

MAREK NAHOTKO
Instytut Informacji Naukowej i Bibliotekoznawstwa
Uniwersytet Jagielloński
e-mail: marek.nahotko@uj.edu.pl

WSPÓLDZIAŁANIE METADANYCH W CHMURZE



Dr hab. Marek Nahotko jest adiunktem w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego. Specjalizuje się w zagadnieniach metadanych dokumentów elektronicznych i wykorzystania Internetu w komunikacji naukowej i działalności informacyjnej. Jest autorem książek: *Komunikacja naukowa w środowisku cyfrowym* (Warszawa, 2010), *Naukowe czasopisma elektroniczne* (Warszawa, 2007), *Opis dokumentów elektronicznych. Teoretyczny model i możliwości jego aplikacji* (Kraków 2006), *Metadane: sposób na uporządkowanie Internetu* (Kraków, 2004) oraz licznych artykułów publikowanych m.in. na łamach „Przeglądu Bibliotecznego”, „Zagadnień Informacji Naukowej” oraz „Biuletynu EBIB”.

SŁOWA KLUCZOWE: Współdziałanie metadanych. Linked Data. Systemy biblioteczne nowej generacji.

ABSTRAKT: Teza/cel artykułu – W artykule podjęto próbę przedstawienia najnowszych kierunków prac dla zapewnienia współdziałania metadanych bibliotecznych. **Metody badań** – Wykorzystano metodę analizy piśmiennictwa dotyczącego współdziałania systemów informatycznych oraz narzędzi służących zapewnieniu współdziałania metadanych. **Wyniki** – Prace służące standaryzacji w obszarze metadanych bibliotecznych, realizowane od XIX w., obecnie zmierzają w kierunku dostosowania rozwiązań do potrzeb i możliwości sieci rozległych. W tym środowisku szczególnego znaczenia nabiera możliwość współdziałania (ang. *interoperability*) metadanych. W artykule przedstawiono problemy współdziałania systemów informacyjnych oraz narzędzia zapewniające współdziałanie metadanych. Najnowsze prace w tym kierunku związane są z zastosowaniem idei Linked Data. Obecnie trwa przenoszenie słowników metadanych stosowanych w bibliotekarstwie do RDF, w celu udostępnienia ich jako zasobów Linked Data. Trwają również prace nad zastosowa-

niem Linked Data w zintegrowanych systemach bibliotecznych. **Wnioski** – Dostosowanie standardów używanych w bibliotekach do rozwiązań stosowanych powszechnie w sieciach rozległych daje nowe perspektywy w zakresie współdziałania metadanych, a pośrednio – funkcjonowania systemów informacyjno-wyszukiwawczych.

WSTĘP

Współczesne systemy informatyczne, w szczególności służące wyszukiwaniu informacji, także naukowej, bardzo rzadko funkcjonują w odosobnieniu, bez powiązań (przynajmniej organizacyjnych) z innymi systemami. Stwierdzenie to jest tym bardziej prawdziwe w odniesieniu do systemów w globalnej sieci – Internecie, który powstał po to, aby ułatwiać współpracę między ludźmi i komputerami. Współpraca ta odbywa się na wszystkich poziomach systemów informacyjnych, a niebagatelną rolę w tym zakresie odgrywają metadane i ich współdziałanie¹ (ang. *interoperability*). 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 wiedzy (Sosińska-Kalata, 2013, s. 124). Decydują o tym techniczne uwarunkowania przetwarzania zasobów wiedzy i informacji utrwalonych w formie cyfrowej i udostępnianych w Internecie. Efektywna organizacja wiedzy w sieciach rozproszonych wymaga stosowania metadanych tworzonych tak, aby były czytelne dla programów komputerowych, o których istnieniu twórca metadanych czasem nawet nie wie i zgodne ze standardami zapewniającymi możliwie pełną przekładalność w globalnej sieci danych.

W artykule przedstawione zostaną sposoby uzyskiwania współdziałania metadanych w kontekście współdziałania systemów informacyjnych. Krótko wymienione zostaną narzędzia stosowane do zapewnienia współdziałania metadanych. Bardziej szczegółowo przedstawię prace nad jednym z nowszych standardów (a właściwie zestawem standardów), służącym zapewnieniu współdziałania metadanych w skali sieci globalnych, zwanym Linked Data oraz jego zastosowaniem w najnowszych aplikacjach bibliotecznych, funkcjonujących w tzw. chmurze.

WSPÓŁDZIAŁANIE SYSTEMÓW INFORMACYJNYCH I METADANYCH

Współdziałanie jest bardzo istotnym zagadnieniem związanym z funkcjonowaniem systemów informacyjnych. Dotyczy ono możliwości współpracy różnych systemów (oraz szerzej – organizacji użytkujących te systemy) na wszystkich płaszczyznach. Dążenie do współdziałania wymaga łącze-

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

nia we wspólnym działaniu różnych standardów, dzięki czemu możliwa jest agregacja wielu zasobów danych, pozwalająca na tworzenie nowych, lepszych produktów i usług. Może być rozumiane bardzo różnie w zależności od poziomu, na którym rozpatrywana jest współpraca. Jedną z wielu definicji współdziałania została przedstawiona w polskiej ustawie o informatyzacji podmiotów publicznych. Interoperacyjność, bo taki termin jest tam stosowany, to „zdolność różnych podmiotów oraz używanych przez nie systemów teleinformatycznych i rejestrów publicznych do współdziałania na rzecz osiągnięcia wzajemnie korzystnych i uzgodnionych celów, z uwzględnieniem współdzielenia informacji i wiedzy przez wspierane przez nie procesy biznesowe realizowane za pomocą wymiany danych za pośrednictwem wykorzystywanych przez te podmioty systemów teleinformatycznych”. Ustawa ta wprowadza także Krajowe Ramy Interoperacyjności, czyli „zestaw wymagań semantycznych, organizacyjnych oraz technologicznych dotyczących interoperacyjności systemów teleinformatycznych i rejestrów publicznych” (*Ustawa...*, 2010, s. 3508).

Dla zapewnienia współdziałania systemów informacyjnych na wielu poziomach stosowane są metadane, chociaż są to zwykle struktury inne niż metadane opisowe, tworzone przez bibliotekarzy. W środowisku informatyków nazywane są metadanymi technicznymi lub administracyjnymi (Nahotko, 2004, s. 25). Na poziomie syntaktycznym zawierają one opis formatów danych, zasad kodowania, budowy struktur. Dotyczą danych związanych z funkcjonowaniem systemu, opisujących stan sieci oraz danych wejściowych i wyjściowych dla realizowanych procesów. Na poziomie semantycznym metadane te pozwalają na odszukanie i dostęp do słownictwa ujednolicającego znaczenie, zazwyczaj gromadzonego w ontologiach lub tezaurusach. Na przykład w systemie można przetwarzać dane o zakresie dokumentów uzyskiwane z różnych źródeł, w formie symboli klasyfikacji UKD i KDD. Dzięki temu, że w systemie umieszczono metadane opisujące syntaktykę i semantykę danych, istnieje możliwość wnioskowania o relacjach podobieństwa zakresów danych istniejących w zbiorze. Metadane techniczne mogą być także stosowane na poziomie pragmatycznym do odszukania interfejsów usług, co ułatwia ich właściwe wykorzystanie. Na przykład, jeżeli z zewnętrznego systemu przesyłane jest żądanie wykonania określonego zadania, to system musi posiadać metadane opisujące interfejs służący realizacji tego zadania. Metadane służą także do opisu stanu systemu i sieci po każdej zmianie tego stanu, co pozwala na określenie ich dynamiki. Według Manso-Callejo, Wachowicz i Bernabé-Poveda metadane wspomagające współdziałanie semantyczne służą również współdziałaniu dynamicznemu i organizacyjnemu (Manso-Callejo & Wachowicz & Bernabé-Poveda, 2009). Współdziałanie systemów oznacza więc m.in. współdziałanie metadanych, tworzonych i stosowanych w tych systemach.

Tak więc metadane w systemach informacyjnych stosowane są do wielu celów; opis danych i informacji przechowywanych w tych systemach, służący do ich identyfikacji i odszukania, jest tylko jednym z nich, ale ważnym z punktu widzenia funkcjonowania systemów informacyjnych. W praktyce tylko tego rodzaju metadane dostrzegane są przez użytkownika końcowego korzystającego z usług informacyjnych, na przykład przy okazji wyświetlania ich podczas prezentacji wyników wyszukiwania.

