

TERMINY METADANYCH DCMI – NOWE ZALECENIE DLA METADANYCH DUBLIN CORE

Agnieszka Brachfogel
Biblioteka Narodowa

Metadane, Dublin Core, DCMI

Metadane Dublin Core, od momentu ich opracowania i opublikowania przez *Dublin Core Metadata Initiative*¹ w 1995 r. jako *Dublin Core Metadata Element Set v.1.1*², stały się jednym z najpopularniejszych zestawów elementów wykorzystywanych w Internecie do opisu dokumentów udostępnianych w repozytoriach cyfrowych³. Podstawowy zestaw zawiera piętnaście elementów, który można było uszczegóławiać za pomocą tzw. kwalifikatorów (*qualifiers*)⁴, co zestawiono w tabeli (str. 58).

Od 1998 r., gdy zestaw piętnastu elementów Dublin Core zyskał postać międzynarodowego standardu, pojęcia sieci semantycznej (*Semantic Web*)⁵ uległy rozszerzeniu o zakresy (*ranges*) i dziedziny (*domains*), za pomocą których jednoznacznie umiejscawia się poszczególne elementy w porządku schematu danych RDF (*Resource Description Framework*)⁶. Zalecenia Konsorcjum WWW (W3C) skutkowały tym, że w 2008 r. zaimplementowano je w strukturze metadanych Dublin Core, na nowo porządkując cały zestaw oraz uzupełniając o nowe potrzebne składniki. W efekcie powstało kolejne zalecenie DCMI, opublikowane jako *Terminy metadanych DCMI*⁷. Funkcjonuje ono równolegle do *DCMES*

¹ Dublin Core Metadata Initiative to organizacja odpowiedzialna za utrzymywanie i upowszechnianie metadanych Dublin Core – więcej patrz <http://dublincore.org/about-us/>.

² *Dublin Core Metadata Element Set (DCMES v.1.1)*, praktycznie zastąpiony przez *Terminy metadanych DCMI*, nadal obowiązuje jako zalecenie DCMI.

³ 15-elementowy zestaw metadanych Dublin Core (*DCMES v.1.1*) wydano także w formie standardów IETF RFC 5013, ANSI/NISO Z39.85-2007 oraz ISO 15836-21003 (polskie tłumaczenie tej normy to PN-ISO 15836:2006).

⁴ Dokument *Dublin Core Qualifiers* (<http://dublincore.org/documents/2000/07/11/dcm-es-qualifiers/>) został zastąpiony w 2008 r. przez *Terminy metadanych DCMI*. Zawierał trzy dodatkowe elementy: Odbiorcy – *Audience*, Pochodzenie – *Provenance*, Właściciel praw – *Rights Holder*, oraz dwadzieścia cztery uszczegółowienia (tzw. *refinements*, także zaliczane do kwalifikatorów) elementów podstawowego zestawu Dublin Core (patrz tabela str. 58).

⁵ Podstawowe informacje o sieci semantycznej patrz Wikipedia, hasło *Semantic Web* (http://pl.wikipedia.org/wiki/Semantic_Web).

⁶ *Resource Description Framework* (<http://www.w3.org/TR/rdf-primer/>).

⁷ *Terminy metadanych DCMI* (<http://www.bn.org.pl/download/document/1253606451.pdf>). Oryg.: DCMI Metadata Terms (<http://dublincore.org/documents/2008/01/14/dcmi-terms/>). Warto zwrócić uwagę na dość istotną zmianę w oficjalnym nazewnictwie: już nie operuje się określeniem „element”, ale „termin”. W pewnym stopniu, to co w *DCMES v. 1.1* określano jako „element”, w *Terminach metadanych DCMI* odpowiada *własnościom (properties)*, o których mowa w dalszej części artykułu.

