głównego priorytetu. Strona oferuje linki afiliacyjne do jednej z największych księgarń internetowych na świecie. Pomaga to fundować projekt, a odbiorcom zapewnia skrócenie ścieżki nabycia własnych kopii opisywanych dzieł. Dostępna wyłącznie przez interfejs strony internetowej, pozwala na przeszukiwanie zawartości. Użytkownicy mogą również poza dodawaniem i sugerowaniem zmian w treści skontaktować się z twórcą i przekazać informację zwrotną. Strona posiada słownik działów matematyki i opisy związku danego dzieła z matematyką.

Na moją prośbę w listopadzie 2022, związaną z Borgesowskim *datasetem*, autor życzliwie przystał na podzielenie się plikiem do odczytania maszynowego. Stąd mogę zaktualizować liczby względem tego, co widoczne w interfejsie: otrzymałem listę dokładnie 1572 rekordów. Oprócz niej, osobno trzy słowniki: 9 rodzajów mediów, 27 motywów, a także 12 tematów. Wszystkie cztery pliki zapisano w formacie csv, który można dowolnie przetwarzać jako tabelę. Główny plik zawierał kolumny: tytuł, autorzy, rok publikacji, sygnatura<sup>421</sup>, medium, gatunek, motyw, tematy, średnia mathvote, liczba mathvote, średnia litvote i liczba litvote (podaję przetłumaczone nazwy polskie). Cztery ostatnie kolumny odnoszą się do ocen, jakie użytkownicy strony mogą wystawiać prezentowanym dziełom za ich wartość literacką (*litvote*), jak i za ukazanie matematyki w danym utworze (*matvote*).

Obróbka polegała tutaj na imporcie tych danych w programie MS Excel do pliku xlsx, przy użyciu narzędzi transformacji danych w trakcie importu do pozbycia się pustych wierszy. Zdecydowałem się na arkusz kalkulacyjny z uwagi na mały wolumen danych. Następnie, ze względu na użycie w liście rekordów skrótów, znajdujących odwołania w plikach ze słownikami, posłużyłem się makrem MS Excel, aby je rozkodować<sup>422</sup>. Makro rozpoznaje wiele wartości w jednej komórce pliku i podstawia tekst ze wskazanej tabeli do substytucji. W ten sposób dla *Powrotu z gwiazd* Stanisława Lema, posiadającego w kolumnie „motif” wartość „fie”, mogłem podstawić w osobnej kolumnie tekst pola motyw: „Math as Beautiful/Exciting/Useful”, zgodnie z odpowiadającą jej pozycją. Działa to także dla wielu wartości (stąd konieczność użycia dodatkowego makra, a nie już dostępnych w programie formuł substytucji lub wyszukania pionowo – w zupełności wystarczających do podmiany

---

<sup>421</sup> Sygnatura jest widoczna także w adresach (URL) podstron serwisu, poświęconych konkretnym dziełom.

<sup>422</sup> MacDougall, John. „How To Find And Replace Multiple Text Strings Within A Text String”. How to Excel. Dostęp 2023-03-15. <https://www.howtoexcel.org/how-to-find-and-replace-multiple-text-strings-within-a-text-string/>.

na zasadzie 1:1), np. *Złoty żuk* Edgara Allana Poe został oznaczony w polu „temat” jako „cc+gtt”, co sprowadziłem do wartości „Computers/Cryptography, Geometry/Topology/Trigonometry”. Jak widać na tym przykładzie, kolejną istotną transformacją była zamiana ciągów znaków bez spacji ze znakiem dodawania jako separatorem tagów na przecinek i spację, dla oddzielenia kolejnych charakterystyk. Mając tabelę odwoławczą, można również dokonywać tłumaczeń zbioru danych na inny język. Kody poszczególnych tagów można podstawić do tabeli z nazwami przetłumaczonymi.

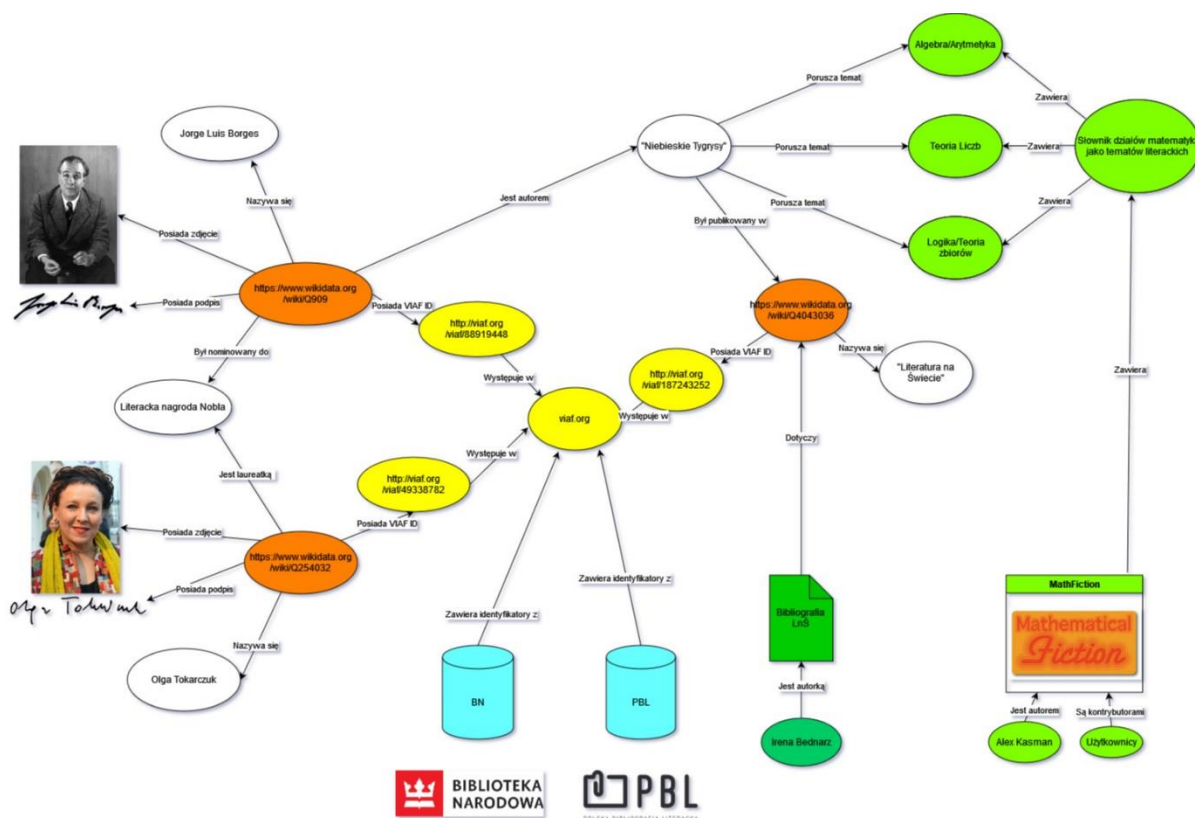
Znów nie obyło się bez mediowania między poziomem przekształceń automatycznych a pracą manualną, gdy podmiana znaków według słownika uwzględniła ten sam ciąg znaków w dwóch tagach bazy: m.in. dla *Podróży Gulliwera* Johnatana Swifta w polu „motyw” stworzyło to potwora w postaci „AntisociAlien Mathematicians” zamiast „Antisocial Mathematicians”. Tak samo rzecz się miała w przypadku przywołanego tematu „Computers/Cryptography”, gdzie „mp” oznaczało również „Mathematical Physics”; pierwsze przekształcenie dało wynik: „CoMathematical Physicsuters/Cryptography”. Wystarczyło jednak zamienić masowo te powtarzalne pomyłki w obrębie całej kolumny, a łatwo dało się wysledzić ich regularność. Skłania to przy okazji do refleksji nad operatywnością używania tego rodzaju kluczy – istniałoby małe ryzyko błędu, gdyby media, gatunki, motywy i tematy posiadały kody alfanumeryczne zamiast alfabetycznych skrótów <sup>423</sup>. Oczywiście w grę wchodzi społecznościowa baza hobbystyczna, której nie projektuje się z myślą o przechowywaniu i deduplikacji *big data*, ani masowym przetwarzaniu danych. Stąd nie ma takiej potrzeby, aby „przeprojektować” jej system ponad miarę, tzn. za bardzo rozbudowywać metodologię opisu wobec dostępnej treści. Istotnym czynnikiem jest także tekstowy charakter zawartości MathFiction i brak uniwersalnych identyfikatorów. Autorzy i tytuły są oznaczani ciągiem znaków (*string*) i nie mamy tutaj np. ID VIAF, Wiki Data czy numerów ISBN, aby oprzeć na nich identyfikację. Dla dostępnych mi danych byłem w stanie je jednak wzbogacić o identyfikatory kontroli autorytatywnej osób tam, gdzie to było potrzebne. Dowodzi to wzajemnego wpływu źródeł na siebie: każde z nich coś wnosi do wspólnego zbioru i jednocześnie może zostać wzbogacone przez pozostałe.

---

<sup>423</sup> Były także inne metody rozwiązania tego problemu zamiast makra do zamiany tekstu, np. rozbicie wszystkich wylistowanych wartości do osobnych kolumn i podmiana względem słownika wartości tak, aby w każdej kolumnie znajdowała się jedna wartość.

Wizja połączonego zbioru danych

Ostateczny *dataset* tworzą rekordy bibliografii LnŚ, BN i PBL, wzbogacone o metadane dotyczące tematyki utworów, pochodzące z MathFiction. Przedstawiam graf wiedzy, z podziałem źródeł i wskazaniem wzajemnych powiązań między nimi, aby zobrazować ścieżkę koncepcji punktu dojścia w opracowywaniu całościowego *datasetu*<sup>424</sup>.



Ryc. 6 Koncepcyjny graf wiedzy: możliwe powiązania między danymi

Na ilustracji u dołu znajdują się źródła (zbiory i informacje oznaczone na niebiesko oraz zielono). Umownym punktem dostępu do danych o osobie lub tytule jest tutaj identyfikator Wikidata (kolor pomarańczowy), a niezbędne dla połączenia wszystkich informacji gwarantują identyfikatory kontroli autorytatywnej VIAF (kolor żółty). Elementy są powiązane ze sobą na zasadzie RDF (*Resource Description Framework*), czyli „trójek” ontologii: podmiot-predykat-obiekt, tak, aby można było stwierdzić między nimi powiązania, np. „Irena Bednarz jest autorką Bibliografii LnŚ” albo „[identyfikator] <https://www.wikidata.org/wiki/Q909> nazywa się [oznacza osobę] Jorge Luis Borges”.

<sup>424</sup> Na etapie opracowywania połączonego *datasetu* koncepcje pracy i wizualizacja w postaci grafu wiedzy zostały zaprezentowane przeze mnie na konferencji WIAD23, organizowanej 5 marca 2023 online przez UMK w Toruniu z okazji Światowego Dnia Architektury Informacji. Referat pt. „Architektura informacji a dane badawcze: studium przypadku czasopisma *Literatura na Świecie* w »Polskiej Bibliografii Literackiej«”.

Identyfikację stałych odniesień umożliwia nadanie struktury schema (Schema.org) – odwołanie do konkretnych pól schematu metadanych np. oznaczeń elementów bibliografii. Można do tego celu użyć także systematyki Dublin Core. Jako punkt końcowy pod względem formatu określiłem na początku zapisanie danych jako JSON, konwersja z *dataframe* do *datasetu* możliwego do włączenia do szerszej całości. Bazowo satysfakcjonujący jest już jednak plik csv, ze względu na możliwość przetwarzania bez konieczności użycia dedykowanego oprogramowania.

Istotny był wybór metody udostępnienia badań, dla stałego dostępu. Zgodnie z przedstawionymi wcześniej rekomendacjami najlepsze okazało się tutaj repozytorium dostępne w sieci. Zapewnia to możliwość dotarcia bez przeszkód i spełniania dodatkowych warunków dostępu. Spośród otwartych repozytoriów najlepsze dla tego celu jest Zenodo, z którego korzysta wiele osób prowadzących badania humanistyki cyfrowej.

Plan działania dla prezentacji danych obejmował trzy kroki: 1) Publikacja surowych danych: stan oryginalny (osobne pliki do ponownego wykorzystania); 2) Publikacja zbiorcza w formacie tabelarycznym jako dane połączone (zawierająca określenia własne ale też identyfikatory baz ogólne i uniwersalne ID jak VIAF; nagłówki kolumn odwołują się także do znaczników Schema czyli schema.org, co zapewnia możliwość odniesienia do sieci semantycznej); 3) Otwarte dane połączone prezentowane jako ontologia – preferowany format JSON. W tym kroku w wartościach tabeli zamiast ciągu znaków (*string*), czyli formy prezentacji danych, której się nie preferuje do odczytu maszynowego, można użyć kodów, oznaczających te same informacje, ale w uporządkowany sposób (np. identyfikatory Wikidata dla gatunków piśmiennictwa itp.).

Wyzwania kompilacji źródeł

Punkt odniesienia stanowił zbiór LnŚ, który można uznać za kompletny i wzorcowy. Warto w tym miejscu nadmienić raz jeszcze, że to wycinek bibliografii o dużo większym potencjale (spośród wszystkich literatur narodowych wybrałem jedną, a spośród nich jednego autora). Otwiera to pole do dalszych poszukiwań. Ramy pracy doktorskiej nieubłaganie nałożyły jednak ograniczenie związane z kontrolą jakości *datasetu*: posłużenie się większym zbiorem naraziłoby na ryzyko zbyt wielu błędów i nieścisłości.