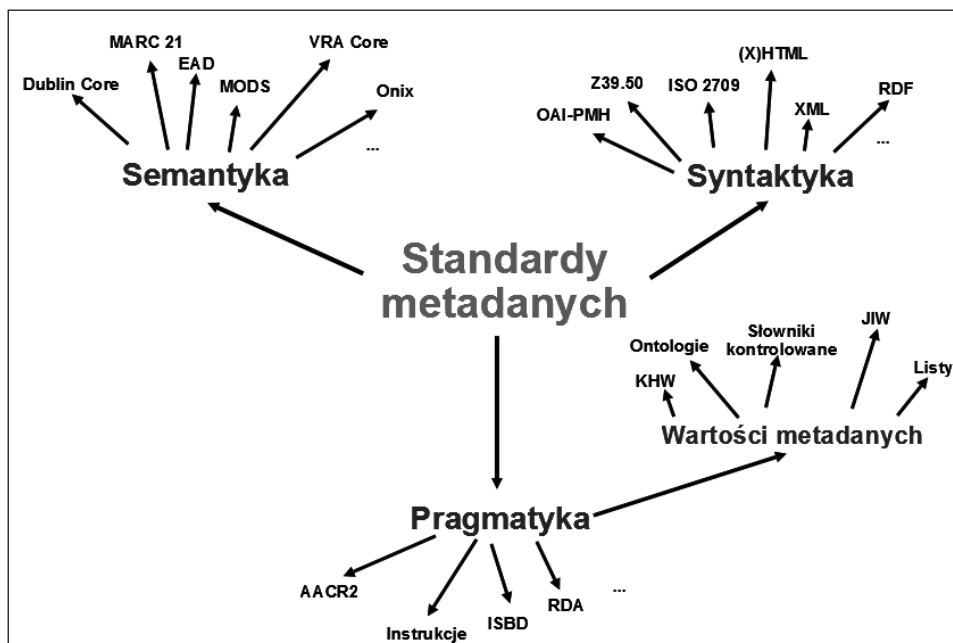
Współdziałanie metadanych oznacza możliwość ich wymiany pomiędzy systemami informacyjnymi bez utraty (lub z akceptowalnie niewielką stratą) ich funkcjonalności na poziomie syntaktycznym, semantycznym i/lub pragmatycznym (por. Nahotko, 2012, s. 364). Dzięki współdziałaniu metadane opisowe (dalej zajmę się tylko tego rodzaju metadanymi), pozyskane z jednego systemu mogą być scalane (łączone) z metadanymi w innym systemie, a zawartość rekordu bibliograficznego jest swobodnie mieszana z zawartością innych rekordów, zawierających komplementarne dane. W efekcie metadane opisujące określony obiekt mogą być pobierane, umieszczane i rozumiane przez różne systemy, co umożliwia ich agregację lub integrację z metadanymi z innych źródeł i/lub opisujących inne obiekty.

Syntaktyka metadanych dotyczy głównie problemów kodowania struktur metadanych i wystąpień elementów metadanych (poziom semantyczny i pragmatyczny) umożliwiającego zapis i odczyt do/z nośnika oraz „rozumienie” przez oprogramowanie podczas przetwarzania. Szczegółowe specyfikacje elementów, tworzone na poziomie semantycznym tutaj uzupełniane są o opis sposobu ich kodowania oraz tworzenia powiązań między elementami. Na tym poziomie wykorzystywane są takie standardy, jak języki kodowania danych (X) HTML, XML, RDF(S) oraz protokoły wymiany metadanych Z39.50 i OAI-PMH.

Semantyka dotyczy głównie definicji znaczeń i funkcjonalności elementów metadanych, a więc także – pośrednio – całego schematu. W efekcie definiowania znaczeń powstaje schemat metadanych, czyli zestaw elementów o precyzyjnie zdefiniowanej semantyce. Semantyka schematu jest definiowana przez znaczenie jego elementów. Na tym poziomie elementy schematu metadanych są nazywane przy pomocy etykiet i określane są sposoby znaczeniowego wiązania elementów metadanych, na przykład z zastosowaniem ontologii.

Praktyczne stosowanie metadanych o określonej semantyce wymaga skodyfikowania ich pragmatyki, czyli tzw. zasad performancji, warunków funkcjonowania metadanych w komunikacji pomiędzy ich użytkownikami. Dla każdego zdefiniowanego elementu metadanych ustalane są zasady tworzenia wartości elementu. Współdziałanie na wszystkich poziomach uzyskuje się m.in. dzięki standaryzacji metadanych (zob. rys. 1).

Współdziałanie metadanych może być zapewnione w bardzo różny sposób, na różnych etapach ich tworzenia i stosowania oraz z użyciem wielu narzędzi. Na rys. 2 przedstawione zostały różne narzędzia służące uzyski-



Rys. 1. Trzy poziomy współdziałania metadanych (por. Nahotko, 2012, s. 364).

Współdziałanie metadanych					
Przed utworzeniem rekordu			Po utworzeniu rekordu		
Ujednoczenie	Współdziałanie schematów		Współdziałanie rekordów		Współdziałanie repozytoriów
Wspólne standardy: Np. MARC, Dublin Core, MODS	Pochodne zestawy: MODS oparty na MARC, Rozszerzenie DC	Profil aplikacyjny: DC-Library AP, DC-Government AP	Konwersja: Np. z MARC do DC, z DC do MODS Podczas konwersji z bogatszego schematu do uboższego następuje utrata danych	Linked Data: Pozwala na wielokrotne stosowanie danych i zdecentralizowaną integrację Np. METS, RDF	Protokoły wymiany: Np. Z39.50, OAI-PMH
	Tablice przejścia: Mapowanie elementów jednego schematu na elementy innego schematu	Schematy przełącznikowe: Mapowanie pomiędzy wieloma schematami jednocześnie			
	Struktura nadrzędna: Np. ADN, OAIS	Rejestry metadanych: Np. UKOLN, OMR, LOV			
<i>Niepraktyczne w środowisku heterogenicznym</i>	<i>Konwersja stralna sprowadzająca metadane do najniższego wspólnego mianownika. utrata semantyki danych zamiast jej wzbogacenia</i>			<i>Modułowe rozwiązania integrujące różne standardy w jedną strukturę. metadane w różnych standardach dla różnych zasobów mogą być ze sobą łączone</i>	

Rys. 2. Narzędzia uzyskiwania współdziałania metadanych.

Na podstawie: <http://commons.wikimedia.org/wiki/File:Metadata-interoperability.png>.

waniu współdziałania metadanych przed (Chan & Zeng, 2006) i po utworzeniu rekordu metadanych (Zeng & Chan, 2006). Zaznaczono narzędzia Linked Data, które zostaną omówione w dalszej części artykułu.

Najprostszym, a przy tym najbardziej efektywnym sposobem zapewnienia współdziałania metadanych, jest stosowanie tego samego schema

tu wraz z zasadami performancji przez wszystkie strony zainteresowane wymianą metadanych. Jeżeli we wszystkich współpracujących systemach (np. w katalogach bibliotecznych lub bibliotekach cyfrowych) stosowany jest ten sam schemat (w środowisku użytkowników MARC nazywany też formatem) metadanych, to używane słowniki i tworzone rekordy mogą być łatwo wymieniane, chociaż nawet w takim przypadku nie obywa się bez sytuacji kłopotliwych.

Bibliotekarze od dawna dobrze rozumieli zalety standaryzacji. W środowisku sieci globalnych trzeba było pogodzić się jednak z rzeczywistością, w której funkcjonuje wiele struktur metadanych opisowych tworzonych w różnych zastosowaniach, nie tylko przez bibliotekarzy. W takim przypadku współdziałanie metadanych możliwe jest na trzech poziomach: schematów metadanych, repozytoriów (bibliotek cyfrowych) oraz rekordów metadanych (por. Chan & Zeng, 2006; Zeng & Chan, 2006).

