

Hipermediacja „Polskiej Bibliografii Literackiej” – w kierunku bibliograficznej infrastruktury badawczej

Przybliżając w rozdziale trzecim koncepcję bieżącej i retrospektywnej bibliografii literackiej Stefana Vrtela-Wierczyńskiego, zwróciłam uwagę na to, że intencją badacza było stworzenie wykazu wszystkich publikacji potrzebnych do badań z zakresu historii literatury. W ten sposób miała powstać bibliografia prezentująca całościowy obraz historii literatury, która równocześnie stałaby się inspiracją dla naukowców do poszukiwania nowych problemów badawczych. Twórca projektu pragnął także efektywnie wykorzystać nakład pracy bibliografów i zainicjować proces wielokrotnego wykorzystania raz przygotowanych materiałów bibliograficznych. „Polska Bibliografia Literacka” zarówno w postaci drukowanych tomów, jak i w formie bibliograficznej bazy danych była i jest nadal źródłem informacji bibliograficznych dla twórców innych bibliografii literackich oraz słowników biobibliograficznych. Materiały bibliograficzne publikowane w PBL były wykorzystywane między innymi w publikacjach: *Dawni pisarze polscy. Od początku piśmiennictwa do Młodej Polski. Przewodnik biograficzny i bibliograficzny*, t. 1–5¹, *Współcześni pol-*

¹ *Dawni pisarze polscy. Od początku piśmiennictwa do Młodej Polski. Przewodnik biograficzny i bibliograficzny*, t. 1, A–H, koordynacja całości Roman Loth, Wydawnictwa Szkolne i Pedagogiczne Spółka Akcyjna, Warszawa 2000; t. 2, I–ME, koordynacja całości Roman Loth, Wydawnictwa Szkolne i Pedagogiczne Spółka

scy pisarze i badacze literatury. Słownik biobibliograficzny, t. 1–10², *Bibliografia polska Stanisława Lema. 1945–2000* Bożeny Bednarek-Michalskiej³, *Polscy pisarze i badacze literatury przełomu XX i XXI wieku*, t. 1–3⁴, *Bronisława Wajs Papusza (17.08.1908–08.02.1987). Zestaw bibliograficzny*⁵. Z pewnością bibliografii, słowników, baz

Akcyjna, Warszawa 2001; t. 3, MIA–R, koordynacja całości Roman Loth, Instytut Badań Literackich, Fundacja Akademia Humanistyczna, Warszawa 2002; t. 4, S–T, koordynacja całości Roman Loth, Instytut Badań Literackich, Fundacja Akademia Humanistyczna, Warszawa 2003; t. 5, U–Ż, uzupełnienia, indeksy, koordynacja całości Roman Loth, Instytut Badań Literackich, Fundacja Akademia Humanistyczna, Warszawa 2004.

² *Współcześni polscy pisarze i badacze literatury. Słownik biobibliograficzny*, t. 1, A–B, t. 2, C–F, t. 3, G–J, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1994; t. 4, K, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1996; t. 5, L–M, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1997; t. 6, N–P, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1999; t. 7, R–Sta, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne Spółka Akcyjna, Warszawa 2001; t. 8, Ste–V, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne Spółka Akcyjna, Warszawa 2003; t. 9, W–Z, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Wydawnictwa Szkolne i Pedagogiczne Spółka Akcyjna, Warszawa 2004; t. 10, Ż i uzupełnienia do t. 1–9, oprac. zespół pod red. Jadwigi Czachowskiej i Alicji Szałağan, Instytut Badań Literackich, Fundacja Akademia Humanistyczna, Warszawa 2007.

³ Bożena Bednarek-Michalska, *Bibliografia polska Stanisława Lema. 1945–2000*, Uniwersytet Mikołaja Kopernika, Toruń 2006.

⁴ *Polscy pisarze i badacze literatury przełomu XX i XXI wieku*, t. 1, oprac. zespół pod red. Alicji Szałağan, Fundacja Akademia Humanistyczna, Instytut Badań Literackich PAN, Warszawa 2011; t. 2, oprac. zespół pod red. Alicji Szałağan, Fundacja Akademia Humanistyczna, Instytut Badań Literackich PAN, Warszawa 2014; *Polscy pisarze i badacze literatury przełomu XX i XXI wieku*, t. 3, oprac. zespół pod red. Alicji Szałağan, Instytut Badań Literackich PAN, Warszawa 2016. Słownik dostępny jest również w postaci bazy danych: <https://pisarzeibadacze.ibl.edu.pl/> (dostęp: 18.09.2025).

⁵ *Bronisława Wajs Papusza (17.08.1908–08.02.1987). Zestaw bibliograficzny*, oprac.

danych i publikacji naukowych korzystających z zasobów PBL jest o wiele więcej. Użytkownicy jednak nie mają obowiązku informowania zespołu autorskiego o tym, że wykorzystują w swoich pracach materiały bibliograficzne pochodzące z PBL. Dlatego podanie pełnej listy takich projektów jest niemożliwe.

W 2020 roku zespół Pracowni Bibliografii Bieżącej IBL PAN wyszedł z inicjatywą przygotowywania zestawów danych bibliograficznych pochodzących z bazy danych PBL i zasobów partnerów Pracowni z międzynarodowych projektów dla naukowców zainteresowanych taką usługą. W ciągu sześciu ostatnich lat dane bibliograficzne pochodzące między innymi z PBL zostały wykorzystane na potrzeby publikacji naukowych, referatów, grantów badawczych, rozpraw doktorskich. W sumie – w siedmiu projektach indywidualnych, sześciu projektach krajowych i sześciu projektach międzynarodowych. Były to między innymi projekty: „Nowy Korbut. Adam Mickiewicz – bibliografia przedmiotowa”, „Metody i narzędzia lingwistyki korpusowej w badaniach bibliografii polskich wydawnictw zwartych z lat 1997–2017”, „Cyfrowa infrastruktura badawcza dla humanistyki i nauk o sztuce DARIAH-PL”, „Czech Literary Bibliography – Czech Literary Internet. Data, Analysis, Research”, „Global Trajectories of Czech Literature Since 1945”, „Cultures of Dissent in Eastern Europe (1945–1989): Research Approaches in the Digital Humanities”, „Basics of Coding for Data Processing”, „TRIPLE (Transforming Research Through Innovative Practices for Linked Interdisciplinary Exploration)”⁶. Takie zastosowanie danych bibliograficznych PBL wpisuje się w wizję Stefana Vrtela-Wierczyńskiego i współczesny trend w nauce związany z ponownym wykorzystaniem danych przygotowanych podczas realizacji projektu naukowego. W ten sposób bibliografowie

Anna Sokółka, Danuta Zielińska, Wojewódzka i Miejska Biblioteka Publiczna im. Zbigniewa Herberta w Gorzowie Wielkopolskim, Gorzów Wielkopolski 2018.

⁶ *Ankiety dotyczące zakresu i sposobu wykorzystania infrastruktury i danych „Polskiej Bibliografii Literackiej”*, Materiały Pracowni Bibliografii Bieżącej IBL PAN w Poznaniu. [Dokument dostępny w Pracowni Bibliografii Bieżącej].

wykonali także krok w kierunku przekształcenia PBL z narzędzia pomocniczego literaturoznawstwa w bibliograficzną infrastrukturę badawczą udostępniającą dane badawcze oraz świadczącą usługi na rzecz naukowców. W tym rozdziale przedstawię potrzeby użytkowników i środowiska naukowego, a także trendy obowiązujące obecnie w komunikacji naukowej, które mają wpływ na to, jak zmieniają się strona internetowa i baza danych PBL. Zaprezentuję również zagadnienie danych badawczych i przebieg odbywającej się obecnie modernizacji PBL.

Potrzeby użytkowników bibliograficznych baz danych

Wyszukiwarki internetowe zapewniające łatwy i szybki dostęp do informacji ukształtowały oczekiwania internautów w zakresie wyszukiwania informacji. Ich zaletami, które cenią sobie użytkownicy, są: „prostota obsługi, pewność wyniku, język naturalny jako podstawa wyszukiwania, natychmiastowy dostęp do pełnych tekstów, czy szerzej – do zawartości obiektu”⁷. Bibliograficznym bazom danych niepodążającym za zmianami w sposobach wyszukiwania wiadomości może być trudno zainteresować użytkowników swoimi zasobami. Nawet jeśli są one uporządkowane, rzetelne, opracowywane przez specjalistów, którzy gwarantują jakość doboru materiału uwzględnionego w bazie danych. Dlatego tak ważna jest świadomość oczekiwań, jakie wobec zasobów wiedzy mają użytkownicy poszukujący informacji o publikacjach.

Z punktu widzenia użytkownika korzystającego z bazy danych ważne jest, aby wyposażona była ona w funkcjonalności, które po-

⁷ Małgorzata Jaskowska, *Potrzeba, trudności i formy przekazywania wiedzy i umiejętności w zakresie bibliograficznych systemów informacyjno-wyszukiwawczych studentom – przedstawicielom „pokolenia Google”*, w: *Bibliografi@. Źródła, standardy, zasoby. Praca zbiorowa*, pod red. Jerzego Franke, Stowarzyszenie Bibliotekarzy Polskich, Warszawa 2013, s. 311.

mogą w szybkim i precyzyjnym wyszukiwaniu publikacji. Internauci doskonalą swoje umiejętności wyszukiwawcze, używając wyszukiwarek internetowych i serwisów typu *discovery*, czyli wyszukiwarek oferujących zaawansowane wyszukiwanie oparte na bogatych, zindeksowanych metadanych (na przykład elektronicznych katalogów bibliotecznych i sklepów internetowych). Przyzwyczajeni do znanych stamtąd sposobów wyszukiwania oczekują, że przeszukiwanie innych zasobów wiedzy będzie wyglądało podobnie. Dlatego bibliograficzna baza danych również powinna posiadać elementy takie jak: wyszukiwanie proste i zaawansowane, wyszukiwanie przy pomocy języka naturalnego, zawężanie wyników wyszukiwania za pomocą filtrów, sortowanie wyników wyszukiwania, linki do pełnych tekstów, możliwość pobierania opisów bibliograficznych do menedżerów bibliografii, na przykład Zotero, Citavi, Mendeley.

