

**„NATIONAL BIBLIOGRAPHIES AND CATALOGS CURATION AND
RESEARCH: DIGITAL HUMANITIES OUTLOOK”
DARIAH Bibliographical Data WG Meeting: an open
online workshop
(30 września 2021 r.)**

Celem Digital Research Infrastructure for the Arts and Humanities (DARIAH), która została powołana w sierpniu 2014 r. jako konsorcjum na rzecz europejskiej infrastruktury badawczej (ang. European Research Infrastructure Consortium, ERIC), jest wspieranie wykorzystania nowych technologii w badaniach z zakresu humanistyki. W jej pracach uczestniczy dwudziestu członków, jeden obserwator, a kilka instytucji z sześciu kolejnych państw to partnerzy współpracujący z DARIAH (DARIAH-EU, b.d.-b). W ramach konsorcjum funkcjonuje obecnie 20 grup roboczych, które koncentrują się na wybranych zagadnieniach badawczych (DARIAH-EU, b.d.-c). Jedną z takich grup, nazwaną Bibliographical Data Working Group (BiblioData WG), została stworzona, aby „wspierać współpracę pomiędzy wszystkimi stronami zaangażowanymi w cykl życia danych bibliograficznych (zwłaszcza producentami danych, menedżerami/kuratorami, badaczami oraz teoretykami dokumentacji i bibliografii), co przyczyni się do większego i bardziej zróżnicowanego wykorzystania danych bibliograficznych w naukach humanistycznych” (DARIAH-EU, b.d.-a) Pracami grupy kierują Tomasz Umerle (Instytut Badań Literackich Polskiej Akademii Nauk) oraz Vojtěch Malínek (Instytut Literatury Czeskiej Akademii Nauk Republiki Czeskiej). Wśród podejmowanych inicjatyw znajduje się również organizacja warsztatów i spotkań poświęconych problemom związanym z bibliografią.

Jedno z takich spotkań odbyło się 30 września 2021 r. w formie zdalnej¹. Było ono podzielone na trzy części. Podczas pierwszej z nich zaprezentowano dwa uzupełniające się spojrzenia na dane bibliograficzne: przedstawiciele grupy roboczej DARIAH oraz Sekcji ds. Bibliografii IFLA. Następnie pracownicy sześciu bibliotek narodowych (czeskiej, fińskiej, niemieckiej, polskiej, szwedzkiej i węgierskiej) zaprezentowali wybrane zagadnienia związane z bibliografiami narodowymi, a w części ostatniej odbył się panel dyskusyjny z ich udziałem.

Uczestników spotkania powitała Francesca Morselli (DARIAH-ERIC). Jego moderatorem był Vojtěch Malínek. W otwierającym pierwszą część warsztatów wystąpieniu Tomasz Umerle i Mikko Tolonen (DARIAH Bibliographical Data Working Group) przedstawili zakres działań podejmowanych w ramach grupy roboczej DARIAH, poświęconej danym bibliograficznym. Zaprezentowali również podstawowe założe-

¹ Nagrania i prezentacje mają zostać zamieszczone odpowiednio w serwisach YouTube i Zenodo.

nia „bibliographic data science”². Zdaniem Tolonena dane pochodzące z katalogów bibliotecznych mają duży potencjał badawczy, szczególnie w kontekście prac dotyczących okresu używania ręcznej prasy drukarskiej. Chociaż opisy bibliograficzne nie są przystosowane do analiz ilościowych, to wstępne przetwarzanie danych (kontrola jakości) pozwala na ich użycie w badaniach. Prelegent podkreślił, że konieczne jest stworzenie otwartego ekosystemu danych bibliograficznych, w ramach którego będą one mogły być łączone z innymi materiałami, np. pełnymi tekstami dokumentów. Kolejną prelegentką była Mathilde Koskas (Biblioteka Narodowa Francji, przewodnicząca Sekcji ds. Bibliografii IFLA), która wraz z Pat Rivą (Uniwersytet Concordia) przygotowała prezentację ukazującą stan bibliografii narodowych w oparciu o ich rejestr prowadzony przez IFLA (<https://www.ifla.org/g/bibliography/national-bibliographic-register/>). Spis ten jest opracowywany na podstawie ankiet wypełnianych przez biblioteki. Koskas podkreśliła, że rejestr uwzględnia przede wszystkim bibliografie europejskie, dodała, że respondenci odpowiadający na pytania nie zawsze rozumieli je w jednakowy sposób.

