

Marcin Roszkowski
Instytut Informacji Naukowej i Studiów Bibliologicznych
Uniwersytet Warszawski
m.roszkowski@uw.edu.pl

Kartoteki nazw osobowych w środowisku sieciowym

Streszczenie: *Celem tekstu jest charakterystyka zmian w roli kartotek haseł wzorcowych dla nazw własnych oraz wskazanie podstaw takich modyfikacji. Główną tezę postawioną w publikacji jest podkreślenie roli, jaką mogą pełnić prowadzone przez biblioteki kartoteki haseł wzorcowych jako wartościowe i wiarygodne źródła informacji w kontekście rozwoju sieci semantycznej i publikowania w modelu Linked Data.*

Słowa kluczowe: *kartoteki haseł wzorcowych, Linked Data, sieć semantyczna*

Wprowadzenie

Od kilku lat w środowisku bibliotekarzy mówi się, że sieć semantyczna czai się tuż za rogiem i powinniśmy być gotowi na wdrożenie jej technologii, jeśli chcemy zachować pozycję liczącego się dostawcy usług informacyjnych. Rzeczywistość wygląda jednak tak, że do końca nie wykorzystaliśmy potencjału technologii sieci drugiej generacji, lub jej adaptacja nie przebiega równomiernie¹. We wrześniu tego roku opiniotwórcze konsorcjum New Media Consortium opublikowało tegoroczne wydanie raportu z serii *NMC Horizon Report* poświęcone zastosowaniu technologii informacyjnych w bibliotekach naukowych². Wskazano tam m.in. na sześć obszarów adaptacji nowych technologii w bibliotekach naukowych:

1. Zwiększenie dostępności do treści naukowych.
2. Rewizja pojęcia przestrzeni bibliotecznej.
3. Rewizja pojęcia publikacji naukowej jako „nośnika” treści naukowych w obiegu informacji w nauce.
4. Zwrócenie uwagi na zarządzanie danymi badawczymi jako nowym zakresem kompetencji bibliotekarzy.
5. Zwiększenie interaktywności cyfrowych usług informacyjnych świadczonych przez biblioteki na podstawie badań w zakresie tzw. doświadczeń użytkownika (ang. User Experience).
6. Nadanie wysokiego priorytetu mobilności w odniesieniu do usług informacyjnych.

1BYRNE, G., GODDARD, L. The Strongest Link: Libraries and Linked Data. *D-Lib Magazine* [on-line] 2010, Vol. 16, no. 11/12 [dostęp 20.09.2015]. Dostępny w: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>.

2JOHNSON, L., ADAMS BECKER, S., ESTRADA, V., FREEMAN, A. *NMC Horizon Report: 2015 Library Edition* [on-line] Austin, Texas: The New Media Consortium, 2015 [dostęp 20.09.2015]. ISBN 978-0-9962832-4-3. Dostępny w: <http://cdn.nmc.org/media/2015-nmc-horizon-report-library-EN.pdf>.

W obszarach pierwszym i trzecim istotną rolę mają odegrać technologie sieci semantycznej oraz publikowanie danych w modelu Linked Data, dla których czas szacunkowy adaptacji w bibliotekach określono na dwa do trzech lat, czyli do 2018 r. Stwierdzono również, że biblioteki naukowe znajdują się w wyjątkowo korzystnym położeniu w kontekście realnych zysków płynących z zastosowania technologii semantycznych do publikowania i udostępniania swoich danych zarówno bibliograficznych, jak i wzorcowych³. Istotą jest tutaj stosowanie standardów sieciowych oraz myślenie w kategoriach interoperacyjności z istniejącymi dostawcami treści, zarówno profesjonalnymi, jak i społecznościowymi (np. Wikipedia). Dobitnie stanowisko to zmanifestowali Mirna Willer i Gordon Dunsire w książce pt. *Bibliographic Information Organization in the Semantic Web*⁴:

Nowe technologie zawsze będą podstawą przyszłych generacji katalogów bibliotecznych. Biblioteki chcąc utrzymać pozycję w dostarczaniu informacji, obsłudze użytkowników oraz w realizacji misji instytucji dziedzictwa kulturowego, muszą dokonać skoku technologicznego – standardy bibliograficzne, modele danych i usługi informacyjne muszą być zgodne ze standardami sieci semantycznej.

Istotną rolę w rozwoju sieci semantycznej odgrywają standardy (np. HTTP, RDF, URI) oraz koncepcja zbiorów słownictwa (ang. *vocabularies*). Poprzez termin ten rozumie się tutaj zarówno schematy metadanych, jak i zbiory ich potencjalnych wartości, czyli to, co w nomenklaturze biblioteczno-bibliograficznej określamy mianem kartotek wzorcowych. Główną tezą tego tekstu jest stwierdzenie, że to, co biblioteki mogą dać sieci semantycznej i co może stanowić jej siłę napędową, to kartoteki haseł wzorcowych, z zastrzeżeniem, że sposób ich prezentacji będzie zgodny ze standardami sieciowymi. Podstawowe warunki, jakie muszą zostać spełnione, to **opis zawartości KHW z wykorzystaniem języka RDF – lingua franca Semantic Web, oraz stosowanie identyfikatorów sieciowych.**

Celem artykułu jest przedstawienie potencjału kartotek haseł wzorcowych w kontekście rozwoju sieci semantycznej. Treści związane z tym przedmiotem zostaną omówione przede wszystkim na przykładzie kartotek haseł wzorcowych dla nazw osobowych.

Funkcje kartotek wzorcowych

Truizmem jest stwierdzenie, że rola kartotek wzorcowych jest podporządkowana celom systemów informacyjnych, w ramach których te z założenia mają funkcjonować. Podstawowe funkcje katalogu bibliotecznego zostały zidentyfikowane przez Charlesa Ammi Cuttera w ważnym dla współczesnego bibliotekarstwa i bibliografii dziele pt. *Rules for a Printed Dictionary Catalog* z 1876 r. Cutter stwierdza m.in., że katalog biblioteczny powinien umożliwić użytkownikowi odnalezienie książki według nazwiska jej autora, formy tytułu czy też na określony temat. Wskazuje również, że kontrola form językowych dla

³ Tamże.

