

MAREK NAHOTKO
(Uniwersytet Jagielloński, Kraków)

WSPÓLDZIAŁANIE METADANYCH DLA INTEGRACJI INFORMACJI NA POZIOMIE GLOBALNYM

WSTĘP

Biblioteki utraciły dawny monopol na tworzenie katalogów – zbiorów metadanych. Każdy może stworzyć własny katalog; kiedyś, gdy katalogi stanowiły zbiór fiszek, ze względu na mały zasięg takich prac nie miały one znaczenia. Teraz, w formie elektronicznej każdy zbiór metadanych może być dostępny dla całego świata, może być scalany z innymi, może też szybko powiększać swoją objętość dzięki współpracy rozproszonych internautów. Powstaje wiele standardów na wszystkich poziomach metadanych, służących różnym celom, w tym dziesiątki schematów metadanych wraz ich z lokalnymi mutacjami. W jaki sposób ten heterogeniczny model funkcjonowania metadanych pogodzić można z globalnym zasięgiem Sieci?

Prace służące rozwiązaniu tych problemów podejmowane były od dawna, wiążą się one z poszukiwaniem sposobów zapewnienia tak zwanego współdziałania systemów informacyjnych, określanego także z angielska interoperacyjnością¹ (*interoperability*). Współczesne systemy informatyczne, w szczególności służące wyszukiwaniu informacji, w tym także naukowej, bardzo rzadko funkcjonują całkowicie odrębnie, bez powiązań (przynajmniej organizacyjnych) z innymi systemami. Stwierdzenie to jest prawdziwsze w odniesieniu do systemów informacyjnych dostępnych w Internecie, który powstał po to, aby ułatwiać współpracę między ludźmi wykorzystującymi w swoich działaniach komputery w celu uzyskiwania dostępu do danych odległych i rozproszonych. Współpraca odbywa się na wszystkich poziomach systemów informacyjnych, a bardzo ważną rolę w tym zakresie odgrywa wspomniane współdziałanie w zakresie tworzenia i stoso-

¹ Współdziałanie często w literaturze informatycznej i informatologicznej nazywane bywa interoperacyjnością, co jest niezbyt szczęśliwą kalką terminu z języka angielskiego. W artykule używam polskiego terminu współdziałanie, które nie tylko jest poprawne językowo, lecz także właściwie oddaje istotę zagadnienia.

wania metadanych pochodzących z wielu systemów, zróżnicowanych pod względem semantycznym i syntaktycznym. Jak ważne jest to zagadnienie, świadczy opinia Barbary Sosińskiej-Kalaty, która pisze, że metadane i ich współdziałanie są jednym z kilku zasadniczych pól badawczych współczesnej organizacji wiedzy². Decydują o tym techniczne uwarunkowania przetwarzania zasobów wiedzy i informacji, utrwalonych w formie cyfrowej i udostępnianych w Internecie. Efektywna organizacja informacji w sieciach rozproszonych wymaga stosowania metadanych tworzonych w taki sposób, aby były możliwe do stosowania wraz z programami komputerowymi, o których istnieniu twórca metadanych czasem nawet nie wie.

Według Getaneha Alemu [i innych] w tym zakresie możliwe są dwa podejścia³. W środowisku tradycyjnym (katalogów kartkowych), a nawet pierwszych bibliotecznych systemów komputerowych jedynym naprawdę możliwym do praktycznej realizacji było działanie bazujące na ontologicznym punkcie widzenia oparte na obiektywizmie, gdzie prace dotyczące metadanych są autorytatywne, hierarchiczne i zmierzające do realizacji jednego i jedynie słusznego rozwiązania. We współczesnym, otwartym środowisku sieci globalnych coraz większą rolę odgrywa punkt widzenia związany z kierunkiem filozoficznym zwanym społecznym konstrukttywizmem, w którym znaczenia dla zjawisk i obiektów świata otaczającego są wyznaczane przez nas; nie istnieją niezależnie od nas, od naszej świadomości⁴. Wówczas współdziałanie metadanych zapewniane jest nie poprzez narzucanie standardów, lecz przez współpracę w różnorodności.

WSPÓLDZIAŁANIE SYSTEMÓW INFORMACYJNYCH

O współdziałaniu metadanych trudno mówić bez wspomnienia o współdziałaniu na wyższym poziomie – systemów informacyjnych. Według Kazimierza Subiety współdziałanie systemów informacyjnych to „możliwość wymiany danych pomiędzy wieloma systemami, z różnym sprzętem i oprogramowaniem, strukturami danych i interfejsami, z minimalną utratą treści i funkcjonalności”⁵. Zgodnie z tą definicją, współdziałanie jest inherentną cechą systemów komputerowych (informatycznych). Każdy system komputerowy składa się przynajmniej z dwóch współdziałających elementów: sprzętu i oprogramowania. Wraz ze wzrostem liczby współdziałających systemów powstaje potrzeba współdziałania na wielu poziomach.

² B. S o s i ń s k a - K a l a t a, *Ewolucja paradygmatu badań organizacji wiedzy*, [w:] *Nauka o informacji w okresie zmian*, red. B. S o s i ń s k a - K a l a t a, E. C h u c h r o, Warszawa 2013, s. 124.

³ G. A l e m u, B. S t e v e n s, P. R o s s, *Towards a conceptual framework for user-driven semantic metadata interoperability in digital libraries. A social constructivist approach*, „New Library World” vol. 113, 2012, nr 1/2, p. 40.

⁴ T. D u f f y, D. J o n a s s e n, *Objectivist and constructivist conceptions of learning and instruction*, [w:] *Constructivism and the technology of instruction. A conversation*, Hillsdale 1992, p. 3.