Wyszedłem od bibliografii autorstwa Bednarz i dobudowałem do niej informacje z PBL. Zaskoczenie stanowiła niekompletność i względnie mała część wspólna zbiorów. Przyczyn upatruję w luce opisu PBL (zakres czasowy przypada na roczniki niezdigitalizowane,

niedostępne do pobrania z bazy, ok. 1971-1989). Ma to dobrą i złą stronę. Wadą tej sytuacji jest brak możliwości porównania i agregacji pomiędzy źródłami. Pozytywny aspekt polega na tym, że to z połączenia wszystkich źródeł powstaje wartość dodana. Każda składowa wnosi coś do zbioru i buduje całość. Widzę tutaj pole do dalszych aktywności, kiedy roczniki wcześniejsze pisma staną się dostępne do pobrania z bazy Oracle.

Niektóre aspekty zupełnie umykają uwadze przy obróbce informacji. Z rozbawieniem zauważyłem, że podczas przetwarzania bibliografii LnŚ na tabelę nie potrzebowałem posługiwać się nazwą czasopisma. Występuje ona z kolei w BN i PBL, różnicując pozostałe źródła. Wspominałem o literówkach w obszernym zasobie opracowanym przez Bednarz, ale poprawki działają też w drugą stronę. W jednym z tekstów (*Czy Borges stworzył „szkołę”?*) doszukałem się niepoprawnego zapisu nazwiska tłumacza w PBL: „Igor Komornicki” zamiast „Igor Komarnicki”, czyli wersją jaka występuje w przewodniku LnŚ i potwierdza ją VIAF, wraz z kwerendą internetową.

Inny był poziom granulacji danych, tzn. Bednarz rejestruje publikacje tak, jak redakcja LnŚ podaje pozycje w spisie treści. Z kolei dla osób tworzących PBL podstawową jednostką opisu stanowi utwór. Miało to znaczenie w przypadku grup wierszy – jak wspominałem, z przewodnika LnŚ trudno rozbić zapisy zbiorcze, nie znając numeracji stron (tzn. bez badania z autopsji). Sposób opisu PBL daje taką możliwość, ponieważ taka jest przyjęta metoda bibliografów i bibliografek PBB: w opisywanym przypadku adresy na stronach publikacji są podane z dokładnością do pojedynczego utworu z grupy. Dla takich utworów przyjęto również inne konwencje zapisu: PBL grupuje w jednym rekordzie z nr 5/6 z 1995 r. wszystkie utwory *Tanka*, a Bednarz podaje dodatkowo ich *incipity*.

Daje to wyobrażenie doboru metody bibliograficznej do opracowania materiału i kolejny raz wskazuje na charakterystykę pracy z *Bibliodata*: przedmiot badań tylko pozornie jest obiektywny. Badacz ma spore pole manewru, jeśli chodzi o kryteria opisu i wybory w obrębie rejestracji faktów bibliograficznych.

Dla LnŚ jako tytułu ujednoliciłem także nazwy wydawcy i miejsce wydania, zgodnie z materiałami na temat pisma. Kolejne wydawnictwa to: Wydawnictwo Współczesne (RSW „Prasa-Książka-Ruch”, do 1990), Agencja Autorska (1990-1994, na zlecenie Ministerstwa Kultury i Sztuki), Biblioteka Narodowa (1994-2010), Instytut Książki (2010-obecnie). Zakresy nie są łatwe do ustalenia, zatem punkt zmiany wydawcy względem roku wydania również jest

arbitralny przy niepełnej informacji (w zbiorze występuje jeden rekord z 2010 opisany jako wydany przez Instytut Książki).

Obróbka MathFiction dostarczyła słownik tematu matematycznego, natomiast zakres tego przedsięwzięcia nie pozwolił na znaczące pokrycie względem danych z pozostałych zbiorów. Obecny zbiór danych zostanie poszerzony. W pierwszej iteracji z PBL zostały dodane rekordy LnŚ – w kolejnej zbiór poszerzą informacje o publikacjach w innych czasopismach, po harmonizacji całości. Pojawią się również o dane BN dotyczące Borgesa jako tematu dzieł, dla dopełnienia obrazu recepcji twórcy w Polsce. Wymaga to powtórzenia analizy MARC 21, jak to miało miejsce z publikacjami autorskimi dzieł twórcy, co będzie rozwinięciem efektu niniejszej pracy.

#### Charakterystyka *datasetu*

Gotowy zbiór obejmuje następujące dane: autor, tytuł, typ publikacji (książka lub czasopism), miejsce i rok wydania, dla tekstów z czasopism: numery bieżące i ciągłe oraz strony, na których znajdują się teksty. Odnotowani są tłumacze i inni współpracownicy. Zawsze jeden wiersz tabeli odpowiada jednemu utworowi, a jedna kolumna przekazuje informację jednego typu (lub inaczej: wyraża dane na konkretny temat). Przynależność do literatury argentyńskiej i narodowość pisarza zaznaczam w odpowiednich kolumnach, używając zarówno oznaczeń słownych, jak i kodowania. Jest to wyrażone za pomocą konwencji przyjętych w standardzie ISO 3166-1, gdzie kraje oznacza się dwuliterowymi skrótami. Tak samo sformatowana treść występuje w przypadku wskazania kierunku tłumaczenia dla części rekordów, gdzie ta informacja była dostępna (język publikacji i język oryginału). Język publikacji odnosi się do Schema.org.

Tam, gdzie były dostępne, pojawiają się adnotacje bibliografów i bibliografek PBL, a także adnotacje z pola 500 MARC 21, pochodzące z BN. Uwzględniłem je, ponieważ mimo formy ciągu znaków przekazują autentyczną informację od osób badających fizyczne egzemplarze i odnoszą się bezpośrednio do treści. To z adnotacji PBL dowiadujemy się o przedruku wiersza pt. *Księżyc* ze sprostowaniem błędów redakcji. Mamy też okazję wglądu w metodykę badawczą, np. kiedy bibliografowie BN umieszczają informacje o skonsultowaniu danych bibliograficznych ze stroną internetową wydawcy dzieła. Nie są to dane przeznaczone do odczytań maszynowych, ale upatruję ich istotności w budowaniu narracji na temat losów wydania tekstów literackich. Odwołuję się także do gatunków i zakresu tematycznego dzieł. Z Math Fiction pochodzi odniesienie do działów matematyki.

Dane są przystosowane do tworzenia ontologii poprzez łączenie ich z innymi zbiorami. Zastosowana struktura odwołuje się do metadanych Schema.org, zbioru oznaczeń powszechnie stosowanego dla publikacji na cele otwartych danych połączonych. Nagłówki kolumn przybierają postać np. „schema:author” dla oznaczenia autora pracy. Dzięki temu możliwe jest przetwarzanie i ponowne użycie tych informacji.

Niektóre nagłówki kolumn stanowią odstępstwo od konwencji nazewnictwa: przewidziano miejsce dla maksymalnie czwórki współpracowników. W tym wypadku nazewnictwo z nimi związane uwzględnia także numer porządkowy, aby mitygować ryzyka przy przetwarzaniu tych danych: „schema:contributor(1)” odróżnia się od „schema:contributor(2)”. Udostępnienie danych, w których występuje więcej niż jedna kolumna o tej samej nazwie byłoby kardynalnym błędem, dlatego zdecydowałem się na rozwiązanie nieeleganckie, ale poprawne w odniesieniu do celu. Stałe identyfikatory osób, pochodzące z VIAF, wyrażono jako linki (URL), zatem treść *datasetu* odsyła bezpośrednio do zasobu.

Zachowana zostaje unikatowość rekordów: dane są zdeduplikowane, informacja o jednym dziele nie powtarza się w zbiorze jako więcej niż jeden rekord. Występują identyfikatory indywidualnych źródeł, czyli ID PBL (wewnętrzny identyfikator zapisu) i BN (wartość pola 001 w MARC 21). Dla kopii bibliografii Bednarz w formie tabeli zachowane zostają wygenerowane wcześniej identyfikatory, by zachować ciągłość operacji na danych. W wyjściowym pliku podstawowej tabeli, służącej za podstawę docelowego csv, zastosowano też osobną numerację opartą na kolejności dodawania *datasetów* cząstkowych.

Podaję też przy rekordach oznaczenie proveniencji (aby śledzić występowanie tych samych dzieł w zbiorach częściowych, przecięcia między nimi i miejsca osobne). Realizuje się to poprzez użycie kolumn z unikatowymi identyfikatorami zasobów częściowych. W osobnej kolumnie jako „Dataset\_ID” jest także identyfikator w formie liczby porządkowej rekordu. Na całym zbiorze stale prowadziłem walidację duplikatów, aby zachować unikatowość zapisów. Ze względu na wspomniane różnice, czyli mapowanie jednego zapisu LnŚ do wielu w bibliografii Bednarz, w kilku przypadkach było konieczne dodanie oznaczeń w formie liczb porządkowych, np. „\_1” dla wyrażenia tej informacji.

Numeracja dla rozróżnienia pól i numeracja identyfikatorów drugiego stopnia mogły być rozwiązane na inne sposoby, przykładowo poprzez dodanie wierszy i odziedziczenie tej samej informacji. Wtedy jeden rekord byłby reprezentowany przez więcej niż jeden wiersz. Tak jak wspominałem, przyjąłem jednak prymat unikatowości materiału bibliograficznego,



co wpływa na ścisły kształt tabeli. Na tym oparłem również metodę deduplikacji. Arkusz kalkulacyjny to narzędzie dwuwymiarowe, dlatego przyjmując to początkowe założenie, można było rozbudowywać tabelę jedynie „w bok”, ale nie „w dół”.

Kierowałem się tutaj zbiorem zasad FAIR, myśląc także o przyszłości zbioru: jego rozwoju w większą całość, możliwości oddania do dyspozycji naukowcom zajmującym się *Bibliographic Data Science* jako danych badawczych. Zbiór jest w zamierzeniu czytelnie wersjonowany. Dostęp do *datasetu* gwarantuje „adres” w sieci, czyli stały identyfikator DOI<sup>425</sup>. W chwili publikacji stał się otwartymi danymi bibliograficznymi, przeznaczonymi dla celów badawczych. W kolejnej iteracji przewidziano budowę ontologii na ich bazie, z użyciem przetwarzania danych zapisanych w formacie csv do struktury sieci semantycznej jako graf wiedzy, wykorzystujące struktury RDF. W tym celu zawiązuje się współpraca z zespołem z Wydziału Elektroniki Politechniki Wrocławskiej, dysponującym narzędziem technicznym przeznaczonym do wsparcia publikowania danych tabelarycznych w sieci Web. Narzędzie nie zostało jeszcze upublicznione (stan na październik 2023) i działa w sieci wewnętrznej, a powstało w ramach inicjatywy Dariah.lab rozwijającej cyfrową infrastrukturę badawczą dla humanistyki i nauki o sztuce.

### Obserwacje

Intencją powstania *datasetu* Borgesowskiego było przede wszystkim oddanie go w ręce innych badaczy, to znaczy wypracowanie produktu zdatnego do użycia i przetwarzania. Postaram się poniżej streścić wstępne spostrzeżenia, które odnotowałem w kontakcie z tymi danymi, wzorem badaczy *Bibliodata* i BTH.

Zbiór liczy w momencie publikacji 357 rekordów, na które składają się książki i twórczość Borgesa wydawana w czasopiśmie, jak i artykuły oraz opracowania w czasopiśmie. Kilka pozycji, jak wspomniana już fotografia czy materiały pochodzące z archiwów ludzi kultury, w tym np. tłumaczenia wierszy Borgesa i innych autorów z archiwum Zbigniewa Herberta.

Dzięki połączeniu źródeł zgromadzono zapisy z lat 1971-2021 (bibliografia Bednarz miała zasięg czasowy 1971-2014, PBL odnotowuje lata 1989-2013, a pozostałe informacje uzupełnia BN. 278 zapisów posiada typ; 79 nie posiada przypisanego typu w źródłach, z których zostały pobrane i wymagają dalszej obróbki ręcznej. Rozkład materiałów pod względem typu to:

---

<sup>425</sup> Jakub Eichler, „Works of Jorge Luis Borges in Polish Bibliographic Sources”, 2023-11-05, <https://doi.org//zenodo.10073461>.

171 materiałów z czasopism, 102 książki, 62 artykuły, 2 e-booki, oraz po jednym oznaczeniu: nuty, poezja, proza (poezja to wspomniane materiały archiwalne Herberta, proza – prawdopodobny błąd zapisu).

Wstępne dane na temat działalności tłumaczy mogą zobrazować, kto najczęściej tłumaczył dzieła Borgesa i związane z nim teksty krytyczne. Arbitralnie przyjąłem tu kryterium powyżej 5 przetłumaczonych prac. Największa produkcja przekładowa przedstawia się następująco: Sobol-Jurczykowski, Andrzej: 87, Chądzyńska, Zofia: 45, Elbanowski, Adam: 14, Engelking, Leszek: 13; Ziętara, Maciej: 9, Stachura, Edward: 8; Walasek-Elbanowska, Dorota: 7, Partyka, J.: 6, Kalicki, Rajmund: 5, Rodowska, Krystyna: 5. W miarę rozwoju datasetu te wartości mogą ulec zmianie, a tego typu analiza frekwencyjna powinna jeszcze ulec pogłębieniu, np. z uwzględnieniem, czy to samo tłumaczenie nie było wznawiane lub opublikowane w różnych tytułach czy antologizowane jako reedycja (można w takim wypadku rozważyć, czy liczy się jednostkowe tłumaczenie, czy wznowienia także powinny się zliczać do sumy produkcji translatorskiej; dodatkowo, czy np. uwzględniać sytuacje edycji i wznowień poprawionych, o ile istnieje na ten temat ślad w danych; wkraczamy tutaj w obszar określania adekwatności danych i dobierania kryterium opisu).