⁴ WILLER, M., DUNSIRE, G. *Bibliographic Information Organization in the Semantic Web* 2013, s. XXVII. ISBN 978-1-84334-731-6.

stosowanych nazw jest jednym ze środków realizacji funkcji katalogu⁵. Zmieniło się techniczne otoczenie systemów bibliotecznych, zaadaptowaliśmy nowe technologie informacyjne w katalogach bibliotecznych, ale wydaje się, że funkcje wskazane przez Cuttera nie zdezaktualizowały się. Wraz z opublikowaniem w 1998 r. *Functional Requirements for Bibliographic Records* (FRBR) – modelu funkcjonalnych wymagań dla rekordów bibliograficznych – zmieniła się perspektywa jednostki opisu (dzieło-realizacja-materializacja-egzemplarz) oraz zmodyfikowano terminologię (dokument – zasób informacyjny, hasło opisu bibliograficznego – punkt dostępu). W dokumentacji tego modelu stwierdza się, że zarówno zasady katalogowania, jak i funkcjonalności systemu informacyjnego powinny umożliwić użytkownikowi wykorzystanie danych:

- do znalezienia materiałów, które odpowiadają kryteriom postawionym w zapytaniu przez użytkownika (np. w kontekście wyszukiwania wszystkich dokumentów na dany temat lub poszukiwania nagrania według określonego tytułu);
- do identyfikacji jednostki (np. potwierdzenie, że dokument opisany w rekordzie odpowiada dokumentowi odnalezionemu przez użytkownika lub odróżnienie dwóch tekstów lub nagrań o tym samym tytule);
- do wyboru jednostki, która spełnia oczekiwania użytkownika (np. wybór tekstu w języku zrozumiałym dla użytkownika lub wybór wersji oprogramowania, która jest kompatybilna ze sprzętem i systemem operacyjnym dostępnym dla użytkownika);
- w celu nabycia lub uzyskania dostępu do opisywanej jednostki (np. złożenie zamówienia na jednostkę, złożenie zamówienia w celu wypożyczenia egzemplarza książki z kolekcji bibliotecznej lub uzyskanie dostępu do czasopisma elektronicznego przechowywanego na zdalnym komputerze)⁶.

W tym kontekście rola kartotek wzorcowych polega na zapewnianiu spójności danych bibliograficznych poprzez kontrolę językową stosowanych form (m.in. nazw własnych, geograficznych, elementów opisu rzeczowego), usprawnienie procesu katalogowania i ułatwienie zarządzania danymi bibliograficznymi⁷. W założeniu kartoteki wzorcowe haseł formalnych miały zapewniać możliwość rozróżniania osób, ciał zbiorowych itd. poprzez przyjęcie jednej formy dla nazwy preferowanej, wskazanie na jej warianty językowe oraz zapewnienie kontekstu poprzez system odsyłaczy. Ich charakter polegał więc na identyfikacji osób, instytucji itd. poprzez stosowanie spójnego systemu nazw preferowanych, podpartego zasadami ich doboru. Tak jak rekord bibliograficzny, również rekord wzorcowy doczekał się swojej formalnej reprezentacji w MARC 21. Analizując zmiany wprowadzane w strukturze rekordu wzorcowego MARC 21, można dojść do konkluzji, że obecna funkcja identyfikacji KHW zostaje rozszerzona i kartoteki będą mieć teraz również charakter informacyjny, deskryptywny. Tabela 1. przedstawia nowe pola wprowadzone w strukturze rekordu wzorcowego w MARC 21 odnoszące się do nazw osobowych.

⁵TILLET, B. Authority Control: State of the Art and New Perspectives. *Cataloging and Classification Quarterly* 2004, Vol. 38, no. 3–4, s. 23.

⁶IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. *Functional requirements for bibliographic records: final report*. München: K.G. Saur, 1998. ISBN 978-3-598-11382-6.

⁷TILLET, B. dz. cyt., s. 26.

Tab. 1. Nowe pola wprowadzone w strukturze rekordu wzorcowego w MARC 21 odnoszące się do nazw osobowych

Numer pola	Nazwa	Charakterystyka	Data wprowadzenia
046	Daty specjalnie kodowane	Np. oznaczenie dat związanych z opisywanymi osobami (np. data urodzin, śmierci, daty aktywności)	2010
368	Pozostałe atrybuty związane z osobą i ciałem zbiorowym	Np. oznaczenie tytułów związanych z osobą	2011
371	Adres	Np. adres zamieszkania, pobytu osoby, państwo, kod pocztowy	2009
372	Dziedzina działalności	Np. obszar aktywności, nazwa specjalizacji danej osoby	2009
373	Powiązana grupa	Informacja na temat grupy, instytucji, stowarzyszenia, itp. powiązanej z jednostką wskazaną w polu 1XX. Dla osoby może to być np. afiliacja przy jednostce naukowej	2009/2011 (zmiana nazwy)
374	Zawód	Informacja o wykonywanym zawodzie oraz możliwość wskazania dat jego wykonywania	2009
375	Płeć	Informacja na temat płci osoby, wraz z datami (jeżeli jest to uzasadnione)	2009
376	Informacje o rodzinie	Dodatkowe informacje na temat rodziny	2009
378	Pełniejsza forma nazwy osobowej	Np. stosowane do prezentacji pełnej nazwy osobowej w przypadku wystąpienia inicjałów w nazwie preferowanej	2011

Źródło: KRYNICKA, M. *Format MARC 21 – stan prac i perspektywy rozwoju* [on-line] 2011 [dostęp 20.09.2015]. Dostępny w: <http://www.bn.org.pl/download/document/1321438599.ppt>.