⁵ K. S u b i e t a, *Słownik terminów z zakresu obiektywności*, Warszawa 1999, s. 218.

W publikacjach informatycznych spotyka się wiele tekstów na temat współdziałania systemów informacyjnych. Na przykład rozbudowany model współdziałania systemów informacyjnych przedstawił Miguel Manso-Callejo z zespołem. Na podstawie dostępnej literatury wyróżnione zostały następujące poziomy współdziałania:

- Techniczny: bity i bajty, pliki, przetwarzanie, protokoły komunikacyjne na poziomie sprzętu i oprogramowania;
- Syntaktyczny: uzgodnione formaty i struktury danych, języki takie jak XML;
- Semantyczny: dotyczy znaczenia informacji, obejmuje wspólne słowniki terminów, umożliwiające interpretację ich znaczenia; współdziałanie semantyczne wspomaga ją standardy i specyfikacje definiujące schematy wymiany informacji;
- Pragmatyczny: metody i procedury dla obsługi wymienianych danych, definicje interfejsów;
- Dynamiczny: nadzór transferu danych, monitoring sieci i innych systemów;
- Konceptualny: modelowanie systemu i danych w standardowej dokumentacji, bez względu na zastosowany rodzaj modelu;
- Organizacyjny: cele biznesowe, polityka dostępu i wykorzystania danych, modelowanie procesów, wzorce postępowania, wzorce odpowiedzialności przydatne podczas oceny systemu⁶.

Specyficzną cechą współdziałania systemów informacyjnych jest istotna rola metadanych w tym zakresie, znajdują one swoje miejsce w każdym modelu opisującym współdziałanie systemów. W strukturze przedstawionej powyżej będą one funkcjonować przede wszystkim na poziomie semantycznym, chociaż na innych poziomach także znajdują zastosowanie.

WSPÓLDZIAŁANIE METADANYCH: PODEJŚCIE OBIEKTYWISTYCZNE

Współdziałanie metadanych można więc uważać za część ogólniejszego zagadnienia współdziałania systemów informacyjnych. Mikael Nilsson przedstawił następującą definicję współdziałania metadanych: jest to „zdolność dwóch lub więcej systemów lub ich składników do wymiany danych opisowych o rzeczach oraz do interpretowania danych opisowych, podlegających wymianie w sposób spójny z interpretacją twórcy danych”⁷.

Według Stuarta Weibela wymiana metadanych pomiędzy systemami informacyjnymi wymaga porozumień na trzech poziomach:

⁶ M. Manso-Callejo, M. Wachowicz, M. Bernabé-Poveda, *Automatic metadata creation for supporting interoperability levels of spatial data infrastructures*, [w:] *GSDI 11 World Conference Spatial Data Infrastructure Convergence* Rotterdam, 15–19 June 2009, [online] <http://www.gsdi.org/gsdi-conf/gsdi11/papers/pdf/194.pdf> [dostęp: 25.05.2016].

⁷ M. Nilsson, *From interoperability to harmonization in metadata standardization*, Stockholm 2010, p. 13.

- Semantyki: znaczenie nadawane wyrażeniom budowanym przy pomocy metadanych. Semantyka jest domeną ludzi. Uzgodnianie semantyki polega na budowie uzgodnionej struktury elementów metadanych (pól i podpól), takiej na przykład, jak DCMES⁸.
- Syntaktyki: metody kodowania elementów metadanych w taki sposób, aby można je było przesyłać pomiędzy komputerami, a po przesłaniu odkodowywać i automatycznie przetwarzać, w tym co najmniej wyświetlać dla użytkowników-ludzi. W trakcie tego przetwarzania semantyka nie powinna ulegać zmianom. W tym celu stosowane są standardy, takie jak ISO 2709, RDF i XML.
- Struktur, czyli pragmatyki: wymienić można niektóre decyzje niezbędne do podejmowania na tym poziomie w celu uzyskania poprawnie zbudowanych metadanych:
 - Określenie granic dla rekordów metadanych;
 - Powtarzalność elementów metadanych, ograniczenia powtarzalności;
 - Struktura i sposób zapisu nazw własnych, w tym głównie osobowych;
 - Sposób obsługi zagnieżdżenia danych;
 - Sposób zapisu dat;
 - Sposób identyfikacji schematów kodowania (standardów) dla powyższych;
 - Sposób identyfikacji przez schemat kodowania (taki jak LCSH lub UKD) wartości metadanych, z którego pobierane mają być te wartości. Decyzja o obligatoryjności schematu.
 - Identyfikacja wartości metadanych przez odnośnik (URI) lub ciąg znaków (literał)⁹.

Aspekty semantyki i syntaktyki metadanych dotyczą każdej formy ich współdziałania. Oba te aspekty są jednakowo ważne i powszechnie używane w celu rozróżnienia zagadnień w rzeczywistości występujących łącznie. W praktyce współdziałanie semantyczne uważane jest za ważniejsze i wymagające bardziej wyszukanych rozwiązań niż syntaktyczne. Jednak współdziałanie semantyczne nie może być uzyskane bez osiągnięcia współdziałania syntaktycznego¹⁰.

Z innego punktu widzenia uzgodnienia muszą być realizowane tak, aby efekty współdziałania widoczne były na różnych poziomach:

- Poziom schematu: prace dotyczą elementów schematów, niezależnie od konkretnych aplikacji. Efektem prac są zazwyczaj schematy przełącznikowe, zestawy pochodne, tablice przejścia, profile aplikacyjne i rejestry elementów;
- Poziom rekordu: działania służą integracji rekordów metadanych poprzez mapowanie elementów na podstawie ich semantyki. W efekcie często powstają skonwert-

⁸ DCMES – Dublin Core Metadata Element Set.