Potrzeby środowiska naukowego i najnowsze trendy w komunikacji naukowej

Badacze stosujący w swoich pracach metody cyfrowe potrzebują dostępu do otwartych danych badawczych (*Open Research Data*). Dane badawcze są to

dokumenty w formie cyfrowej, inne niż publikacje naukowe, które są gromadzone lub produkowane w ramach działalności badawczo-naukowej i są wykorzystywane jako dowody w procesie badawczym bądź też są powszechnie akceptowane w środowisku naukowym jako konieczne do weryfikacji poprawności ustaleń i wyników badań⁸.

⁸ Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/1024 z dnia 20 czerwca 2019 r. w sprawie otwartych danych i ponownego wykorzystania informacji sektora publicznego, <https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:32019L1024&from=EN> (dostęp: 18.09.2025).

Zgodnie z obecną polityką naukową dane powinny być otwarte, czyli zarchiwizowane i udostępnione za darmo do ponownego wykorzystania. Można je przetwarzać, rozpowszechniać z poszanowaniem prawa i wykorzystywać do innych celów niż te, do których były gromadzone wcześniej.

Zagadnienie to łączy się z koncepcją *Linked Open Data*, czyli połączonych otwartych danych. Zakłada ona tworzenie ustrukturyzowanych danych z wykorzystaniem otwartych standardów, na przykład RDF (*Resource Description Framework*), w postaci umożliwiającej ich odczytanie przez maszyny (programy komputerowe). Takie dane połączone są za pomocą standardu URI (*Uniform Resource Identifier*) umożliwiającego identyfikację zasobów internetowych. Dzięki temu możliwe jest łączenie danych z danymi pochodzącymi z innych zbiorów oraz wyszukiwanie kolejnych danych za pomocą tych wcześniej połączonych. W ten sposób powstaje globalna sieć zawierająca różnego rodzaju informacje.

Badacze oczekują, że udostępniane dane będą odpowiednie pod względem merytorycznym i dostępne w postaci umożliwiającej ich przetworzenie za pomocą programów komputerowych. W przypadku danych bibliograficznych ważne jest, aby dane były ustandaryzowane. Na ten aspekt tworzenia danych bibliograficznych zwrócił uwagę Marcin Roszkowski, który pisał, że „standaryzacja procesu tworzenia metadanych odnosi się zarówno do ich zawartości (zasady katalogowania, kartoteki haseł wzorcowych), jak i do formy prezentacji (format danych bibliograficznych), co w założeniu zapewnia ich wysoką jakość i wiarygodność”⁹. Do jego uwag należy dodać również stosowanie słowników kontrolowanych, czyli ustalonych list nazw osób, instytucji, terminów służących do klasyfikacji materiałów, oraz łączenie rekordów tworzących wspólny zasób za pomocą relacji. Istotne jest ponadto, aby dane powstawały zgodnie z zasadami FAIR.

⁹ Marcin Roszkowski, „*Bibliographic data science*” – konceptualizacja obszaru badawczego, „Przegląd Biblioteczny” 2022, t. 90, nr 1, s. 6.

Zasady FAIR to wskazówki dla twórców danych dotyczące tego, jak prawidłowo tworzyć dane i przygotowywać je do udostępnienia. Twórcy danych badawczych powinni uwzględnić te zasady na każdym etapie przygotowania danych. Zgodnie z tymi zasadami dane powinny być: łatwe do znalezienia przez ludzi i komputery (*findable*); dostępne (*accessible*); interoperacyjne, czyli zintegrowane z innymi danymi, dostępne w postaci umożliwiającej wykorzystanie w aplikacjach służących do analizy, przechowywania i przetwarzania danych (*interoperable*); nadające się do ponownego wykorzystania, czyli dobrze opisane (*reusable*)¹⁰. W tym celu w bazie danych trzeba między innymi: używać powszechnie stosowanych i trwałych identyfikatorów (na przykład identyfikatorów międzynarodowych haseł wzorcowych VIAF (*Virtual International Authority File*)¹¹, identyfikatorów badaczy ORCID (*Open Researcher and Contributor ID*)¹² czy identyfikatorów publikacji elektronicznych DOI¹³); tworzyć bogate metadane; używać formalnego, dostępnego, współdzielonego i szeroko stosowanego języka służącego do prezentacji wiedzy; korzystać z kwalifikowanych odniesień do innych danych, udostępniać dane na licencji umożliwiającej ich użytkowanie. Tak przygotowane dane znajdą zastosowanie w badaniach naukowych.

Dane badawcze to zagadnienie, którym w ostatnim czasie zajmują się między innymi naukowcy działający na rzecz otwartej komunikacji naukowej. Metadane i wyzwania z nimi związane były tematem raportu *Guidelines on the Research Data in the Humanities* zawierającego wytyczne dotyczące danych humanistycznych¹⁴. Powstał

¹⁰ FAIR Principles, <https://www.go-fair.org/fair-principles/> (dostęp: 18.09.2025).

¹¹ Virtual International Authority File, <https://viaf.org/> (dostęp: 18.09.2025).

¹² ORCID, <https://orcid.org/> (dostęp: 18.09.2025).

¹³ DOI, <https://www.doi.org/> (dostęp: 18.09.2025).

¹⁴ Tomasz Umerle, Marta Błaszczczyńska, Magdalena Wnuk, Mateusz Franczak, Jadranka Stojanovski, Cezary Rosiński, Nikodem Wołczuk, Agnieszka Mikołajczyk-Bareła, Agnieszka Karlińska, Maciej Ogrodniczuk, Piotr Pęzik, Bianca Kramer, Silvio Peroni, -Luca De Santis, *TRIPLE Deliverable 8.5: Guidelines on the Research Data in the Humanities*, <https://zenodo.org/record/7568873#.zDlWqHZBxD8> (dostęp: 18.09.2025).

on w 2023 roku w ramach międzynarodowego projektu „TRIPLE”¹⁵. Informacje zawarte w raporcie mogą być wskazówką dla twórców baz danych chcących tworzyć lub modernizować swoje bazy zgodnie z ideami otwartej nauki i przetwarzania danych.

Badacze przyjrzeni się bliżej czterem typom metadanych: trwałym identyfikatorom, słowom kluczowym, abstraktom i cytatom. Nakreślili wyzwania, którym należy sprostać, aby wykorzystać potencjał omawianych metadanych. Skupię się na trzech pierwszych rodzajach metadanych, które moim zdaniem mają realne zastosowanie w bibliograficznej bazie danych takiej jak PBL. Trwały identyfikator to „sekwencja znaków umożliwiająca jednoznaczne i trwałe identyfikowanie obiektu, z którym jest powiązana”¹⁶. Służy do identyfikowania oraz wyszukiwania różnych zasobów. Jest unikatowy w skali globalnej i ma nieskończoną żywotność¹⁷. Stanowi gwarancję długoterminowej ochrony danych naukowych i ich dostępności¹⁸. Autorzy raportu podkreślili zalety poprawnego stosowania trwałych identyfikatorów. Pomagają one w znajdowaniu zasobów danych badawczych, cytowaniu, ponownym wykorzystaniu danych, interoperacyjności bazy danych, nawiązywaniu współpracy między instytucjami opracowującymi zasoby, dyscyplinami i krajami oraz badaniu wpływu¹⁹. Trwałe identyfikatory mają za zadanie zwiększyć zaufanie do używającej ich bazy danych. Pomagają w realizacji dwóch zasad FAIR: *findable* (dane

¹⁵ Efektem realizacji projektu jest także platforma GoTriple, dostępna pod adresem: <https://gotriple.eu/>, będąca agregatorem danych badawczych z nauk społecznych i humanistycznych, w którym metadane pochodzące od różnych dostawców są łączone, ujednolicane, wzbogacane i udostępniane do ponownego przetworzenia za pośrednictwem interfejsów programistycznych API.

¹⁶ *Instrukcja. Polityka NCN dotycząca otwartego dostępu do publikacji*, s. 4, https://www.ncn.gov.pl/sites/default/files/pliki/2021_10_instrukcja_open_access_ncn.pdf (dostęp: 18.09.2025).

¹⁷ TRIPLE *Deliverable 8.5: Guidelines on the Research Data in the Humanities...*, dz. cyt., s. 11.

¹⁸ Tamże.

¹⁹ Tamże.

powinny być możliwe do znalezienia przez ludzi i komputery) oraz *interoperable* (czyli interoperacyjności)²⁰.

Kolejnym elementem metadanych omówionym w raporcie są słowa kluczowe. Mają one dwa zastosowania: opis i klasyfikację²¹. W pierwszej sytuacji służą przedstawieniu treści artykułu. Wówczas są nadawane przez autora lub osoby indeksujące, zaczerpnięte z tekstu, nie tworzą słownika kontrolowanego, nie są powiązane z zewnętrznym źródłem informacji i nie służą do wyszukiwania czy filtrowania wyników²².

Z punktu widzenia bibliograficznej bazy danych istotne jest zastosowanie drugiego typu słów kluczowych, które tworzą słownik kontrolowany. Używane w takiej postaci informują o głównych tematach publikacji, pomagają w klasyfikacji i kierują użytkowników do odpowiednich zasobów²³. W tym sensie stanowią narzędzie komunikacji między naukowcami a społecznościami, które są zainteresowane wynikami ich badań, i dlatego odgrywają kluczową rolę w naukowych systemach wyszukiwawczych²⁴. Autorzy raportu zwrócili uwagę na zalety stosowania słowników kontrolowanych, takie jak: łączenie wszystkich wystąpień słowa kluczowego, a przez to ułatwienie wyszukiwania i filtrowania, poprawa wyników wyszukiwania, uporządkowanie i strukturyzowanie danych²⁵. Słowa kluczowe funkcjonujące w taki sposób są także wykorzystywane jako „punkt dostępu”²⁶. Wówczas przeglądanie zawartości zasobu bibliograficznego zaczyna się od słowa kluczowego i możliwe jest zapoznanie się z pełną listą publikacji na dany temat.