Przy porównaniu rekordów pochodzących z czasopism, nie można jeszcze przesądzać na podstawie analizy frekwencji. Wstępne porównanie, przy uwzględnieniu około 400 dodatkowych rekordów z PBL, dotyczących innych czasopism niż LnŚ, ukazuje następujące sumy publikacji za okres 1971-2022 (przy odcięciu tytułów z 10 i mniej publikacjami): *Literatura na Świecie*: 109, *Odra*: 59, *Przegląd Filozoficzno-Literacki*: 38, *Literatura*: 22, *Tytuł*: 19, *Twórczość*: 17, *Wiadomości Kulturalne*: 15, *Kwartalnik Artystyczny*: 14, *Tygodnik Powszechny*: 13, *Więź*: 12. Można zauważyć czasopisma o długiej tradycji wydawniczej jako te, które autora Fikcji publikowały częściej od innych. Są to jednak wstępne, szacunkowe dane, przed połączeniem z dodatkowym źródłem danych.

Perspektywy i kolejne kroki

Przeznaczenie *datasetu* to umożliwienie badań innym, a powyższe ustalenia opisują prototypowy przypadek testowego użycia powszechnie stosowanych metod (*proof of concept*). Co daje *dataset*? Na pewno jest to automatyzacja ścieżki, którą przeszedłem i nadal przechodzę teraz, by wydobyć informacje. Jako literaturoznawca, musiałem rozwinąć kompetencje cyfrowe, aby wytworzyć ten produkt. Z efektu można korzystać bez tych umiejętności lub niewielkim wysiłkiem.

Udało się przede wszystkim przetworzyć treść pliku PDF z przewodnikiem LnŚ, możliwego wcześniej do odczytania tylko przez człowieka, na formę tabelaryczną i przystosowaną do użycia ontologii, a zatem nadającą się do przetwarzania maszynowego, co można uznać za sukces remediacyjny. Z pewnością polem rozwoju jest rozszerzenie zakresu tego działania, to znaczy wejście na skalę literatur narodowych, języków publikacji i wreszcie całego badanego zbioru. Podobnie w przypadku PBL i BN, istnieje potencjał na śmieiej zakrojone narracje związane z danymi, a pilotażowy charakter moich działań (uczących pokory i studzenia zapału) nie tylko budzi więcej badawczego apetytu, ale mam nadzieję, że okaże się pomocny dla innych humanistów, szczególnie stawiających pierwsze kroki w obszarze humanistyki cyfrowej i poszukujących danych treningowych.

Projekt łączy źródła powstałe z pasji bibliografii (Ireny Bednarz) oraz społecznościowego katalogowania tematycznego (Kasman i użytkownicy Math Fiction), jak i zachowujące dziedzictwo krajowego piśmiennictwa i realizujące misję udostępniania informacji naukowej oraz kulturalnej (PBL i BN). Bez takiej „danowej interwencji” podobna kompilacja by nie mogła powstać.

Po dodaniu częściowego datasetu BN dotyczącego recepcji (Borges jako temat dzieła), obraz tej mozaiki różnorodnych źródeł stanie się kompletniejszy i pozwoli na formułowanie narracji na podstawie danych co mam nadzieję zaprezentować w formie odrębnej publikacji, zmierzającej ku własnej próbie zastosowania narzędzi BTH (wypada zaznaczyć, że wolumen danych nie zasługuje jeszcze na B w nazwie, możemy mówić o *little data* i szczegółowym zbiorze tematycznym). Porównanie literatury krytycznej i publikacji książkowych da asumpt do refleksji nad tym, jak Borges wszedł do systemu literatury polskiej: skutek tłumaczeń twórczości oryginalnej czy jej omówień? Czy w przypadku tekstów autorskich były to od razu publikacje książkowe, czy wyprzedziły je fragmenty lub krótkie formy w czasopismach? Wstępne rozpoznanie, wynikające z dostępnych informacji, wskazuje na przewagę produkcji czasopiśmienniczej, ale jestem zmuszony wstrzymać się z kategorycznymi wnioskami do czasu uzyskania pełniejszej perspektywy. Liczę na to, że uda się w miarę rozbudowy kolejnych wersji datasetu odpowiedzieć na pytanie jak proces przyswajania przez polską kulturę pisarstwa autora *Alefa* pokazują dane bibliograficzne.

Już obecny zestaw danych może być natomiast przydatny dla osób, które chcą formułować narracje o literaturze na podstawie danych bibliograficznych. Widzę w obecnie spreparowanej informacji punkt wyjścia do badań w obszarach: a) korpusy i bazy bibliograficzne; b) recepcja

twórczości autorów; c) historia przekładu d) aktywność tłumaczy e) wpływy jednych literatur na inne f) wizualizacje geograficzne.

Inny wątek inspirowane użycie linków afiliacyjnych do księgarni internetowej przez Kasmana w MathFiction. Nie były to dostępne dla mnie dane, ale mając taką informację można też śledzić rynkowe losy publikacji, wchodzących w obręb *datasetu*. Są to zagadnienia przesuwające obszar bardziej w stronę relacji literaturoznawstwa z ekonomią i socjologią kultury.

## Podsumowanie

Ostateczny zbiór składa się z trzech głównych źródeł bibliograficznych i jednego dodatkowego, w postaci bazy fikcji matematycznej. Praca przebiegała w kolistym ruchu między „dużymi” i „małymi” danymi, tzn. od całości baz bibliograficznych i pełnotekstowej zawartości bibliografii Bednarz, do poszczególnych zapisów. Efektem jest ucyfrowienie wycinka tego zbioru w kontekście przenikania literatur obcych. Mamy do czynienia z pewnym wglądem w praktyki zastosowania i interpretacji danych. Mimo niewielkiej próbki w tym zbiorze już drzemią nowe hipotezy. Było to także ćwiczenie ukazujące różne sposoby przygotowania danych dotyczących tego samego przedmiotu, co wnosi walor porównawczy względem każdego ze źródeł.

Aby uniknąć pułapek braku obiektywizmu – i pozornego obiektywizmu – starałem się tutaj jak najdokładniej przedłożyć dokumentację badań. Dane są ogólnodostępne w internecie, a proces opracowania transparentnie oddany w ręce czytelników.

Publikacja jest hostowana na Zenodo jako otwarty projekt, który ma możliwość wersjonowania<sup>426</sup>, o stałym i otwartym dostępie dla wszystkich zainteresowanych, w tym niezalogowanych użytkowników. Gotowy *dataset* posiada dokumentację w języku angielskim w formie planu zarządzania danymi badawczymi (Data Management Plan).

Udostępnienie tych danych spełnia funkcję zasilenia większych źródeł do masowego przetwarzania (tak jak to zostało przedstawione m. in. w kontekście *Distant reading* i makroanaliz), a jednocześnie emancypacyjną kulturowo, tzn. swego rodzaju „wstrzyknięcia” polskiego *datasetu* do ekosystemu danych światowych, zgodnie z proponowanym podejściem BTH. Działanie to potrzebne, choć mam świadomość, jak wiele jeszcze można zrobić na tym

---

<sup>426</sup> Dzięki temu pozostawiam przestrzeń dla rozbudowy takiego zbioru w przyszłości, aby *Proof of Concept* mógł się stać projektem o większej skali.

polu. Po projektowaniu na wstępie prac właśnie auterek opracowania dotyczącego BTH jako person badawczych czy odbiorczyń *datasetu*, sądzę, że prezentowane w nim informacje z powodzeniem mogłyby zostać użyte w takich analizach, jakie prowadzą.

# Zakończenie

W semestrze letnim 2022/23 władze WFPiK UAM, wspaniałomyślnie i ryzykownie, zezwoliły mi na poprowadzenie kursu dla studentów kierunku Filologia Polska i innych, które się studiują na tym Wydziale (Filmoznawstwo, Groznawstwo). W zamierzeniu te zajęcia miały dać osobom z pierwszych roczników szansę na rozmowy bliżej związane z ich zainteresowaniami kulturą współczesną: dać wytchnienie od historycznych kanonów i surowych prawideł języka. Pozwolić na swobodniejszą refleksję i wypowiedź. W kursie *Współczesna kultura literacka - dyskusje, interpretacje, nawiązania*, toczyliśmy debaty na temat „profesjonalnej” lektury tekstów literackich, współczesnej literatury gatunków i najżywotniejszych konwencji fabularnych, adaptacji literatury na inne media, odróżnianiu przekazu reklamowego od krytycznego, cechach dobrej recenzji, czy zjawisku cenzury oraz tego, jak rozumiemy ją dzisiaj. Jednym z moich pomysłów na aktywizowanie grupy było stworzenie wspólnego pliku z „bazą danych” o wydawnictwach, nagrodach i imprezach literackich: w jakich miastach się znajdują lub odbywają, jaka instytucja nimi zawiaduje; kto ostatnio został nagrodzony; gdzie znaleźć ich strony w internecie. Część z tych danych wizualizowaliśmy później na mapie (śmiałem się, że może to pomóc uczestnikom zaplanować wakacyjne wyjazdy). W jednej z rozmów na zajęciach o życiu literackim, pokazując grupie przewodnik i kalendarium literatury i krytyki dla okresu od lat 70. XX w. do przełomu wieków<sup>427</sup>, pytałem studentów, jak powinniśmy opisywać kulturę dzisiaj, aby osiągnąć podobny efekt dokumentalistyczny. „Jak dziś dokumentować kulturę?” – był to jeden z tych momentów (odmiennie do proponowania haseł jak „Wiedźmin”, „Gry wideo” czy „RPG”), kiedy na sali zapadła cisza. W tej charakterystycznej ciszy można przemyśleć jakiś cytat, np. taki (wyróżnienia moje – JE):

**Sporządzony przez nas układ danych** – zarówno w obrębie jednego roku, jak i w zestawach porównawczych – dostarcza wiedzy o głębokich prawidłowościach układu literackiego. Mówi więc o pisarzach stwarzanych przez państwo ideologiczne (co uwidacznia się na listach bestsellerów za lata siedemdziesiąte i osiemdziesiąte), dzisiaj zaś kreowanych w błyskawicznym tempie przez środki masowego przekazu (co pokazują listy z lat dziewięćdziesiątych);

---

<sup>427</sup> Chodzi o: Przemysław Czapliński i Piotr Śliwiński, *Literatura polska 1976-1998: przewodnik po prozie i poezji*, Wyd. 1., dodruk (Kraków: Wydawn. Literackie, 1999). Przemysław Czapliński, red., *Kalendarium życia literackiego 1976-2000: wydarzenia, dyskusje, bilanse*, Wyd. 1 (Kraków: Wydawn. Literackie, 2003).

o wydolności rynku wydawniczego (dopiero największe bestsellery pokazują rzeczywistą chłonność naszego rynku); o rosnącej liczbie tytułów i malejącej liczbie łącznych nakładów; o tym, że u schyłku lat dziewięćdziesiątych Polska stała się krajem poezji (...) i o tym, że pod względem liczby książek przypadających na jednego mieszkańca zajmujemy jedno z ostatnich miejsc w Europie. **Informacje te, rozmaicie zestawiane, układają się we frapujące opowieści, których odczytania podjąć się może każdy** – odbiorca przygodny, pasjonat, zainteresowany dziejami naszego czytelnictwa, student, nauczyciel. Tkwią jednak w tych informacjach opowieści dla innych jeszcze odbiorców – dla socjologów rynku książki, krytyków literackich, a także dla historyków preferencji lekturowych – tych preferencji, które pomagają Sienkiewiczowi trwać na listach bestsellerów mimo zmieniających się ustrojów (...) Kalendarium nie powstałoby bez pomocy wielu osób (...) dziękujemy (...) osobom, które pożyczały nam brakujące numery i roczniki pism, nie pytając, kiedy oddamy; znajomym rozsianym po kraju, którzy odpowiadając na nasze rozpaczliwe apele, cierpliwie kopiowali strony z rozmaitych czasopism, później zaś ślali je do nas szybką pocztą...<sup>428</sup>

Studenci intuicyjnie rozumieją odpowiedzi na pytania, ale często aż wstyd mówić coś oczywistego. W takich sytuacjach znajduje się w końcu jakaś życzliwa dusza, która ratuje prowadzącego odpowiedzią i na tendencyjne pytanie odpowie, że „kulturę należy dziś dokumentować za pomocą baz danych”. Trochę naiwny gest początkującego nauczyciela uzmysławia mi, że do życia zawodowego weszli już dawno *digital natives* i dla dzisiejszego studenta świat „nowych” mediów jak rozumiał je Manovich jest po prostu światem mediów. Jeszcze nie umiałem wyrazić tej myśli wtedy, ale wydaje się, że statyczna, „bazodanowa” humanistyka stale próbuje nadążyć za przemianami technologicznymi, niczym Achilles za żółwiem – późny start, małe szanse. Ale jednocześnie, szczególnie jako humanistyka *cyfrowa*, rozumie i posiada dyspozycje by objąć poznawczo tę rzeczywistość medialną i ją opisać. Pozostaje wyrazić życzenie, by mogła na nią także choćby w minimalnym stopniu wpływać.

---

<sup>428</sup> Czapliński, *Kalendarium życia literackiego 1976-2000*, XI-XIII.

Niniejsza praca nieśmiało próbuje budować pomost od refleksji teoretycznoliterackiej, osadzonej w kształceniu sztuki interpretacji, do analizy danych o kulturze – nie tracąc wrażliwości na otaczający je kontekst. Sam produkt danowy jest bardzo skromny, a gdyby miał sprostać początkowym ambicjom, wymagałby zainicjowania projektu, prowadzonego przez zespół badawczy. Stanowi niemniej pierwszy krok do uprawiania Big Translation History, z myślą o przyszłych wykorzystaniach tego rodzaju zasobów w szerszym kontekście.