⁹ S. W e i b e l, *Metadata. Semantics; structure; syntax*, [online] <http://weibel-lines.typepad.com/weibel-lines/2008/02/metadata-semant.html> [dostęp: 15.05.2016].

¹⁰ L. C a n d e l a [i in.], *The DELOS digital library reference model. Foundations for digital libraries*, [online] http://delosw.isti.cnr.it/files/pdf/ReferenceModel/DELOS_DLReferenceModel_096.pdf [dostęp: 14.09.2016].

wane rekordy lub nowe rekordy powstałe z agregacji wartości rekordów istniejących;

- Poziom systemu informacyjnego: prace odbywają się na zintegrowanych rekordach z różnych źródeł, służą mapowaniu ciągów wartości związanych z określonymi elementami metadanych. W efekcie możliwe staje się wyszukiwanie w wielu zasobach jednocześnie.

Porozumienia te mogą być osiągnięte w różny sposób, na każdym z wymienionych poziomów odrębnie lub na kilku jednocześnie. Lois Chan i Marcia Zeng pisały o dwóch zasadniczych etapach zapewnienia współdziałania metadanych: przed utworzeniem rekordu metadanych (opisu) i na etapie istnienia rekordu, po jego utworzeniu z zastosowaniem wybranych standardów, ułatwiających współdziałanie¹¹. Autorki wyróżniły następujące techniki:

- Przed utworzeniem rekordu:
 - Ujednolicenie: stosowanie we współpracujących systemach informacyjnych tych samych standardów metadanych, najlepiej globalnych, takich jak MARC 21 lub Dublin Core;
 - Współdziałanie na poziomie schematu:
 - Tworzenie zestawów pochodnych, na przykład z MARC 21 powstały MODS, MARCXML oraz MARCLite; różne wersje językowe Dublin Core również zaliczyć można do efektów stosowania tej metody;
 - Profile aplikacyjne stanowiące deklarację stosowania elementów metadanych przez organizację, zasób informacyjny, aplikację lub społeczność użytkowników; zawierają zazwyczaj elementy metadanych pobrane z jednego lub kilku standardowych schematów metadanych;
 - Tablice przejścia zawierające mapowania elementów (struktury), semantyki i syntaktyki z jednego schematu metadanych do innego. Jest to jedna z najczęściej stosowanych metod zapewnienia współdziałania pomiędzy schematami metadanych; mapowanie może dotyczyć takich cech, jak:
 - Semantycznych definicji elementów,
 - Opcjonalności/obowiązkowego występowania elementu,
 - Powtarzalności elementu,
 - Ograniczeń wartości elementu (np. słowniki kontrolowane)
 - Możliwość stosowanie elementów definiowanych lokalnie.
 - Schematy przełącznikowe: jeden, wybrany schemat jest stosowany jako mechanizm przełącznikowy pomiędzy wieloma innymi schematami, dzięki cze-

¹¹ L. C h a n, M. Z e n g, *Metadata interoperability and standardization – a study of methodology. Part I. Achieving interoperability at the schema level*, „D-Lib Magazine” vol. 12, 2006, nr 6, [online] <http://www.dlib.org/dlib/june06/chan/06chan.html> [dostęp: 1.06.2015] oraz M. Z e n g, L. C h a n, *Metadata interoperability and standardization – a study of methodology. Part II. Achieving interoperability at the record and repository levels*, „D-Lib Magazine” vol. 12, 2006, nr 6, [online] <http://www.dlib.org/dlib/june06/zeng/06zeng.html> [dostęp: 15.05.2016].

mu zamiast tworzyć wiele tablic przejścia dla par schematów metadanych, każdy z nich jest mapowany tylko do schematu przełącznikowego;

- Struktura nadrzędna, rozumiana jako parasol, pod którym integrowane są różne obiekty dla danego rozwiązania; często niezbędna w procesach cyfrowej archiwizacji, czego przykładem może być OAIS¹²;
 - Rejestry metadanych służące gromadzeniu danych dotyczących schematów metadanych, ułatwiają wielokrotne stosowanie istniejących terminów metadanych, co wspomaga standaryzację.
- Po utworzeniu rekordu:
 - Współdziałanie rekordów:
 - Konwersja rekordów metadanych, podczas której największym problemem są: utrata oraz zniekształcenia danych (w szczególności podczas konwersji ze schematu bardziej rozbudowanego do uboższego). Jedną z ich przyczyn są różnice między specyfikacją schematu metadanych (często stanowiącego podstawę konwersji) a rzeczywistymi metadanymi;
 - Linked Data – nowe technologie pozwalające na wielokrotne stosowanie danych w miejscach, gdzie są potrzebne, oraz ich zdecentralizowaną integrację, model danych identyfikujący, opisujący, linkujący i budujący relacje pomiędzy ustrukturyzowanymi elementami danych, przeznaczony do działania w skali Web. Instytucje, takie jak biblioteki nie muszą porzucać wcześniej stosowanych schematów metadanych, wystarczy dostosować je do zasad Linked Data.
 - Współdziałanie systemów informacyjnych:
 - Protokoły wymiany metadanych, takie jak OAI-PMH¹³, służące tworzeniu zasad współdziałania niezależnych od aplikacji, stosowane przez społeczności publikujące w Web. Narzędzie OAI pozwala na przeszukiwanie zasobów Web dzięki możliwości publikowania, rozpowszechniania i archiwizowania metadanych;
 - Agregacja – każdy rekord metadanych zawiera zestaw stwierdzeń o opisywanym źródle, dzięki czemu metadane z różnych źródeł mogą być agregowane dla utworzenia pełniejszego opisu tego źródła.