W omawianym raporcie opisano jeszcze jedną metodę tworzenia słów kluczowych, która jest odpowiedzią na wyzwania, z jakimi mierzą się twórcy zasobów bibliograficznych, a jakie wiążą się z wciąż

²⁰ Tamże, s. 13.

²¹ Tamże, s. 25.

²² Tamże, s. 25–26.

²³ Tamże, s. 25.

²⁴ Tamże.

²⁵ Tamże, s. 26.

²⁶ Tamże.

rosnącą liczbą publikacji naukowych. Tą metodą jest generowanie słów kluczowym przy pomocy sztucznej inteligencji. Autorzy podkreślili, że tak powstające słowa kluczowe są najefektywniejszym sposobem opisywania treści publikacji, ponieważ powstają na masową skalę i są w stanie obsłużyć dużą liczbę tekstów naukowych²⁷. Naukowcy zwrócili również uwagę na związane z nimi problemy – trzeba brać je pod uwagę, myśląc o wdrożeniu takiego rozwiązania. Słowa kluczowe tworzone przez sztuczną inteligencję są trudne do zweryfikowania i zapamiętania, ograniczone modelami językowymi, a także nieprecyzyjne²⁸.

Trzecim elementem metadanych omawianym w raporcie są abstrakty. Zdaniem autorów raportu odgrywają one kluczową rolę w komunikacji naukowej. Pomagają użytkownikom poznać tematykę publikacji, zawierają jej najważniejsze wnioski i podsumowanie, dzięki uwzględnianiu ich w bibliografiach wpływają na zwiększenie widoczności tekstu, co ułatwia jego odkrywanie²⁹. Dla komunikacji naukowej ważne jest, aby były w otwartym dostępie i nadawały się do ponownego wykorzystania³⁰. Abstrakty mogą mieć zastosowanie także w bibliograficznych bazach danych. Mogą tam nie tyle zastępować adnotacje dotyczące treści publikacji, ile dostarczać dodatkowych informacji. To ważne w sytuacji, gdy badacz poszukuje szczegółowych wiadomości, które pozwolą mu ocenić, czy wyszukiwany materiał będzie dla niego użyteczny. Udostępnianie abstraktów może przyczynić się do oszczędności czasu poświęcanego przez użytkownika na kwerendę. Dodatkowo udostępnianie abstraktów w postaci umożliwiającej ich przetworzenie wraz z innymi metadany może rozszerzyć zakres badań, które da się przeprowadzić przy użyciu danych bibliograficznych.

²⁷ Tamże.

²⁸ Tamże.

²⁹ Tamże, s. 41.

³⁰ Tamże.

Tak scharakteryzowane potrzeby wynikają z dotychczasowych doświadczeń środowiska naukowego. Wprowadzenie zmian w sposobie przygotowania danych bibliograficznych wpłynie na częstsze wykorzystanie ich w projektach naukowych. Zwracam uwagę na użyte przeze mnie sformułowanie „częstsze wykorzystanie ich w projektach naukowych”, ponieważ dane bibliograficzne już mają zastosowanie w pracach z zakresu humanistyki. Przejdę teraz do charakterystyki danych literaturoznawczych i omówienia przykładów wykorzystania danych badawczych w pracach z zakresu literaturoznawstwa oraz historii książki. Zaprezentuję także publikacje dotyczące polskiego życia literackiego, w których były analizowane dane bibliograficzne pochodzące z „Polskiej Bibliografii Literackiej”.

Dane badawcze w badaniach literaturoznawczych

Dane badawcze do tej pory raczej nie kojarzyły się z naukami humanistycznymi. Maciej Maryl, Marta Błaszczczyńska, Bartłomiej Szleszyński i Tomasz Umerle w artykule *Dane badawcze w literaturoznawstwie*³¹ zaprezentowali zagadnienie danych badawczych, uwzględniając specyfikę badań literaturoznawczych. Zdefiniowali dane badawcze jako „wszystkie materiały i źródła, które uczeni zbierają, wytwarzają, wzbogacają i wykorzystują na wszystkich etapach procesu badawczego”³². Autorzy artykułu zaproponowali następującą typologię danych badawczych w literaturoznawstwie:

1. **Tekst kultury** – konkretny egzemplarz utworu, wyrażonego znakami słownymi czy graficznymi, takiego jak wydane dzieło, rękopis, afisz teatralny, dwu- lub trójwymiarowa reprezentacja artefaktu (na przykład. skan), nagranie.
2. **Metadane** – opis fizyczny i informacje o danym egzemplarzu

³¹ Maciej Maryl, Marta Błaszczczyńska, Bartłomiej Szleszyński, Tomasz Umerle, *Dane badawcze w literaturoznawstwie*, „Teksty Drugie” 2021, nr 2, s. 13–44.

³² Tamże, s. 16.

utworu bądź dokumentu, jego wydaniu, czy też dane techniczne pliku cyfrowego.

3. **Adnotacje** – notatki, komentarze, uwagi, aparat krytyczny czy znaczniki TEI (*Text Encoding Initiative*) dodane do tekstu w procesie analizy i interpretacji przez badaczy lub wygenerowane mechanicznie (np. dane frekwencyjne, biogramy, n-gramy, nazwy własne, geolokacja miejsc w utworze). Adnotacje tym się różnią od metadanych, że wynikają z interpretacji własnej badacza i przyjętych przez nich założeń metodologicznych, a nie wyłącznie ze standardu opisu np. bibliograficznego.
4. **Dane kultury literackiej** – zestawienia informacji o życiu literackim, takie jak kalendaria, listy osób, zdarzeń, dat, zestawienia statystyczne (np. nakładów, recepcji), wyniki ankiet, słowniki terminów i pojęć literackich.
5. **Literatura przedmiotu** – interpretacje, podręczniki, świadectwa odbioru i inne opracowania lub wypowiedzi naukowe dotyczące przedmiotu badań.
6. **Dokumentacja procesu badawczego** – metodologia, notatki, systemy klasyfikacji pozwalające zrozumieć dane projektowe i sposób ich wytworzenia, treść kwestionariuszy, notatki ze spotkań, lista członków zespołu³³.

Dane bibliograficzne według tej typologii to metadane. Wydawać się może, że zestaw opisów bibliograficznych, metadanych w pracy naukowej może nie mieć innego zastosowania jak wykorzystanie go w przypisach i bibliografii załącznikowej. Jednak dzięki dostępności narzędzi cyfrowych do przetwarzania danych może być on elementem wchodzącym w skład publikacji naukowej. Należy przy tym zaznaczyć, że wykorzystanie danych badawczych w badaniach humanistycznych nie jest celem samym w sobie, dane nie zastąpią interpretacji. Przetworzone dane badawcze są dla interpretacji podstawą. Jak pisali Maryl, Błaszczczyńska, Szleszyński i Umerle, służą one „pogłębieniu kontekstu badań (np. jako suplement artykułu) lub dal-

³³ Tamże, s. 21–22.

szemu wykorzystaniu przez innych badaczy materiałów opracowanych w ramach projektu³⁴. W dalszej części rozdziału zaprezentuję dwa podejścia badawcze: czytanie zdystansowane i *bibliographic data science*, wykorzystujące dane badawcze w literaturoznawstwie, oraz dwie publikacje, których autorzy: Tomasz Umerle, Cezary Rosiński oraz Maciej Maryl wykorzystali dane bibliograficzne pochodzące z PBL do badania polskiego życia literackiego.

Franco Moretti jest twórcą koncepcji czytania zdystansowanego (*distant reading*), w której metody i narzędzia obliczeniowe wykorzystywane są do analizy tekstów literackich³⁵. Podstawowym założeniem tej koncepcji jest twierdzenie, że można badać historię literatury bez konieczności czytania poszczególnych tekstów. Uważne czytanie (*close reading*) zastępuje się tu analizą dużego zbioru danych. Czytanie zdystansowane stosowane jest w badaniach systemu literackiego, które zakładają wyjście poza kanon tekstów. W tej koncepcji dystans „jest warunkiem wiedzy; pozwala on skupić się na jednostkach mniejszych lub większych od pojedynczego tekstu – na środkach stylistycznych, tematach, tropach, gatunkach literackich i systemach”³⁶. Moretti zaproponował, aby badając literaturę, analizę znaczenia tekstów zastąpić rozpoznawaniem wzorców³⁷. Dzięki czytaniu zdystansowanemu wyłaniają się nowe perspektywy badawcze. Daje ono podstawę do zrozumienia przyczyn i procesów zmian zachodzących w systemie literackim. Włoski literaturoznawca stosuje nie tylko nowoczesne metody analizy tekstów, ale także prezentacji wyników badań. Teksty interpretacyjne uzupełnia modelami zapożyczonymi z innych dyscyplin. W swojej książce dotyczącej historii literatury światowej prezentował rozwój gatunków literackich oraz rynku książki na wykresach, wykorzystywał także mapy literackie

³⁴ Tamże, s. 20.

³⁵ Franco Moretti, *Distant Reading*, Verso, London – New York 2013.

³⁶ Tenże, *Przypuszczenia na temat literatury światowej*, przeł. Przemysław Czapliński, „Teksty Drugie” 2014, nr 4, s. 135.

³⁷ Tenże, *Wzorce i interpretacja*, przeł. Ewa Rajewska, „Rocznik Komparatystyczny” 2018, nr 9, s. 13.

i drzewa ewolucyjne³⁸. Badanie literatury z wykorzystaniem narzędzi cyfrowych do analizy zbiorów danych nie ogranicza się do analizowania tekstów literackich. Nowym obszarem badawczym jest *bibliographic data science*, w którym analizowane są dane bibliograficzne pochodzące z katalogów bibliotecznych i bibliografii.