Czy, idąc za spostrzeżeniem opisywanych przez Andrew Pipera słuchaczy wystąpień cyfrowych humanistów, wiedzieliśmy już to wszystko wcześniej? Myślę, że poniekąd tak. Wysiłek projektu cyfrowego, ma nas tylko utwierdzić w tej wiedzy za pomocą zdobycia punktów danych jako referencji, wspólnego punktu odniesienia. Jednocześnie, będzie to dodatkową wartością, jeśli powstanie przy takiej okazji nowa wiedza lub okazja do zrewidowania powszechnych założeń. Potencjał zestawień tego rodzaju tkwi bowiem w ich łączliwości z innymi zbiorami danych, dzięki czemu właściwie każdy badacz na świecie, posługując się przyjętymi narzędziami, protokołami i standardami, może nie tyle replikować, tj. odtworzyć przebieg rozumowania, powtórzyć eksperyment czy zebrać takie same dane, ale reutilizować, czyli budować na zastanej wiedzy nowe informacje w nowych kontekstach, wizualizować, oświetlać inne problemy badawcze za pomocą uzyskanego *datasetu*.

Ze świadomością zagrożeń, opisywanych przez badaczy osadzonych w humanistycznej tradycji, jak Nycz i Pawlicka, w środowisku cyfrowym naukowcy różnych dziedzin pragnący obiektywizacji spojrzenia na system literacki mogą dużo efektywniej podejmować trud „liczenia okien w gmachu Kafki”, jak to robił Walser. Otwiera to zupełnie nowe, trzymając się jego metafory, „urbanistyczne” ujęcia: perspektywy porównań budynków sąsiednich na ulicy literatury narodowej i aspekty planowania przestrzennego w systemie literatury światowej jako mieście. A także, z pomocą rozwijających się prężnie narzędzi przetwarzania automatycznego, spojrzenia na całe regiony, kontynenty... w obszarze połączonych danych, nie tylko związanych z kulturą, ale także indywidualnymi gustami użytkowników sieci czy kulturą materialną. I pod warunkiem, że samo zliczanie okien nie przeszkodzi literaturoznawcy przynajmniej co jakiś czas zajrzeć do środka owego gmachu, mediując między bazodanowym ogółem, a szczegółem treści, będzie to owocować nowymi sposobami opisu tego umiłowanego przedmiotu, jakim pozostaje twórczość literacka wszystkich kultur i epok, w odmianach znanych tradycji oraz tych, które dopiero nadejdą.



# Bibliografia

Bibliografia przedmiotowa

Aljas, A. *Transforming Culture in the Digital Age International Conference in Tartu 14-16 April 2010*. Tartu: Estonian National Museum, 2010.

Arnold, Taylor, Nicolas Ballier, Paula Lissón, i Lauren Tilton. „Beyond Lexical Frequencies: Using R for Text Analysis in the Digital Humanities”. *Language Resources and Evaluation* 53, nr 4 (grudzień 2019): 707–33. <https://doi.org/10.1007/s10579-019-09456-6>.

Bachmann, Christian A., i Stephanie Heimgartner. „Book - Material - Text (Download Book)”. *Book - Material - Text*, 2017-01-09, 1-78 Pages. <https://doi.org/10.13154/BMT.1.2016.1-78>.

Bail, Christopher A. „The Cultural Environment: Measuring Culture with Big Data”. *Theory and Society* 43, nr 3–4 (lipiec 2014): 465–82. <https://doi.org/10.1007/s11186-014-9216-5>.

Bartmiński, Jerzy. „Droga naukowa Anny Wierzbickiej. Od składni polskiej prozy renesansowej do semantyki międzykulturowej.” *Teksty Drugie*, 2011.

Bechhofer, Sean, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, i in. „Why Linked Data Is Not Enough for Scientists”. *Future Generation Computer Systems* 29, nr 2 (luty 2013): 599–611. <https://doi.org/10.1016/j.future.2011.08.004>.

Benjamin, Walter. „Dzieło sztuki w epoce jego reprodukowalności technicznej”. W *Twórca jako wytwórca: eseje i rozprawy*, przetłumaczone przez Robert Reszke, 23–51. Warszawa: Wydawnictwo KR, 2011.

Berry, David. „The Computational Turn: Thinking about the Digital Humanities”. *Culture Machine* 12 (2011-02-18). <https://sro.sussex.ac.uk/id/eprint/49813/>.

Berry, David M., i Anders Fagerjord, red. *Digital humanities: knowledge and critique in a digital age*. Cambridge, England: Malden, MA: Polity, 2017.

Biasi, Pierre-Marc de. *Genetyka tekstów*. Przetłumaczone przez Filip Kwiatek i Maria Prussak. Filologia XXI 5. Warszawa: Instytut Badań Literackich PAN. Wydawnictwo, 2015.

Biesiada, Jacek. „Pięćdziesiąt lat «Polskiej Bibliografii Literackiej»”. *Nauka*, nr 1 (2004): 137–141.

Black, Michael L. „The World Wide Web as Complex Data Set: Expanding the Digital Humanities into the Twentieth Century and Beyond through Internet Research”. *International Journal of Humanities and Arts Computing* 10, nr 1 (marzec 2016): 95–109. <https://doi.org/10.3366/ijhac.2016.0162>.

Bode, Katherine. *A world of fiction: digital collections and the future of literary history*. Ann Arbor, MI: University of Michigan Press, 2018.

- . *Reading by numbers: recalibrating the literary field*. New York: Anthem Press, 2012.
- . „The Equivalence of “Close” and “Distant” Reading; or, Toward a New Object for Data-Rich Literary History”. *Modern Language Quarterly* 78, nr 1 (2017-03-01): 77–106. <https://doi.org/10.1215/00267929-3699787>.
- Bode, Katherine, i Paul Longley Arthur, red. *Advancing digital humanities: research, methods, theories*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan, 2014.
- Bodenhamer, David J., John Corrigan, i Trevor M. Harris, red. *The spatial humanities: GIS and the future of humanities scholarship*. Spatial humanities. Bloomington: Indiana University Press, 2010.
- Bolter, J. David, i Richard Arthur Grusin. *Remediation: Understanding New Media*. Cambridge (Mass.): MIT Press, 2000.
- Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. First MIT Press paperback edition. Cambridge, Massachusetts London, England: The MIT Press, 2016.
- Boyers, Robert. „George Steiner as Cultural Critic: Confronting America”. *New England Review* 15, nr 2 (1993): 136–56.
- Broeder, Daan, Thorsten Trippel, Emiliano Degl’Innocenti, Roberta Giacomi, Maurizio Sanesi, Mari Kleemola, Katja Moilanen, i in. „SSHOC D3.1 Report on SSHOC (Meta)Data Interoperability Problems”, 2019-12-10. <https://doi.org/10.5281/ZENODO.3569868>.
- Bruckman, Amy S. *Should You Believe Wikipedia?: Online Communities and the Construction of Knowledge*. 1. wyd. Cambridge University Press, 2022. <https://doi.org/10.1017/9781108780704>.
- Burzyńska, Anna, Michał Paweł Markowski, i Anna Burzyńska. *Podręcznik. Teorie literatury XX wieku / Michał Paweł Markowski*. Kraków: Wydaw. Znak, 2009.
- Caronia, Antonio. *The Cyborg: A Treatise on the Artificial Man*. Przetłumaczone przez Robert Booth. DE: meson press, 2015. <https://doi.org/10.14619/007>.
- Cioffi, Alessia, i Silvio Peroni. „Structured References from PDF Articles: Assessing the Tools for Bibliographic Reference Extraction and Parsing”. W *Linking Theory and Practice of Digital Libraries*, zredagowane przez Gianmaria Silvello, Oscar Corcho, Paolo Manghi, Giorgio Maria Di Nunzio, Koraljka Golub, Nicola Ferro, i Antonella Poggi, 13541:425–32. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022. [https://doi.org/10.1007/978-3-031-16802-4\\_42](https://doi.org/10.1007/978-3-031-16802-4_42).
- Culler, Jonathan D. *Teoria literatury. Bardzo krótkie wprowadzenie*. Przetłumaczone przez Maria Bassaj. Wydanie pierwsze. Warszawa: Prószyński i S-ka, 1998.
- Czachowska, Jadwiga, i Roman Loth. *Przewodnik polonisty: bibliografie, słowniki, biblioteki, muzea literackie*. Wyd. 2 uzupełnione. Vademecum Polonisty. Wrocław: Zakł. Narodowy im. Ossolińskich, 1981.

Czapliński, Przemysław, red. *Kalendarium życia literackiego 1976-2000: wydarzenia, dyskusje, bilanse*. Wyd. 1. Kraków: Wydawn. Literackie, 2003.

———. *Powrót centrali: literatura w nowej rzeczywistości*. Kraków: Wydawnictwo literackie, 2007.

Czapliński, Przemysław, Ryszard Nycz, Dominik Antonik, i Agnieszka Dauksza, red. *Nowa humanistyka: zajmowanie pozycji, negocjowanie autonomii*. Nowa Humanistyka (Warszawa), tom XXXVIII. Warszawa: Instytut Badań Literackich PAN Wydawnictwo, 2017.

Czapliński, Przemysław, i Piotr Śliwiński. *Literatura polska 1976-1998: przewodnik po prozie i poezji*. Wyd. 1., Dodruk. Kraków: Wydawn. Literackie, 1999.

Derrida, Jacques. *O gramatologii*. Przetłumaczone przez Bogdan Banasiak. Wyd. 2, zm.Rozszerz. Łódź: Wydawnictwo Officyna, 2011.

Deutsche Forschungsgemeinschaft. „Digitaler Wandel in den Wissenschaften”, 2020-10-28. <https://doi.org/10.5281/ZENODO.4191345>.

Dobson, James E. *Critical Digital Humanities: The Search for a Methodology*. Topics in the digital humanities. Urbana, Illinois: University of Illinois Press, 2019.

Dominas, Konrad. *Internet jako nowa przestrzeń interpretacji literatury antycznej*. Poznań: Pracownia Humanistycznych Studiów Interdyscyplinarnych, 2017.

Drabek, Aneta. „Indeksowanie czasopism w referencyjnych bazach danych”, 2018. <https://doi.org/10.6084/M9.FIGSHARE.5683972>.

Drucker, Johanna. „Blind Spots”. *The Chronicle of Higher Education*, 2009-04-03. <https://www.chronicle.com/article/blind-spots/>.

———. „Humanities Approaches to Graphical Display”. *Digital Humanities Quarterly* 5, nr 1 (2011). <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.

———. *Visualization and interpretation: humanistic approaches to display*. Cambridge, Massachusetts: The MIT Press, 2020.

Dzięglewski, Mariusz, Aldona Guzik, i Marta Juza. „Digitalizacja dziedzictwa kulturowego w polsce - repozytoria cyfrowe jako potencjalne źródło dostępu do zasobów kulturowych: Digitalization of cultural heritage in poland - digital repositories as potential access points to cultural resources”. *Studia Humanistyczne AGH* 16, nr 2 (2017): 89. <https://doi.org/10.7494/human.2017.16.2.89>.

Eco, Umberto, Stefan Collini, i Tomasz Bieroń. *Interpretacja i nadinterpretacja*. Wyd. 2. Kraków: Wydawn. Znak, 2008.

Eder, Maciej. „Metody ścisłe w literaturoznawstwie i pułapki pozornego obiektywizmu – przykład stylometrii”. *IBL PAN, call no. P.I.2524*, nr 2 (2014). <https://doi.org/10.18318/td>.

Edmond, Jennifer, i Erzsébet Tóth-Czifra. „Open Data for Humanists, A Pragmatic Guide”, grudzień 2018. <https://shs.hal.science/halshs-02115443>.

European Commission. Directorate General for Research and Innovation. *Turning FAIR into Reality :Final Report and Action Plan from the European Commission Expert Group on FAIR Data*. LU: Publications Office, 2018. <https://data.europa.eu/doi/10.2777/54599>.

Even-Zohar, Itamar. „Teoria polisystemów”. Przetłumaczone przez Katarzyna Lukas. *Przestrzenie Teorii*, nr 7 (2007-11-15): 347–66. <https://doi.org/10.14746/pt.2007.7.22>.

Fenrich, Wojciech. „Selekcja-i-przygotowanie-danych-badawczych-do-udostępnienia.pdf”. Uniwersytet Warszawski Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, 2019-12-03. <https://drodb.icm.edu.pl/wp-content/uploads/2019/10/Selekcja-i-przygotowanie-danych-badawczych-do-udost%C4%99pnienia.pdf>.

Filiciak, Mirosław. *Media, wersja beta: film i telewizja w czasach gier komputerowych i internetu*. Gdańsk: Wydawnictwo Naukowe Katedra, 2013.

Filimowicz, Michael. *Information Semiotics for Digital Humanities: Towards a New Conception of Quantity and Quality in Culture*. Code Blaze Books, 2020.

Fish, Stanley Eugene, Andrzej Szahaj, Richard Rorty, i Krzysztof Abriszewski. *Interpretacja, retoryka, polityka: eseje wybrane*. Wyd. 2. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, 2008.

Galloway, Scott. *Four Wielka Czwórka: ukryte DNA: Amazon, Apple, Facebook i Google*. Przetłumaczone przez Jolanta Kubiak. Wydanie I. Poznań: Dom Wydawniczy Rebis, 2018.

———. *The Four: The Hidden DNA of Amazon, Apple, Facebook and Google*. London: Bantam Press, 2017.

Garcia-Molina, Hector, Jeffrey D. Ullman, Jennifer Widom, i Tomasz Walczak. *Systemy baz danych: kompletny podręcznik*. Gliwice: Wydawnictwo Helion, 2011.