W zależności od tego, które koncepcje i relacje współdziałają w obszarze systemów informacyjnych, możemy wyróżnić kolejne poziomy współdziałania. System informacyjny może być uważany za zdolny do współdziałania z innym na przykład na poziomie ontologii lub obiektów informacyjnych. Ten ostatni poziom oznacza istnienie współdziałania wyższego stopnia, gdyż zakłada istnienie tego pierwszego.

Problem współdziałania zwykle ma wiele wspólnego z zagadnieniami archiwizacji i wielojęzyczności. Ta ostatnia cecha może być traktowana jako współdziałanie pomię-

¹² OAIS – Open Archival Information System.

¹³ OAI-PMH – Open Archives Initiative-Protocol for Metadata Harvesting.

dzy językami, natomiast archiwizacja może być traktowana jako współdziałanie ponad czasem.

Działania podejmowane z obiektywistycznego punktu widzenia powodują mnożenie wielopoziomowych standardów metadanych, nie umożliwiając jednocześnie efektywnego współdziałania w zakresie jednoczesnego wyszukiwania w wielu serwisach, wymiany danych i integracji informacji. Istnienie kilku międzynarodowych standardów, w połączeniu z mnożeniem standardów lokalnych, powoduje zaostrzenie problemów związanych ze współdziałaniem metadanych. Jednym z problemów jest istnienie wielu standardów służących realizacji podobnych funkcji. Dochodzi do sytuacji, gdy istniejące standardy, stosowane w bibliotekach, wymagają tworzenia nowych standardów, umożliwiających ich współdziałanie¹⁴.

WSPÓLDZIAŁANIE GLOBALNYCH STANDARDÓW METADANYCH: MARC 21 I DUBLIN CORE

Prześledźmy problemy współdziałania metadanych na trzech wymienionych poziomach, używając przykładowo najczęściej stosowane schematy metadanych, to znaczy MARC 21 i Dublin Core. Na poziomie semantycznym schematy te wydają się bardzo różne, co naturalnie utrudnia ich współdziałanie. Już stopień rozbudowy struktury elementów metadanych uwidacznia te różnice. MARC 21 to ponad 250 pól zmiennej długości, w obrębie których mogą funkcjonować podpola i wskaźniki modyfikujące znaczenie danych w polach. Natomiast struktura podstawowego zestawu elementów Dublin Core to 15 elementów, które mogą być uszczegóławiane w warstwie semantycznej przy pomocy tak zwanych kwalifikatorów. Różnice między tymi dwoma schematami nie są jednak tak duże, jakby się wydawało, a w każdym razie zmniejszają się. Oba standardy reprezentują podejście obiektywistyczne. Po pierwsze, schemat Dublin Core jest stale rozbudowywany i obecnie tak zwany DC Metadata Terms zawiera około 100 elementów, służących opisowi źródeł¹⁵, nie licząc profili aplikacyjnych¹⁶. Po drugie, bardzo rozbudowane opisy rzadko są naprawdę niezbędne; według polskich przepisów katalogowania do jednoznacznej identyfikacji opisywanego źródła wystarczy pięć elementów (pierwszy stopień szczegółowości opisu)¹⁷. Podobnego zdania są chyba użytkownicy naszych katalogów komputerowych, którzy w 98,5% przypadków do wyszukiwania stosują proste kryteria typu autor/tytuł¹⁸. Dla zapewnienia współdziałania

¹⁴ G. A l e m u [i in.], dz. cyt., s. 47.

¹⁵ A. B r a c h f o g e l, „Terminy metadanych DCMI” i możliwości ich wykorzystania w opisie rzeczowym, „Zagadnienia Informacji Naukowej” 2010, nr 2(96), s. 69–76.

¹⁶ Profile aplikacyjne to schematy metadanych powstałe w wyniku łączenia elementów z różnych standardów.

¹⁷ M. L e n a r t o w i c z, *Przepisy katalogowania książek. Część 1. Opis bibliograficzny*, Warszawa 1986, s. 3.

¹⁸ T. W o l n i e w i c z, *Katalogi wirtualne na przykładzie polskiej wyszukiwarki KaRo*, [w:] *Rola katalogu centralnego NUKAT w kształtowaniu społeczeństwa wiedzy w Polsce. Międzynarodowa konferencja*

metadanych na tym poziomie (semantycznym) stosowane są techniki mapowania schematów metadanych.

Na poziomie syntaktycznym również trwają prace nad ograniczeniem różnic utrudniających współdziałanie między dwoma schematami. Powszechnie stosowanym standardem syntaktyki metadanych MARC 21 jest struktura zgodna z ISO 2709, pierwotnie dostosowana do wymiany danych na taśmach magnetycznym i praktycznie niestosowana poza środowiskiem bibliotekarskim, co utrudnia współdziałanie. Jednak realizowane są prace nad zastosowaniem współczesniejszych i powszechnie stosowanych standardów syntaktyki, takich jak XML (stąd prace nad schematami MARCXML i MODS¹⁹). Inaczej sytuacja wygląda w przypadku Dublin Core. Tu od początku pracowano nad uniezależnieniem semantyki schematu od warstwy syntaktycznej. Dzięki temu stosowane mogą być wszystkie współczesne standardy syntaktyki metadanych, takie jak HTML, XML, RDF i inne. Problemy z konwersją opisów na poziomie syntaktycznym pojawiają się po utworzeniu rekordów metadanych, chyba że wcześniej uzgodniony został wspólny standard w tym zakresie.