Bibliographic data science to postawa badawcza, zaproponowana przez fińskich badaczy Leo Lahtiego, Jani Marjanena, Hege Roivainena, Mikko Tolonena, Ville Vaary, polegająca na zastosowaniu „algorytmicznie wspomaganých procesów harmonizacji i integracji danych pochodzących z różnych źródeł bibliograficznych na potrzeby badań nad produkcją wiedzy”³⁹. Upraszczając tę definicję, można przyjąć, że cechą tego podejścia jest wykorzystanie metod statystycznych i technologii informatycznych do badania zestawów danych bibliograficznych pochodzących od różnych dostawców danych. Marcin Roszkowski pisał o pięciu założeniach *bibliographic data science*:

1. Przedmiotem badań BDS są duże i heterogeniczne zasoby bibliograficzne.
2. Zasoby bibliograficzne (metadane) są traktowane jako obiekty badawcze.
3. Celem BDS jest wypracowanie metod i narzędzi optymalizacji jakości i kompletności metadanych.
4. BDS jest oparte na ilościowych badaniach informacji (metody statystyczne i probabilistyczne).
5. BDS jest oparte na modelu iteracyjnym⁴⁰.

³⁸ Tenże, *Wykresy, mapy, drzewa. Abstrakcyjne modele na potrzeby historii literatury*, przeł. Tomasz Bilczewski, Anna Kowalcze-Pawlik, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2016.

³⁹ Leo Lahti, Jani Marjanen, Hege Roivainen, Mikko Tolonen, *Bibliographic Data Science and the History of the Book (c. 1500–1800)*, „Cataloging & Classification Quarterly” 2019, t. 57, nr 1, s. 6. Cyt. za: Marcin Roszkowski, „*Bibliographic data science*”..., dz. cyt., s. 8.

⁴⁰ Marcin Roszkowski, „*Bibliographic data science*”..., dz. cyt., s. 10.

Badacz zwrócił uwagę, że w przypadku *bibliographic data science* ważnym problemem przy łączeniu danych z różnych źródeł są jakość i kompletność metadanych, czyli stosowane formaty danych, konwencje zapisu informacji w rekordach wynikające z zasad katalogowania stosowanych w tych źródłach oraz błędne lub brakujące dane⁴¹. Jak słusznie zauważył, wykorzystanie danych bibliograficznych do badań z zakresu *bibliographic data science* oznacza „wykorzystanie zasobów bibliograficznych poza głównym kontekstem ich funkcjonowania (katalog biblioteczny, bibliograficzna baza danych) i ingerencję w ich strukturę oraz zawartość na potrzeby prowadzonych badań”⁴². Należy podkreślić, że zmieniają się oczekiwania użytkowników – badaczy – wobec bibliografii. Dane bibliograficzne muszą być przygotowywane, a także udostępniane w takiej postaci, by można było je pobrać i przetworzyć przy pomocy narzędzi wykorzystujących zaawansowane technologie informatyczne.

Lahti, Marjanen, Roivainen i Tolonen pokazali, że zbiory danych można wykorzystać w badaniach z zakresu historii książki. Użyli ich w artykule dotyczącym rozwoju formatów książek i wernakularyzacji (upisłmiennienia) we wczesnonowożytnej Europie⁴³. Na potrzeby tego tekstu przeprowadzili analizę danych bibliograficznych pochodzących z czterech bibliografii: fińskiej bibliografii narodowej (*Fennica – Suomen kansallisbibliografia*), szwedzkiej bibliografii narodowej (*Svensk bokförteckning*), katalogu centralnego angielskich krótkich tytułów (*The English Short-Title Catalogue*) oraz Bazy Dziedzictwa Drukowanej Książki (*Heritage of the Printed Book*)⁴⁴.

Fińscy badacze pisali o tym, że proces wernakularyzacji był wcześniej badany przez językoznawców i historyków z różnych perspektyw, jednak dzięki analizie dużych zbiorów danych bibliograficznych

⁴¹ Tamże, s. 11.

⁴² Tamże.

⁴³ Leo Lahti, Jani Marjanen, Hege Roivainen, Mikko Tolonen, *Bibliographic Data Science...*, dz. cyt., s. 5–23.

⁴⁴ Tamże, s. 6–7.

autorzy artykułu zastosowali nowatorskie podejście do tego zagadnienia⁴⁵. Przeanalizowali zbiory metadanych zawierające miliony tytułów i powiązanych z nimi informacji bibliograficznych⁴⁶. W ten sposób wprowadzili wcześniej niezbadane źródła, co pozwoliło im prześledzić, w jaki sposób proces wernakularyzacji zmaterializował się w konkretnych publikacjach⁴⁷. Analiza połączonych danych, pochodzących z kilku krajowych bibliografii, pozwoliła na określenie szerszych procesów i trendów w historii książki oraz na inne niż narodowe spojrzenie na badany okres⁴⁸. Badacze wykazali, że formaty książek i język tytułów mogą odgrywać kluczową rolę w formowaniu się sfery publicznej we wczesnonowożytnej Europie⁴⁹. Pokazali także, że podobne krajowe bibliografie mogą służyć nie tylko do badania krajowych tradycji wydawniczych, ale także do badań porównawczych i zmiany perspektywy badawczej z narodowej na globalną⁵⁰.

Czy PBL może stać się dostawcą danych na potrzeby badań z zakresu *bibliographic data science*? Czy dane PBL spełniają oczekiwania badaczy posługujących się metodami cyfrowymi? Odpowiedź na pierwsze pytanie brzmi: tak. PBL to baza danych zawierająca duży zestaw danych, którego można użyć do badań prowadzonych z wykorzystaniem technologii informatycznych. W przypadku pytania drugiego sprawa jest skomplikowana. Dane bibliograficzne PBL, ze względu na to, że tworzone były na wzór opisów bibliograficznych sporządzanych na potrzeby drukowanych tomów bibliografii, które rządziły się swoimi prawami i miały pewne ograniczenia – skróty i pominięcia informacji w celu oszczędności papieru, zaspokajanie podstawowych potrzeb informacyjnych użytkowników – nie spełniają wszystkich oczekiwań badaczy, ale nawet w obecnej postaci

⁴⁵ Tamże, s. 15.

⁴⁶ Tamże.

⁴⁷ Tamże.

⁴⁸ Tamże, s. 7.

⁴⁹ Tamże, s. 17.

⁵⁰ Tamże, s. 18.

pozwalają na przeprowadzenie badań z ich udziałem. Dane bibliograficzne pochodzące z PBL zostały wykorzystane w dwóch artykułach dotyczących polskiego życia literackiego, których autorami są Tomasz Umerle i Cezary Rosiński, współautorzy bazy danych PBL, oraz Maciej Maryl, kierownik Centrum Humanistyki Cyfrowej IBL PAN i współpracownik Pracowni Bibliografii Bieżącej IBL PAN.

Umerle i Rosiński przeprowadzili wstępne badania wpływu, jaki mają przyznanie nagrody literackiej i zwycięstwo w konkursie literackim na dalszą twórczość pisarza⁵¹. PBL jest wartościowym źródłem danych bibliograficznych dla tak sformułowanego pytania badawczego, ponieważ dokumentuje ona wszystkie nagrody i konkursy literackie wraz z ich recepcją, o których informują źródła bibliograficzne (książki i czasopisma) opracowywane na potrzeby bibliografii. Badacze poddali analizie rekordy bibliograficzne z roczników PBL 1990–2001. Przygotowanie materiału badawczego podzielono na trzy etapy. Najpierw wygenerowano listę rekordów wydarzeń, które w bazie danych zdefiniowane są jako konkurs lub nagroda⁵². Drugi etap obejmował „podsumowanie wystąpień poszczególnych osób i wskazanie tego wydarzenia, w którym dana osoba pojawiła się po raz pierwszy; utożsamienie form nazewnictwa pochodzących z indeksu osobowego z literatami obecnymi w kartotece twórców (...)”⁵³. W trzecim kroku wygenerowano listę książek podmiotowych i tekstów literackich zamieszczonych w czasopismach, które publikowane były „w zdefiniowanych dziesięcioletnich przedziałach czasowych (4 lata przed laurem/zwycięstwem do 4 lat po)”, a których autorami byli laureaci nagród i zwycięzcy konkursów⁵⁴. Po zestawieniu tak przygotowanych danych bibliograficznych badacze wyciągnęli następujące wnioski:

⁵¹ Tomasz Umerle, Cezary Rosiński, *Bibliograficzna socjologia literatury...*, dz. cyt.

⁵² Tamże, s. 88.

⁵³ Tamże, s. 88–89.

⁵⁴ Tamże, s. 89.

- a) fakt zdobycia nagrody wiąże się ze wzrostem liczby publikacji, którego dynamika stopniowo maleje;
- b) dla większości twórców (od ponad 50 aż po 90%) ów wzrost jest bardzo znaczący (ponad 90%);
- c) znacząca część laureatów i wyróżnionych nie publikuje w ogóle w danym okresie⁵⁵.

Analiza danych bibliograficznych potwierdziła hipotezy badawcze autorów artykułu. Zwrócili oni także uwagę na to, że uzyskany wynik pokazuje, że badania oparte na danych bibliograficznych pochodzących z PBL „mogą pokazać pełniejszy obraz polskiej kultury literackiej”⁵⁶.

Drugim przykładem wykorzystania danych bibliograficznych pobranych z PBL do badania polskiego życia literackiego jest artykuł Macieja Maryła *Operationalising The Change. Dispersion of Polish Literary Life (1989–2002)*. Badacz śledził rozproszenie i ponowną centralizację polskiego życia literackiego po 1989 roku. W badaniu wykorzystał trzy rodzaje danych bibliograficznych: książki podmiotowe twórców polskich i obcych (przekłady na język polski wydane w Polsce); artykuły zdefiniowane jako teksty literackie publikowane w polskich czasopismach literackich; recepcję, czyli literaturę wtórną, na przykład recenzje, wywiady, streszczenia, polemiki itp.⁵⁷ Skorygował ponadto błędy i wzbogacił dane bibliograficzne o dodatkowe informacje⁵⁸, na przykład takie badawcze typologie jak: wydawnictwo duże i wydawnictwo małe oraz czasopismo duże i czasopismo małe.

Zasięg czasowy badanego materiału został ograniczony do lat 1989–2002 ze względu na to, że w czasie pracy nad artykułem tylko te roczniki PBL były dostępne w bazie danych⁵⁹. Marył przedsta-

⁵⁵ Tamże.

⁵⁶ Tamże, s. 97.