Współdziałanie na poziomie schematów metadanych musi być zapewnione jeszcze przed rozpoczęciem tworzenia rekordów i z natury rzeczy dotyczy współdziałania semantycznego. Prace te są najbardziej zróżnicowane metodologicznie. Jedną ze stosowanych metod jest tworzenie tzw. *zestawów pochodnych* (ang. *derivation*), czyli schematów metadanych budowanych na podstawie wcześniej istniejących, najczęściej bardziej rozbudowanych (np. takich jak MARC). Zestawy pochodne powstają w wyniku takich procesów, jak: adaptacja, modyfikacja, rozszerzenie, tłumaczenie. W ten sposób powstał MODS na podstawie formatu MARC oraz tłumaczenia na wiele języków oryginalnego schematu Dublin Core². *Profile aplikacyjne* zawierają elementy metadanych pobrane z jednego lub kilku schematów metadanych i połączonych w nowy schemat, bez zmiany ich znaczenia. W ten sposób powstały schematy AVEL³, DC-Lib⁴, NBII⁵ i inne. *Tablice przejścia* (ang. *crosswalks*) służą do mapowania elementów metadanych, ich semantyki oraz syntaktyki z jednego schematu metadanych na elementy innego schematu. Mapowanie odbywa się zwykle na podstawie tablic wskazujących odpowiedniość semantyczną elementów mapowanych formatów względem siebie. Dzięki temu możliwe jest jednoczesne przeszukiwanie kilku heterogenicznych zasobów. Mapowanie może powodować utratę części semantyki. Twórcy większości schematów metadanych zapewniają ich mapowanie do najpopularniejszych formatów, takich jak MARC i DC. Odmianą tablic przejścia są *schematy przelącznikowe*, zapewniające mapowanie pomiędzy wieloma schematami jednocześnie; przykładem może być Vocabulary Map-

² Zob. wykaz wykonanych tłumaczeń na stronie Web: <http://dublincore.org/resources/translations/>.

³ Australasian Virtual Engineering Library (AVEL) Metadata Set składa się z 19 elementów: 14 pochodzi z Dublin Core, po jednym z AGLS i EDNA i 3 elementów administracyjnych, utworzonych lokalnie.

⁴ DC Library Application Profile (DC-Lib) składa się głównie z elementów DC i MODS. Służy stosowaniu DCMES w bibliotekach i zastosowaniach podobnego rodzaju (Clayphan, Guenther, 2004).

⁵ National Biological Information Infrastructure (NBII) posiada Biological Data Profile, wzorowany na Content Standard for Digital Geospatial Metadata (CSDGM).

ping Framework⁶, w którym wiele schematów metadanych umieszczono w matrycy, co ułatwia mapowanie. *Struktury nadrzędne* (ang. *frameworks*) stanowią rodzaj schematu służącego integracji różnych obiektów w określonym celu. Schemat taki może być tworzony przed powstaniem schematów metadanych i systemów, które mają współdziałać lub po ich powstaniu. Przykładem jest OAIS⁷, stanowiący strukturę nadrzędną, zawierającą terminy i pojęcia mogące służyć standaryzacji w zakresie archiwizacji cyfrowej. Wreszcie *rejestry metadanych* służą gromadzeniu danych o schematach metadanych. Stanowią narzędzie do identyfikacji i wyszukiwania istniejących schematów i profili aplikacyjnych, co może być przydatne podczas stosowania wielu sposobów zapewnienia współdziałania metadanych (np. mapowania, profili aplikacyjnych, zestawów pochodnych). Przykładem może być rejestr SCHEMAS⁸, tworzony przez UKOLN (obecnie nieaktywny); rejestr elementów jednego schematu – Dublin Core (DCMI Registry⁹); OMR¹⁰ zawierający wykazy elementów, m.in.: RDA, ISBD, FRBR i MARC 21 w RDF; oraz najnowszy LOV¹¹.

Współdziałanie na poziomie repozytoriów (bibliotek cyfrowych) realizowane jest zazwyczaj w trakcie wyszukiwania, a więc po utworzeniu rekordów metadanych. Po przeszukaniu wielu zasobów za pomocą wyszukiwarki zazwyczaj otrzymujemy zestaw niespójnych metadanych w różnych formatach. Ich ujednoczenie to zadanie z zakresu współdziałania syntaktycznego. Dla rozwiązania tego problemu systemy z bazami metadanych, powstałymi w wyniku ich agregacji z różnych zasobów, stosują spójne i niezawodne narzędzia służące integracji metadanych. Wybór odpowiedniej metody działania zależy od tego, czy w scalonym zasobie metadane mają zachować oryginalny format, czy mają być skonwertowane. W pierwszym przypadku należy zapewnić możliwość wyszukiwania w zbiorze heterogenicznych metadanych. W drugim trzeba zapewnić poprawne działanie narzędzi konwersji i integracji metadanych w wybranym formacie. Współdziałanie na tym poziomie osiągane jest poprzez: *zbieranie* (ang. *harvesting*) metadanych (np. z użyciem protokołu OAI-PMH lub Z39.50), *stosowanie wielu formatów bez konwersji rekordów* (np. DLESE Collection System¹²), *agregację elementów* metadanych i ich wartości pobieranych z różnych źródeł, *tablice przejścia*, mapowanie wartości metadanych dla wyszukiwania w wielu kolekcjach (np. haseł przedmiotowych w MACS¹³)

⁶ Vocabulary Mapping Framework, <http://www.doi.org/VMF/>.

⁷ OAIS – Open Archival Information System.

⁸ SCHEMAS: Forum for Metadata Schema Implementers, <http://www.ukoln.ac.uk/metadata/schemas/>.

⁹ The Dublin Core Metadata Registry, <http://dublincore.org/dcregistry/>.

¹⁰ Open Metadata Registry, <http://metadataregistry.org>.

¹¹ Linked Open Vocabularies, <http://lov.oktn.org/dataset/lov>.

¹² Digital Library for Earth System Education (<http://www.dlese.org/>)

¹³ Multilingual Access to Subjects (http://www.nb.admin.ch/nb_professionnel/projektarbeit/00729/00733/index.html?lang=en).

i *mapowanie współwystępowania* wartości metadanych (mapowanie zawartości słowników na podstawie współwystępowania tych wyrażen jako wartości elementów w rekordach).

W wielu przypadkach potrzeba współdziałania metadanych pojawia się, gdy zasoby metadanych, tworzone z wykorzystaniem różnych schematów, już istnieją. Rekordy metadanych mogą więc stać się przedmiotem wymiany. Najczęściej wykorzystywaną metodą integracji istniejących zasobów rekordów metadanych jest ich konwersja. Zazwyczaj efektem jest zbiór skonwertowanych rekordów. Problemem jest utrata semantyki, przekłamania danych, szczególnie podczas konwersji danych opartych na rozbudowanym schemacie do prostszego zestawu elementów; problemy na poziomie pragmatyki stwarza konwersja wartości metadanych pobieranych z różnych słowników kontrolowanych. Inna metoda polega na wzbogacaniu rekordów poprzez ich agregację w wyniku stosowania mechanizmu służącego identyfikacji podobnych pojęć w różnych bazach danych (Linked Data). Elementy metadanych różnego rodzaju, pochodzące z różnych schematów, słowników i aplikacji mogą być łączone w sposób zapewniający współdziałanie, gdyż budowane są według wspólnego modelu RDF. Rekord RDF scala w nowy zestaw (poprzez odnośniki) elementy wielu rekordów, które mogły być tworzone w różnym czasie do różnych celów.

Nowym, i jak się wydaje, przyszłościowym sposobem zapewnienia współdziałania metadanych w heterogenicznym środowisku sieci globalnych jest rozwój narzędzi związanych z ideą, określaną jako dane powiązane (Linked Data). W dalszej części artykułu przedstawię tę koncepcję w kontekście jej zastosowania dla uzyskania współdziałania metadanych i możliwości jej implementacji w systemach informacyjnych.

LINKED (META)DATA

W ostatnich latach wśród bibliotekarzy coraz bardziej popularne stają się rozwiązania dotyczące zarządzania i zapewnienia współdziałania masowych metadanych związane z technologiami tzw. sieci semantycznej (Semantic Web). W szczególności chodzi o zastosowanie RDF i wyników prac dotyczących modelowania metadanych z jego wykorzystaniem (Dunsire et al., 2012, p. 4). Tendencja ta staje się szczególnie widoczna po uproszczeniu wielu technologii sieci semantycznej przez W3C¹⁴, co doprowadziło do powstania Linked Open Data (LOD), praktyki stosowania identyfikatorów używanych w Web, modelowania przepływu danych dla poszczególnych dyscyplin i zastosowań oraz ontologii do projektowania metadanych funkcjonujących na zasadzie grafu. Chociaż Linked Data dotyczy wszel-

¹⁴ WWW Consortium (W3C) to organizacja powołana do tworzenia standardów dla Web. <http://www.w3.org/>.

kiego rodzaju danych (nie tylko bibliograficznych) publikowanych w Web, to ożywienie w ostatnich latach obserwowane w tworzeniu nowych narzędzi, słowników i zasobów danych idea ta zawdzięcza szerokiemu i aktywnemu włączeniu się do prac osób ze środowiska bibliotekarskiego, które wniosły rozwiązania stosowane w zakresie (meta)danych bibliotecznych. To dla przedstawicieli tego środowiska w 2010 r., w ramach konsorcjum W3C, powstał projekt Library Linked Data Incubator Group. Ma on wspomagać rozwój współdziałania metadanych bibliotecznych w sieci (Roszkowski, 2010, s. 61).