Zmiany w strukturze rekordu wzorcowego w MARC 21 zostały wprowadzone z uwagi na nowe angloamerykańskie zasady katalogowania – *Resource Description and Access* (RDA), które znacznie rozszerzyły zakres odwzorowywanych informacji dla haseł formalnych. Mamy więc do czynienia z sytuacją, w której kartoteka wzorcowa haseł formalnych oprócz podstawowej funkcji rozróżniania ma pełnić również funkcję źródła informacji na temat podstawowych danych biograficznych danej osoby. Oczywiście zwiększa się potencjalna wartość informacyjna KHW, ale nie należy zapominać o kosztach związanych z wprowadzaniem nowych informacji, na które składać się będą m.in.: dłuższy czas opracowania rekordu wzorcowego, konieczność wskazania dopuszczalnych źródeł przejmowania danych (kwestia wiarygodności i oceny jakości pozyskanych informacji), konieczność rozszerzenia pragmatyki o nowe elementy metadanych.

Podstawę do zmian w odniesieniu do haseł osobowych w RDA i MARC 21 dał opublikowany w 2009 r. *Functional Requirements for Authority Data* (FRAD) – model funkcjonalnych wymogów dla haseł wzorcowych. Tak jak FRBR przedstawia model uniwersum bibliograficznego, tak FRAD zawiera konceptualizację kartoteki haseł

wzorcowych formalnych za pomocą modelu związków encji. Według ogólnych założeń FRAD **jednostka bibliograficzna** jest opisywana przez **nazwy i/lub identyfikatory**, których podstawą ekspresji są kontrolowane **punkty dostępu**⁸. Według specyfikacji FRAD w kartotekach haseł wzorcowych dla nazw osobowych dopuszcza się odwzorowanie następujących atrybutów:

- daty związane z osobą,
- tytuły osób,
- płeć,
- miejsce urodzenia,
- miejsce śmierci,
- narodowość,
- miejsce pobytu,
- afiliacja,
- adres,
- język, którym posługiwała się osoba w zasadniczym obszarze aktywności,
- dziedzina aktywności,
- zawód,
- nota biograficzna,
- pozostałe informacje związane z osobą (np. tytuły, relacje z innymi osobami).

Jak widać z powyższych analiz zarówno na poziomie koncepcyjnym (FRAD), pragmatycznym (RDA), jak i formalnym (MARC 21) kartoteki haseł wzorcowych dla nazw osobowych mają w założeniu pełnić funkcję źródła informacji o osobie, a nie tylko o nazwach przyjętych w działalności biblioteczej i bibliograficznej do jej identyfikacji. Jest to dosyć istotna zmiana z punktu widzenia zasadniczych funkcji KHW.

Model Linked Data

Linked Data jest to model publikowania danych w środowisku sieciowym, w którym wykorzystuje się określone standardy sieciowe i którego podstawą jest ustanawianie relacji między opisywanymi dokumentami, osobami, pojęciami, wydarzeniami itd. w celu optymalizacji procesów wyszukiwania i automatycznej eksploracji informacji. Sieć semantyczna jako taka jest jedynie wizją, natomiast Linked Data uważa się za jedną z form jej realizacji. Podstawy Linked Data przedstawił Tim Berners-Lee⁹ i można je zamknąć w czterech punktach:

1. Używaj URI (Uniform Resource Identifier) jako nazw i sposobu odróżniania obiektów.
2. Stosuj protokół HTTP, aby uzyskać informacje o opisywanych obiektach.

⁸ PATTON, G.E. (red.). *Functional Requirements for Authority Data – A Conceptual Model* [on-line] München: K.G. Saur, 2009. ISBN 978-3-598-24282-3. [dostęp 20.09.2015]. Dostępny w: <http://www.ifla.org/publications/functional-requirements-for-authority-data>.

⁹ BERNERS-LEE, T. *Linked Data – Design Issues* [on-line] 2006 [dostęp 20.09.2015]. Dostępny w: <http://www.w3.org/DesignIssues/LinkedData.html>.

3. Udostępniaj użyteczne informacje o obiekcie identyfikowanym przez jego URI, za pomocą standardów RDF/XML.
4. Wykorzystuj powiązania z innymi obiektami za pomocą URI, aby zapewnić możliwość eksploracji i odkrywania informacji o innych obiektach.

Podstawowym założeniem w tym modelu jest odwoływanie się do opisywanych obiektów za pomocą ich unikalnych identyfikatorów wyrażonych w standardzie URI, które są transferowalne sieciowo oraz stanowią element istniejącego systemu referencyjnego (tzw. dereferenceable Uniform Resource Identifier). Ostatni warunek zapewnia uzyskanie opisu danego obiektu poprzez wywołanie jego URI w ramach usługi HTTP. *Lingua franca* Linked Data to język reprezentacji wiedzy – RDF (Resource Description and Framework), w którym stosuje się trójelementowe deklaracje na temat opisywanego tzw. zasobu informacyjnego oraz relacji, w jakie wchodzi z innymi zasobami. Tzw. trójki RDF zbudowane są z podmiotu (przedmiotu opisu), predykatu (opisywana własność, rodzaj relacji) oraz obiektu (wartość własności, wskazanie na podmiot stanowiący człon relacji), np.