Elementy DCMES	Uszczegółowienia <i>Dublin Core Qualifiers</i>
Tytuł (<i>Title</i>)	Wariant tytułu (<i>Alternative</i>)
Twórca (<i>Creator</i>)	–
Temat (<i>Subject</i>)	–
Opis (<i>Description</i>)	Spis treści (<i>Table of contents</i>) Abstrakt (<i>Abstract</i>)
Wydawca (<i>Publisher</i>)	–
Współtwórca (<i>Contributor</i>)	–
Data (<i>Date</i>)	Utworzenia (<i>Created</i>) Ważności (<i>Valid</i>) Dostępu (<i>Available</i>) Wydania (<i>Issued</i>) Modyfikacji (<i>Modified</i>)
Rodzaj (<i>Type</i>)	–
Format (<i>Format</i>)	Rozmiar (<i>Extent</i>) Nośnik (<i>Medium</i>)
Identyfikator (<i>Identifier</i>)	–
Źródło (<i>Source</i>)	–
Język (<i>Language</i>)	–
Relacja (<i>Relation</i>)	Jest wersją (<i>Is version of</i>) Ma wersję (<i>Has version</i>) Zastąpiony przez (<i>Is replaced by</i>) Zastępuje (<i>Replaces</i>) Jest wymagany przez (<i>Is required by</i>) Wymaga (<i>Requires</i>) Jest częścią (<i>Is part of</i>) Składa się z (<i>Has part</i>) Jest cytowany przez (<i>Is referenced by</i>) Cytuje (<i>References</i>) Jest formatem (<i>Is format of</i>) Jest w formacie (<i>Has format</i>)
Zasięg (<i>Coverage</i>)	Przestrzenny (<i>Spacial</i>) Czasowy (<i>Temporal</i>)
Prawa (<i>Rights</i>)	–
	Elementy Dublin Core Qualifiers
	Odbiorcy (<i>Audience</i>)
	Pochodzenie (<i>Provenance</i>)
	Właściciel praw (<i>Rights holder</i>)

v. 1.1, ale autorzy podkreślają we wstępie, żeby osoby zajmujące się implementacją Dublin Core korzystały już z nowych zaleceń.

Struktura Terminów metadanych DCMI

Terminy metadanych DCMI stanowią najnowszy wykaz wszystkich elementów metadanych utrzymywanych przez *Dublin Core Metadata Initiative (DCMI)*. Lista uporządkowana jest według następujących kategorii danych:

- 1) własności (*properties*),
- 2) słownikowe schematy kodowania (*vocabulary encoding schemes*),
- 3) syntaktyczne schematy kodowania (*syntax encoding schemes*),
- 4) klasy (*classes*).

Dodatkowo, w dwóch osobnych rozdziałach, przedstawiono terminy *Słownika typów DCMI*, którymi można wskazać formę opisywanego dokumentu, oraz dwa terminy związane abstrakcyjnym modelem danych DCMI⁸.

Każdy termin (element) prezentowany jest za pomocą podstawowego zestawu atrybutów (il. str. 60):

- **Nazwa:** jednoznaczne oznaczenie terminu, niepowtarzalne wśród terminów DCMI występujących w przestrzeni nazw⁹,
- **Etykieta:** zrozumiała dla człowieka etykieta terminu,
- **URI:** *Uniform Resource Identifier* stosowany dla jednoznacznej identyfikacji terminu,
- **Definicja:** objaśnienie terminu, jego znaczenie i cechy,
- **Rodzaj terminu:** rodzaj terminu zgodnie z definicją określoną w abstrakcyjnym modelu DCMI (*DCMI Abstract Model*).