Linked Data ma służyć tworzeniu danych dotyczących wszelkich obiektów, nie tylko bibliograficznych, ale także np.: ludzi, organizacji, procesów i pojęć. Zasady Linked Data w zastosowaniu do danych bibliograficznych przedstawił Marcin Roszkowski (Roszkowski, 2010, s. 53 i nast.). Są to dane utworzone i opublikowane w taki sposób, aby umożliwić tworzenie połączeń pomiędzy zestawami danych oraz słownikami. Te ostatnie rozumiane są szerzej niż to było dotąd przyjęte w bibliotekarstwie, gdyż nazwą tą obejmuje się zarówno słowniki stosowane na poziomie pragmatyki, zawierające wartości metadanych, zwane także słownikami kontrolowanymi (KHW, słowniki JIW, wykazy kodów nazw języków, nazw geograficznych itp.), jak również zestawy pojęć stosowane na poziomie semantycznym, dotąd nazywane formatami lub schematami (wykazy elementów metadanych, pól i podpól formatów metadanych). W ten sposób do chmury danych trafiają słowniki zarówno poziomu semantycznego, jak i pragmatycznego metadanych, kodowane przy użyciu języków kodowania, a więc dostępne do bezpośredniego przetwarzania komputerowego (rys. 3). Każdy element słownika jest jednoznacznie identyfikowany za pomocą identyfikatora; na ogół stosowany jest URI¹⁵ ze względu na jego otwartość (może być przydzielany oddolnie, przez użytkownika¹⁶, bez pośrednictwa żadnej organizacji). Dzięki identyfikacji zasoby mogą być swobodnie cytowane przez inne zasoby, przez co metadane są powszechniej dostępne. Relacje pomiędzy identyfikowanymi elementami (dla danych bibliograficznych: dzieła, osoby, instytucje, miejsca, elementy treści itp.) opisywane są za pomocą języków, takich jak RDF¹⁷ i OWL¹⁸, których wyrażenia mogą być przetwarzane bezpośrednio przez komputery. Relacje te pozwalają na nawigację pomiędzy danymi zawartymi w źródłach publikowanych w Web, co powoduje ich integrację ułatwiającą współdziałanie techniczne. Współdziałanie organizacyjne (prawne) uzyskiwane jest dzięki otwarciu danych – mogą być one

¹⁵ URI – Uniform Resource Identifier.

¹⁶ URI oczywiście mogą być i są zarządzane również przez instytucje. Bardzo ważną cechą URI jest jego stabilność, dlatego też użytkownicy są zachęceni do rozsądnego tworzenia własnych przestrzeni nazw dla swoich URI, które powinny być niepowtarzalne i niezmiennie. Te właśnie cechy są łatwiej osiągalne pod nadzorem instytucjonalnym.

¹⁷ RDF – Resource Description Framework.

¹⁸ OWL – Web Ontology Language.

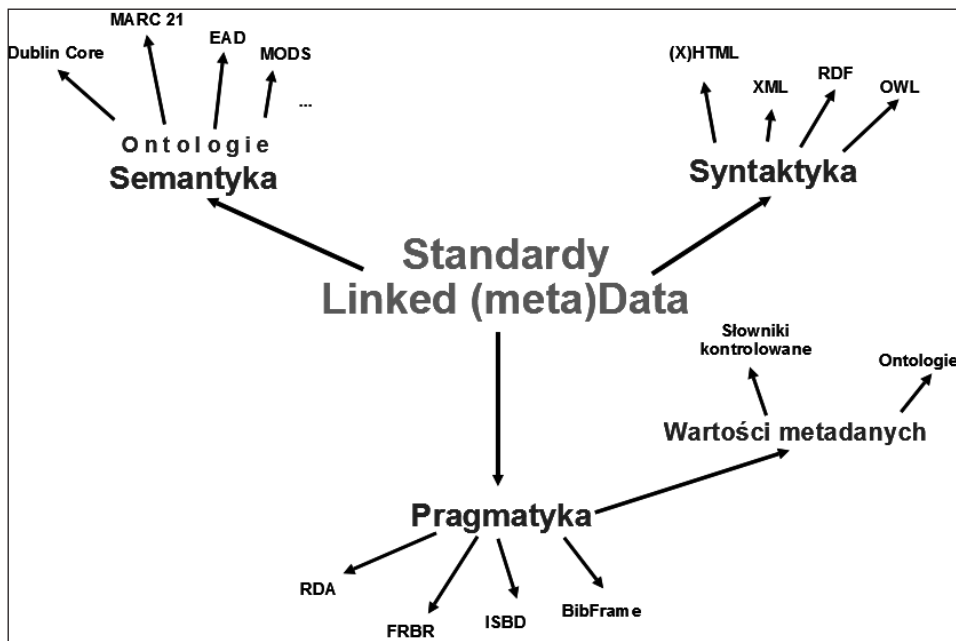
swobodnie stosowane i rozpowszechniane; stąd obok nazwy Linked Data pojawia się także określenie Linked Open Data (LOD). Otwartość danych nie jest niezbędna dla osiągnięcia współdziałania technicznego, ale umożliwia pełną efektywność stosowanych technologii.

Według Thomasa Bakera i grupy współpracowników (Baker et al., 2011) Linked Data ma wiele zalet w porównaniu z dotychczas stosowanymi i rozwijanymi od lat technologiami służącymi tworzeniu i rozpowszechnianiu metadanych bibliotecznych, gdyż stanowi ich rozwinięcie z myślą o zastosowaniu w sieciach globalnych. Dane powiązane są łatwe do rozpowszechniania, rozszerzalne i możliwe do wielokrotnego stosowania w wielu aplikacjach. Przynajmniej częściowo niezależne są od stosowanego języka naturalnego, gdyż URI zbudowany może być z dowolnych symboli, także liczbowych. Prace nad tworzeniem opisów zasobów mogą być realizowane na zasadzie współpracy bibliotek oraz innych instytucji, a także użytkowników. Ten rozproszony model danych pozwala „każdemu mówić cokolwiek o czymkolwiek”, co musi powodować zróżnicowanie opinii i brak zgodności co do znaczenia elementów (Allemang & Hendler, 2008, p. 7). 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).

Twórcy metadanych w Linked Data mogą udostępniać je w małych porcjach, jako indywidualne stwierdzenia („wypowiedzi”). Wypowiedzi te przyjmują formę trójek RDF, składających się z podmiotu (opisywany zasób, np. dzieło: jego identyfikator), predykatu lub własności (relacja, np. „ma autora”) i obiektu (np. nazwa lub identyfikator autora)¹⁹ (Nahotko, 2006, s. 27). Zamiast wymiany całych rekordów, zawierających pełny opis, jak to ma miejsce obecnie, możliwe jest więc dostarczanie tylko pojedynczych stwierdzeń o zasobie (trójek RDF), a wszystkie stwierdzenia o zasobie, mogące pochodzić z wielu źródeł, są agregowane w globalnym grafie. Nowe systemy biblioteczne mogą funkcjonować w oparciu o stwierdzenia istniejące w chmurze, uzupełniane danymi przechowywanymi lokalnie.

Z punktu widzenia współdziałania metadanych tworzonych w tym środowisku ważne jest oddzielenie semantyki od syntaktyki metadanych. W Linked (meta)Data znaczenie metadanych (semantyka) jest oddzielone od ich syntaktyki, dzięki czemu zmiana struktur nie powoduje modyfikacji znaczenia (rys. 3). Tylko pozornie różnica w sytuacji przedstawionej na rysunkach 1 i 3 polega na zastąpieniu jednych standardów innymi. W rzeczywistości nowe standardy mają jedną, ale bardzo ważną nową cechę –

¹⁹ Elementy trójek RDF bywają także nazywane podmiotem-orzeczeniem-dopełnieniem lub obiektem-predykatem (relacją)-wartością cechy.



Rys. 3. Standardy metadanych związane ze stosowaniem Linked Data. Oprac. własne.