⁵⁷ Maciej Marył, *Operationalising The Change...*, dz. cyt., s. 518.

⁵⁸ Tamże.

⁵⁹ Tamże.

wił wybrane zasady metody bibliograficznej PBL, mające wpływ na kształt danych bibliograficznych, które analizował. Zwrócił uwagę na kompletność opracowywanych źródeł bibliograficznych, uwzględnianie w bibliografii pełnej twórczości (także nieliterackiej) wybitnych pisarzy polskich, obowiązującą w PBL definicję literatury, stosowanie słowników kontrolowanych (kartotek haseł wzorcowych) tylko w przypadku niektórych typów informacji i dublowanie niektórych opisów bibliograficznych⁶⁰. Mając świadomość tych ograniczeń, uznał PBL za szeroką, reprezentatywną próbę dotyczącą polskiego życia literackiego i stanowiącą podstawę do analizy⁶¹.

Badacz wykorzystał metody statystyczne, geograficzne i analizę sieciową. Statystyki dotyczące publikacji (książek i tekstów literackich zamieszczanych w czasopismach), aktywnych czasopism, wydawnictw i autorów publikujących książki lub utwory oraz recepcji literatury potwierdziły tezy innych badaczy dotyczące ponownej centralizacji życia literackiego. Dane pokazały, że w 1989 roku na rynku książki dominowały publikacje dużych wydawców, i tychże publikacji dotyczyła większość recepcji⁶². Następnie po 1991 roku rynek zdominowali mali wydawcy, a około 1995 roku trend się odwrócił i w ciągu dwóch lat znów najwięcej było omówień książek dużych wydawców⁶³. Maryl wykorzystał informacje o miejscach działalności wydawnictw oraz czasopism zawarte w rekordach bibliograficznych i zaprezentował przemiany życia literackiego w wymiarze geograficznym. Analiza pokazała, że w kolejnych latach analizowanego okresu stale rosła liczba małych ośrodków wydawniczych, co około roku 2002 skutkowało osiągnięciem podobnego poziomu w przypadku wydawców działających w dużych i małych miastach⁶⁴. Natomiast wnioski dotyczące czasopism były inne. Początkowo liczba czaso-

⁶⁰ Tamże, s. 517.

⁶¹ Tamże.

⁶² Tamże, s. 521.

⁶³ Tamże.

⁶⁴ Tamże, s. 526.

pism wydawanych w małych miastach trochę wzrosła, a później nastąpił jej spadek⁶⁵. W końcowej części tekstu Maryl wykorzystał narzędzia analizy sieciowej do zbadania życia literackiego jako sieci relacji i współzależności. Na podstawie danych bibliograficznych pochodzących z PBL odtworzył relacje łączące autorów z wydawcami, czasopismami i recenzentami. Sieć relacji między pisarzami i sieć relacji instytucjonalnych potwierdziły tezę o rozproszeniu i ponownej centralizacji życia literackiego⁶⁶.

Maryl nie ograniczył się jedynie do prezentowania statystyk na wykresach i ich omówienia. Aby lepiej zrozumieć przemiany zachodzące w polskim życiu literackim, wykorzystał miary z ekonomii, takie jak współczynnik Giniego, zastosowany tutaj do zbadania rozkładu krytycznej uwagi poświęcanej autorom, i indeks Herfindahla-Hirschmana, użyty do opisu koncentracji rynku. W ten sposób rozszerzył zakres zagadnień, które mogą być badane przy pomocy danych bibliograficznych.

Autor artykułu swoim badaniem udowodnił, że dane bibliograficzne pochodzące z PBL są bogate i pozwalają na przeprowadzenie analiz uwzględniających różne aspekty życia literackiego, ale mogą być też punktem wyjścia do bardziej pogłębionych badań. Pokazał, że analiza wykorzystująca dane PBL, wzbogacone o dodatkowe typy informacji, rozszerza horyzont badawczy. Ponadto potwierdził słuszność wykorzystania metod ilościowych do badania życia literackiego. Ich zastosowanie pozwala na analizę wymiarów tego zagadnienia wymykających się historykom literatury, którzy swoje rozważania opierają na wiedzy, obserwacji i doświadczeniu wynikającym z aktywnego uczestnictwa w życiu literackim.

Omówione przeze mnie przykłady badań wykorzystujących dane PBL pokazują, że dane te już teraz spełniają oczekiwania badaczy wykorzystujących metody cyfrowe, ale nie wszystkie. Dlatego twórcy PBL podejmują działania mające na celu polepszenie jakości zgro-

⁶⁵ Tamże.

⁶⁶ Tamże, s. 530.

madzonych danych i uzupełnienie bazy danych o rekordy pochodzące z retrokonwersji drukowanych tomów za lata 1944/1945–1988, a także o nowe rekordy z roczników dzielących ostatni opracowany rok (2012) i rok bieżący. Poza tym wydaje się, że przyszłością PBL jest wpisanie jej w krąg zagadnień z zakresu humanistyki cyfrowej. W tym celu podjęte zostały prace związane z przekształceniem PBL w bibliograficzną infrastrukturę badawczą. Kierunek tych zmian przedstawię w dalszej części rozdziału.

Hipermediacja „Polskiej Bibliografii Literackiej”

Powstanie pierwszej bazy danych „Polskiej Bibliografii Literackiej” zbiegło się w czasie z upowszechnieniem w Polsce internetu, dlatego warto oglądać przebieg remediacji PBL na tle przemian sieci internetowej. Za początek internetu w Polsce uznawany jest 17 sierpnia 1991 roku, kiedy nawiązano połączenie w protokole TCP/IP (*Transmission Control Protocol/Internet Protocol*) między Wydziałem Fizyki Uniwersytetu Warszawskiego a Uniwersytetem w Kopenhadze⁶⁷. Wówczas wysłany został pierwszy list elektroniczny⁶⁸. Na początku lat 90. XX wieku internet był w Polsce elitarny i miał niewielu użytkowników. Tworzyli go i korzystali z niego przede wszystkim naukowcy oraz pracownicy techniczni związani z uczelniami⁶⁹. Początkowo działające w Polsce sieci komputerowe oferowały niewiele usług, głównie pocztę elektroniczną i powiązane z nią listy dyskusyjne oraz system wymiany plików FTP (*File Transfer Protocol*)⁷⁰. Zdaniem Marty Juzy, badaczki zajmującej się początkami internetu w Polsce i analizą jego społecznych aspektów, prawdopodobnie największe

⁶⁷ Marta Juza, *Kultura Internetu w Polsce. Od akademickich początków do upowszechnienia zjawiska*, Wydawnictwo Antykwa, Kraków 2012, s. 13.

⁶⁸ Tamże.

⁶⁹ Tamże, s. 19.

⁷⁰ Tamże.

znaczenie miało pojawienie się systemu stron internetowych www (*World Wide Web*) w 1993 roku⁷¹. W tym czasie, określanym jako Web 1.0, możliwe było przede wszystkim odczytywanie treści stron internetowych⁷². Były one tworzone tak, by odczytywać mogli je ludzie, a nie maszyny (czyli programy komputerowe)⁷³. To okres, kiedy internet nadal miał niewielu użytkowników. W Polsce pierwsze badania na ten temat przeprowadzono w 1995 i 1996 roku na zlecenie Naukowej i Akademickiej Sieci Komputerowej⁷⁴. Wykazały one, że społeczność ta liczyła około 500 000 osób⁷⁵. W takich warunkach – niewielkiego jeszcze wykorzystania internetu w Polsce, który miał niewiele zastosowań i niewielu użytkowników – pod koniec 1997 roku bibliografowie wraz z programistami przystąpili do tworzenia bazy danych PBL i zapoczątkowali proces remediacji bibliografii.

Zespół autorski PBL postrzegał potencjalnego użytkownika bazy danych PBL jako użytkownika bibliografii rozumianej w sposób tradycyjny, z takimi samymi potrzebami informacyjnymi, oczekiwaniami wobec bibliografii i umiejętnościami wyszukiwania informacji⁷⁶, a nie jako użytkownika internetu czy bazy danych. Bibliografowie nie byli do końca świadomi możliwości wyszukiwawczych, jakie dawał wówczas i mógł dawać w przyszłości internet. W ten sposób powstała strona internetowa, która była efektem przeniesienia do przestrzeni cyfrowej tradycyjnej bibliografii, z jej układem, sposobem klasyfikowania materiałów bibliograficznych oraz ich wyszukiwania. Nie wykorzystano wszystkich narzędzi i funkcjonalności, które dawała baza danych PBL. Stworzono stronę internetową podobną do innych

⁷¹ Tamże, s. 21.

⁷² Oskar Pacelt, *Historia Internetu. Od Web 1.0 do Web 3.0 i 4.0*, <https://botland.com.pl/blog/historia-internetu-od-web-1-0-do-web-3-0-i-4-0/> (dostęp: 18.09.2025).

⁷³ Tamże.

⁷⁴ Marta Juza, *Kultura Internetu w Polsce...*, dz. cyt., s. 22.

⁷⁵ Tamże.

⁷⁶ Beata Domosławska, Zyta Szymańska, „*Polska Bibliografia Literacka*” on-line..., dz. cyt.

powstających w tamtym czasie – statyczną, umożliwiającą jedynie odczytanie jej treści.

W okresie nazywanym Web 1.0 treści internetowe tworzone z myślą o indywidualnym użytkowniku. Jednak wraz z rozwojem technologii zmieniły się możliwości, jakie oferował internet, nowy wymiar zyskały też potrzeby samych internautów. Kolejnym etapem w historii internetu jest Web 2.0, nazywany także drugą generacją sieci internetowej⁷⁷. Za jego początek w Polsce uznaje się rok 2004. Wtedy zaczęły powstawać strony internetowe współtworzone przez internautów, na przykład Wikipedia. Wokół portali internetowych kiełkowały społeczności ich użytkowników⁷⁸. Zdobywali oni wpływ na kształt stron internetowych, a dzięki nawiązywaniu kontaktów między internautami możliwa była wymiana informacji, także na temat oczekiwań wobec stron internetowych, ich architektury i funkcjonalności. Cechami sieci informatycznej w tym okresie stały się współpraca oraz współuczestnictwo⁷⁹.