Większość terminów uzupełniono o dodatkowe atrybuty:

- **Komentarz:** dodatkowe informacje o terminie lub jego użyciu,
- **Zob.:** skierowanie do dokumentów źródłowych powiązanych z terminem,
- **Odniesienie:** źródło wskazane w *Komentarzu* lub *Definicji*,
- **Uszczegóławia:** własność, względem której opisywany termin jest "podwłasnością",
- **Szerszy niż:** klasa, względem której opisywany termin jest "nadklasą",
- **Węższy niż:** klasa, względem której opisywany termin jest "podklasą",
- **Należy do dziedziny:** klasa, której przypadek stanowi źródło opisywane danym terminem,
- **Należy do zakresu:** klasa, której przypadek stanowi wartość opisana danym terminem,
- **Składnik:** wykaz źródeł (słownikowy schemat kodowania danych), którego elementem jest opisywany termin,
- **Przypadek:** klasa, której przypadkiem jest opisywany termin,
- **Wersja:** szczegóły dotyczące historii powstania terminu.

⁸ Abstrakcyjny model danych DCMI (*DCMI Abstract Model* <http://dublincore.org/documents/2007/06/04/abstract-model/>) jest opracowaniem, w którym przedstawiono w postaci diagramów i według założeń schematu RDF zależności między elementami metadanych Dublin Core.

⁹ Przestrzeń nazw (ang. *namespace*) identyfikowane są przez referencje IRI lub URI i pozwalają na stosowanie wielu języków opartych na XML-u w jednym pliku XML. Zostały stworzone przez W3C, aby zapobiec problemom spowodowanym przez powtarzanie się tych samych elementów w różnych językach. Przykład: element <title/> jest używany i w XHTML-u i w Atomie, ale dzięki temu że XHTML ma przestrzeń nazw <http://www.w3.org/1999/xhtml>, a Atom – <http://www.w3.org/2005/Atom>, programy interpretujące XML mogą je rozróżnić (źródło: Wikipedia. [online]. [dostęp: 10.12.2009]. Dostępny w World Wide Web: <http://pl.wikipedia.org/wiki/Przestrze%C5%84_nazw>).

Nazwa terminu: abstract	
URI:	http://purl.org/dc/terms/abstract
Etykieta:	Abstrakt
Definicja:	Streszczenie zasobu.
Rodzaj terminu:	<u>Property</u>
Uszczegóławia:	http://purl.org/dc/elements/1.1/description
Uszczegóławia:	http://purl.org/dc/terms/description
Wersja:	http://dublincore.org/usage/terms/history/#abstract-003

Przykład prezentacji terminu w *Terminach metadanych DCMi*

Dzięki tym informacjom można rozeznac się w strukturze terminów oraz w zakresie ich stosowania, a także w zachodzących pomiędzy nimi ewentualnych zależnościach.

Własności (*properties*)

Na początku wykazu wymienione są własności¹⁰. Są to kluczowe metadane Dublin Core, bo to za ich pomocą opisuje się dany zasób. Rozdzielono je na dwie kategorie:

- 1) własności w przestrzeni nazw (*terms*),
- 2) własności w dawnej przestrzeni nazw (*elements/1.1*).

We wstępie do *Terminów metadanych DCMi* znajdujemy następujące wytłumaczenie, dlaczego dokonano tego podziału:

„Dla zachowania zgodności pomiędzy istniejącymi implementacjami „prostego Dublin Core” w RDF nie precyzowano dziedzin i zakresów dla piętnastu własności *dc: namespace* (<http://purl.org/dc/elements/1.1/>). Tych piętnaście elementów zostało utworzonych w *dcterms: namespace* (<http://purl.org/dc/terms/>) jako nowe własności o nazwach identycznych jak te wymienione w DCMES w wersji 1.1 (Dublin Core Metadata Element Set). Zostały one zdefiniowane jako „podwłasności” odpowiednich własności DCMES wersji 1.1 z odpowiednio przypisanymi dziedzinami i zakresami [...].