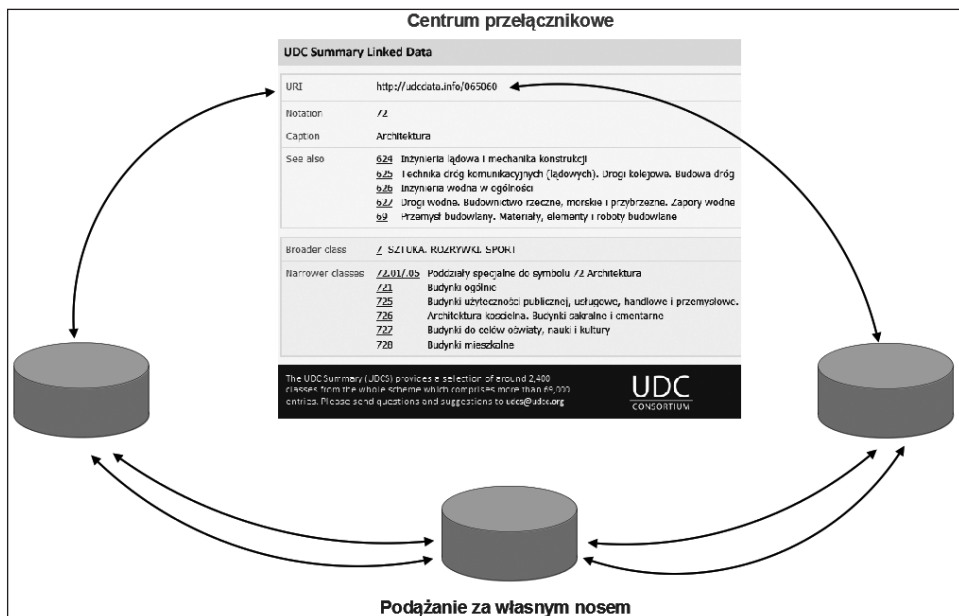
zarówno semantyka, jak i pragmatyka kodowane są za pomocą tej samej syntaktyki (głównie XML/RDF, choć istnieją inne serializacje), a co więcej, jest to syntaktyka, która umożliwia automatyczne przetwarzanie danych (dzięki jej „rozumieniu” przez komputery), bez udziału człowieka. Opasłe, drukowane tablice UKD przestają być potrzebne, powstaje ich wersja w Linked Data z użyciem SKOS²⁰ (na razie dostępne są tablice skrócone) (Hys, Kwiatkowska, 2013, s. 6). 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 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 wprowadzić kwalifikowany odnośnik, czyli taki, który nazywa relację występującą pomiędzy połączonymi elementami.

Tak zorganizowane środowisko metadanych umożliwia dwojakiego rodzaju współdziałanie metadanych. Pierwszy sposób na zapewnienie współdziałania związany jest z istnieniem słowników dwóch opisanych rodzajów, przydatnych szczególnie, gdy chcemy łączyć dane z różnych obszarów (dziedzin, zastosowań). Użytkownik Web może przeglądać jego zasoby

²⁰ SKOS – Simple Knowledge Organization System.

bez potrzeby znajomości 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 formatach (Bermes, 2011).

Zasobami danych tego rodzaju są także zawartości słowników obu wcześniej wymienionych rodzajów. Umożliwiają one współdziałanie dzięki temu, że działają jak centrum przełącznikowe, umieszczone w chmurze danych, łączące dane wyrażone zgodnie z różnymi semantykami danych. Takie centrum umożliwia w Linked Data nawigację od jednego zbioru metadanych do innego, przez podążanie za odnośnikami, czyli URI, nawet jeżeli dane podłączone do centrum są zróżnicowane. Wyobraźmy sobie na przykład, że istnieją dwa zasoby, w których znajdują się m.in. symbole UKD. Bez względu na to, w jakich miejscach różnych struktur metadanych (polach, elementach) umieszczone zostały te symbole, odsyłają one za pomocą URI do ontologii UKD, wykonanej w SKOS i dostępnej w Internecie (rys. 4). Dzięki temu, po pierwsze, następuje ujednoczenie znaczenia (semantyki) danych, po drugie, możliwe jest wyszukiwanie podobnych danych poprzez proste podążanie za odnośnikami URI. Sposób ten pozwala na unikanie niejednorodności dzięki porozumieniu ontologicznemu (Heath & Bizer, 2011, p. 24).



Rys. 4. Dwa modele współdziałania w Linked Data (na przykładzie UKD). Oprac. własne.

Co więcej, 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 danych 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 jednostki identyfikującej osobę można na przykład przyłączyć obiekty, miejsca, inne osoby itp. za pomocą relacji bycia autorem, pracownikiem, ojcem/matką, mieszkańcem, fanem i dowolnej innej. Taka forma aktywności bywa nazywana także „toURIs’m’em”. Odnośniki pomiędzy serwisami bibliotecznymi oraz innymi, takimi jak Wikipedia, GeoNames, MusicBrainz, BBC scalają lokalne zasoby w globalne uniwersum informacji w Web.

Twórcy danych w Linked Data powinni przede wszystkim sprawdzić, czy potrzebne im terminy nie zostały już zdefiniowane w jednym z istniejących słowników; w takim przypadku należy ich użyć do reprezentacji danych. Jeżeli odpowiednich terminów brak, to należy utworzyć własny słownik, którego terminy zostaną połączone z zasobami istniejących słowników. Gdyby w późniejszym czasie okazało się, że jakiś inny słownik zawiera te same terminy, które zostały umieszczone w naszym słowniku, należy utworzyć odnośnik RDF pomiędzy URI terminu w obu słownikach, ze wskazaniem na ekwiwalentność łączonych terminów za pomocą relacji przewidzianych w OWL, RDFS lub SKOS.

Już dostosowanie słowników używanych dotąd w bibliotekach do zasad Linked Data i udostępnienie ich do powszechnego stosowania umożliwi współdziałanie na podstawie tak powstałych centrów przełącznikowych od razu, bez żadnych dodatkowych prac. Według zasad stosowanych w Linked Data, jeżeli dwa zasoby danych, stosujących różne formaty, używają tego samego słownika wartości metadanych (jak w powyższym przykładzie – tablice UKD), to istnieje możliwość przechodzenia od jednego zbioru danych do innego poprzez odnośniki od tych zasobów do tego słownika. Jedynym wymogiem jest stosowanie identyfikatorów URI.

W bibliotekach zasada ta od dawna stosowana jest we wszystkich zintegrowanych systemach bibliotecznych, w których istnieją powiązania pomiędzy danymi bibliograficznymi i khw, często budowane przy użyciu kodów identyfikujących poszczególne wartości elementów khw (na wzór URI). Prace nad przystosowaniem funkcjonalności OPAC i zasad katalogowania do nowych modeli stosowanych na poziomie semantyki i pragmatyki metadanych, takich jak FRBR, FRAD, FRSAD i BIBFRAME oraz przepisów RDA implikują identyfikację większej ilości informacji, niż mogła być rozpo-

wszechniana pomiędzy rekordami OPAC za pomocą odnośników. Jednostki (ang. *entity*) wyróżnione w wymienionych modelach mogą być uważane za węzły dla odnośników łączących różne obiekty: dzieła, realizacje, materializacje, osoby, ciała zbiorowe, hasła przedmiotowe i inne. Stosowanie tych jednostek może być łatwiej rozpowszechniane w społecznościach poza bibliotekami, dzięki czemu dane biblioteczne znajdują nowe zastosowania.

Szczególnie interesującym modelem jest BIBFRAME, produkt Library of Congress, utworzony na podobnych zasadach co FRBR²¹ (oba modele reprezentują typ modeli ER), ale specjalnie z myślą o Linked Data (Miller et al., 2012, p. 9). Twórcy modelu wyszli naprzeciw postulatowi uproszczenia modeli stosowanych w środowisku bibliotecznym tak, aby były one zrozumiałe także poza tym środowiskiem, co również ułatwia współdziałanie. Dzięki jego prostocie osobom stosującym ten model nie trzeba tłumaczyć, czym różni się Dzieło od Realizacji lub co to jest Materializacja; model jest więc mniej hermetyczny dla osób spoza środowiska bibliotekarskiego niż FRBR. Zgodnie z zasadami Linked Data wszystkie jednostki, ich atrybuty i relacje między nimi są identyfikowane za pomocą URI (Roszkowski, 2013, s. 32). Model taki może ułatwić współdziałanie na poziomie semantyki metadanych (schematy metadanych) stanowiąc dla nich rodzaj centrum przełącznikowego. Jeżeli dwa elementy struktury różnych słowników Linked Data są łączone odnośnikami z elementem BIBFRAME, traktowanym jako struktura nadrzędna, to połączenie takie wskazuje na ich podobieństwo semantyczne.

Model BIBFRAME składa się z niewielu elementów. Są to:

1. Dzieło (*creative work*) – zasób odzwierciedlający treści konceptualne opisywanej pozycji.

2. Wystąpienie dzieła (egzemplarz – *instance*) – zasób odzwierciedlający indywidualne, materialne ucieleśnienie dzieła.

3. Hasło (*authority*) – zasób odzwierciedlający główną koncepcję pochodzenia informacji, która zdefiniowała relacje przedstawione w dziele i jego wystąpieniu. Przykładami haseł są nazwy osób, miejsc, przedmiotów (dzieł), organizacji itp.