Rozwiązania stosowane na poziomie pragmatycznym metadanych związane są z tym, co funkcjonuje na poziomie semantycznym. Jako elementy pragmatyki stosowane są przepisy i instrukcje wypełniania danymi wcześniej utworzonych struktur pól i elementów (czyli poziomu semantycznego). Tutaj sytuacja przedstawia się o wiele korzystniej dla MARC 21, dla którego przygotowano (lub wykorzystano w nim) bardzo rozbudowane narzędzia tworzone na różnym poziomie, od międzynarodowego (AACR2, ISBD) po krajowy i lokalny (np. zestawy szczegółowych instrukcji tworzonych w Centrum NUKAT). Użytkownicy Dublin Core nie mogą liczyć na taką standaryzację, więc pozostaje im tworzenie własnych zasad pragmatyki na najniższym poziomie poszczególnych implementacji (np. bibliotek cyfrowych). W rezultacie rozwiązania te są dalece niestandardowe (choć czasem oparte na ISBD) i bywa, że sprzeczne z podstawowymi zasadami przyjętymi dla Dublin Core, na przykład z tak zwaną zasadą Dumb Down²⁰. Jest to przyczyną wielu kłopotów, utrudniających współdziałanie²¹. Zmiana sytuacji może wkrótce wyniknąć z przejścia nowych standardów pragmatyki przez oba środowiska; chodzi tu głównie o RDA²².

z okazji 5-lecia istnienia katalogu centralnego NUKAT, Warszawa 23–25 stycznia 2008 roku, [online] http://www.nukat.edu.pl/konf2008/mater/54_wolniewicz.pdf [dostęp: 12.05.2016].

¹⁹ MODS – Metadata Object Description Schema (Library of Congress).

²⁰ Zasada Dumb Down dotyczy kwalifikowania elementów metadanych.

²¹ L. D e r f e r t - W o l f, *Jak posługiwać się biblioteką cyfrową?* [w:] *Cyfrowy świat dokumentu. Wydawnictwa, biblioteki, muzea, archiwa*, red. H. H o l l e n d e r, Warszawa 2011, s. 203.

²² RDA – Resources Description and Access.

PODEJŚCIE KONSTRUKTYWISTYCZNE: WSPÓLDZIAŁANIE W LINKED DATA

Jak już wspomniano, monopol na tworzenie i udostępnianie danych bibliograficznych biblioteki utraciły bezpowrotnie. Oznacza to między innymi konieczność współpracy pomiędzy bardzo różnymi serwisami sieciowymi, zarządzającymi metadanymi. Gdyby na przykład serwisy takie, jak Scopus, WorldCat i Goodreads współdziałały w zakresie metadanych, wymiana metadanych pomiędzy nimi umożliwiłaby badaczom określenie poziomu postrzegania ich publikacji jako kulturalnej wartości dodanej oraz stanu miary ich naukowego wpływu i stopnia widzialności wśród czytelników mediów społecznościowych²³. Problem z realizacją tej wizji polega na tym, że interesariusze tych serwisów bibliograficznych mają sprzeczne interesy. W podanym przykładzie Elsevier tworzący Scopus jest firmą komercyjną, WorldCat jest światowym, bibliotecznym katalogiem centralnym, a Goodreads prywatną platformą społecznościową, współpracującą między innymi z Google, Amazonem, Microsoftem i WorldCat. Każdy z nich stosuje własny standard metadanych, więc współdziałanie wymaga prac opisanych w poprzednich częściach artykułu. Zasadniczym problemem są jednak odmienne podstawy filozoficzne działania tych serwisów. O ile WorldCat i Scopus to przykłady serwisów funkcjonujących na zasadach obiektywistycznych, to Goodreads opiera się na poglądach konstruktywistycznych.

Interesującym nowym zastosowaniem dla zasad współdziałania metadanych, które pozwala na użycie innego, bardziej konstruktywistycznego punktu widzenia tego zagadnienia, jest Linked Data. Jest to praktyka stosowania identyfikatorów dla obiektów informacyjnych, udostępnianych w Web, modelowania przepływu danych dla poszczególnych dyscyplin i zastosowań oraz ontologii do projektowania metadanych funkcjonujących na zasadzie grafu. Chociaż Linked Data dotyczy wszelkiego rodzaju danych (nie tylko bibliograficznych) publikowanych w Web, to ożywienie obserwowane w ostatnich latach w zakresie tworzenia nowych aplikacji i zasobów danych idea ta zawdzięcza włączeniu się do prac osób ze środowiska bibliotekarskiego, które wniosły rozwiązania stosowane od dawna w obszarze metadanych bibliotecznych, w tym słowniki metadanych.

W Linked Data znaczenie metadanych (semantyka) jest oddzielone od ich syntaktyki, dzięki czemu zmiana struktur nie powoduje modyfikacji znaczenia. Nowe standardy mają jedną bardzo ważną nową cechę – zarówno semantyka, jak i pragmatyka kodowane są przy pomocy tej samej syntaktyki (głównie XML/RDF, choć istnieją inne serializacje), a w dodatku jest to syntaktyka, która umożliwia automatyczne przetwarzanie danych (gdyż „rozumiana” jest przez komputery) bez udziału człowieka. Dzięki tym pracom wielotomowe zestawy drukowanych tablic UKD przestają być potrzebne, gdyż powstaje ich wersja w Linked Data z użyciem SKOS²⁴ (na razie dostępne są tablice skró-

²³ A. Zuccala, R. Cornacchia, *Data matching, integration and interoperability for a metric assessment of monographs*, „Scientometrics” vol. 108, 2016, nr 1, p. 467.

²⁴ SKOS – Simple Knowledge Organization System jest modelem danych dla rozpowszechniania Systemów Organizacji Wiedzy (SOW).

cone²⁵). Dawne schematy metadanych nadal są używane (łącznie ze strukturą pól i podpól MARC 21), ale struktury te są kodowane w RDFS lub OWL (zmianie uległa syntaktyka), przez co mogą stać się ontologiami umieszczanymi w chmurze zlinkowanych danych. Do każdego symbolu UKD i do każdej relacji wyrażonej strukturą MARC (a więc do etykiety każdego pola tego formatu) oraz między tymi symbolami i etykietami można poprowadzić odnośnik (link) kwalifikowany, czyli identyfikujący relację występującą pomiędzy połączonymi elementami.