W drugiej części spotkania swoje wystąpienia zaprezentowali przedstawiciele bibliotek narodowych. Rozpoczęło ją wystąpienie Osmy Suominena (Biblioteka Narodowa Finlandii) przygotowane wspólnie z Mikko K.A. Lappalainenem i Mirją Anttilą. Było ono poświęcone publikacji słownictwa kontrolowanego i opisów z fińskiej bibliografii narodowej zgodnie z modelem danych powiązanych. Podjęte prace polegały na przekształceniu rekordów zapisanych w formacie MARC 21 w dane powiązane. Prelegent zaprezentował model danych przyjęty w projekcie. Przedstawił także relacje z innymi zbiorami opublikowanymi w sieci WWW oraz możliwości języka zapytań SPARQL. W następnej prezentacji Angela Vorndran (Niemiecka Biblioteka Narodowa) przedstawiła niemiecką bibliografię narodową oraz różne sposoby dostępu do danych bibliograficznych przygotowane przez Niemiecką Bibliotekę Narodową. Jedną z usług przygotowanych specjalnie dla badaczy jest DNBLab (https://www.dnb.de/DE/Professionell/Services/WissenschaftundForschung/DNBLab/dnblab_node.html), który umożliwia pobieranie tematycznych zestawów danych takich jak: spisy treści, rozprawy lub artykuły. W kolejnym wystąpieniu Kamil Pawlicki (Biblioteka Narodowa) zaprezentował zintegrowaną wyszukiwarkę OMNIS, której celem jest zapewnienie dostępu do zasobów polskich bibliotek. Przedstawiciel polskiej księżnicy narodowej omówił sposób transformacji rekordów zgodnie z modelem FRBR-LRM. Dane bibliograficzne i wzorcowe dostępne są poprzez serwis API (<https://data.bn.org.pl/>).

² Termin ten nie ma jeszcze polskiego odpowiednika. Pojawił się w publikacjach fińskich badaczy Leo Lahtiego i Mikko Tolonena w 2019 r. (zob. np. Lahti i in., 2019).

W 2021 r. odnotowano dotychczas od 1 do 2,6 mln zapytań i od 50 do 85 tys. unikatowych użytkowników miesięcznie.

Temat wystąpienia Marii Haškovcovej (Biblioteka Narodowa Republiki Czeskiej) odbiegał od dotychczas przedstawionych prezentacji. Czeska bibliotekarka zaprezentowała prace nad archiwizacją czeskich stron internetowych. Prawo w Republice Czeskiej umożliwia zapis czeskich stron jedynie w celu ich archiwizacji. W związku z tym całość archiwum jest dostępna dla użytkowników wyłącznie na miejscu w bibliotece. Archiwizowane zasoby są katalogowane przy użyciu specjalnego narzędzia WA-KAT, które umożliwia ich półautomatyczne opracowanie. Następną prezentację przedstawili Szabina Ilácsa i Márton Németh (Państwowa Biblioteka Széchényiego), którzy skupili się na semantycznym wzbogacaniu metadanych archiwizowanych zasobów internetowych. Tę część spotkania zakończyło wystąpienie Ylvy Sommerland (Biblioteka Królewska w Sztokholmie) poświęcone możliwości wykorzystania i wizualizacji danych. Przedstawione przykłady dotyczyły m.in. języków, na które tłumaczone są prace autorów szwedzkich oraz oryginalnych języków publikacji wydawanych w Szwecji.

Na zakończenie spotkania przedstawiciele bibliotek narodowych wzięli udział w panelu dyskusyjnym, podczas którego uczestnicy odpowiedzieli na trzy pytania. Pierwsze z nich, dotyczyło obecnego stanu i ograniczeń związanych z wykorzystaniem danych bibliograficznych. K. Pawlicki podkreślił dużą liczbę zapytań kierowanych do API Biblioteki Narodowej oraz wymienił przykłady użycia danych przez twórców oprogramowania dla bibliotek publicznych. Zainteresowanie rekordami wykazują również wydawcy, księgarze oraz hurtownicy. Przedstawiciel Biblioteki Narodowej wymienił dwa ograniczenia: równoległe funkcjonowanie w Polsce dwóch języków informacyjno-wyszukiwawczych (DBN i KABA) oraz sposób przesyłania egzemplarza obowiązkowego przez wydawców, który przyczynia się do opóźnień w tworzeniu opisów. Szabina Ilácsa pokreśliła, że na Węgrzech użycie danych bibliograficznych, poza kontekstem bibliotecznym, jest niewielkie. Podobnie jak Pawlicki wspomniała o problemach z przesyłaniem egzemplarza obowiązkowego. Stwierdziła ponadto, że dane dostarczane przez wydawców do węgierskiej biblioteki narodowej są niskiej jakości. Poważnym problemem jest również brak wiedzy na temat wartości i jakości metadanych przygotowywanych przez bibliotekę. Zdaniem Ilácsy biblioteki powinny nie tylko dostarczać danych, ale również pracować nad nimi, pokazując, jak mogą zostać użyte. Z kolei Peter Leinen (Niemiecka Biblioteka Narodowa) wśród elementów wpływających na ograniczone wykorzystanie danych, tworzonych przez niemiecką bibliotekę, wymienił ich dużą ilość oraz nieznaną ich strukturę wśród badaczy.