Osoba X ? nazwa ◇ *Douglas Adams*

Kluczowym elementem w takiej ekspresji jest zastąpienie każdego z elementów trójki odpowiednim identyfikatorem, który spełniałby przedstawione wyżej warunki. Nie wszystkie elementy opisu można w ten sposób wyrazić (np. formy tytułów dokumentów), ale im bardziej będziemy unikać tzw. literałów (ciągi znaków), tym większa możliwość automatycznego przetwarzania takiego opisu, np.

http://przykladowe.uri/Douglas_Adams ◇ http://xmlns.com/foaf/spec/#term_name ◇
Douglas Adams

W powyższym przykładzie do wyrażenia informacji zastosowano przykładowy identyfikator URI dla opisywanego pisarza, a element metadanych „nazwa” wyrażono za pomocą wykładnika metadanych ze schematu metadanych FOAF, który jest prostą specyfikacją danych osobowych. Rekonstrukcja rekordu bibliograficznego czy wzorcowego za pomocą RDF prowadzi do jego wyrażenia za pomocą możliwych formalnych ekspresji, tzw. serializacji RDF. Może to być serializacja za pomocą języka XML czy np. tzw. składni N3 (tab. 2.). Istotne jest, że prezentacje te są zgodne ze standardami sieciowymi.

Tab. 2. Przykłady serializacji trójki RDF

Serializacja RDF/XML	Serializacja N3
<pre><?xml version="1.0"?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:foaf="http://xmlns.com/foaf/0.1/" <rdf:Description</pre>	<pre>@prefix foaf: < http://xmlns.com/foaf/0.1/> . @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> . @prefix xml: <http://www.w3.org/XML/1998/namespace> . @prefix xsd: <http://www.w3.org/2001/XMLSchema#> . <http://przykladowe.uri/Douglas_Adams > foaf:name "Douglas Adams" .</pre>

<pre>rdf:about=" http://przykladowe.uri/Douglas_Adams "> <foaf:name>Douglas Adams</ foaf:name > </rdf:Description> </rdf:RDF></pre>	
--	--

Sam język RDF oferuje niewiele wykładników metadanych, a jedynie pewną strukturyzację wypowiedzi. Każdy z elementów reprezentowanych za pomocą URI jest transferowalny sieciowo, tzn. istnieje możliwość „odpytania” danej usługi sieciowej, w ramach której on funkcjonuje i pobrania dodatkowych informacji. Idea danych powiązanych, to ustalanie powiązań z zewnętrznymi zasobami sieciowymi wyrażonymi zgodnie z modelem Linked Data. W tym przypadku można wyrazić informacje o miejscu urodzenia Douglasa Adamsa (Cambridge, Anglia) nie tyle poprzez ciąg znaków reprezentujących tę nazwę miejscową, ile poprzez URI do tego zasobu np. z bazy Geonames.org¹⁰. Można więc zastosować wykładnik metadanych dla „miejsca urodzenia” z odpowiedniego schematu, a jego wartość reprezentować za pomocą URI – <http://sws.geonames.org/2653941/>. Można również ustalić relację tożsamości między rekordem dla Douglasa Adamsa w naszej bazie a rekordem go opisującym w zewnętrznym źródle. Wykorzystuje się do tego wykładnik owl:sameAs, który jest elementem języka opisu ontologii – OWL (Web Ontology Language), np. z bazą VIAF – <http://viaf.org/viaf/113230702> czy DBpedią (semantyczną wersją Wikipedii) – http://dbpedia.org/resource/Douglas_Adams. Tworzymy wówczas formalnie specyfikowaną sieć relacji, która pozwala na eksplorację danych.

Takie podejście do publikowania danych bibliograficznych w środowisku sieciowym określa się mianem Library Linked Data, a biblioteki od kilku lat coraz chętniej włączają się do procesu współtworzenia zasobów sieci semantycznej. Warto odnotować kilka inicjatyw, szczególnie bibliotek narodowych, zmierzających do udostępniania danych bibliograficznych i wzorcowych w standardach sieciowych:

- Biblioteka Narodowa Francji: projekt <http://data.bnf.fr/>,
- Biblioteka Narodowa Hiszpanii: projekt <http://datos.bne.es/>,
- Biblioteka Narodowa Niemiec: <http://www.dnb.de/EN/lds>,
- Bibliografia narodowa Wielkiej Brytanii: <http://bnb.data.bl.uk/>.

Ostatnie statystyki związane z wielkością chmury danych powiązanych pochodzą z sierpnia 2014 r¹¹. Zidentyfikowano wówczas 1014 zbiorów danych, z czego największy odsetek stanowią tzw. dane rządowe i dane administracji publicznej (18,5%). Na drugim miejscu pod względem wielkości jest kategoria tzw. publikacji, która obejmuje również dane bibliograficzne i wzorcowe udostępniane przez biblioteki. W badaniu uwzględniono 96 takich zbiorów, co stanowi 9,47% całości.

10 Baza danych geograficznych zawierająca informacje o obiektach geograficznych, dane demograficzne oraz identyfikację za pomocą współrzędnych geograficznych. Geonames funkcjonuje jako serwis internetowy i usługa sieciowa umożliwiająca pobieranie danych (np. poprzez API oraz pobranie plików).

11SCHMACHTENBERG, M., BIZER, Ch., PAULHEIM, H. State of the LOD Cloud 2014 [on-line] 2014 [dostęp 20.09.2015]. Dostępny w: <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/#toc4>.

Linked Open Vocabularies

Projekt Linked Open Vocabularies (LOV – <http://lov.okfn.org/dataset/lov/>)¹² powstał w 2011 r. jako odpowiedź na problem dostępności do istniejących schematów metadanych i ontologii, które mogłyby zostać wykorzystane do publikowania danych również w modelu Linked Data. Celem projektu LOV jest gromadzenie i udostępnianie informacji na temat istniejących schematów metadanych. Tym samym narzędzie to ułatwia prace nad wyborem odpowiedniego schematu, który może mieć zastosowanie do reprezentacji całej kolekcji lub tylko wybranej grupy atrybutów. Obecnie w bazie zarejestrowano 520 schematów metadanych. Co ciekawe LOV rejestruje również poszczególne wersje schematów. Dzięki tej usłudze sieciowej mamy możliwość wyszukania poszczególnego atrybutu (np. tytuł – rys. 1.) i uzyskania informacji, w którym schemacie metadanych został on zarejestrowany, eksplorowania schematów metadanych według dziedzin wiedzy oraz wyszukiwania według języka, w którym wyrażono odpowiedniki słowne atrybutów.