4. Adnotacja (*annotation*) – zasób, który opisuje inny zasób BIBFRAME zawierając o nim dodatkowe informacje. Przykładami adnotacji są dane o znaku miejsca (sygnaturze) w bibliotece, wizerunek okładki, recenzja (Miller et al., 2012, p. 8).

Do zadań organizacji zarządzających standardami metadanych bibliotecznych należeć będzie tworzenie powiązań pomiędzy elementami modelu BIBFRAME i odpowiednich słowników, takich jak Dublin Core, FOAF, zasobów kodowanych w SKOS (tezaury, klasyfikacje) i słowników, które powstaną w przyszłości w celu obsługi różnych aspektów działalności bibliotecznej.

²¹ Por. też (Śnieżko, 2013, s. 7).

Implementacja modelu Linked Data stwarza oczywiście wiele problemów, dla których wciąż poszukuje się rozwiązań:

- Konieczność zmiany istniejącego modelu metadanych na reprezentację RDF, co stwarza problemy związane po części z trudnościami przedstawienia niektórych informacji w postaci trójek, a częściowo z nowymi możliwościami, jakie daje RDF w stosunku do używanych dotąd modeli. Stosunkowo łatwo można zastosować RDF w implementacji FRBR i BIBFRAME, z różnych powodów wciąż niezbyt często stosowanych w praktyce informacyjnej (nad tym ostatnim prace wciąż trwają).

- Poprawne zarządzanie URI; niezbędne jest przypisanie identyfikatora do każdego elementu danych przez organizację obsługującą standard. W przypadku braku takiego działania pojawiać się mogą nieautoryzowane URI, nadawane przez przypadkowe osoby, co powoduje, że temu samemu elementowi przypisano wiele URI. Dodatkowe problemy stwarza konieczność podejmowania decyzji, czy tworzyć jak najwięcej własnych URI, czy tam, gdzie to możliwe, korzystać z już istniejących. Konieczne decyzje mogą też dotyczyć sposobów zapewnienia długotrwałości identyfikatora oraz zasad tworzenia (np. czy ma być czytelny dla człowieka).

- Wiele problemów dotyczy także współdziałania słowników. Oprócz słowników zawierających wykazy elementów i zestawy wartości metadanych istnieją także zasady, np. tworzenia dat, jak również systemy organizacji wiedzy o obiektach świata materialnego, takich jak osoby. Dla opanowania tej sytuacji tworzone są specjalne struktury danych, takie jak VIAF²², gdzie informacja o osobie i o etykiecie traktowane są jak odrębne elementy.

- Poważnym problemem jest tworzenie odnośników pomiędzy elementami słowników. Oznacza to ich mapowanie względem siebie. Mapowanie symboli UKD do KDD, a tych do haseł LCSH nigdy nie było trywialnym zadaniem, ale też w zamkniętym środowisku bibliotek tradycyjnych miało ograniczone zastosowanie. W środowisku Linked Data wszystko ma być powiązane ze wszystkim i to za pomocą oznaczonych relacji. Powiązania mi tymi trzeba zarządzać, np. modyfikować wraz ze zmianami dokonywanymi w słownikach.

LINKED DATA W SYSTEMACH BIBLIOTECZNYCH

Opisane rozwiązania mogą stać się podstawą tworzenia przez dostawców oprogramowania bibliotecznego zintegrowanych systemów bibliotecznych nowej generacji (Wilson, 2012, p. 110). W tym celu realizowane są eksperymenty dla dostosowania do środowiska Linked Data tzw. Discovery Systems, systemów bibliotecznych nowej generacji, pozwalających

²² The Virtual International Authority File, <http://viaf.org/>.

na wyszukiwanie zintegrowane zasobów drukowanych i elektronicznych (Nahotko, 2011, s. 204). Zasadniczą częścią każdego z nich jest system przetwarzania metadanych w chmurze. Realizowanych jest kilka takich rozwiązań. Produkt firmy Ex Libris nazywa się Alma²³, SerialsSolution oferuje system Intota²⁴, OCLC – WorldShare Management Services²⁵. Swoją ofertę przedstawił także VTLS (system Open Skies)²⁶ oraz Kual Foundation (system open source OLE²⁷). Ich zasadniczą cechą jest integracja nie tylko rozumiana tak jak dotychczas (integracja modułów oprogramowania i procesów realizowanych w ramach systemu bibliotecznego), ale także jako integracja systemu bibliotecznego ze światem zewnętrznym. Według Marshalla Breedinga wkrótce systemy tego typu pozwolą na tworzenie większości metadanych jako LOD, przy czym każdy sukces osiągnięty w praktycznym zastosowaniu Linked Data będzie powodował dalszy postęp w rozwoju dostępu bibliotek i użytkowników do metadanych i zasobów przy jednoczesnej redukcji uzależnienia od rozwiązań komercyjnych (Breeding, 2012, p. 14).

Każdy z wymienionych systemów ma zastępować co najmniej kilka narzędzi bibliotecznych, takich jak zintegrowany system biblioteczny, system zarządzania zasobami elektronicznymi²⁸, baza wiedzy²⁹, resolver OpenURL, lista A-Z. Powoduje to konieczność stosowania zintegrowanego modelu danych. Dzięki temu użytkownik za pomocą jednego interfejsu może wykonywać w jednolity sposób wiele zadań w różnych miejscach systemu.

Większość wymienionych systemów funkcjonuje w modelu SaaS³⁰ lub PaaS³¹, co oznacza, że oprogramowanie to udostępniane jest w chmurze danych. Dostawcy oprogramowania oferują oprogramowanie i usługi w chmurze, udostępniając je użytkownikom poprzez interfejsy Webowe. Dzięki temu znikną także problemy z instalacją nowych wersji systemów: będą one aktualizowane bez żadnego udziału biblioteki i przerw w jej działaniu (tzw. *agile development technology*). Z drugiej strony biblioteka traci kontrolę nad swoim oprogramowaniem³².

Z punktu widzenia tworzenia i współdziałania metadanych interesujące rozwiązania oferują twórcy systemu Alma (Ex Libris). Jego integralną czę-

²³ Aleph Polska, <http://www.aleph.pl/produkty/alma/>.

²⁴ Serials Solutions, <http://www.serialsolutions.com/en/services/intota>.

²⁵ OCLC, <http://www.oclc.org/webscale/default.htm>.

²⁶ VTLS, <http://www.vtls.com/openskies>.

²⁷ Kual, <http://www.kuali.org/ole>.

²⁸ Ang. Electronic resource management system (ERMS).

²⁹ Baza wiedzy jest tu dość specyficznie rozumiana, jako połączenie danych o zasobach elektronicznych biblioteki i linków, za pomocą których uzyskuje się dostęp do tych zasobów z uwzględnieniem posiadanych licencji. Baza wiedzy służy też obsłudze procesów związanych z zasobami elektronicznymi.

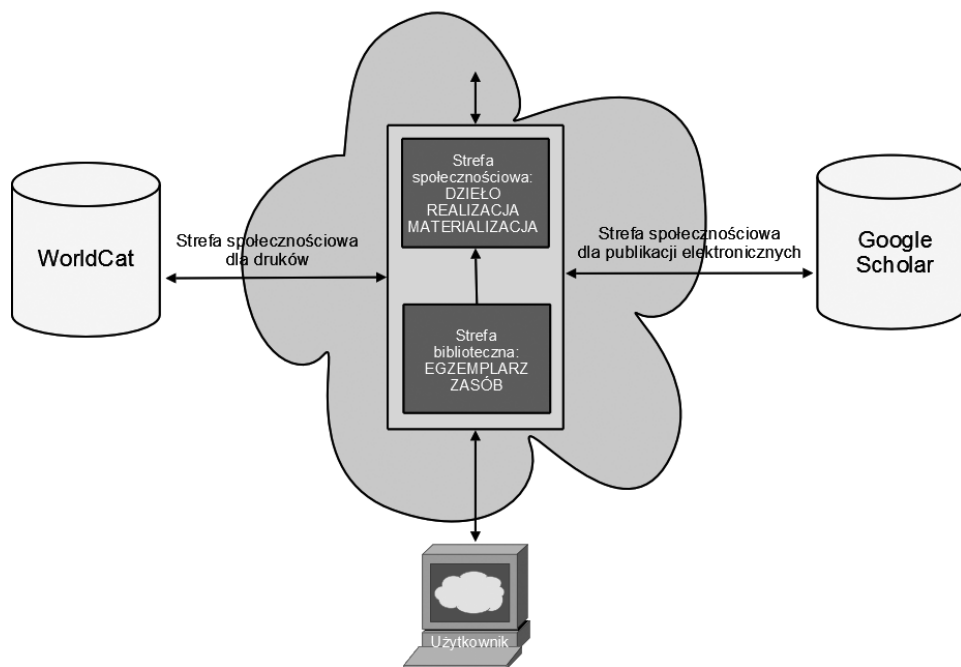
³⁰ SaaS – Software as a Service, oprogramowanie jako usługa (dostępna w Internecie).