Gaszyńska-Magiera, Małgorzata. *Recepcja przekładów literatury iberoamerykańskiej w Polsce w latach 1945-2005 z perspektywy komunikacji międzykulturowej*. Varia Culturalia. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2011.

Gągolewski, Marek, Maciej Bartoszek, i Anna Cena. *Przetwarzanie i analiza danych w języku Python*. Warszawa: Wydawnictwo Naukowe PWN, 2016. <https://ksiegarnia.pwn.pl/Przetwarzanie-i-analiza-danych-w-jezyku-Python,634359876,p.html>.

Gerrish, Sean, i Filip Fierek. *Jak myślą inteligentne maszyny*. Wydanie I. Warszawa: PWN, 2020.

Gondek, Elżbieta, Zbigniew Żmigrodzki, i Stowarzyszenie Bibliotekarzy Polskich, red. *Bibliografia: metodyka i organizacja*. Nauka, Dydaktyka, Praktyka = Science, Didactics, Practice 38. Warszawa: Wydaw. SBP, 2000.

Grübler, Arnulf. „Time for a Change: On the Patterns of Diffusion of Innovation”. *Daedalus* 125, nr 3 (1996): 19–42.

- Grusin, Richard. „The Dark Side of Digital Humanities: Dispatches from Two Recent MLA Conventions”. *Differences* 25, nr 1 (2014-05-01): 79–92. <https://doi.org/10.1215/10407391-2420009>.
- Gumbrecht, Hans Ulrich. *Making Sense in Life and Literature*. Theory and History of Literature, v. 79. Minneapolis: University of Minnesota Press, 1992.
- . *Our Broad Present: Time and Contemporary Culture*. New York: Columbia University Press, 2014.
- Haraway, Donna Jeanne. *Manifestly Haraway*. Posthumanities 37. Minneapolis: University of Minnesota Press, 2016.
- Hauptert, Michael. „The Impact of Cliometrics on Economics and History”. *Revue d'économie politique* 127, nr 6 (2017): 1059. <https://doi.org/10.3917/redp.276.1059>.
- Hubar, Patryk, i Marcin Giersz. „European Literary Bibliography: Aggregation and Harmonization of Literary Bibliographical Data”, 2023-03-07. <https://doi.org/10.5281/ZENODO.7705026>.
- Hughes, Lorna, Panos Constantopoulos, i Costis Dallas. „Digital Methods in the Humanities: Understanding and Describing Their Use across the Disciplines”. W *A New Companion to Digital Humanities*, zredagowane przez Susan Schreibman, Ray Siemens, i John Unsworth, 150–70. Chichester, UK: John Wiley & Sons, Ltd, 2015. <https://doi.org/10.1002/9781118680605.ch11>.
- Hylewski, Marcin, i Tomasz Burdzik. „Teoria krytyczna szkoły frankfurckiej jako krytyka kultury masowej”, 2014. <https://doi.org/10.6084/M9.FIGSHARE.1050567>.
- Hyvönen, Eero. „Using the Semantic Web in digital humanities: Shift from data publishing to data-analysis and serendipitous knowledge discovery”. Zredagowane przez Pascal Hitzler i Krzysztof Janowicz. *Semantic Web* 11, nr 1 (2020-01-31): 187–93. <https://doi.org/10.3233/SW-190386>.
- Ingarden, Roman. „Formy poznawania dzieła literackiego”. *Pamiętnik Literacki: czasopismo kwartalne poświęcone historii i krytyce literatury polskiej* 33, nr 1–4 (1936): 163–92.
- Jannidis, Fotis. *Digitale Literaturwissenschaft: DFG-Symposion 2017*. Berlin: J.B. Metzler, 2022.
- Jasiewicz, Justyna, i Michalina Stopnicka. „Data literacy. Kompetencje w zakresie korzystania z danych w odniesieniu do przemian komunikacji naukowej”. *Kultura Popularna* 4, nr 50 (2017-05-31): 36–49. <https://doi.org/10.5604/01.3001.0010.0042>.
- Jaworska, Karolina. „Jak powstaje «Polska Bibliografia Literacka»? Znaczenie współpracy z bibliotekami naukowymi”. W *Biblioteka naukowa: czy jeszcze naukowa?*, zredagowane przez Jazdon, 167–78. Poznań: Biblioteka Uniwersytecka w Poznaniu, 2018.
- . „Projektowanie przestrzeni informacyjnej – przykład nowej strony internetowej «Polskiej Bibliografii Literackiej»”. 2018. <https://depot.ceon.pl/handle/123456789/15438>.

Jaworska, Karolina, i Beata Koper. „Polska Bibliografia Literacka – nowe wyzwania”. W *Krajowe bibliograficzne bazy danych - przegląd*, T. 25. Kraków, 2017.  
[https://ebibojs.pl/index.php/Mat\\_konf/article/view/611](https://ebibojs.pl/index.php/Mat_konf/article/view/611).

Jockers, Matthew Lee. *Macroanalysis: digital methods and literary history*. Topics in the digital humanities. Urbana: University of Illinois Press, 2013.

———. *Macroanalysis: digital methods and literary history*. Topics in the digital humanities. Urbana: University of Illinois Press, 2013.

Kamińska, Anna. „Cyfrowa humanistyka – koncepcja, kierunki i stan rozwoju oraz powiązania z informatologią”. Doktorat, Uniwersytet Warszawski, 2020. Doktoraty Wydziału Dziennikarstwa Informacji i Bibliologii (Archiwum).  
<https://depotuw.ceon.pl/handle/item/3853>.

Kandziora, Jerzy, Zyta Szymańska, i Krystyna Tokarzówna. *Bez cenzury 1976-1989: literatura, ruch wydawniczy, teatr: bibliografia*. Warszawa: IBL, 1999.

Karlińska, Agnieszka, Cezary Rosiński, Jan Wieczorek, Patryk Hubar, Jan Kocoń, Marek Kubis, Stanisław Woźniak, Arkadiusz Margraf, i Wiktor Walentynowicz. „Towards a contextualised spatial-diachronic history of literature: mapping emotional representations of the city and the country in Polish fiction from 1864 to 1939”, październik 2022.  
<https://aclanthology.org/2022.latechclfl-1.14/>.

Karpińska, Hanna. „Bibliografia przekładów literatury bułgarskiej w Polsce w latach 1990—2006 (uzupełnienia)”. *Przekłady Literatur Słowiańskich* 4, nr 2 (2014): 497–503.

Kasperski, Marek. *Sztuczna Inteligencja*. Helion, 2015.

Keen, Andrew. *The cult of the amateur: how today's internet is killing our culture*. 1st ed. New York: Doubleday/Currency, 2007.

Klejdzysz, Natalia. „Andrea Dworkin – najbardziej zniechędzona feministka”. *Środkowoeuropejskie Studia Polityczne*, nr 1 (2006): 159–72.

Kłosiński, Michał. *Hermeneutyka gier wideo: interpretacja, immersja, utopia = Hermeneutics of video games: interpretation, immersion, utopia*. Lupa obscura. Warszawa: Instytut Badań Literackich PAN. Wydawnictwo, 2018.

Koper, Beata, Cezary Rosiński, Barbara Wachek, Maciej Maryl, i Tomasz Umerle. „iPBL – supplementing literary bibliography with internet sources. Collaborative cataloguing”, 2023-06-30. <https://doi.org/10.5281/ZENODO.8228490>.

———. „iPBL – Supplementing Literary Bibliography with Internet Sources. Collaborative Cataloguing”, 2023-06-30. <https://doi.org/10.5281/ZENODO.8107864>.

Koper, Beata, i Tomasz Umerle. „Nowa bibliografia dla nowej humanistyki? : o przemianach funkcji bibliografii naukowej”. W *Nowa humanistyka: zajmowanie pozycji, negocjowanie autonomii*, 499–511. Warszawa, 2017.

Kubiński, Piotr. *Gry wideo: zarys poetyki*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, 2016.

- Kulecka, Alicja, red. *Archiwistyka cyfrowa i nauki pomocnicze historii w edukacji archiwalnej. Problemy dydaktyki archiwistyki*. Warsaw University Press, 2016. <https://doi.org/10.31338/uw.9788323521259>.
- Kurzweil, Ray. *The singularity is near: when humans transcend biology*. New York: Viking, 2005.
- Lahti, Leo, Niko Ilomäki, i Mikko Tolonen. „A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470–1800”. *LIBER Quarterly: The Journal of the Association of European Research Libraries* 25, nr 2 (2015-12-04): 87–116. <https://doi.org/10.18352/lq.10112>.
- Lahti, Leo, Jani Marjanen, Hege Roivainen, i Mikko Tolonen. „Bibliographic Data Science and the History of the Book (c. 1500–1800)”. *Cataloging & Classification Quarterly* 57, nr 1 (2019-01-02): 5–23. <https://doi.org/10.1080/01639374.2018.1543747>.
- Langmead, Alison, Jessica M. Otis, Christopher N. Warren, Scott B. Weingart, i Lisa D. Zilinski. „Towards Interoperable Network Ontologies for the Digital Humanities”. *International Journal of Humanities and Arts Computing* 10, nr 1 (marzec 2016): 22–35. <https://doi.org/10.3366/ijhac.2016.0157>.
- Latour, Bruno, i Steve Woolgar. *Laboratory life: the construction of scientific facts*. Princeton, N.J: Princeton University Press, 1986.
- Leão, Delfim, Marina Angelaki, Andrea Bertino, Suzanne Dumouchel, i Frédéric Vidal. „Operas Multilingualism White Paper”, 2018-07-30. <https://doi.org/10.5281/ZENODO.1324026>.
- Lears, Jackson. „Uneasy Courtship: Modern Art and Modern Advertising”. *American Quarterly* 39, nr 1 (1987): 133. <https://doi.org/10.2307/2712634>.
- Leonowicz, Iwona, Magdalena Rowińska, Paweł Rygiel, Leszek Śnieżko, Małgorzata Wielek-Konopka, i Iwona Wiśniewska. *Format MARC 21 rekordu bibliograficznego dla książki*. Wyd. 4. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 2017.
- Lindberg, Susanna. „Being with Technique—Technique as Being-with: The Technological Communities of Gilbert Simondon”. *Continental Philosophy Review* 52, nr 3 (wrzesień 2019): 299–310. <https://doi.org/10.1007/s11007-019-09466-9>.
- Lorenzi, Jean-Hervé, Mickaël Berrebi, i Justyna Nowakowska. *Przyszłość naszej wolności: czy należy rozmontować Google'a... i kilku innych?* Wydanie pierwsze. Biblioteka Myśli Współczesnej. Warszawa: Państwowy Instytut Wydawniczy, 2019.
- Manovich, Lev. „Avant-garde as Software”. *Artnodes* 0, nr 2 (2002-12-01). <https://doi.org/10.7238/a.v0i2.681>.
- . „Database as Symbolic Form”. W *Database Aesthetics. Art in the Age of Information Overflow*, zredagowane przez Victoria Vesna, 20:39–60. Electronic Mediations, b.d. Dostęp 2023-03-01.
- . *Software Takes Command: Extending the Language of New Media*. International Texts in Critical Media Aesthetics 5. New York, NY: Bloomsbury, 2013.

- . *The Language of New Media*. 8. print. A Leonardo Book. Cambridge, Mass.: MIT Press, 2001.
- . „The Science of Culture? Social Computing, Digital Humanities, and Cultural Analytics”, 2015. <http://manovich.net/index.php/projects/cultural-analytics-social-computing>.
- Marecki, Piotr, red. *Liternet.pl*. Wyd. 1. Kraków: Rabid, 2003.
- Maryl, Maciej. „Co to jest humanistyka cyfrowa – odpowiedź na ankietę Czasu Kultury”. *Czas Kultury* 185, nr 10 (2015): 10–11.
- . „F5: odświeżanie filologii”. *Maciej Maryl F5 : odświeżanie filologii Teksty Drugie : teoria literatury, krytyka, interpretacja* 146, nr 2 (2014): 9–20.
- . „Kim są polscy humaniści cyfrowi?” *Teksty Drugie* 1 (2017): 286–300. <https://doi.org/10.18318/td.2017.1.24>.
- . „Operationalising the Change. Dispersion of Polish literary life (1989–2002)”. W *Digitale Literaturwissenschaft*, zredagowane przez Fotis Jannidis, 509–34. Germanistische Symposien. Stuttgart: J.B. Metzler, 2022. [https://doi.org/10.1007/978-3-476-05886-7\\_21](https://doi.org/10.1007/978-3-476-05886-7_21).
- . „Operationalising the Change: Dispersion of Polish literary life (1989-2002)”, b.d.
- . „Subnarracje metadanych”. *Teksty Drugie*, nr 3 (2014): 7–11.
- Maryl, Maciej, Marta Błaszczynska, Bartłomiej Szleszyński, i Tomasz Umerle. „Dane badawcze w literaturoznawstwie”. *Teksty Drugie* 2 (2021): 13–44. <https://doi.org/10.18318/td.2021.2.2>.
- Maryl, Maciej, i Piotr Wciślik. „Remediations of Polish Literary Bibliography: Towards a Lossless and Sustainable Retro-Conversion Model for Bibliographical Data”. W *Digital humanities 2016: Conference Abstracts*, 621–23. Jagiellonian University & Pedagogical University, 2016. <https://rcin.org.pl/dlibra/publication/84542/edition/65718>.
- McGann, Jerome J., Paweł Bem, Łukasz Cybulski, Olga Mastela, i Jan Prussak. *Nowa Respublica Litteraria: pamięć i nauka w wieku cyfryzacji*. Warszawa: Instytut Badań Literackich PAN. Wydawnictwo, 2015.
- McLuhan, Marshall. *Galaktyka Gutenberga: tworzenie człowieka druku*. Przetłumaczone przez Andrzej Wojtasik i Grzegorz Godlewski. Wydanie III. Myśl o Kulturze. Warszawa: Narodowe Centrum Kultury, 2021.
- Meer, Jan IJ. van der. „The Literary System of the Stanisław Age in Poland: Its Participants”. *Russian Literature* 49, nr 4 (maj 2001): 449–88. [https://doi.org/10.1016/S0304-3479\(01\)80047-8](https://doi.org/10.1016/S0304-3479(01)80047-8).
- Meyer-Heß, Anke, i Jupp Rothen. „What happens if you publish the National Bibliography under a CC0license –Experiences of the German National Library (DNB).pdf”, b.d.
- Mills, Mark P. *Digital cathedrals*. First American edition. New York, New York: Encounter Books, 2019.
- Moretti, Franco. *Atlas of the European Novel: 1800-1900*. London New York: Verso, 1998.