The screenshot shows the LOV search interface. At the top, there are navigation tabs: VOCABS, TERMS, AGENTS, and SPARQL/DUMP. Below the search bar, the term 'title' is entered. The results are displayed in a table with three entries:

Results	Term	Score
793 results	dcterms:title (dcterms) 13,420,023 occurrences in 132 LOD datasets http://purl.org/dc/terms/title rdfs:label Title @en localName title	0.745
	dce:title (dce) 7,626,720 occurrences in 107 LOD datasets http://purl.org/dc/elements/1.1/title rdfs:label Title @en localName title	0.679
	npg:title (npg) n/a (use in LOD) http://ns.nature.com/terms/title skos:definition The title property specifies a title to the main @en skos:prefLabel Title @en localName title	0.556

Rys. 1. Fragment rezultatów wyszukiwania w bazie LOV dla atrybutu tytuł (ang. title)
Źródło: *Linked Open Vocabularies* [on-line] [dostęp 20.09.2015]. Dostępny w:
<http://lov.okfn.org/dataset/lov/terms?q=title>.

Dostęp do danych

Dane opublikowane w standardzie RDF są przechowywane w charakterystycznych dla modelu Linked Data systemach informacyjnych nazywanych *triplestore*. Tego rodzaju aplikacje pozwalają na zarządzanie dużymi zbiorami danych w RDF. Przykładowe rozwiązania programistyczne w tym zakresie to: AllegroGraph

¹²Wszystkie odesłania do stron internetowych przedstawiają wersję aktualną w dn. 19.10.2015 r.

(<http://franz.com/agraph/allegrograph/>), Dydra (<https://dydra.com/>), Jena (<http://jena.apache.org/>) czy też Sesame (<http://rdf4j.org/>). Bezpośredni dostęp do danych zapewnia usługa nazywana SPARQL Endpoint. Jest to element architektury systemu odpowiedzialny za wyszukiwanie i aktualizację danych przechowywanych w bazie. Językiem zapytań oficjalnie uznanym przez Konsorcjum World Wide Web za standard dla sieci semantycznej jest język SPARQL (SPARQL Protocol and RDF Query Language). Jest to stosunkowo prosty język zapytań, ale dający możliwość pełnego wykorzystania modelowania danych za pomocą RDF. Składnia SPARQL wykorzystuje trójelementowe podejście do formułowania deklaracji obecne w RDF. Przykładowy interfejs SPARQL można znaleźć np. dla bibliografii narodowej British Museum – <http://bnb.data.bl.uk/flint-sparql> lub dla DBpedii – <http://dbpedia.org/sparql>. Inne formy udostępniania danych to publikowanie na serwerach plików zawierających opis kolekcji za pomocą RDF oraz oferowanie przekierowania do „widoku” danych w RDF z interfejsu graficznego przeznaczonego dla użytkownika.

Tabela 3. zawiera wyniki badań przeprowadzone przez zespół pod kierownictwem Ioannisa Papadakis¹³ dotyczące sposobów udostępniania danych przez biblioteki realizujące projekty wpisujące się w koncepcję Linked Data.

Tab. 3. Sposoby udostępniania danych w modelu Linked Data przez wybrane biblioteki

Biblioteka	SPARQL Endpoint	Pliki	Przekierowanie
British Library (Bibliografia narodowa)	x		
Biblioteka Narodowa Francji (Rameau)		x	x
Biblioteka Narodowa Niemiec (GND)		x	x
Biblioteka Narodowa Hiszpanii	x		
Biblioteka Królewska Szwecji (LIBRIS)	x		
Biblioteka Kongresu (LCSH)		x	x
Biblioteka Narodowa Węgier	x		

Źródło: PAPANAKIS, I., KYPRIANOS, K., STEFANIDAKIS, M. Linked Data URIs and Libraries: The Story So Far. *D-Lib Magazine* [on-line] maj 2015, Vol. 21, no. 5/6 [dostęp 20.09.2015].
Dostępny w: <http://www.dlib.org/dlib/may15/papadakis/05papadakis.html>.

KHW w środowisku sieciowym

O ile na poziomie schematów metadanych środowisko bibliotek musi zaakceptować nieprzystawalność MARC 21 do współczesnego sposobu myślenia o reprezentacji informacji w środowisku sieciowym oraz do jego wymagań formalnych, o tyle w przypadku

¹³ PAPANAKIS, I., KYPRIANOS, K., STEFANIDAKIS, M. Linked Data URIs and Libraries: The Story So Far. *D-Lib Magazine* [on-line] 2015, Vol. 21, no. 5/6 [dostęp 20.09.2015]. Dostępny w: <http://www.dlib.org/dlib/may15/papadakis/05papadakis.html>.

kartotek haseł wzorcowych jest duża szansa na wypełnienie niszy w szeroko rozumianej sieci semantycznej. Kluczem jest tutaj odpowiednia warstwa prezentacji (RDF) oraz co równie istotne – stosowanie systemu stałych identyfikatorów (URI) zarówno do opisu zasobów informacyjnych, jak i jednostek kartotek haseł wzorcowych¹⁴.

VIAF

Sztandarowym przykładem globalnego podejścia do koncepcji sieciowej kartoteki haseł wzorcowych jest VIAF (Virtual International Authority File – <https://viaf.org/>) – wirtualna internetowa kartoteka haseł wzorcowych. Zawartość VIAF współtworzy 35 bibliotek z 30 państw, w tym z Polski – Centrum NUKAT oraz Biblioteka Narodowa. VIAF z założenia został zaprojektowany jako narzędzie bibliotek do współtworzenia zawartości sieci semantycznej. Na koniec 2014 r. VIAF rejestrował ponad 45 mln haseł wzorcowych oraz ponad 105 mln rekordów bibliograficznych. W przypadku haseł wzorcowych ponad 77% stanowiły hasła osobowe. Dostęp do danych VIAF realizowany jest za pomocą interfejsu graficznego, ale również za pośrednictwem API (<https://platform.worldcat.org/api-explorer/VIAF>) oraz poprzez możliwość pobierania plików z danymi. Na chwilę obecną VIAF nie udostępnia SPARQL Endpoint. Każdy rekord posiada dodatkowo możliwość eksportu w RDF.