Obecnie trwa trzecia generacja internetu, zwana Web 3.0. Jej celem jest tworzenie otwartych i połączonych ze sobą stron internetowych, których treść będzie rozumiana przez maszyny⁸⁰. W tym celu powstają sieci semantyczne (*semantic web*), które mają za zadanie stworzenie i rozpowszechnienie standardów opisywania treści w internecie, aby programy komputerowe mogły przetwarzać zgromadzone tam dane.

W przypadku PBL możemy mówić o pominięciu drugiej generacji internetu. Baza danych i strona internetowa powstały pod koniec lat 90. XX wieku, z wykorzystaniem możliwości, które wówczas oferowała sieć internetowa. Kolejne zmiany związane z powstaniem i rozwojem portali społecznościowych, współtworzeniem treści internetowych przez użytkowników, a także przetwarzaniem danych zostały dostrzeżone przez zespół autorski PBL. Bibliografowie

⁷⁷ Oskar Pacelt, *Historia Internetu...*, dz. cyt.

⁷⁸ Tamże.

⁷⁹ Tamże.

⁸⁰ Tamże.

i Maciej Matysiak, twórca oraz administrator – z firmy Advis – bazy danych PBL, mieli plany modernizacji bazy danych i strony, zakładające przekształcenie jej w portal naukowo-społecznościowy⁸¹. Celem autorów tej koncepcji było między innymi przebudowanie bazy danych PBL w sposób umożliwiający analizowanie treści bibliografii i pozyskiwanie nowej wiedzy za pomocą metod automatycznej eksploracji danych (*data mining*), dodanie do bazy odnośników do pełnych tekstów dokumentowanych publikacji dostępnych w zasobach cyfrowych, przekształcenie PBL w portal naukowo-badawczy oferujący powiadomienia o nowych publikacjach i wydarzeniach kulturalnych, możliwość uzupełniania PBL o autorskie materiały, personalizację treści portalu⁸². Zgodnie z zamysłem autorów projektu PBL miała stać się platformą umożliwiającą osobom zainteresowanym humanistyką nawiązywanie nowych kontaktów, współpracę oraz współtworzenie omawianej bibliografii. Autorom koncepcji nie udało się jednak pozyskać finansowania na realizację tego projektu naukowego⁸³.

Przez ponad dwadzieścia lat strona internetowa PBL oraz baza danych dostępne były dla użytkowników i pracowników w postaci nadanej im pod koniec lat 90. Strona internetowa PBL (<https://pbl.ibl.poznan.pl/dostep/>) w dalszym ciągu jest stroną statyczną, bez możliwości nawiązywania interakcji i pobierania danych bibliograficznych.

⁸¹ „Polska Bibliografia Literacka” – rozbudowa bazy i szkolenie nowych kadr, digitalizacja archiwalnych roczników oraz budowa zintegrowanego systemu informatycznego do udostępniania bazy danych w środowisku internetowego portalu naukowo-społecznościowego, Materiały Pracowni Bibliografii Bieżącej IBL PAN w Poznaniu. [Dokument dostępny w Pracowni Bibliografii Bieżącej].

⁸² Tamże.

⁸³ Pracownia Bibliografii Bieżącej IBL PAN w 2012 roku w ramach „Narodowego Programu Rozwoju Humanistyki” złożyła wniosek o przyznanie środków na realizację projektu „Polska Bibliografia Literacka» – rozbudowa bazy i szkolenie nowych kadr, digitalizacja archiwalnych roczników oraz budowa zintegrowanego systemu informatycznego do udostępniania bazy danych w środowisku internetowego portalu naukowo-społecznościowego”. Projekt nie uzyskał finansowania.

Dopiero w 2015 roku zapoczątkowano proces przekształcania tej bibliografii w bibliograficzną infrastrukturę badawczą, wiążący się z wykorzystaniem narzędzi do importu i przetwarzania danych oraz technik uczenia maszynowego na potrzeby tworzenia bazy danych PBL, ale także na potrzeby badaczy chcących wykorzystywać dane bibliograficzne w swojej pracy naukowej. Częścią tego procesu jest również wdrożenie międzynarodowych standardów dotyczących danych. W ramach tej transformacji powstała nowa strona internetowa PBL (<https://pbl.ibl.waw.pl/>) spełniająca współczesne wymogi. Obecnie obie strony internetowe PBL funkcjonują równolegle. Projektowanie nowej strony internetowej PBL poprzedziły badania ankietowe użytkowników.

W lutym 2015 roku na stronie internetowej PBL została umieszczona ankieta, której celem było stworzenie profilu użytkownika bazy danych PBL i pozyskanie informacji na temat jego potrzeb⁸⁴. Internauci zgłosili uwagi dotyczące tego, w jaki sposób zbudowana jest baza danych PBL i jakie możliwości wyszukiwawcze oferuje. Beata Koper i Tomasz Umerle w artykule *Nowa bibliografia dla nowej humanistyki. O przemianach funkcji bibliografii naukowej*⁸⁵ omówili wyniki tego sondażu po uzyskaniu odpowiedzi 434 respondentów. Pisali o tym, że

analiza ankiety pozwoliła określić istotne z perspektywy użytkowników wady systemu (nieczytelny i nieintuicyjny interfejs użytkownika, problemy z wyszukiwarką) oraz najistotniejsze potrzeby użytkowników (zaawansowana wyszukiwarka, przekierowania do pełnotekstowych wersji publikacji, pobieranie rekordów w odpowiednim stylu cytowania, przekierowania do katalogów bibliotecznych)⁸⁶.

⁸⁴ Beata Koper, Tomasz Umerle, *Nowa bibliografia dla nowej humanistyki...*, dz. cyt., s. 506.

⁸⁵ Tamże.

⁸⁶ Tamże.

Już wówczas wśród innych potrzebnych funkcji wymienionych przez użytkowników pojawiały się także: narzędzia do pobierania danych, narzędzie do zapisywania wyników wyszukiwania, narzędzia bibliometryczne i narzędzie do raportowania błędów, jakkolwiek wskazywała na nie mniejszość ankietowanych⁸⁷. Badanie to było wstępem do zmian, jakie później zaplanowano w PBL. Uwagi ankietowanych zostały uwzględnione w czasie pracy nad nową stroną internetową PBL.

Dotychczasowa strona internetowa PBL oferuje możliwość przeprowadzenia prostego wyszukiwania i zawężenia wyników do określonego rocznika lub zakresu roczników. W takiej formie nie jest atrakcyjna dla użytkowników przyzwyczajonych do funkcjonalności, jakie dają wyszukiwarki internetowe i elektroniczne katalogi biblioteczne. Dlatego nowa strona internetowa omawianej bibliografii posiada opcje takie jak: wyszukiwanie proste i zaawansowane, zawężanie wyników wyszukiwania za pomocą filtrów, sortowanie wyników wyszukiwania według zdefiniowanych kategorii. Zmieniła się także postać, w jakiej prezentowane są wyniki wyszukiwania. Długą listę wyników w układzie odpowiadającym schematom redakcji działów osobowych i rzeczowych zastąpił wykaz rekordów uszeregowanych według trafności wyszukiwania i podzielonych na mniejsze części (do wyboru strony po 25, 50 i 100 rekordów). Plany na przyszłość obejmują wzbogacenie rekordów bibliograficznych o linki do pełnych tekstów.

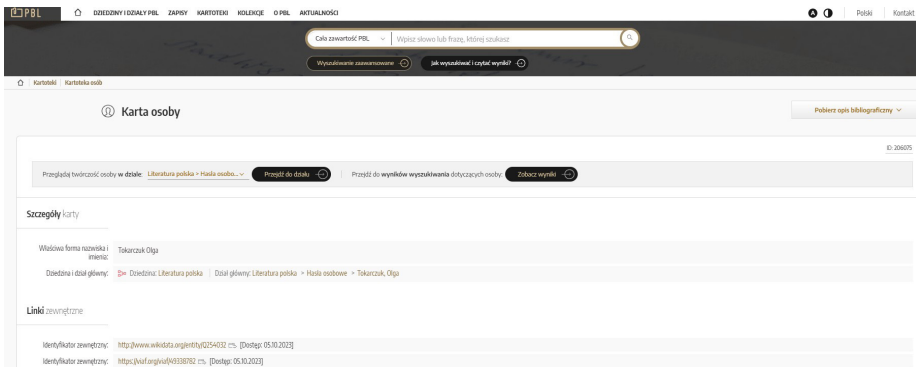
Baza danych PBL powstała w połowie 1998 roku, ale zmiany w jej funkcjonowaniu wprowadzane były w kolejnych latach i dzieje się to nadal. W 2015 roku zmienił się jednak charakter tych zmian. Wcześniej zespół autorski PBL i administrator opracowywali i wdrażali ulepszenia, które ułatwiały i przyspieszały pracę bibliografów, ale nie wykorzystywali oni w pełni możliwości, jakie dawały baza danych i nowy nośnik – internet. Dopiero w ramach wspomnianego kilkakrotnie w tej książce projektu „»Polska Bibliografia Literacka« –

⁸⁷ *Ankieta PBL – raport SPSS*, Materiały Pracowni Bibliografii Bieżącej IBL PAN w Poznaniu. [Dokument dostępny w Pracowni Bibliografii Bieżącej].

laboratorium wiedzy o współczesnej kulturze polskiej” realizowanego w latach 2015–2018 zapoczątkowana została kolejna transformacja PBL. Wówczas też rozpoczęła się refleksja dotycząca bibliograficznej bazy danych jako zbioru danych badawczych. Konsekwencją tego projektu było także wykorzystanie narzędzi cyfrowych w pracach dokumentacyjnych.