- . *Distant reading*. London ; New York: Verso, 2013.
- . *Graphs, Maps, Trees: Abstract Models for a Literary History*. Paperback ed. London: Verso, 2007.
- . „Pamflet 6, grudzień 2013. «Operacjonalizowanie» albo funkcja pomiaru we współczesnej teorii literatury”. Przetłumaczone przez Borys Szumański. *Forum Poetyki* Jesień 2017 (b.d.).
- Ngai, Sianne. *Our Aesthetic Categories: Zany, Cute, Interesting*. Cambridge, Massachusetts London, England: Harvard University Press, 2015.
- Nieżgódka, Marek, Dominika Czerniawska, Karol Leszczyński, Jakub Szprot, Wojciech Fenrich, Paweł Bartecki, i Krzysztof Siewicz. „Wdrożenie i promocja otwartego dostępu do treści naukowych i edukacyjnych. Praktyki światowe a specyfika polska. Przewidywane koszty, narzędzia, zalety i wady”. ICM, 2011. <http://depot.ceon.pl/handle/123456789/1545>.
- Nowak, Jakub. „Andrew Keen. 2007. Kult amatora. Jak Internet niszczy kulturę. Warszawa: Wydawnictwa Akademickie i Profesjonalne, ss. 198; Bohdan Jung (red.). 2010. Wokół mediów ery Web 2.0. Warszawa: Wydawnictwa Akademickie i Profesjonalne, ss. 263”. *Nowe Media* 0, nr 1 (2010-07-01): 230. <https://doi.org/10.12775/NM.2010.015>.
- Nycz, Ryszard. „Nowa humanistyka w Polsce: kilka bardzo subiektywnych obserwacji, koniektur, refutacji”. *Teksty Drugie* 1 (2017): 18–39. <https://doi.org/10.18318/td.2017.1.2>.
- . „The New Humanities in Poland: A Few Subjective Observations, Conjectures, and Criticisms”. Przetłumaczone przez David Schaufli. *Er(r)go. Teoria - Literatura - Kultura*, nr 43 (2021-12-30): 315–38. <https://doi.org/10.31261/errgo.11711>.
- O'Reilly, Tim. „What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software”. *Communications & Strategies* 65, nr 1st quarter 2007 (2007): 17–37.
- Parikka, Jussi, Mateusz Borowski, i Małgorzata Sugiera. *Owady i media*. Interpretacje 78. Kraków: Księgarnia Akademicka, 2017.
- Parikka, Jussi, i Tony D. Sampson, red. *The spam book: on viruses, porn, and other anomalies from the dark side of digital culture*. Hampton Press communication series : Communication alternatives. Cresskill, N.J: Hampton Press, 2009.
- Pawlicka, Urszula. „Humanistyka: pracownia, centrum czy laboratorium?” *Teksty Drugie* 1 (2017): 314–33. <https://doi.org/10.18318/td.2017.1.26>.
- Pawlicka-Deger, Urszula. „The Laboratory Turn: Exploring Discourses, Landscapes, and Models of Humanities Labs”. *Digital Humanities Quarterly* 014, nr 3 (2020-09-25).
- Peroni, Silvio, Alexander Dutton, Tanya Gray, i David Shotton. „Setting Our Bibliographic References Free: Towards Open Citation Data”. *Journal of Documentation* 71, nr 2 (2015-03-09): 253–77. <https://doi.org/10.1108/JD-12-2013-0166>.
- Piotrowski, Piotr. *Awangarda w cieniu Jałty: sztuka w Europie Środkowo-Wschodniej w latach 1945-1989*. Wyd. 1. Poznań: Dom Wydawniczy Rebis, 2005.

Piper, Andrew. *Enumerations: data and literary study*. Chicago ; London: The University of Chicago Press, 2018.

*Polska bibliografia literacka 1979, I za rok 1979, część 1. - (1986)*. Warszawa: Państwowe Wyd. Naukowe, b.d.

Pressner, Todd, i Chris Johanson. „The Promise of Digital Humanities”, 2009-03-01.

Próchniak, Paweł. „Czasami jestem”. W *Promieniowanie tła: szkice o wierszach i czytaniu*, Wydanie I., 25–37. Wołowiec: Wydawnictwo Czarne, 2020.

———. „Czasami jestem (czytając wiersz McDonald’s)”. W *Mistrz świata. Szkice o twórczości Marcina Świetlickiego*, zredagowane przez Piotr Śliwiński, 149–57. Poznań, 2011.

———. „Paweł Próchniak: zyciorys naukowy z elementami autoreferatu”. Ośrodek „Brama Grodzka - Teatr NN”. Dostęp 2023-06-15. <https://teatrnn.pl/leksykon/artykuly/pawel-prochniak-zyciorys-naukowy/>.

Próchniak, Paweł, i Georges Bataille. „Bataille: słoneczna plama (na marginesie Martwego)”. W *Martwy*, 39–54. Kraków, 2010.

Raley, Rita. „Digital Humanities for the Next Five Minutes”. *Differences* 25, nr 1 (2014-05-01): 26–45. <https://doi.org/10.1215/10407391-2419991>.

Ramsay, Stephen. *Reading machines: toward an algorithmic criticism*. Topics in the digital humanities. Urbana: University of Illinois Press, 2011.

Raymond, Eric S. *The cathedral and the bazaar: musings on Linux and Open Source by an accidental revolutionary*. Rev. ed. Beijing ; Cambridge, Mass: O’Reilly, 2001.

Rodley, Chris, i Andrew Burrell. „On the Art of Writing with Data”. W *The Future of Writing*, zredagowane przez John Potts. Palgrave Macmillan, 2014. <https://doi.org/10.1057/9781137440402.0012>.

Roig-Sanz, Diana, i Laura Fólica. „Big Translation History: Data Science Applied to Translated Literature in the Spanish-Speaking World, 1898–1945”. *Translation Spaces* 10, nr 2 (2021-11-08): 231–59. <https://doi.org/10.1075/ts.21012.roi>.

Rosenfeld, Louis, Peter Morville, i Jorge Arango. *Architektura informacji w serwisach internetowych i nie tylko*. Wydanie IV. Gliwice: Wydawnictwo Helion, 2017.

Rosiński, Cezary, Agnieszka Karlińska, Marek Kubis, Patryk Hubar, i Jan Wieczorek. „Towards Metadata-Enriched Literary Corpora in Line with FAIR Principles: 19/20MetaPNC”, 2023-06-30. <https://doi.org/10.5281/ZENODO.8107574>.

Roszkowski, Marcin. „Bibliographic Data Science – konceptualizacja obszaru badawczego”. *Przegląd Biblioteczny* 90, nr 1 (2022-10-31): 5–26. <https://doi.org/10.36702/pb.914>.

———. „Każda reprezentacja jest interpretacją – w stronę hermeneutycznej koncepcji architektury informacji”. *Zagadnienia Informacji Naukowej - Studia Informacyjne* 57, nr 2(114) (2018-12-05). <https://doi.org/10.36702/zin.455>.

Rybicka, Elżbieta. *Geopoetyka. Przestrzeń i miejsce we współczesnych teoriach i praktykach literackich*. S.l.: Universitas, 2014.

Sandecki, Jan. *Stefan Vrtel-Wierczyński: w kręgu bibliografii i bibliotekarstwa*. Nauka, Dydaktyka, Praktyka = Science, Didactics, Practice 92. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 2007.

Schreibman, Susan, Ray Siemens, i John Unsworth, red. *A New Companion to Digital Humanities*. 1. wyd. Wiley, 2015. <https://doi.org/10.1002/9781118680605>.

Schreibman, Susan, Raymond George Siemens, i John Unsworth, red. *A companion to digital humanities*. Blackwell companions to literature and culture 26. Malden, MA: Blackwell Pub, 2004.

Science Europe. „Practical Guide to the International Alignment of Research Data Management - Extended Edition”, 2021-01-27. <https://doi.org/10.5281/ZENODO.4915861>.

Selden, Raman, Peter Widdowson, i Peter Brooker. *A reader's guide to contemporary literary theory*. 5th ed. Harlow, England ; New York: Pearson Longman, 2005.

Shieh, Jackie, i Terry Reese. „The Importance of Identifiers in the New Web Environment and Using the Uniform Resource Identifier (URI) in Subfield Zero (\$0): A Small Step That Is Actually a Big Step”. *Journal of Library Metadata* 15, nr 3–4 (2015-10-02): 208–26. <https://doi.org/10.1080/19386389.2015.1099981>.

Sikorski, Marcin. *Interakcja człowiek - komputer*. Dodr. Warszawa: Wydawnictwo Polsko-Japońskiej Wyższej Szkoły Technik Komputerowych, 2013.

Sosińska-Kalata, Barbara, Maria Przystek-Samokowa, i Zuzanna Wiorogórska. *Nauka o informacji w okresie zmian: informatologia i humanistyka cyfrowa : praca zbiorowa*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, 2016.

Stala, Marian. „Jedzenie hamburgerów”. W *Blisko wiersza: 30 interpretacji*, 254–61. Kraków: Społeczny Instytut Wydawniczy Znak, 2013.

Starakiewicz, Maja, i Jakub Woynarowski. *Model i metafora: komunikacja wizualna w humanistyce*. Wyd. 1. Kraków: Korporacja „Ha!art”, 2019.

Steiner, George. *Rzeczywiste obecności: czy w tym, co mówimy, coś w ogóle jest?* Przetłumaczone przez Olga Kubińska. Wydanie drugie zmienione. Gdańsk: słowo/obraz terytoria, 2023.

Stoll, Clifford. *Krzemowe remedium. Garść rozważań na temat infostrady*. Przetłumaczone przez Tomasz Hornowski. Poznań: Dom Wydawniczy Rebis, 2000.

———. *Kukulcze jajo*. Przetłumaczone przez Tomasz Hornowski. Poznań: Dom Wydawniczy Rebis, 1998.

———. „Stalking the Wily Hacker”. *Communications of the ACM* 31, nr 5 (maj 1988): 484–97. <https://doi.org/10.1145/42411.42412>.

———. *The cuckoo's egg: tracking a spy through the maze of computer espionage*. 1st ed. New York: Doubleday, 1989.

- . „Why the Web Won't Be Nirvana”. *Newsweek*, 1995-02-26.  
<https://www.newsweek.com/clifford-stoll-why-web-wont-be-nirvana-185306>.
- Sumpter, David J. T, i Radosław Kosarzycki. *Osaczeni przez liczby: o algorytmach, które kontrolują nasze życie : od Facebooka i Google'a po fake newsy i bańki filtrujące*. Wydanie I. Kraków: Copernicus Center Press, 2019.
- Szymańska, Zyta. „Bibliografia literacka w Instytucie Badań Literackich Polskiej Akademii Nauk (1948-2010) – geneza, dokonania, perspektywy”. *Biblioteka* 14, nr 23 (2010).
- . *Funkcje bibliografii literackich*. Uniwersytet Wrocławski, Wydział Filologiczny, Instytut Informacji Naukowej i Bibliotekoznawstwa, 2011.
- Świetlicki, Marcin. *Nieoczywiste: 77 wierszy religijnych Marcina Świetlickiego*. Zredagowane przez Wojciech Bonowicz. Seria Czarna. Kraków: Wydawnictwo EMG, 2007.
- Terras, Melissa. „Crowdsourcing in the Digital Humanities”. W *A New Companion to Digital Humanities*, zredagowane przez Susan Schreibman, Ray Siemens, i John Unsworth. Wiley, 2015.
- Tolonen, Mikko, i Umerle, Tomasz. „An Analysis of the Current Bibliographical Data Landscape in the Humanities. A Case for the Joint Bibliodatas Agendas of Public Stakeholders - Video Presentation”, 2022-10-11. <https://doi.org/10.5281/ZENODO.7118150>.
- Torres Parejo, Úrsula, Jesús R Campaña, M Amparo Vila, i Miguel Delgado. „A Survey of Tag Clouds as Tools for Information Retrieval and Content Representation”. *Information Visualization* 20, nr 1 (styczeń 2021): 83–97. <https://doi.org/10.1177/1473871620966638>.
- Torres-Ronda, Lorena, Emma Beanland, Sarah Whitehead, Alice Sweeting, i Jo Clubb. „Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis”. *Sports Medicine - Open* 8, nr 1 (grudzień 2022): 15.  
<https://doi.org/10.1186/s40798-022-00408-z>.
- Troszyński, Marek, i Juliusz Słowacki. *Alchemia rękopisu „Samuel Zborowski” Juliusza Słowackiego*. Filologia XXI 11. Warszawa: Instytut Badań Literackich PAN Wydawnictwo, 2017.
- Umerle, Tomasz. „Combining Research and Cultural Data to Understand Complex Phenomena”, 2022-09-30. <https://doi.org/10.5281/ZENODO.7326103>.
- Umerle, Tomasz. „Rethinking the Potential of Documentation of Culture as a Data Gathering Practice”. *Proceedings from the Document Academy* 4, nr 2 (2017-12-12).  
<https://doi.org/10.35492/docam/4/2/15>.
- Umerle, Tomasz, Błaszczńska, Marta, Wnuk, Madgalena, Franczak, Mateusz, Stojanovski, Jadranka, Rosiński, Cezary, Wołczuk, Nikodem, i in. „TRIPLE Deliverable 8.5: Guidelines on the Research Data in the Humanities”, 2023-01-20.  
<https://doi.org/10.5281/ZENODO.7568873>.
- Umerle, Tomasz, Beata Koper, i Cezary Rosiński. „Dokumentacja kultury a cyfrowe zasoby archiwalne. Przypadek Polskiej Bibliografii Literackiej (PBL) i Archiwum Telewizji