Przykładowy rekord wzorcowy dla Douglasa Adamsa zawiera również odesłania do zewnętrznych źródeł ze wskazaniem na relację tożsamości z osobą tam opisywaną. Są to:

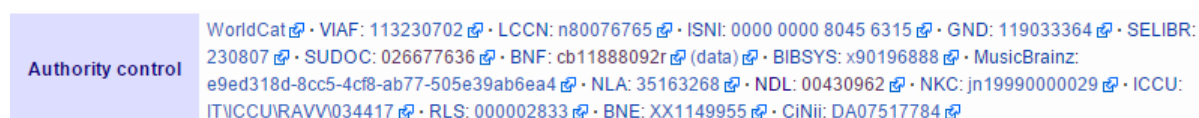
- Projekt data.bnf.fr – <http://data.bnf.fr/ark:/12148/cb11888092r#foaf:Person>,
- DBpedia – http://dbpedia.org/resource/Douglas_Adams,
- Kartoteka wzorcowa Biblioteki Kongresu – <http://id.loc.gov/authorities/names/n80076765>,
- Kartoteka wzorcowa Biblioteki Narodowej Japonii – <http://id.ndl.go.jp/auth/entity/00430962>,
- Bazy ISNI – międzynarodowego znormalizowanego identyfikatora nazwy <http://isni.org/isni/0000000080456315>,
- Kartoteka wzorcowa w katalogu LIBRIS – <http://libris.kb.se/resource/auth/230807>,
- Kartoteka wzorcowa w Sudoc <http://www.idref.fr/026677636/id>,
- Bazy Wikidata – <http://www.wikidata.org/entity/Q42>.

Wikipedia

W rekordzie dla Douglasa Adamsa w bazie data.bnf.fr znajdziemy również odesłania do noty biograficznej tego pisarza w Wikipedii. Co ciekawe również ze strony Wikipedii widoczne jest zainteresowanie ustanawianiem relacji z rekordami wzorcowymi z baz bibliotecznych. Powołano inicjatywę kontroli wzorcowej Wikipedii

¹⁴ BIZER, Ch., HEATH, T. *Linked Data: Evolving the Web into a Global Data Space* [on-line] Morgan & Claypool Publishers, 2011 [dostęp 20.09.2015]. ISBN 978-1-60845430-3. Dostępny w: <http://linkeddatabook.com/editions/1.0/>.

(https://en.wikipedia.org/wiki/Wikipedia:Authority_control_integration_proposal/RFC) i na podstawie wniosków zdecydowano się tutaj na osobny szablon do wprowadzania danych za pośrednictwem identyfikatorów sieciowych do kartotek wzorcowych i innych źródeł o charakterze normalizacyjnym. Opracowano i wdrożono aplikację VIAFbot, której zadaniem było automatyczne łączenie haseł Wikipedii odpowiednimi zapisami we VIAF i ustalanie dwukierunkowych powiązań. Dzięki tej aplikacji uzyskano ponad 250 tys. takich zapisów w Wikipedii. W przypadku hasła Douglas Adams (https://en.wikipedia.org/wiki/Douglas_Adams) w anglojęzycznej Wikipedii znajdziemy 17 identyfikatorów sieciowych (rys. 2.) m.in. do bazy WorldCat, BnF, Sudoc, VIAF czy ISNI.



Rys. 2. Dodatkowe informacje o identyfikatorach w strukturze hasła w Wikipedii
Źródło: Douglas Adams. *Wikipedia: The Free Encyclopedia*. [on-line], [dostęp 20.09.2015]. Dostępny z: https://en.wikipedia.org/wiki/Douglas_Adams

DBpedia

DBpedia (<http://wiki.dbpedia.org>) to projekt, którego celem jest automatyczna ekstrakcja danych z Wikipedii i ich konwersja do modelu Linked Data. Obecna wersja DBpedii zawiera bazę wiedzy, którą tworzy ponad 38 mln haseł w 125 językach. Podstawą konwersji jest anglojęzyczna Wikipedia. Do bazy DBpedii trafiło 144 500 rekordów dla nazw osobowych. Polska wersja DBpedii (<http://pl.dbpedia.org/mediawiki>) jest rozwijana przez Katedrę Informatyki Ekonomicznej Uniwersytetu Ekonomicznego w Poznaniu. Dostęp do danych DBpedii jest realizowany przez SPARQL Endpoint, np. <http://dbpedia.org/sparql>, <http://dbpedia.org/snorql/>. W rekordach dla nazw osobowych znajdziemy odwołania przede wszystkim do VIAF. Obecnie w DBpedii znajduje się 268 429 identyfikatorów VIAF¹⁵ na 2 548 134 haseł osobowych (<http://dbpedia.org/services-resources/datasets/dataset-statistics>), co stanowi tylko ok. 10%. W ontologii DBpedii wprowadzono nawet odrębną własność – posiadanie identyfikatora VIAF: <http://dbpedia.org/ontology/viafId>. W DBpedii w hasłach osobowych obecne są również informacje o lokalnych identyfikatorach Biblioteki Kongresu – LCCN. Tutaj również atrybut ten wprowadzono do ontologii w postaci dwóch wykładników <http://dbpedia.org/ontology/lccnId> oraz <http://dbpedia.org/property/lccn>. W pierwszym przypadku występuje on w 22 158 hasłach osobowych w DBpedii, w drugim – 22 338 razy.