Pytanie drugie, dotyczyło doświadczeń związanych z użyciem danych z bibliografii i katalogów w badaniach. W swojej odpowiedzi Mathilde Koskas wymieniła użycie tradycyjnej, XIX-wiecznej bibliografii narodowej przez historyków książki. Podkreśliła, że stopień wykorzystania współczesnych danych jest trudny do oszacowania ze względu na brak bezpośredniego kontaktu między biblioteką a osobami korzystającymi z udostępnionych zasobów. Kontakt następuje jedynie w przypadku bardziej złożonych potrzeb. Zapytania te mają głównie charakter statystyczny. Y. Sommerland wymieniła literaturoznawców i językoznawców jako odbiorców danych bibliograficznych. Według niej wielu z nich nie posiada nawet podstawowych umiejętności pracy z tego typu zbiorami. Inną, istotną przeszkodę w wykorzystaniu danych bibliograficznych, poza środowiskiem bibliotecznym, wymienił Osmo Suominen, który podkreślił odmienne podejście do danych bibliograficznych wśród bibliotekarzy i badaczy. Wiąże się to z różnymi celami obu grup. Pewne elementy opisu, istotne i prawidłowe z punktu widzenia bibliotekarzy, niekoniecznie są tak postrzegane przez naukowców. Ostatnie pytanie dotyczyło przyszłości współpracy ze społecznością badawczą i oczekiwań bibliotekarzy w stosunku do tej grupy użytkowników. S. Ilácsa stwierdziła, że katalogi powinny zostać wzbogacone w wyniku prac badawczych i zawierać więcej informacji, np. kontekstowych, takich jak cytowania i inne efekty pojawienia się danej publikacji. Z kolei M. Koskas podkreśliła, że oczekiwałyby od naukowców informacji zwrotnej, w jaki sposób bibliotekarze i ich praca może być dla nich przydatna. Spotkanie zakończył krótkim podsumowaniem T. Umerle.

Wydaje się, że współpraca między bibliotekarzami a naukowcami w kontekście wykorzystania danych bibliograficznych do celów badawczych jest na bardzo wstępnym etapie. Potwierdziły to wypowiedzi niemal wszystkich panelistów. Pojawiające się pomysły są mało konkretne i trudno stwierdzić czy mają szanse na realizację (i czy w ogóle jest taka potrzeba). Dotyczy to np. dalszego wzbogacania katalogów biblioteczných. Spotkania takie są potrzebne i tylko dalszy dialog między dwiema wymienionymi społecznościami może przybliżyć nas do odpowiedzi na pytanie o możliwości użycia danych tworzonych przez biblioteki w pracach badawczych. Inicjatywa powstania i działalności grupy badawczej poświęconej tej tematyce jest cenna i warta kontynuacji.

BIBLIOGRAFIA

DARIAH-EU. (b.d.-a). *Bibliographical Data (BiblioData)* [online]. Digital Research Infrastructure for the Arts and Humanities. [dostęp: 5.10.2021]. Dostępny w WWW: <<https://www.dariah.eu/activities/working-groups/bibliographical-data-bibliodata/>>.

- DARIAH-EU. (b.d.-b). *DARIAH in a Nutshell* [online]. Digital Research Infrastructure for the Arts and Humanities. [dostęp: 5.10.2021]. Dostępny w WWW: <<https://www.dariah.eu/about/dariah-in-nutshell/>>.
- DARIAH-EU. (b.d.-c). *Working Groups* [online]. Digital Research Infrastructure for the Arts and Humanities. [dostęp: 5.10.2021]. Dostępny w WWW: <<https://www.dariah.eu/activities/working-groups-list/>>
- Lahti, L.; Vaara, V.; Marjanen, J. & Tolonen, M. (2019). Best Practices in Bibliographic Data Science. In: *Proceedings of the Research Data and Humanities (RDHUM) 2019 Conference: Data, Methods and Tools*. Eds. J. H. Jantunen, S. Bruni, N. Kunnas, S. Palviainen, & K. Västi. Oulu: University of Oulu, pp. 57–65 [dostęp: 5.10.2021]. Dostępny w WWW: <<http://urn.fi/urn:isbn:9789526223216>>

Bartłomiej Włodarczyk
Wydział Dziennikarstwa, Informacji i Bibliologii
Uniwersytet Warszawski

Tekst wpłynął do Redakcji 6 października 2021 r.