- Polskiej”. *Archiwa - Kancelarie - Zbiory*, nr 9(11) (2018-12-31): 109.  
<https://doi.org/10.12775/AKZ.2018.006>.
- Umerle, Tomasz, Beata Koper, i Marcin Werla. „Nowa bibliografia literacka dla nowej humanistyki”. 2016. <https://depot.ceon.pl/handle/123456789/14863>.
- Uniłowski, Krzysztof. „Czy gatunki i konwencje ewoluują (dzisiaj)?” *Teksty Drugie*, nr 3 (2011): 52–63.
- Verwayen, Harry. „Could AI and data mining technologies overcome issues in cultural heritage?” 2018-09-25. <https://pro.europeana.eu/post/could-ai-and-data-mining-technologies-overcome-copyright-issues-in-cultural-heritage>.
- Vimr, Ondřej, i Cezary Rosiński. „Mapping Czech literature in translation on a large scale (1820-2020)”. *Česká literatura* 70, nr 6 (luty 2023): 711–34.  
<https://doi.org/10.51305/cl.2022.06.03>.
- Walser, Martin. *Opis formy - studium o Kafce*. Przetłumaczone przez Edmund Misiołek. Biblioteka Krytyki Współczesnej. Warszawa: Państwowy Instytut Wydawniczy, 1972.
- Wardrip-Fruin, Noah, i Nick Montfort, red. *The New Media Reader*. Cambridge, Mass: MIT Press, 2003.
- Werla, Marcin. „Serwis internetowy “Polska Bibliografia Literacka”. Analiza danych z Google Analytics za okres 7 miesięcy: od 1 kwietnia 2016 r. do 31 października 2016 r.” Poznań: PCSS, 2016-11-02. <https://lib.psn.pl/dlibra/publication/edition/781?id=781>.
- Werla, Marcin, i Maciej Maryl. „Humanistyczne projekty cyfrowe w Polsce”. Poznań-Warszawa, 2014. <https://rcin.org.pl/ibl/publication/84544>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, i in. „The FAIR Guiding Principles for Scientific Data Management and Stewardship”. *Scientific Data* 3, nr 1 (grudzień 2016): 160018.  
<https://doi.org/10.1038/sdata.2016.18>.
- Wilkowski, Marcin. *Wprowadzenie do historii cyfrowej*. Gdańsk: Instytut Kultury Miejskiej, 2013.
- Winiecka, Elżbieta. *Poszerzanie pola literackiego: studia o literackości w internecie*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas, 2020.
- „Wstęp”. W *Polska bibliografia literacka 1979,1 za rok 1979, część 1. - (1986)*, IV–V. Warszawa: Państwowe Wyd. Naukowe, b.d.
- Zhang, Bohui, Filip Ilievski, i Pedro Szekely. „Enriching Wikidata with Linked Open Data”, 2022. <https://doi.org/10.48550/ARXIV.2207.00143>.

## Źródła audiowizualne i elektroniczne

„African American History”.

[https://www.loc.gov/collections/?fa=subject\\_topic:african+american+history](https://www.loc.gov/collections/?fa=subject_topic:african+american+history). Dostęp: 2023-09-16.

„Andrea Dworkin” W Filmweb. <https://www.filmweb.pl/person/Andrea+Dworkin-1348773>. Dostęp: 2023-09-23.

„Andrea Dworkin” W *Lubimy czytać*, <https://lubimyczytac.pl/autor/52950/andrea-dworkin>. Dostęp: 2023-07-15.

„Andrea Dworkin” W Wikipedia, [https://en.wikipedia.org/wiki/Andrea\\_Dworkin](https://en.wikipedia.org/wiki/Andrea_Dworkin). Dostęp: 2023-05-28.

„Andrea Dworkin” W *Wikipedia*, [https://pl.wikipedia.org/wiki/Andrea\\_Dworkin](https://pl.wikipedia.org/wiki/Andrea_Dworkin) dostęp: 2023-05-28.

„APIs for LoC.gov”. <https://www.loc.gov/apis/>. Dostęp: 2023-09-16.

„Barbara Wachek” W GitHub. <https://github.com/barbara-wachek>. Dostęp 2023-09-20.

„BazHum” <https://www.bazhum.pl/>. Dostęp: 2023-07-15.

Bednarek-Michalska, Bożena. „Zarządzanie danymi badawczymi – informacje ogólne i plany zarządzania danymi” W *Youtube*. <https://www.youtube.com/watch?v=iopHvHKFD-A>. Dostęp: 2020-03-01

„Centrum Humanistyki Cyfrowej IBL PAN” W *Facebook*. <https://www.facebook.com/Centrum.Humanistyki.Cyfrowej>. Dostęp 2023-09-20.

„cezary-rosinski” W GitHub. <https://github.com/cezary-rosinski>. Dostęp 2023-09-20.

„Ciało, płęć, seksualność - wymiar społeczno-kulturowy”, [https://usosweb.uni.lodz.pl/kontroler.php?\\_action=katalog2/przedmioty/pokazPrzedmiot&kod=0600-PS149C](https://usosweb.uni.lodz.pl/kontroler.php?_action=katalog2/przedmioty/pokazPrzedmiot&kod=0600-PS149C). Dostęp: 2023-09-20.

„CLARIN-PL”. <https://clarin-pl.eu>. Dostęp: 2023-09-16.

„Creative Commons”. <https://creativecommons.org/>. Dostęp: 2023-09-16.

Czwordon-Lis, Paulina, David Österle, Cezary Rosiński, i Tomasz Umerle. „Dataset of European Literary Awards”. Zenodo, 2022-11-30. <https://doi.org/10.5281/ZENODO.7380507>.

„DARIAH | Digital Research Infrastructure for the Arts and Humanities”. <https://www.dariah.eu>. Dostęp: 2023-09-23.

„DCMI: Home”, <https://www.dublincore.org/>.

Dib, Firas. „Regex101”. Regex101. <https://regex101.com/>. Dostęp 2022-12-12.

„DMPTool”, <https://dmptool.org/>. Dostęp: 2023-09-16.

*Druga fala feminizmu, czyli prywatne jest polityczne*. Nikt nas nie pytał, ale i tak się wypowiemy!, 2020. <https://www.estrada.poznan.pl/series/nikt-nas-nie-pytal/?i=9>.

„Dworkin Andrea” W *Biblionetka*, <https://www.biblionetka.pl/author.aspx?id=118819>. Dostęp: 2023-07-15.

Eichler, Jakub. „Works of Jorge Luis Borges in Polish Bibliographic Sources”, 2023-11-05. <https://doi.org/zenodo.10073461>.

Fanciullo, Lapo. „Men at Arms (1993)”. W *Mathematical Fiction*. <https://kasmana.people.cofc.edu/MATHFICT/mfview.php?callnumber=mf1415>. Dostęp 2023-03-15.

„Fandom”, <https://www.fandom.com/>. Dostęp: 2023-09-23.

„Federacja Bibliotek Cyfrowych”. <https://fbc.pionier.net.pl>. Dostęp 2023-07-15.

„Gender, Sexuality and Politics - Przedmioty – USOSweb” [https://usosweb.uni.lodz.pl/kontroler.php?\\_action=katalog2/przedmioty/pokazPrzedmiot&prz\\_kod=0600-WW728B](https://usosweb.uni.lodz.pl/kontroler.php?_action=katalog2/przedmioty/pokazPrzedmiot&prz_kod=0600-WW728B). Dostęp: 2023-09-20.

„GIS & Mapping - Digital Humanities Tools and Resources” W *LibGuides at University of Texas at Austin*. <https://guides.lib.utexas.edu/digitalhumanities/gis-mapping>. Dostęp: 2023-09-16.

„GIS for Humanities” W *Esri*, <https://learn.arcgis.com/en/paths/gis-for-humanities/>. Dostęp: 2023-09-16.

„GIS Mapping - Introducing the Digital Humanities” W *Research Guides at Temple University*. <https://guides.temple.edu/c.php?g=78518&p=506481>. Dostęp: 2023-09-16.

„Globální trajektorie české literatury od roku 1945”, <https://clb.ucl.cas.cz/projekty/aktualni-projekty/globalni-trajektorie-ceske-literatury-od-roku-1945/>. Dostęp: 2023-09-16.

*Hans Ulrich Gumbrecht (Stanford) delivers Annual Prokhorov Lecture HQ*, 2016. <https://youtu.be/88Ij4C8fakM?si=CMTE4psUFtsuurIs>. Dostęp 2023-09-16.

Hodge, Amy. „Guides: Data Best Practices and Case Studies: Name Files”. <https://guides.library.stanford.edu/data-best-practices/name-files>. Dostęp 2023-01-11.

"Homepage | CLARIN ERIC". <https://www.clarin.eu/>. Dostęp: 2023-09-16.

„ID.LOC.GOV - Linked Data Service”, <https://id.loc.gov/>. Dostęp: 2023-09-16.

IMDb Developer. „Sample Queries”. <https://developer.imdb.com/documentation/api-documentation/sample-queries/>. Dostęp 2023-09-05.

„Instytut Badań Literackich” PAN W *Facebook*. <https://www.facebook.com/InstytutBadanLiterackichPAN>. Dostęp 2023-09-20.

Instytut Badań Literackich Polskiej Akademii Nauk - Pracownicy. „Jacek Biesiada”. <https://ibl.waw.pl/pl/o-instytucie/pracownicy/biesiada-jacek>. Dostęp 2020-07-13.

Introna, Lucas. „Phenomenological Approaches to Ethics and Information Technology”. W *The Stanford Encyclopedia of Philosophy*, zredagowane przez Edward N. Zalta, Fall 2017.

Karlińska, Agnieszka, Cezary Rosiński, Jan Wieczorek, Patryk Hubar, Marek Kubis. „Towards Metadata-enriched Literary Corpora in Line with FAIR Principles”. b.d. <https://tinyurl.com/literary-corpora-workflow>.

Kasman, Alex. „Mathematical Fiction”. *Mathematical Fiction*. <https://kasmana.people.cofc.edu/MATHFICT/>. Dostęp 2023-03-15.

„Katalog Bibliotek Uniwersytetu Śląskiego i Uniwersytetu Ekonomicznego w Katowicach - ProLib Integro”. <https://integro.ciniba.edu.pl/>. Dostęp: 2023-07-15.

*Larry Flynt: The Right to Be Left Alone*. Documentary. Midtown Films, 2007. Dostęp: 2023-09-16.

„Library of Congress”. <https://www.loc.gov>. Dostęp 2020-03-20.

„Library of Congress Catalog”, <https://catalog.loc.gov/>. Dostęp 2020-03-20.

„Linked Open Data - W3C eGovernment Wiki”. [https://www.w3.org/egov/wiki/Linked\\_Open\\_Data](https://www.w3.org/egov/wiki/Linked_Open_Data). Dostęp: 2023-09-23.

MacDougall, John. „How To Find And Replace Multiple Text Strings Within A Text String”. *How to Excel*. Dostęp 2023-03-15. <https://www.howtoexcel.org/how-to-find-and-replace-multiple-text-strings-within-a-text-string/>.

"Macroanalysis" W *Matthew L. Jockers*, <https://www.matthewjockers.net/macroanalysisbook>. Dostęp 2023-09-16.

"Marcin Wilkowski", <https://wilkowski.org>. Dostęp: 2023-09-23.

Metaphysics Research Lab, Stanford University, 2017. <https://plato.stanford.edu/archives/fall2017/entries/ethics-it-phenomenology/>.

„My name is Andrea” W Filmweb. <https://www.filmweb.pl/film/My+Name+Is+Andrea-2022-10037764>. Dostęp: 2023-09-23.

„My name is Andrea” W *Tofifest*. [https://tofifest.mojfestiwal.pl/movies/ZsAEisu6jE\\_8LcCjpbMzSA](https://tofifest.mojfestiwal.pl/movies/ZsAEisu6jE_8LcCjpbMzSA). Dostęp: 2023-09-23.

„Nikodem Wołczuk” w GitHub, <https://github.com/nwolczuk>. Dostęp 2023-09-20.

„Nowa Panorama Literatury Polskiej”. <https://nplp.pl/>. Dostęp: 2023-09-16.

„Nowa Panorama Literatury Polskiej” W *Facebook*. <https://www.facebook.com/nowa.panorama.literatury.polskiej>. Dostęp 2023-09-20.

„open-access.network: Open Access Green and Gold”. <https://open-access.network/en/information/open-access-primers/green-and-gold>. Dostęp: 2023-09-23.

„OpenCitations – Home”. <https://opencitations.net/>. Dostęp: 2023-09-23.

OPERAS-PL. „Nie ma niewinnych danych”, 2021-12-09. <https://operas.pl/2021/12/09/nie-ma-niewinnych-danych/>.

"Otwarte dane i zasoby" W *Ministerstwo Cyfryzacji - Portal Gov.pl*. <https://www.gov.pl/web/cyfryzacja/otwarte-dane-publiczne>. Dostęp: 2023-09-23.