W 2019 roku w bazie danych PBL wdrożono importer służący do pobierania opisów bibliograficznych książek wchodzących w zakres PBL z bazy danych Biblioteki Narodowej w Warszawie przy pomocy interfejsu programistycznego API. Tak pozyskane dane są automatycznie dodawane do bazy danych PBL wykorzystującej oprogramowanie Oracle i poddane wstępnej klasyfikacji, zgodnie z siatką działów osobowych i rzeczowych PBL. Wymagają one wprowadzenia dodatkowych informacji i akceptacji przez bibliografów. Następnie dane są publikowane na starej stronie internetowej PBL. Importer został stworzony przez Macieja Matysiaka z Advis, współtwórcę i administratora bazy danych PBL. Natomiast w procesie tworzenia koncepcji tego narzędzia i testowaniu rozwiązań informatycznych udział brał Cezary Rosiński z Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN. Zaletami tego rozwiązania są oszczędność czasu (wcześniej informacje z „Przewodnika Bibliograficznego” wprowadzano do bazy danych ręcznie) i zmniejszenie liczby błędów. Prace nad automatyzacją pozyskiwania rekordów bibliograficznych z bazy danych Biblioteki Narodowej kontynuowano dzięki pozyskanej w 2020 roku dotacji na utrzymanie specjalnej infrastruktury informatycznej (SPUBi)⁸⁸. Ich efektem jest opracowanie przez Poznańskie Centrum Superkomputerowo-Sieciowe i wdrożenie uniwersalnego importera rekordów bibliograficznych z bazy danych Biblioteki Na-

⁸⁸ Projekt „Infrastruktura bazodanowa Polskiej Bibliografii Literackiej – utrzymanie” zrealizowany ze środków Ministerstwa Nauki i Szkolnictwa Wyższego w ramach dotacji na utrzymanie specjalnej infrastruktury informatycznej (SPUBi/SP/458272/2020) w latach 2020–2022. Koordynator: dr Tomasz Umerle (w latach 2020–2021), dr Cezary Rosiński (w latach 2021–2022).



Fot. 16. Widok hasła wzorcowego Olgi Tokarczuk
 Źródło: Materiały własne

rodowej, który dostarcza rekordy bibliograficzne bezpośrednio na nową stronę internetową PBL bez pośrednictwa bazy danych PBL i starej strony internetowej PBL⁸⁹. W pracach nad koncepcją tego narzędzia i w testowaniu rozwiązań informatycznych ponownie brał udział Cezary Rosiński. Kolejnym ułatwieniem pracy bibliografów jest wykorzystanie na większą skalę języka zapytań SQL. Służy on zespołowi do przygotowywania zestawów danych badawczych i wyszukiwania w bazie nieprawidłowości. Często powtarzające się błędy są usuwane poprzez nadpisywanie poprawnych informacji. Dzięki temu w przypadku niektórych błędów autorzy bibliografii nie muszą ręcznie wprowadzać poprawek. Usprawniło to pracę nad rocznikami.

Do nowej strony internetowej PBL dodawane są nowe funkcjonalności, dzięki którym użytkownicy będą mogli zestawiać i wizualizować dane bibliograficzne, a dostęp do nich będzie łatwiejszy.

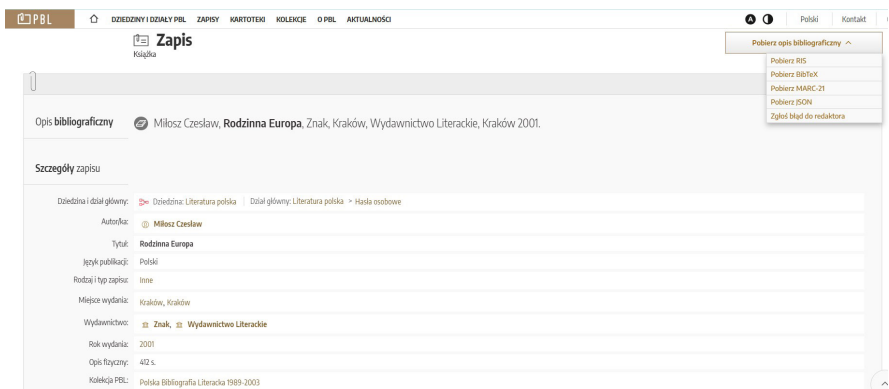
⁸⁹ Beata Koper, Tomasz Umerle, *Bibliografie cyfrowe*, w: *Panorama literaturoznawstwa cyfrowego*, red. Maciej Maryl, Bartłomiej Szleszyński, Tomasz Umerle, Marta Błaszczyńska, Instytut Badań Literackich PAN, Warszawa 2023, <https://nplp.pl/panorama-literaturoznawstwa-cyfrowego/bibliografie-cyfrowe/> (dostęp: 18.09.2025).

W ramach wspomnianego już projektu realizowanego dzięki dotacji na utrzymanie specjalnej infrastruktury informatycznej (SPUBi) na nowej stronie internetowej PBL zaimplementowano oprogramowanie Kibana służące do analizy i wizualizacji danych. Obecnie prowadzone są prace w kolejnym projekcie⁹⁰, które mają na celu między innymi udostępnienie danych bibliograficznych PBL za pomocą interfejsu programistycznego API.

Transformacja PBL w bibliograficzną infrastrukturę badawczą wiąże się także z wdrożeniem międzynarodowych standardów dotyczących danych. Jednym z nich jest wzbogacenie danych bibliograficznych o trwałe identyfikatory. W 2023 roku hasła wzorcowe osób dostępne na nowej stronie internetowej PBL zostały wzbogacone o trwałe identyfikatory VIAF oraz identyfikatory z Wikidanych, czyli otwartego repozytorium zbierającego uporządkowane dane na potrzeby Wikipedii. Przykładem jest hasło wzorcowe Olgi Tokarczuk, widoczne na fotografii 16.

Jest to krok w kierunku włączenia bazy danych PBL do globalnego obiegu danych. Aby dane PBL były atrakcyjne dla badaczy, bibliografowie muszą zadbać również o to, by rekordy bibliograficzne powstawały zgodnie z zasadami FAIR. Zatem powinny być łatwe do znalezienia, dostępne, interoperacyjne oraz nadające się do ponownego wykorzystania. Krokiem w tym kierunku jest wspomniane już wzbogacenie PBL o trwałe identyfikatory. Wartościowym byłoby także dodanie w przyszłości identyfikatorów VIAF instytucji oraz identyfikatorów wydarzeń (jeśli dostępne będą odpowiednie repozytoria). Realizacja drugiej zasady, mówiącej o dostępności danych, odbywa się poprzez udostępnienie danych PBL w ramach otwartej licencji. Dane bibliograficzne PBL są danymi otwartymi. Bibliografowie przeprowadzają kwerendy i przygotowują zestawy danych dla badaczy

⁹⁰ Projekt „Infrastruktura bazodanowa Polskiej Bibliografii Literackiej – utrzymanie” realizowany ze środków Ministerstwa Nauki i Szkolnictwa Wyższego w ramach dotacji na utrzymanie specjalnej infrastruktury informatycznej (SPUBi/SP/597223/2024) w latach 2024–2026. Koordynator: dr Cezary Rosiński.



Fot. 17. Widok pojedynczego rekordu książki Czesława Miłosza Rodzina Europa
 Źródło: Materiały własne

oraz prowadzą prace nad wdrożeniem na nowej stronie internetowej PBL mechanizmu służącego do udostępniania danych bibliograficznych PBL za pomocą interfejsu programistycznego API. Działania na rzecz interoperacyjności danych PBL skupiają się na udostępnieniu danych w postaci umożliwiającej ich przetworzenie w zewnętrznych programach komputerowych. I tu ważne jest stosowanie powszechnie używanego formatu danych. PBL posiada indywidualny format danych, ale w ramach projektu realizowanego z dotacji na utrzymanie specjalnej infrastruktury informatycznej (SPUBi) „zaimplementowano mapowanie oraz eksport danych w formatach MARC 21 oraz formatach cytowani RIS i Bibtex” i dodano możliwość eksportu danych w formacie JSON-LD⁹¹. Na zdjęciu 17 pokazany został przycisk z rozwijaną listą formatów, w których można pobrać rekord.

Ostatnia zasada mówi o tym, że dane powinny nadawać się do ponownego wykorzystania. By tak było, muszą być dobrze opisane. Dlatego ważne jest, aby nową stroną internetową PBL uzupełnić o opis stosowanej metody bibliograficznej – tak, by użytkownicy

⁹¹ Beata Koper, Tomasz Umerle, *Bibliografie cyfrowe...*, dz. cyt.

mieli dostęp do wiedzy na temat tego, jakie dane gromadzi PBL, jakich w bibliografii nie ma, w jaki sposób powstają.

Transformacja PBL, którą opisałam, spełnia definicję hipermediacji, czyli jednej ze strategii remediacji, polegającej na tym, że bibliografia jest prezentowana na nowym nośniku (nadal chodzi o internetową bazę danych), ale w większym stopniu wykorzystuje możliwości oferowane przez nowe medium. Proces ten jeszcze nie dobiegł końca. PBL jest na początku drogi do przekształcenia bazy danych – wykorzystującej metodę sformułowaną w pierwszych latach opracowywania PBL, prezentującej bibliografię w postaci znanej z jej poprzedniej, drukowanej wersji i spełniającej funkcję informacyjną na takim samym poziomie jak tradycyjna bibliografia – w bibliograficzną infrastrukturę badawczą. Ale modernizacja bazy danych PBL i nowa strona internetowa to nie wszystko. Zespół Pracowni Bibliografii Bieżącej IBL PAN podejmuje też działania mające na celu integrację danych PBL z innymi zasobami bibliograficznymi.

Dane bibliograficzne „Polskiej Bibliografii Literackiej” w międzynarodowej bazie danych European Literary Bibliography / Europejskiej Bibliografii Literackiej

Międzynarodowa baza danych European Literary Bibliography / Europejska Bibliografia Literacka powstaje w ramach projektu „Literary Bibliography Research Infrastructure” (LIBRI) realizowanego wspólnie przez Instytut Badań Literackich PAN i Instytut Literatury Czeskiej Akademii Nauk Republiki Czeskiej. Od 2021 roku dostępna jest wersja beta portalu. Grant ma na celu gromadzenie, łączenie i udostępnianie na wspólnym portalu internetowym danych bibliograficznych dotyczących literatury i literaturoznawstwa pochodzących z różnych krajów. Powstaje baza danych, w której poprzez dane bibliograficzne prezentowany będzie współczesny europejski system

literacki⁹². Twórcy projektu planują również rozwijać współpracę między instytucjami tworzącymi zasoby bibliograficzne i badaczami wykorzystującymi dane bibliograficzne w badaniach naukowych⁹³. W ramach tego zadania udoskonalana będzie infrastruktura dla badań opartych na danych badawczych.