Zasadniczą różnicą pomiędzy poprzednimi rozwiązaniami stosowanymi w skomputeryzowanych bibliotekach a opisywanymi są połączenia (linki), w szczególności połączenia pomiędzy elementami metadanych będących produktem różnych społeczności. Tego rodzaju połączenia pozwalają na transformację aktywności informacyjnej z zamkniętego świata danych bibliotecznych do otwartego świata semantycznego Web. To z kolei oznacza konieczność dokonania połączeń pomiędzy danymi bibliotecznymi i danymi pochodzącymi z innych środowisk, bez względu na to, czy pochodzą one z badań naukowych, prac jednostek administracji różnego szczebla (w tym rządu), organizacji komercyjnych, a nawet zasoby powstałe dzięki tak zwanej mądrości tłumów. Biblioteki stają się w ten sposób częścią bardzo heterogenicznego środowiska, zarówno z punktu widzenia różnorodności zasobów, jak i ich jakości. Jak pisze Karen Coyle, zamknięty świat kontroli bibliograficznej, który pokolenia bibliotekarzy budowały do ostatnich lat, nie jest już częścią naszej przyszłości²⁶. Istnieje potrzeba nie tylko umieszczenia katalogów bibliotecznych w Web, ale także takie ich dostosowanie do funkcjonujących tam standardów, aby stały się w pełni webowe, co oznacza, że podział na dane wewnętrzne (biblioteczne) i zewnętrzne traci znaczenie.

Stworzenie narzędzi pozwalających na mediacje pomiędzy tym heterogenicznym środowiskiem informacji bibliograficznej a użytkownikiem wymaga:

- Transformacji opisów tekstowych w zestawy kodowanych danych przydatnych do przetwarzania maszynowego;
- Tworzenia elementów danych unikatowo identyfikowanych w obrębie kontekstu informacyjnego Web;
- Posiadania danych kompatybilnych z technologiami i standardami Web;
- Stosowania języka rzeczywiście współdziałającego w całym Web²⁷.

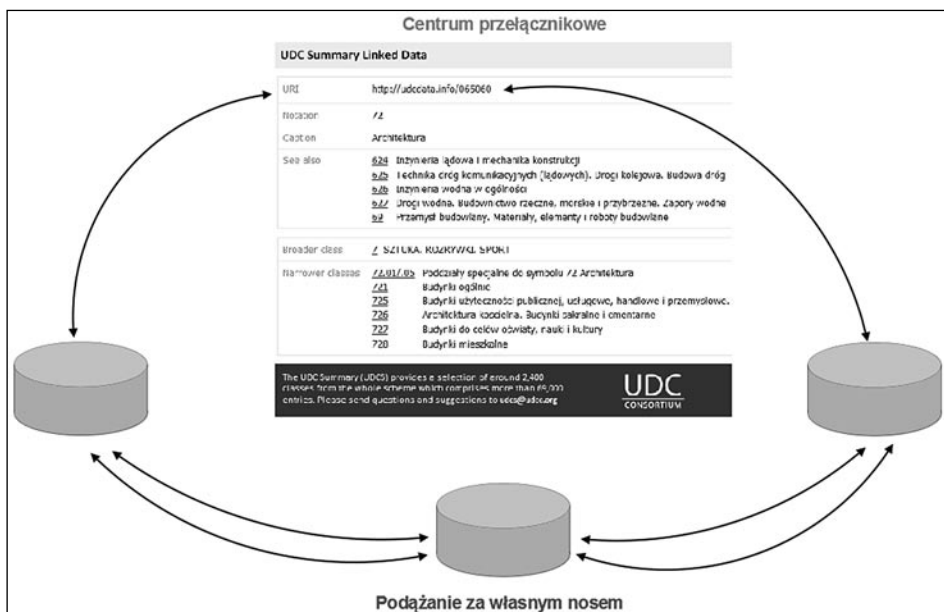
W ten sposób zorganizowane środowisko metadanych umożliwi dwójakiego rodzaju współdziałanie metadanych. Pierwszy sposób na zapewnienie współdziałania związany jest z istnieniem słowników dwóch rodzajów (terminologia tworzona i gromadzona w ramach schematów metadanych, takich jak MARC 21 oraz SOW, takich jak UKD), przydatnych szczególnie, gdy chcemy łączyć dane z różnych obszarów (dziedzin, zasto-

²⁵ J. Hys, J. Kwiatkowska, *UDC Summary*, „Bibliotekarz” 2013, nr 4, s. 6.

²⁶ K. Coyle, *Library linked data: an evolution*, „Italian Journal of Library and Information Science” vol. 4, 2013, nr 1, p. 58.

²⁷ M. Guerrini, T. Possemato, *Linked data. A new alphabet for the semantic web*, „Italian Journal of Library and Information Science” vol. 4, 2013, nr 1, p. 78.

sowań). Użytkownik Web może przeglądać jego zasoby bez potrzeby znajomości zastosowanych technologii i wynikających z nich struktur, stanowiących podstawę działania Sieci. Podczas przeglądania swobodnie przekraczane są granice między zasobami, bez względu na ich fizyczne oddalenie. Na tej samej zasadzie można w Linked Data przeglądać zbiory danych, podążając za odnośnikami od jednego zasobu do innego, nawet gdy są fizycznie umieszczone w różnych miejscach i zapisane w różnych formatach²⁸.