³¹ PaaS – Platform as a Service, platforma jako usługa.

³² Przy okazji zmienia się rozumienie „własności” oprogramowania bibliotecznego: biblioteki raczej otrzymują do niego licencjonowany dostęp niż go kupują.

ścią jest Strefa Społecznościowa (ang. *Community Zone*), obejmująca społecznościowy katalog zawierający rekordy metadanych (*Metadata Management System*) wszystkich bibliotek, w których używana jest Alma oraz centralną bazę wiedzy (Wang & Dawes, 2012, p. 82). MMS obsługuje kilka formatów metadanych (MARC 21, Dublin Core, MODS) bez potrzeby dokonywania mapowania. Formaty te są częścią MMS. Katalog społecznościowy służy jako źródło metadanych opisowych, pozwalając bibliotekom na pracę bezpośrednio w środowisku wspólnego zarządzania metadanymi. Bibliotekarze mogą dodawać rekordy zasobu bezpośrednio do wspólnych rekordów metadanych, kopiować je do lokalnych katalogów lub edytować i modyfikować rekordy wspólne. Katalog społecznościowy Almy współdziała z innymi źródłami, dostarczonymi przez instytucje „trzeciej strony”, używanymi przez bibliotekę, takimi jak WorldCat lub krajowe katalogi centralne. Bibliotekarze zachęceni są do umieszczania metadanych w tym otwartym środowisku bez ograniczania warunków wykorzystania tych rekordów. Do wyszukiwania rekordów metadanych z zewnętrznych baz danych i kopiowania ich do własnego katalogu stosowane są protokoły Z39.50 lub SRU. Wygląda to na stosowanie Z39.50 do realizacji celów, do których stosowany jest raczej protokół OAI.

Alma obsługuje centralną bazę wiedzy, zarządzaną i aktualizowaną przez Ex Libris. Jest ona częścią Strefy Społecznościowej i zawiera informacje



Rys. 5. Zasoby metadanych w chmurze. Oprac. własne.

o pakietach bezpłatnych i licencjonowanych zasobów elektronicznych dostępnych u różnych dostawców. Jest udostępniana wszystkim członkom społeczności Alma.

Zasoby bibliograficzne przechowywane w chmurze w Strefach Społecznościowych, pozwalają na współzarządzanie metadanymi pochodzącymi z wielu źródeł. Strefy te tworzone są z myślą o bibliotekach współużytkujących to samo oprogramowanie komercyjne, lecz można sobie wyobrazić również Strefę znacznie szerszej, jako zasoby metadanych przydatnych użytkownikom bibliotek. Uwolni to biblioteki od potrzeby tworzenia własnych, zamkniętych zasobów metadanych. W takim przypadku Strefą Społecznościową dla metadanych o drukach mógłby być WorldCat, natomiast indeks Google Scholar mógłby służyć jako Strefa Społecznościowa dla dokumentów elektronicznych (zob. rys. 5).

http://lobid.org/resource/HT002948556	
Titel	With reference to reference
Autor	http://d-nb.info/gnd/135539897
Erscheinungsjahr	1983
Typ	< http://purl.org/ontology/bibo/Book >
Typ	< http://purl.org/vocab/frbr/core#Manifestation >
Sprache	http://id.loc.gov/vocabulary/iso639-2/eng
Erscheinungsort	Indianapolis u.a.
Verlag	Hackett
Format	print
Umfang	VIII, 200 S.
Schlagwort	Referenz <Linguistik>
Schlagwort	Semantik
ISBN-10	0915145529
ISBN-10	0915145537
ISBN-13	9780915145539
ISBN-13	9780915145522
Exemplar	http://lobid.org/item/HT002948556%3AAAE1772
Typ	< http://purl.org/vocab/frbr/core#Item >
Zugehöriger Titel	http://lobid.org/resource/HT002948556
Besitzer	Rheinisch-Westfälische TH, Hochschulbibliothek

Rys. 6. Biblioteczny opis bibliograficzny z zastosowaniem Linked Data w systemie lobid.org. Oprac. własne na podstawie <http://lobid.org>

Sposób funkcjonowania środowiska informacyjnego, przedstawiony na rys. 5 pozwala bibliotece na pełnienie funkcji interfejsu wyszukiwawczego do wszystkich zasobów niezbędnych jej użytkownikom. Bez względu na formę dokumentów i ich umiejscowienie użytkownik otrzyma o nich możliwie pełną metainformację, która także dostępna będzie w chmurze, jako powiązane wyrażenia Linked Data. Bibliotekarze, zamiast katalogować swoje zbiory w tradycyjny sposób, będą wykonywali czynności nazywane „katalinkowaniem” (Kaschte, 2013, p. 321). Przykładowy rekord metadanych, pochodzący z systemu lobid.org (Linking Open Bibliographic Data), z odnośnikami URI do słowników Linked Data przedstawiony jest na rys. 6. Spowoduje to udostępnienie zasobów bibliotek w Web na tych samych zasadach, co innych serwisów; metadane biblioteczne będą w pełni dostępne dla systemów zewnętrznych, ale, co równie istotne, biblioteki będą także pobierać elementy rekordów z zewnątrz. Rozwiązania w tym zakresie funkcjonują na razie w fazie eksperymentów.

ZAKOŃCZENIE

Przedstawione tu zagadnienia współdziałania metadanych obrazują ich rozwój, którego dotychczasowym zwieńczeniem jest Linked Data. Od XIX wieku rozwijane i doskonalone są standardy dotyczące zarówno semantyki, pragmatyki, jak i syntaktyki metadanych. Obecne rozwiązania zmierzają w kierunku dostosowania standardów do potrzeb środowiska Web. Dzięki stosowaniu RDF słowniki metadanych są łatwe do zdefiniowania oraz samoopisujące się, co ułatwia współdziałanie, a to z kolei pozwala na dokonywanie ich rekombinacji: do budowy opisu stosowane są elementy metadanych z wielu różnych słowników i zasobów. Tworzenie narzędzi dla Linked Data dało w środowisku bibliotekarskim silny impuls do tworzenia powiązań pomiędzy narzędziami (słownikami) dotychczas stosowanymi odrębnie, chociaż oczywiście prace służące współdziałaniu, takie jak mapowanie, realizowane były zawsze, jednak nie na taką skalę. Praca ta oczywiście nigdy się nie skończy; słowniki ulegają stałym modyfikacjom i powiązania między nimi także muszą być aktualizowane. Ostatycznym efektem może być większa standaryzacja słowników. Pozostaną tylko te, które będą stale rozwijane, inne zostaną wchłonięte lub porzucone.

Warto także wspomnieć, że Linked Data to nie jedyna inicjatywa mająca na celu zwiększenie semantyzacji procesów wyszukiwania informacji. Chodzi o takie, czasem bardzo różne inicjatywy, jak Knowledge Graph Google'a, Satori Microsoftu oraz Wolfram Alpha. Knowledge Graph (oraz mniej zaawansowany Satori) to nowa technologia służąca dostarczaniu danych (faktów) o ludziach, miejscach i rzeczach, wyświetlanych obok tradycyjnych wyników wyszukiwania przedstawianych przez wyszukiwarę. W ten sposób inżynierowie Google wskazują na nowe możliwo-

ści wyszukiwawcze, gdzie wynikiem nie są strony Web odpowiadające zapytaniu, w których trzeba szukać potrzebnej informacji, ale obiekty („jednostki”) opisywane terminami zawartymi w zapytaniu. W tym celu są gromadzone fakty dotyczące ludzi, miejsc i rzeczy materialnych oraz tworzone są relacje pomiędzy nimi na podstawie zachowań użytkowników wyszukiwarki. Dzięki temu, pytając o znaną osobę, użytkownik otrzyma na przykład informację o jej rodzinie, znajomych, dziełach i osobach zajmujących się pokrewną działalnością. Podobny efekt, ale w zupełnie inny sposób osiągnąć można stosując Wolfram Alpha³³. Wyszukiwarka ta w ogóle nie przeszukuje stron Web; odpowiada na zapytania przedstawiając fakty, ale nie wyszukuje ich w sieci, tylko w specjalnie przygotowanej (przez ludzi) i komputerowo przetwarzanej bazie wiedzy, zawierającej wyselekcjonowane w procesach intelektualnych dane, algorytmy, modele i metodologie. Tu również, pytając na przykład o znaną osobę otrzymamy jej wizerunek, daty i miejsca urodzin i śmierci oraz podstawowe informacje o działalności. Informacje te są przedstawiane w oparciu o relacje ontologiczne między pojęciami lub danymi gromadzonymi w bazie wiedzy Wolfram Alpha. Biorąc pod uwagę te przykłady oraz to, co wcześniej napisałem o Linked Data, można zauważyć stałą dychotomię rozwiązań służących wyszukiwaniu informacji: obok procesów intelektualnych wykonywanych przez ludzi realizowane są komputerowo procesy zalgorytmizowane.