„Patryk Hubar” W GitHub. <https://github.com/patthub>. Dostęp 2023-09-20.

„PBL | Strona główna”. Dostęp 2023-10-10. <https://pbl.ibl.waw.pl/>.

„PBL Lab”, <http://chc.ibl.waw.pl/pl/projekty/pbl-lab/>. Dostęp: 2023-09-23.

„Polska Bibliografia Literacka (PBL)”. <https://pbl.ibl.poznan.pl/>. Dostęp 2023-10-10.

„Polska Bibliografia Literacka” W *Facebook*.  
<https://www.facebook.com/Polska.Bibliografia.Literacka>. Dostęp: 2023-09-16.

„Polska Bibliografia Literacka”. W *Wikipedia, wolna encyklopedia*. Dostęp 2020-03-13.  
[https://pl.wikipedia.org/wiki/Polska\\_Bibliografia\\_Literacka](https://pl.wikipedia.org/wiki/Polska_Bibliografia_Literacka).

„Pracownia Bibliografii Bieżącej”. <https://ibl.waw.pl/pl/o-instytucie/pracownie-i-zespoły/pracownia-bibliografii-biezacej>. Dostęp 2023-09-20.

„Protests Against Racism Web Archive”, <https://www.loc.gov/collections/protests-against-racism-web-archive/>. Dostęp: 2023-09-16.

„Raport z literatury”. <https://www.raportzliteratury.pl/>. Dostęp 2023-07-31.

„Sample Queries”, IMDb Developer, <https://developer.imdb.com/documentation/api-documentation/sample-queries/>. Dostęp 2023-09-05.

„Semantic Web Standards”. [https://www.w3.org/2001/sw/wiki/Main\\_Page](https://www.w3.org/2001/sw/wiki/Main_Page). Dostęp: 2023-09-16.

„Schema.org”, <https://www.schema.org>. Dostęp: 2023-09-16.

„Stack Overflow”, <https://stackoverflow.com>. Dostęp: 2022-10-01.

"Techsty - literatura i nowe media", <http://techsty.art.pl>. Dostęp: 2023-09-23.

„Towards Metadata-enriched Literary Corporain Line with FAIR Principles”. b.d.  
<https://tinyurl.com/literary-corpus-workflow>.

Walsh, Melanie. Introduction to Cultural Analytics & Python.  
<https://melaniewalsh.github.io/Intro-Cultural-Analytics/welcome.html>. Dostęp: marzec 2022.

„Working with Geospatial Data - Digital Humanities – Introduction” W *LibGuides at University of Exeter*. <https://libguides.exeter.ac.uk/digitalhumanities/geospatial/>. Dostęp: 2023-09-16.

„Wydawnictwo Znak” <https://www.znak.com.pl/>. Dostęp: 2023-09-16.

„Zielona granica” w *Filmweb*, <https://www.filmweb.pl/film/Zielona+granica-2023-10039040>  
dostęp: 2023-09-10

Amadeo, Ron. „Google+ shuts down April 2, all consumer data will be deleted”. *Ars Technica*, 2019-01-31. <https://arstechnica.com/gadgets/2019/01/google-shuts-down-april-2-all-data-will-be-deleted/>. Dostęp 2023-09-16.

Bartkiewicz, Artur. „Oxford Dictionaries uznały «post-prawdę» za słowo 2016 roku”. *Rzeczpospolita*, 2016-11-16. <https://www.rp.pl/spoleczenstwo/art3137761-oxford-dictionaries-uznaly-post-prawde-za-slowo-2016-roku>. Dostęp 2023-09-16.

Borejza, Tomasz. „Stanisław Lem przewidział, co internet zrobi z ludzkim mózgiem”. *Krowoderska.pl* (blog), 2020-05-26. <https://krowoderska.pl/stanislaw-lem-przewidzial-co-internet-zrobi-z-ludzkim-mozgiem/>. Dostęp 2023-09-16.

Chow, Andrew R., i Billy Perrigo. „The AI Arms Race Is Changing Everything”. *Time*, 2023-02-16. <https://time.com/6255952/ai-impact-chatgpt-microsoft-google/>. Dostęp 2023-09-16.

Fortuna, Piotr. „Więcej niż propaganda”. *Dwutygodnik* 299, nr 1 (2021). <https://www.dwutygodnik.com/artukul/9329-wiecej-niz-propaganda.html>. Dostęp 2023-09-16.

Ghaffary, Shirin. „Author Rebecca Solnit Found ‘Half’ Her Books in a Dataset Used to Develop AI”. *Bloomberg*, 2023-10-05. <https://www.bloomberg.com/news/newsletters/2023-10-05/rebecca-solnit-found-half-her-books-in-a-dataset-used-to-develop-ai#xj4y7vzkg>. Dostęp 2023-09-16.

Goal.com. „Footballers don't 'wear bras' - sporting reasons for under-shirt clothing explained”. <https://www.goal.com/en/news/footballers-dont-wear-bras-sporting-reasons-under-shirt-clothing-explained/1aak15v6271f814s624c5ws52t>. Dostęp: 2023-09-16.

Houghton, Chris. „Guest Post — Three Challenges (and Solutions) to Expand Digital Humanities”. *The Scholarly Kitchen* (blog), 2022-09-26. <https://scholarlykitchen.sspnet.org/2022/09/26/guest-post-three-challenges-and-solutions-to-expand-digital-humanities/>.

Huang, Junting. „Programming Without Coding: Orange for Digital Humanities”. Dostęp 2023-07-15. <https://medium.com/@juntinghuang/programming-without-coding-orange-for-digital-humanities-39b60078cd27>. Dostęp 2023-09-16.

Jaggi, Maya. „George and his dragons”. *Guardian*, 2021-03-17. <https://www.theguardian.com/books/2001/mar/17/arts.highereducation>. Dostęp 2023-09-16.

Kantrowitz, Alex. „Tim O’Reilly Coined Web 2.0. He Thinks Web3 Hype Is Naive.” *OneZero* (blog), 2022-03-01. <https://onezero.medium.com/tim-oreilly-coined-web-2-0-he-thinks-web3-hype-is-naive-ce8cd575bb44>. Dostęp 2023-09-16.

Levine, Jon. „That Time in 1995 Newsweek Predicted This ‘Internet’ Thing Wouldn’t Catch On”, 2018-02-06. <https://www.thewrap.com/newsweek-predicted-internet-wouldnt-catch-on-1995/>. Dostęp 2023-09-16.

Los Angeles Times. „Actually, That «off Target» 1995 Anti-Internet Column Was Amazingly on-Target”, 2015-03-02. <https://www.latimes.com/business/hiltzik/la-fi-mh-actually-that-offbase-20150227-column.html>. Dostęp 2023-09-16.

„Mathematical fiction”. W *Wikipedia*, b.d.

[https://en.wikipedia.org/wiki/Mathematical\\_fiction](https://en.wikipedia.org/wiki/Mathematical_fiction). Dostęp 2023-09-16.

McCulloch, Gretchen. „Coding Is for Everyone—as Long as You Speak English”, *Wired*, 2019-04-08. <https://www.wired.com/story/coding-is-for-everyone-as-long-as-you-speak-english/>. Dostęp 2023-09-16.

Medium. „Microsoft shuts down ‘Tay’ AI”. 2019-01-31.

<https://medium.com/collections/2dca30d0ee6a>. Dostęp 2023-09-16.

Piper, Andrew. „Did we already know that?” W *.txtlab @ mcgill*.

<https://txtlab.org/2018/09/did-we-already-know-that/> Dostęp: 2023-09-23.

Ryciak, Robert. „Danetyka, czyli o polskim tłumaczeniu Data Science”. *Kodolamacz* (blog), 2019-10-03. <https://www.kodolamacz.pl/blog/data-science-po-polsku/>. Dostęp 2023-09-16.

Soundcloud. „Czas Kultury”. Dostęp 2023-07-31. <https://soundcloud.com/user-80673682>. Dostęp 2023-09-16.

Stackpole, Thomas. „What Is Web3? Your guide to (what could be) the future of the internet.” *Harvard Business Review*, 2022-05-10. <https://hbr.org/2022/05/what-is-web3>. Dostęp 2023-09-16.

Stańczak, Agnieszka. „Cyfrowa humanistyka”. *Głos Nauki*.

<https://www.umcs.pl/pl/odcinki,20240,odc-5-cyfrowa-humanistyka,96829.chtm>. Dostęp 2023-07-15.

Wilkowski, Marcin. „Nie potrzebujesz Web 2.0 w historii webu”. *Web96* (blog), 2023-03-18. <https://web96.pl/2023/03/18/web-2-0.html>. Dostęp 2023-09-16.

Wrocławski Portal Matematyczny. „Matematyka w literaturze”.

<http://www.matematyka.wroc.pl/book/matematyka-wokol-nas/matematyka-w-literaturze>. Dostęp 2023-06-20.

Pozostałe źródła

Aro, Jessikka. *Trolle Putina*. Przetłumaczone przez Marta Laskowska. Wydanie I. Kraków: Wydawnictwo Sine Qua Non, 2020.

Baldwin, Sandy. *The Internet unconscious: on the subject of electronic literature*. International texts in critical media aesthetics, vol. 9. New York, NY: Bloomsbury Academic, an imprint of Bloomsbury Publishing Inc, 2015.

Barba, José Juan. „Kurt Schwitters. Avant-Garde and Advertising”. *Metalocus*, 2014-11-27. <https://www.metalocus.es/en/news/kurt-schwitters-avant-garde-and-advertising>.

Cambridge Digital Humanities. <https://www.cdh.cam.ac.uk/>. Dostęp 2023-09-16.

Czczot, Katarzyna, Agnieszka Mroziak, Katarzyna Nadana-Sokołowska, Anna Nasiłowska, Monika Rudaś-Grodzka, Ewa Maria Serafin, Barbara Smoleń, Kazimiera Szczuka, i Agnieszka Wróbel, red. *Encyklopedia gender: płeć w kulturze*. Warszawa: Wydawnictwo Czarna Owca, 2014.

Dziadek, Adam. „Anagramy Ferdynanda de Saussure’a: historia pewnej rewolucji”. *Teksty Drugie : teoria literatury, krytyka, interpretacja* 71, nr 6 (2001): 109–25.

Fish, Stanley Eugene. *Interpretacja, retoryka, polityka: eseje wybrane*. Zredagowane przez Andrzej Szahaj. Przetłumaczone przez Krzysztof Abriszewski. Wyd. 2. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”, 2008.

Krzaczkowski, Paweł, i Mateusz Kwaterko, red. *Przewodnik dla dryfujących: antologia sytuacjonistycznych tekstów o mieście*. Bęc. Seria Kieszonkowa. Warszawa: Fundacja Nowej Kultury Bęc Zmiana : Austriackie Forum Kultury, 2016.

Literary and Linguistic Computing Centre (LLCC). <https://www.repository.cam.ac.uk/handle/1810/131666>. Dostęp 2023-09-16.

MIT Press. „Autographic Design”, <https://mitpress.mit.edu/9780262547024/autographic-design/>. Dostęp 2023-10-10.

Pomerantsev, Peter. *To nie jest propaganda: przygody na wojnie z rzeczywistością*. Przetłumaczone przez Aleksandra Paszkowska. Wydanie pierwsze. Warszawa: Wydawnictwo Krytyki Politycznej, 2020.

Schlossberg, Tatiana. *Inconspicuous consumption: the environmental impact you don't know you have*. First edition. New York: Grand Central Publishing, 2019.

Szahaj, Andrzej. *Teoria krytyczna szkoły frankfurckiej: wprowadzenie*. Współczesna Myśl Humanistyczna. Polskie Spojrzenia. Warszawa: Wydawnictwa Akademickie i Profesjonalne. Grupa Kapitałowa WSiP, 2008.

Vydas, Saulius. „Cyborg: Evolution of the Superman”. *JAMA: The Journal of the American Medical Association* 194, nr 4 (1965-10-25): 474. <https://doi.org/10.1001/jama.1965.03090170152043>.

Wylie, Christopher. *Mind\*ck: Cambridge Analytica, czyli jak popsuć demokrację*. Przetłumaczone przez Michał Strąkow. Kraków: Insignis, 2020.

Ziomek, Jerzy. *Literatura Odrodzenia*. Wyd. 3 popr. i Uzupełn. Dzieje Literatury Polskiej : synteza uniwersytecka. Warszawa: Wydaw. Naukowe PWN, 1999.

———. *Renesans*. Wyd. 11. Wielka historia literatury polskiej. Warszawa: Wydawnictwo Naukowe PWN, 2012.

Zuboff, Shoshana. *Wiek kapitalizmu inwigilacji: walka o przyszłość ludzkości na nowej granicy władzy*. Przetłumaczone przez Alicja Unterschuetz. Wydanie I. Poznań: Zysk i S-ka Wydawnictwo, 2020.

## Spis ilustracji

Ryc. 1 Interfejs wprowadzania danych dla pracowników PBB .....	149
Ryc. 2 Interfejs użytkownika serwisu w domenie ibl.poznan.pl; zaznaczono identyfikator rekordu .	149
Ryc. 3 Nowy serwis PBL.....	150
Ryc. 4 Zapisy dotyczące Borgesa w pierwszym zbiorze danych LnŚ (rok i numer pisma).....	220
Ryc. 5 Zapisy dotyczące Borgesa wyrażone w formie wykresu frekwencji.....	220
Ryc. 6 Koncepcyjny graf wiedzy: możliwe powiązania między danymi .....	236

## Spis tabel

Tabela 1. Porównanie klasyfikacji odbioru sygnałów przez obserwatora w ujęciu Filimowicza .....	45
Tabela 2. Pola MARC 21, opracowane na potrzeby uzyskania informacji do datasetu.....	223