Rys. 1. Dwa modele współdziałania w Linked Data (na przykładzie tablic UKD)

Zasobami danych tego rodzaju są także zawartości słowników obu wcześniej wymienionych rodzajów. Umożliwiają one współdziałanie dzięki temu, że używane są jako tak zwane centra przełącznikowe, umieszczane w chmurze danych, łączące dane wyrażone zgodnie z różnymi, lokalnie definiowanymi semantykami danych. Centrum przełącznikowe umożliwia w Linked Data nawigację od jednego zbioru metadanych do innego przez podążanie za odnośnikami, czyli URI, nawet, jeżeli dane połączone do centrum są heterogeniczne. Wyobraźmy sobie na przykład, że istnieją dwa zasoby metadanych, w których znajdują się między innymi symbole UKD. Bez względu na to, jakie miejsca w różnych strukturach metadanych (polach, elementach) przeznaczone zostały dla tych symboli, znajdują się w nich odesłania przy pomocy URI do ontologii UKD (wykonanej w SKOS i dostępnej w Internecie) we wspomnianej chmurze danych (Rys. 1).

²⁸ E. B e r m e s, *Convergence and interoperability. A Linked Data perspective*, [w:] *World Library and Information Congress: 77th IFLA General Conference and Assembly*, Puerto Rico, 13–18 August 2011, [online] <http://conference.ifla.org/past/ifla77/149-bermes-en.pdf> [dostęp: 28.05.2016].

Dzięki temu, po pierwsze, następuje ujednoczenie znaczenia (semantyki) metadanych, po drugie, możliwe jest wyszukiwanie podobnych danych poprzez proste podążanie za odnośnikami URI. Sposób ten pozwala na unikanie niejednorodności dzięki porozumieniu ontologicznemu²⁹.

Jako centrum przełącznikowe w Linked Data może funkcjonować dowolny zasób udostępniony dla użytkowników. Nie istnieje jedno, główne centrum, ale wiele takich centrów, połączonych ze sobą odnośnikami. Podążanie od centrum do centrum za odnośnikami łączącymi dane, które tam się znajdują, pozwala na odnajdywanie nowej informacji w sposób intuicyjny. Jest to druga metoda zapewnienia współdziałania metadanych, nazwana „podążaniem za własnym nosem”. Suma zasobów metadanych stosujących RDF i URI stanowi globalny graf informacyjny, który może być w sposób nieograniczony przeglądany przez użytkowników (ludzi) i automatyczne aplikacje podążające za odnośnikami URI. Do opisu (rekordu) identyfikującego osobę można na przykład przyłączyć obiekty, miejsca, inne osoby (a dokładnie ich URI) przy pomocy takich relacji, jak na przykład miejsce urodzenia, studiowania, bycie autorem, pracownikiem, ojcem/matką, mieszkańcem, fanem i innych. Stwarza to nowe, interesujące możliwości zapewniania współdziałania zasobom metadanych rozproszonych w Internecie. Powoduje także częściowe odchodzenie od ontologii rozumianej jako światowa „reprezentacja wiedzy” w kierunku uwolnienia obfitości danych zlinkowanych, choć niedoskonałych, co prowadzi do ich szerokiego udostępnienia w sposób ustandaryzowany, a to z kolei umożliwia przemienienie Sieci w udoskonaloną infrastrukturę wiedzy³⁰.

Tworzenie struktur bibliograficznych funkcjonujących w modelu Linked Data jest celem inicjatywy o nazwie BIBFRAME. Ich zadaniem jest między innymi przejście od rozwiązań opartych na MARC 21 w kierunku standardów Linked Data. Format MARC 21 ma być całkowicie zastąpiony, zarówno jako format wymiany danych, jak i format katalogowania, a także jako wewnętrzny format zintegrowanych systemów bibliotecznych. Model ten ma być niezależny od jakichkolwiek zasad katalogowania, ma być za to odpowiedni zarówno dla rozwiązań „płaskich”, opartych na rekordach bibliograficznych, jak i dla wysoko relacyjnych rozwiązań opartych na FRBR. Sam BIBFRAME nie jest oparty na FRBR, lecz zawiera własne jednostki (encje). Dla zachowania współdziałania w środowisku Linked Data niezbędne jest posiadanie podobnego modelu conceptualnego, który zbudowany został w oparciu o analizę istniejących danych bibliograficznych³¹.

Prace nad tworzeniem opisów zasobów Linked Data mogą być realizowane na zasadzie współpracy każdego z każdym: bibliotek oraz innych instytucji, a także użytkowników. Ten rozproszony model danych pozwala „każdemu mówić cokolwiek o czym-

²⁹ T. H e a t h, Ch. B i z e r, *Linked Data. Evolving the Web into a Global Data Space*, San Rafael, CA 2011, p. 24.

³⁰ D. W e i n b e r g e r, *Too big to know. Rethinking knowledge now that the facts aren't the facts, experts are everywhere, and the smartest person in the room is the room*, New York 2012, p. 188.

³¹ L. S t e v e n s s o n, *Are current bibliographic models suitable for integration with the Web?* „Information Standards Quarterly” vol. 25, 2013, nr 4, p. 12.

kolwiek”, co musi powodować zróżnicowanie opinii i brak zgodności co do znaczenia elementów³². Przypomina to sytuację raczej Wikipedii, niż katalogu bibliotecznego. Z drugiej strony dzięki gęstej sieci połączeń (linków) do danych komplementarnych, tworzonych przez wiarygodne źródła (biblioteki) wzrasta wartość danych ponad to, co oferować może prosta suma zasobów traktowanych indywidualnie (bez powiązań). Z pewnością jednak w świecie cyfrowym zarówno twórcy/właściciele informacji, jak i jej użytkownicy mogą tworzyć własne kategorie metadanych na niezliczoną ilość sposobów, ułatwiając sobie odnajdywanie obiektów cyfrowych. Często przywoływanym przykładem takiej organizacji informacji jest Amazon³³.