Deutsche Biographie

Deutsche Biographie (<http://www.deutsche-biographie.de/>) jest serwisem internetowym zawierającym elektroniczną wersję słownika biograficznego Neue Deutsche Biographie. Są tam dostępne biogramy osób nieżyjących związanych z Niemcami. Baza zawiera ponad 21 800 biogramów i jest rozbudowywana. Baza dostępna jest również w formacie

¹⁵ Dane uzyskano na podstawie zapytania SPARQL poprzez <http://dbpedia.org/snorql/> dotyczącego liczby wszystkich unikalnych wystąpień atrybutu <http://dbpedia.org/ontology/viafId> w DBpedii w dniu 26.09.2015 r.

RDF, ale nie oferuje API oraz aktywnego SPARQL Endpoint. Z punktu widzenia relacji między bazą Deutsche Biographie a KHW widoczne są odwołania do Gemeinsame Normdatei (GND) – zintegrowanej kartoteki haseł wzorcowych Niemieckiej Biblioteki Narodowej oraz VIAF. W rekordzie w formacie RDF dla Friedricha Schillera (<http://www.deutsche-biographie.de/sfz45355.html>) znajdziemy tutaj odwołania do identyfikatora tej osoby w GND oraz szereg metadanych wyrażonych za pomocą ontologii zaprojektowanej na potrzeby GND. W Deutsche Biographie stosuje się własny system stałych identyfikatorów sieciowych (URI), a odesłania do GND i VIAF mają charakter odsyłaczy typu zobacz też, które są specyfikowane w RDF Schema (rdfs:seeAlso). Mamy tutaj więc odsyłacz do rekordu dla Friedricha Schillera o identyfikatorze: <http://data.deutsche-biographie.de/Person/sfz45355> do rekordu wzorcowego w GND o identyfikatorze <http://d-nb.info/gnd/118607626>.

Obszary do eksploracji

Prace w ramach VIAF rozpoczęto od haseł wzorcowych dla nazw osobowych – obecnie jest to kluczowa inicjatywa naszego środowiska wpisująca się we współtworzenie zasobów sieci semantycznej za pomocą kartotek wzorcowych. Pozostałe rodzaje haseł wzorcowych (geograficzne, korporatywne) stanowią nadal mały odsetek kolekcji VIAF, ale z pewnością i ten obszar będzie rozwijany. Konwersja do standardów sieciowych i udostępnienie na odpowiednich licencjach lokalnych kartotek haseł wzorcowych oraz słowników stosowanych języków informacyjno-wyszukiwawczych to szansa nie tyle na zaistnienie w sieci, ile na stanie się jej częścią. To także możliwość wykorzystania naszych danych przez podmioty zewnętrzne spoza kontekstu katalogu bibliotecznego.

Z formalnego punktu widzenia zarówno VIAF, GND, jak i KHW LoC stosują odmienne schematy metadanych w ramach RDF do opisu haseł wzorcowych. Istnieje co prawda standard MADS (Metadata Authority Description Schema), który jest specyfikacją XML dla rekordu wzorcowego w MARC 21, ale standard ten nie zwrócił uwagi ani środowiska bibliotecznego, ani tym bardziej podmiotów spoza niego. Gotowe propozycje wkrótce otrzymamy ze strony RDA oraz projektu BIBFRAME. W pierwszym przypadku cały czas trwają prace nad przygotowaniem schematu metadanych, który reprezentowałby punkt widzenia nowych angloamerykańskich zasad katalogowania. Będzie to specyfikacja RDA jako schematu metadanych w RDF/XML. Wyniki prac można śledzić w serwisie Open Metadata Registry (<http://rdvocab.info/>). BIBFRAME (<http://bibframe.org/>) to inicjatywa Biblioteki Kongresu, której celem jest opracowanie alternatywy wobec MARC 21, czyli formatu danych bibliograficznych i wzorcowych przystosowanego do funkcjonowania w ramach modelu Linked Data. Tutaj również nie ma jeszcze gotowego schematu, który mógłby zostać wykorzystany na potrzeby publikowania kartotek wzorcowych w modelu LD. W przypadku słowników języków informacyjno-wyszukiwawczych w 2009 r. Konsorcjum World Wide Web opublikowało standard SKOS (Simple Knowledge Organization System), który uznało za oficjalną rekomendację w zakresie opisu struktury systemów organizacji wiedzy. SKOS zawiera formalną specyfikację pojęć ich własności oraz typologię zachodzących między nimi relacji, które są obecne w wielu rodzajach języków informacyjno-wyszukiwawczych.

Nie ma jako takiej oficjalnej metodyki publikowania danych bibliograficznych i wzorcowych w modelu Linked Data. Natomiast w dokumentacji projektów bibliotecznych oraz z wystąpień konferencyjnych je omawiających wynika, że najlepiej zaczynać tego typu projekty właśnie od kartotek haseł wzorcowych. Zdaniem Larsa G. Svenssona, doradcy ds. usług sieciowych w Bibliotece Narodowej Niemiec, właśnie informacje na temat osób, miejsc obecne w KHW wydają się mieć najszerszy zakres zastosowania¹⁶.

Udostępnianie KHW jako usług sieciowych, to szansa na zainteresowanie naszymi danymi podmiotów zewnętrznych. Z pewnością warte uwagi są m.in. obszary:

- społecznościowego internetu, takie jak serwisy dedykowane czytelnikom, np. polskie Lubimy Czytać (<http://lubimyczytac.pl/>) i światowe Good Reads (<http://www.goodreads.com/>), Library Thing (<https://www.librarything.com/>) oraz Open Library (<https://openlibrary.org/>);
- biograficznych źródeł informacji w internecie, np. Internetowy Polski Słownik Biograficzny (<http://www.ipsb.nina.gov.pl/>);
- wydawców i księgarzy, np. w ramach wymiany informacji z wykorzystaniem specyfikacji ONIX;
- dziedzinowych źródeł informacji w internecie, np. możliwość wykorzystania biblioteczno-bibliograficznych języków informacyjno-wyszukiwawczych do opisu treści.