Projekt ten realizuje ideę ponownego wykorzystania istniejących zasobów bibliograficznych – katalogów bibliotecznych, bibliografii narodowych i bibliografii dziedzinowych⁹⁴. W pierwszym etapie realizacji projektu baza danych została zasilona rekordami pochodzącymi z dwóch bibliografii literackich – czeskiej (Česká literární bibliografie) i polskiej („Polska Bibliografia Literacka”) – jednak zespół realizujący grant chce rozszerzać ją o kolejne zasoby⁹⁵. Do bazy dodano także rekordy z Biblioteki Narodowej Finlandii i Biblioteki Narodowej Hiszpanii.

Scalenie rekordów pochodzących z różnych zasobów, przygotowywanych według różnych metod bibliograficznych, wymagało od twórców bazy European Literary Bibliography / Europejskiej Bibliografii Literackiej wypracowania otwartego modelu danych, który można rozbudować i który pozwoli na unifikację rekordów bibliograficznych zasilających bazę. Rekordy przekazane do omawianej bazy ujednolicono w oparciu o wspólny system klasyfikacji opracowany na potrzeby tego projektu⁹⁶. W tym celu zaprojektowano specjalne narzędzia służące do analizy słowników kontrolowanych, systemów klasyfikacji i trwałych identyfikatorów stosowanych w bazach zasilających European Literary Bibliography / Europejską Bibliografię Literacką⁹⁷. Identyfikacja danych w dużej mierze opierała się na ist-

⁹² Tomasz Umerle, Vojtěch Malínek, *Literarybibliography.eu...*, dz. cyt., s. 583.

⁹³ Tamże, s. 591.

⁹⁴ Tamże, s. 583.

⁹⁵ Tamże, s. 587.

⁹⁶ Tamże, s. 583.

⁹⁷ Tamże.

niejących klasyfikacjach, takich jak: Klasyfikacja Dziesiąta Deweya, Uniwersalna Klasyfikacja Dziesiąta (UKD), oraz na słownikach kontrolowanych stosowanych przez instytucje tworzące poszczególne bibliografie⁹⁸. Ułatwiło to wyodrębnienie danych z zasobów opracowywanych w różnych językach⁹⁹. Harmonizacja danych polegała również na zastosowaniu wspólnego formatu danych. W tym przypadku jest to format MARC 21¹⁰⁰. Dane bibliograficzne zostały wzbogacone o trwałe identyfikatory osób VIAF i miejsc GeoNames oraz tematy stosowane w Bibliotece Kongresu (LOC *Subject Headings*)¹⁰¹. Hasła wzorcowe osób i miejsc połączono z Wikidanyimi¹⁰². W ten sposób hasła wzorcowe uzupełniono o dane ze źródeł zewnętrznych oraz informacje, których nie było w bazach zasilających European Literary Bibliography / Europejską Bibliografię Literacką, na przykład informacje biograficzne czy dotyczące historii miast i ich nazw, zmieniających się na przestrzeni czasu¹⁰³.

Plany na przyszłość obejmują: wprowadzenie do bazy danych nowych typów haseł wzorcowych (instytucji, tematów); implementację modelu danych kompatybilnego z RDF (*Resource Description Framework*), czyli językiem pozwalającym na opisywanie zasobów internetowych służącym ich przedstawieniu tak, by były zrozumiałe dla komputerów i nadawały się do przetworzenia przez programy komputerowe; automatyczne generowanie metadanych; wprowadzenie do bazy rekordów pochodzących z retrokonwersji drukowanych zasobów bibliograficznych¹⁰⁴.

⁹⁸ Tamże, s. 587.

⁹⁹ Tamże.

¹⁰⁰ Tamże.

¹⁰¹ Patryk Hubar, Marcin Giersz, *European Literary Bibliography: Aggregation and Harmonization of Literary Bibliographical Data*, <https://zenodo.org/record/7705026> (dostęp: 18.09.2025).

¹⁰² Tamże.

¹⁰³ Tamże.

¹⁰⁴ Tamże.

Serwis został zbudowany z użyciem otwartoźródłowego oprogramowania VuFind, które dostosowano do potrzeb tego projektu¹⁰⁵. Opracowano nową architekturę bazy danych i interfejs użytkownika. Rozwinęto kartotekę haseł wzorcowych, by w przyszłości mogła być wykorzystywana nie tylko do katalogowania materiałów zasila-
jących bazę danych, ale także by można było zebrane w bazie dane wzbogacać dużą ilością danych ułatwiających identyfikację i interpretację wyników wyszukiwania¹⁰⁶. Takie wykorzystanie kartotek haseł wzorcowych ma pomóc w zrozumieniu systemu literackiego. Dodano także moduł służący do wizualizacji danych¹⁰⁷. Składa się on ze statystyk prezentowanych w postaci tabel i wykresów¹⁰⁸ oraz z map, na których znajdują się informacje geograficzne zawarte w rekordach bibliograficznych. Wyszukiwarka spełnia również oczekiwania użytkowników, którzy chcą przetwarzać dane pochodzące z omawianej bazy danych. Mogą oni tworzyć, edytować i uruchamiać kod w języku programowania Python¹⁰⁹.

Jak już pisałam, PBL jest jednym z zasobów zasilających bazę European Literary Bibliography / Europejską Bibliografię Literacką. Między innymi na podstawie metody bibliograficznej PBL opracowano model danych międzynarodowego agregatora. W związku z tym, że model ten jest wypadkową rozwiązań stosowanych w polskiej i czeskiej bibliografii, dane PBL trzeba było dostosować do potrzeb tego projektu. Konieczne okazało się przede wszystkim przeprowadzenie mapowania formatu danych PBL na format MARC 21. Ze względu na to, że format stosowany w PBL odpowiada na indywidualne potrzeby tej bibliografii i nie istnieją powszechnie dostępne narzędzia służące do automatycznej konwersji formatów danych, jego mapowanie wiązało się z dodatkowym nakładem pracy. Ale

¹⁰⁵ Tomasz Umerle, Vojtěch Malínek, *Literarybibliography.eu...*, dz. cyt., s. 588.

¹⁰⁶ Tamże.

¹⁰⁷ Tamże.

¹⁰⁸ Tamże.

¹⁰⁹ Tamże.

przykład ten pokazuje, że własny format danych nie jest ograniczeniem, po odpowiednim przygotowaniu można łączyć różne zasoby bibliograficzne.

Podsumowanie

Proces hipermediacji „Polskiej Bibliografii Literackiej”, który właśnie się odbywa, związany jest z zastosowaniem narzędzi cyfrowych w pracy bibliografów, stworzeniem nowej strony internetowej PBL posiadającej funkcje ułatwiające precyzyjne i szybkie wyszukiwanie informacji oraz przetworzenie i pobranie danych, wdrożeniem międzynarodowych standardów oraz dostosowaniem danych bibliograficznych PBL do potrzeb badań naukowych. Wszystkie te działania są istotne, ale biorąc pod uwagę kierunek, w którym rozwijają się współczesne zasoby bibliograficzne, należy uznać, że przyszłością PBL jest humanistyka cyfrowa. PBL to bogaty zbiór danych bibliograficznych, których potencjał nie został jeszcze w pełni odkryty przez naukowców. Niektórzy użytkownicy, także badacze, postrzegają ją nadal jako narzędzie pomocnicze literaturoznawstwa. Jednak zaprezentowane przeze mnie w tym rozdziale działania podejmowane przez bibliografów zbliżają PBL do przekształcenia w bibliograficzną infrastrukturę badawczą. Chcę również pokreślić, że bibliografia ta powinna rozwijać się w otoczeniu zagadnień będących w kręgu zainteresowania badaczy uprawiających humanistykę cyfrową.

Dlatego dobrym krokiem jest modernizacja, która ma na celu przekształcenie PBL udostępnianej na starej stronie internetowej, mającej cechy zamkniętego serwisu, który umożliwia jedynie odczyt treści, w bibliografię otwartą, dającą możliwości wyszukiwawcze, takie, jakie mają wyszukiwarki internetowe i katalogi biblioteczne, oraz oferującą zestawy danych badawczych przygotowane w sposób umożliwiający ich pobranie i przetworzenie. Serwisem otwartym jest nowa strona internetowa PBL. Poza tym zespół Pracowni Bibliografii Bieżącej Instytutu Badań Literackich PAN nie może ograniczać swych

działań jedynie do opracowania bibliografii i jej modernizacji. Konieczne są także świadczenie usług na rzecz środowiska naukowego w zakresie przygotowywania danych badawczych, oferowanie szkoleń i promocja wykorzystania danych bibliograficznych w pracach naukowych.

Na koniec chcę zwrócić uwagę na to, że hipermediacja PBL wiąże się nie tylko z modernizacją bazy danych, ale skutkuje także zmianą charakteru pracy bibliografów. Mierzą się oni z nowymi wyzwaniami, które wymagają od nich już nie tylko merytorycznej wiedzy literaturoznawczej i umiejętności obsługi oprogramowania służącego do tworzenia rekordów.

Powinni być otwarci na nowe rozwiązania, porzucić stosowane przez lata schematy postępowania i poszukiwać nowych metod oraz zdobywać kompetencje cyfrowe. Dlatego zespół opracowujący PBL tworzą zarówno bibliografowie zajmujący się wprowadzaniem rekordów bibliograficznych do bazy danych, jak i specjaliści posiadający praktyczne umiejętności programistyczne. Bibliografowie w pewnym stopniu przestają być specjalistami od dokumentacji literatury i stają się specjalistami do spraw danych badawczych. Ich praca nadal ma charakter pomocniczy, ale teraz działają również na rzecz rozwoju infrastruktury badawczych i badań wykorzystujących narzędzia cyfrowe. Sami są badaczami opisującymi system literacki przy pomocy danych bibliograficznych.