Zaangażowanie użytkowników w tworzenie metadanych pozwala na stworzenie bogatszego środowiska metadanych, w którym możliwe jest konstruktywistyczne zróżnicowanie poglądów. Podejście oparte na współpracy, gdzie wykorzystywane są metadane tworzone automatycznie, przez autorów, bibliotekarzy i użytkowników, powoduje powstawanie ekologii metadanych, w której metadane tworzone przez użytkowników przekraczają objętościowo metadane z innych źródeł. Jak pisze Alemu wraz z innymi, optymalne wykorzystanie możliwości Linked Data wymaga podejścia mieszanego, w którym metadane zawierają nie tylko fizyczne opisy obiektów informacyjnych, ale także elementy opisujące ich cechy socjo-kulturalne, przy zastosowaniu takich narzędzi, jak tagi, komentarze, odnośniki, punktowania, rekomendacje³⁴.

WNIOSKI

Z tego, co dotąd napisano, wynika, że współdziałanie systemów we współczesnym, heterogenicznym środowisku informacyjnym jest niezbędne, ale w stu procentach nieosiągalne, tak jak niemożliwe jest pełne uzgodnienie poglądów indywidualnie konstruowanych. Jest to więc stan idealny, do którego należy dążyć, zdając sobie sprawę z nigdy do końca niezadowolających efektów tych starań. W trakcie realizacji prac należy mieć na uwadze zapewnienie praktycznej współpracy systemów informacyjnych na założonym poziomie. Dotyczy ono każdego elementu systemu, w tym również metadanych opisowych na wszystkich poziomach.

Współdziałanie metadanych to zagadnienie wielowymiarowe: pogląd obiektywistyczny wymaga uzgodnienia semantyki, syntaktyki i pragmatyki metadanych na poziomie schematów, rekordów i/lub baz danych. Na poziomie schematu najważniejszą rolę odgrywa semantyka; na poziomie rekordu – pragmatyka; zaś na poziomie baz danych – syntaktyka metadanych.

³² D. A l l e m a n g, J. H e n d l e r, *Semantic Web for working ontologist. Effective modeling in RDFS and OWL*, Amsterdam 2008, p. 7.

³³ D. W e i n b e r g e r, *Everything is miscellaneous*, New York 2007.

³⁴ G. A l e m u [i in.], *Linked Data for libraries. Benefits of a conceptual shift from library-specific record structures to RDF-based data models*, „New Library World” vol. 113, 2012, nr 11/12, p. 552.

Działania związane ze stosowaniem międzynarodowych standardów, takich jak MARC lub Dublin Core, opierają się na obiektywistycznym poglądzie filozoficznym, podczas gdy zasoby informacyjne i interpretacja gromadzonych w nich metadanych, szczególnie w heterogenicznym środowisku sieci globalnych, są zasadniczo zróżnicowane, co sugeruje potrzebę stosowania podejścia interpretacyjnego (konstruktywistycznego). Dotychczas w projektowaniu i rozwoju OPAC faworyzowany był obiektywistyczny punkt widzenia. Powstanie i rozwój aplikacji Web 2.0 i Linked Data powoduje przesunięcie zainteresowania w stronę perspektywy filozoficznej związanej z konstruktywizmem. W taki sposób podstawy filozoficzne, na których wspierają się decyzje o wyborze standardów metadanych, mogą istotnie wpływać na zasady i zakres współdziałania metadanych.

Zastosowanie Linked Data powoduje poważne zmiany w sposobie pracy osób katalogujących i w sposobie funkcjonowania katalogów, co charakteryzuje żartobliwe określenie czynności wykonywanych przez katalogerów jako „katalinkowanie” zamiast katalogowania. Jeżeli bowiem zasady Linked Data znajdą powszechne zastosowanie, opracowanie zbiorów w znacznej mierze polegać będzie na tworzeniu odnośników do słowników umieszczonych w chmurze danych. To z kolei pociąga za sobą problemy związane z koniecznością określania poziomu wiarygodności źródeł informacji i przyjęcia do wiadomości możliwości istnienia wielu poglądów na każdy temat.

THE INTEROPERABILITY OF METADATA FOR THE INTEGRATION OF INFORMATION AT THE GLOBAL LEVEL

SUMMARY

Libraries have lost their former monopoly on creating metadata. Currently, the internet abounds in many non-library-related websites whose work is based on metadata resources that are sometimes created by merging (or aggregating) the resources of the largest libraries. In such a heterogeneous environment, it is of fundamental importance to find a solution to the problem of the interoperability of systems at all levels, including the metadata level. The interoperability of metadata – defined as the ability of two or more systems or their parts (using different hardware and software, with different data structures and interfaces) to exchange data with a minimal loss of content and functionality (NISO) – can be achieved in various ways, depending – among other things – on whether this process is planned before the creation of the metadata, or whether it concerns resources which have already been created in different systems and formats. There are two distinct approaches to metadata cooperation: one is traditional, authoritative and objectivist, while the other is constructivist and relates to the development of new information technologies. Special attention has been paid not only to new possibilities, but also to the importance of the cooperation of metadata that results from the globalisation of the processes of its creation and the sharing of metadata in so-called data clouds (Linked Data).

SŁOWA KLUCZOWE:

integracja informacji, konstruktywizm, Linked Data, standardy metadanych, współdziałanie metadanych.

KEYWORDS:

integration of information, constructivism, Linked Data, metadata standards, interaction of metadata.