Podsumowanie

Eric Miller, współautor specyfikacji RDF oraz kierujący projektem BIBFRAME, wskazał na cztery główne zadania bibliotek w kontekście obecności w sieci semantycznej¹⁷:

1. Udostępnianie kolekcji z wykorzystaniem standardów sieciowych.
2. Udostępnianie kartotek haseł wzorcowych z wykorzystaniem standardów sieciowych. Barbara Tillet stwierdziła, że konwersja tych źródeł informacji do postaci „semantycznej” da nieograniczony potencjał na ich wykorzystanie w wielu kontekstach.
3. Współdzielenie danych – promowanie idei współdzielenia wśród członków społeczności sieci semantycznej.
4. Stanowczość i upór – nacisk na dostawców oprogramowania bibliotecznego, aby było ono zgodne ze standardami sieciowymi, promowanie idei otwartości danych.

Analizując zasady katalogowania RDA odnoszące się do kartotek haseł wzorcowych można zauważyć tendencję do zwiększenia zakresu odwzorowywanych informacji. Wraz

¹⁶ Case study: German National Library publishes its authority and bibliographic records as linked open data [online]. [dostęp 20.09.2015]. Dostępny w: <http://www.theeuropeanlibrary.org/tel4/newsitem/5802>.

¹⁷ BYRNE, G., GODDARD, L. The Strongest Link: Libraries and Linked Data. *D-Lib Magazine* [on-line] 2010, Vol. 16, no. 11/12 [dostęp 20.09.2015]. Dostępny w: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>.

ze zmianą postrzegania funkcji kartoteki haseł wzorcowych, szczególnie w odniesieniu do haseł formalnych, pojawia się realna szansa na zaistnienie tych źródeł informacji w infrastrukturze sieciowej jako wartościowych punktów odniesienia również dla podmiotów spoza kontekstu bibliotecznego. Eksperymenty bibliotek w kontekście publikowania swoich danych w modelu Linked Data pokazują, że jest to jedna z dróg, którą powinniśmy uwzględnić, chcąc zachować pozycję liczącego się gracza w obszarze dostawców usług informacyjnych w sieci. Jednocześnie projekty BIBFRAME i RDA (jako schemat metadanych) dają nadzieję na otrzymanie formalnych narzędzi do realizacji tego celu.

Bibliografia:

1. BERNERS-LEE, T. *Linked Data – Design Issues* [on-line] 2006 [dostęp 20.09.2015]. Dostępny w: <http://www.w3.org/DesignIssues/LinkedData.html>.
2. BIZER, Ch., HEATH, T. *Linked Data: Evolving the Web into a Global Data Space* [on-line] Morgan & Claypool Publishers, 2011 [dostęp 20.09.2015]. ISBN 978-1-60845430-3. Dostępny w: <http://linkeddatabook.com/editions/1.0/>.
3. BYRNE, G., GODDARD, L. The Strongest Link: Libraries and Linked Data. *D-Lib Magazine* [on-line] 2010, Vol. 16, no. 11/12 [dostęp 20.09.2015]. Dostępny w: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>.
4. CASE STUDY: GERMAN NATIONAL LIBRARY PUBLISHES ITS AUTHORITY AND BIBLIOGRAPHIC RECORDS AS LINKED OPEN DATA [online] [dostęp 20.09.2015]. Dostępny w: <http://www.theeuropeanlibrary.org/tel4/newsitem/5802>.
5. IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS. *Functional requirements for bibliographic records: final report*. München: K.G. Saur, 1998. ISBN 978-3-598-11382-6.
6. JOHNSON, L., ADAMS BECKER, S., ESTRADA, V., FREEMAN, A. *NMC Horizon Report: 2015 Library Edition* [on-line] Austin, Texas: The New Media Consortium, 2015 [dostęp 20.09.2015]. ISBN 978-0-9962832-4-3. Dostępny w: <http://cdn.nmc.org/media/2015-nmc-horizon-report-library-EN.pdf>.
7. KRYNICKA, M. *Format MARC 21 – stan prac i perspektywy rozwoju* [on-line] 2011 [dostęp 20.09.2015]. Dostępny w: <http://www.bn.org.pl/download/document/1321438599.ppt>.
8. PAPADAKIS, I., KYPRIANOS, K., STEFANIDAKIS, M. Linked Data URIs and Libraries: The Story So Far. *D-Lib Magazine* [on-line] 2015, Vol. 21, no. 5/6 [dostęp 20.09.2015]. Dostępny w: <http://www.dlib.org/dlib/may15/papadakis/05papadakis.html>.
9. PATTON, G.E. (red.). *Functional Requirements for Authority Data – A Conceptual Model* [on-line] München: K.G. Saur, 2009 [dostęp 20.09.2015]. ISBN 978-3-598-24282-3. Dostępny w: <http://www.ifla.org/publications/functional-requirements-for-authority-data>.
10. SCHMACHTENBERG, M., BIZER, Ch., PAULHEIM, H. *State of the LOD Cloud 2014* [on-line] 2014 [dostęp 20.09.2015]. Dostępny w: <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/#toc4>.
11. TILLET, B. Authority Control: State of the Art and New Perspectives. *Cataloging and Classification Quarterly* 2004, Vol. 38, no. 3–4, s. 23–41.
12. WILLER, M., DUNSIRE, G. *Bibliographic Information Organization in the Semantic Web* 2013. ISBN 978-1-84334-731-6.

Roszkowski, M. Kartoteki nazw osobowych w środowisku sieciowym. *Biuletyn EBIB* [on-line] 2015, nr 7 (160), *Po co bibliotekom normalizacja?* [Dostęp 20.11.2015]. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/380>. ISSN 1507-7187.