W obliczu tych wszystkich zmian i wynalazków dużym modyfikacjom ulega profesja bibliotekarzy. Z jednej strony po raz kolejny okazało się, że w tym środowisku od wielu dziesięcioleci powstawały rozwiązania (w tym przypadku zasoby semantyki i zasady pragmatyki), obecnie gotowe do przeniesienia do nowych standardów i wzbogacenia w kierunku tworzenia z nich ontologii. Z drugiej strony bibliotekarze na rzecz społeczności sieciowej tracą kolejne funkcje, które od zawsze wydawały się immanentnie przypisane do tego zawodu: katalogowanie staje się „katalinkowaniem” dostępnym każdemu. Wszelkie zasoby biblioteczne, informacji, danych i metadanych przenoszone są do sieci, gdzie wiązane są relacjami tworząc zasoby globalnej biblioteki cyfrowej.

BIBLIOGRAFIA

- Allemang, Dean; Hendler, Jim (2008). *Semantic Web for working ontologist: effective modeling in RDFS and OWL*. Amsterdam: Morgan Kaufmann.
- Baker, Thomas [et al.] (2011). *Library Linked Data Incubator Group final report* [online]. Library Linked Data Incubator Group; [dostęp: 28.05.2013]. Dostępny w World Wide Web: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>.

³³ WolframAlpha, <http://wolframalpha.com/>.

- Bermes, Emmanuelle (2011). Convergence and interoperability: a Linked Data perspective. In: *World Library and Information Congress: 77th IFLA General Conference and Assembly* [online]. Puerto Rico, 13-18 August 2011. [dostęp: 28.05.2013]. Dostępny w World Wide Web: <<http://conference.ifla.org/past/ifla77/149-bermes-en.pdf>>.
- Breeding, Marshall (2012). Current and future trends in information technologies for information units. *El profesional de la información*, vol. 21, no. 1, pp. 9-15.
- Chan, Lois; Zeng, Marcia (2006). Metadata interoperability and standardization – a study of methodology. Part I. Achieving interoperability at the schema level. *D-Lib Magazine* [online], vol. 12, no. 6 [dostęp: 1.06.2013]. Dostępny w World Wide Web: <<http://www.dlib.org/dlib/june06/chan/06chan.html>>.
- Clayphan, Robina; Guenther, Rebecca (2004). *DC-Library Application Profile (DC-Lib)* [online]. Dublin Core Metadata Initiative; [dostęp: 25.05.2013]. Dostępny w World Wide Web: <<http://dublincore.org/documents/library-application-profile/>>.
- Dunsire, Gordon [et al.] (2012). Linked Data vocabulary management: infrastructure support, data integration, and interoperability. *Information Standards Quarterly*, vol. 24, no. 2/3, pp. 4-13.
- Heath, Tom; Bizer, Christian (2011). *Linked Data. Evolving the Web into a Global Data Space*. San Rafael, CA: Morgan & Claypool Publ.
- Hys, Jolanta; Kwiatkowska, Joanna (2013). UDC Summary. *Bibliotekarz*, nr 4, s. 4-7.
- Kaschte, Axel (2013). Linked open data on its way into next generation library management and discovery solutions. *Italian Journal of Library and Information Science*, vol. 4, no. 1, pp. 313-323.
- Manso-Callejo, Miguel; Wachowicz, Monica; Bernabé-Poveda, Miguel (2009). Automatic metadata creation for supporting interoperability levels of spatial data infrastructures. In: *GSDI 11 World Conference Spatial Data Infrastructure Convergence* [online]. Rotterdam, 15-19 June 2009 [dostęp: 25.05.2013]. Dostępny w World Wide Web: <<http://www.gsdi.org/gsdiconf/gsd11/papers/pdf/194.pdf>>.
- Miller, Eric [et al.] (2012). *Bibliographic Framework as a Web of data: Linked Data model and supporting services*. Washington: Library of Congress.
- Nahotko, Marek (2012). Struktury danych. W: *Biblioteki cyfrowe*. Red. M. Janiak, M. Krakowska, M. Próchnicka. Warszawa: Wydaw. SBP, s. 362-371.
- Nahotko, Marek (2011). Integracja wyszukiwania w zasobach informacyjnych. *Przegląd Biblioteczny*, z. 2, s. 192-210.
- Nahotko, Marek (2006). *Opis dokumentów elektronicznych. Teoretyczny model i możliwości jego aplikacji*. Kraków: Wydaw. UJ.
- Nahotko, Marek (2004). *Metadane. Sposób na uporządkowanie Internetu*. Warszawa: Wydaw. SBP.
- Roszkowski, Marcin (2013). Od MARC 21 do Semantic Web: reprezentacja metadanych bibliograficznych w środowisku sieciowym. W: *Bibliografi@ : źródła, standardy, zasoby*. Red. Jerzy Franke. Warszawa: Wydaw. SBP, s. 13-37.
- Roszkowski, Marcin (2010). Linked Data – model danych powiązanych w Semantic Web. *Zagadnienia Informacji Naukowej*, nr 2, s. 52-68.
- Sosińska-Kalata, Barbara (2013). Ewolucja paradygmatu badań organizacji wiedzy. W: *Nauka o informacji w okresie zmian*. Red. Barbara Sosińska-Kalata, Ewa Chuchro. Warszawa: Wydaw. SBP, s. 113-127.
- Śnieżko, Leszek (2013). Jak wam się podoba? Czyli pierwsza odsłona następcy MARC-a. *Tytuł Ujednolicony*, nr 9, s. 6-7.
- Ustawa z dn. 12 lutego 2010 r. o zmianie ustawy o informatyzacji działalności podmiotów realizujących zadania publiczne oraz niektórych innych ustaw (2010). *Dz. U.*, Nr 40, poz. 230, s. 3507-3519.
- Wang, Yongming; Dawes, Trevor (2012). The next generation integrated library system: a promise fulfilled? *Information Technology and Libraries*, vol. 31 no. 3, pp. 76-84.

- Wilson, Kristen (2012). Introducing the next generation of library management systems. *Serials Review*, vol. 38, no. 2, pp. 110-123.
- Zeng, Marcia; Chan, Lois (2006). Metadata interoperability and standardization – a study of methodology. Part II. Achieving interoperability at the record and repository levels. *D-Lib Magazine* [online], vol. 12, no. 6 [dostęp: 15.05.2013]. Dostępny w World Wide Web: <<http://www.dlib.org/dlib/june06/zeng/06zeng.html>>.

MAREK NAHOTKO
Institute of Information and Library Studies
Jagiellonian University
e-mail: marek.nahotko@uj.edu.pl

METADATA INTEROPERABILITY IN THE CLOUD

KEYWORDS: Metadata interoperability. Linked Data. Next generation library systems.

ABSTRACT: **Objective** – The paper is an attempt to present the most recent trends for ensuring interoperability of library metadata. **Research methods** - The author uses the analysis of the literature concerning the interoperability of information systems and tools ensuring the interoperability of metadata. **Results** - Various standardization steps in the field of library metadata, implemented since the nineteenth century, currently are adjusted to the needs and capabilities of wide-area networks. In this environment the interoperability of metadata is particularly important. The paper presents problems of interoperability of information systems and tools ensuring metadata interoperability. Recent activities in this field are related to the concept of Linked Data. Metadata vocabularies used in the area of librarianship are being transferred to RDF in order to make them available as Linked Data. There is also an ongoing work on the implementation of Linked Data in integrated library systems. **Conclusions** - The adjustment of standards used in libraries to the solutions commonly used in wide-area networks offers a new perspective for the interoperability of metadata, and, indirectly, for the operation of information retrieval systems.

Artykuł w wersji poprawionej wpłynął do Redakcji 12 lutego 2014 r.