Osoby zajmujące się implementacją Dublin Core mogą swobodnie wybierać pomiędzy stosowaniem piętnastu elementów z wcześniejszej wersji *dc:* (np. <http://purl.org/dc/elements/1.1/creator>) a *dcterms:* (np. <http://purl.org/dc/terms/creator>), stosownie do wymagań aplikacji. W schematach RDF dla przestrzeni

¹⁰ Patrz uwaga w przypisie 5.

nazw DCMI opisano relację „podwłasności”, zachodzącą pomiędzy elementami dcterms: creator a dc:creator, co można wykorzystać w semantycznych aplikacjach w sieci Web. Z czasem jednak będzie się zachęcać do korzystania z bardziej jednoznacznych własności dcterms, jako że pełniej nadszają za wyłaniającymi się podstawami dobrych praktyk w dziedzinie automatycznego przetwarzania metadanych [...]”¹¹.

Własności w dawnej przestrzeni nazw (*elements/1.1*) to w zasadzie wykaz piętnastu elementów Dublin Core, którymi dotychczas operowano, zaś własności w nowej przestrzeni nazw (*terms*) to zaktualizowany i uzupełniony wykaz metadanych Dublin Core, dostosowany do aktualnych wymagań informatycznych.

Własności w przestrzeni nazw (*terms*)

Na własności w nowej przestrzeni nazw (*terms*) składa się pięćdziesiąt pięć terminów¹², które można wykorzystać do opisu dokumentu. Wśród nich znajdują się wszystkie elementy znane z podstawowego zestawu Dublin Core (m.in. Twórca – *Creator*, Tytuł – *Title*, Opis – *Description*, Temat – *Subject*, Zasięg – *Coverage*) oraz dawniejsze kwalifikatory (m.in. Właściciel praw – *Rights Holder*, Pochodzenie – *Provenance*, Odbiorcy – *Audience*, Wariant tytułu – *Alternative Title*, Data modyfikacji – *Date Modified*, Zasięg przestrzenny – *Spatial Coverage*). Poza tym dodano trzynaście zupełnie nowych terminów, za pomocą których opisać można różne aspekty dokumentu, zarówno należące do strefy opisu formalnego, jak i rzeczowego. Są to:

1. Prawa dostępu – *Access Rights*,
2. Licencja – *License*,
3. Metoda uzupełnienia zbioru – *Accrual Method*,
4. Częstotliwość uzupełniania zbioru – *Accrual Periodicity*,
5. Reguły uzupełniania zbioru – *Accrual Policy*,
6. Cytata bibliograficzna – *Bibliographic Citation*,
7. Zgodny z – *Conforms to*,
8. Data przyjęcia – *Date Accepted*,
9. Data copyright – *Date Copyrighted*,
10. Data złożenia – *Date Submitted*,
11. Poziom edukacyjny odbiorcy – *Education Level*,
12. Metoda szkolenia – *Instructional Method*,
13. Pośrednik – *Mediator*.

Zwiększenie ilości elementów Dublin Core może służyć tworzeniu precyzyjniejszych opisów obiektów udostępnianych w repozytoriach. W istniejących bazach już korzysta się z tej możliwości, zastępując ogólne terminy (np. *Zasięg*) elementami o węższej definicji (np. *Zasięg przestrzenny*, *Zasięg czasowy*).

¹¹ *Terminy metadanych DCMI*. [online]. [dostęp: 10.12.2009]. Dostępny w World Wide Web: <<http://www.bn.org.pl/download/document/1253606451.pdf>>.

¹² Liczba wszystkich elementów należących do kategorii *własność* wynosi w sumie siedemdziesiąt jeden, z czego pięćdziesiąt pięć występuje w przestrzeni nazw (*terms*), piętnaście – w dawnej przestrzeni nazw (*elements/1.1*), a jedno – w przestrzeni nazw (*dcam*) (chodzi o własność *Składnik [Member of]*).

Słownikowe schematy kodowania (*vocabulary encoding schemes*)

Ta część opracowania DCMI to wykaz dziewięciu słowników oraz list wykorzystywanych w opisie rzeczowym dokumentów. Są to słowniki haseł i terminów przedmiotowych (MeSH, LSCH), klasyfikacje (UKD, DDC, LCC, NLM) oraz inne przydatne zestawienia (IMT, Słownik typów DCMI, TGN). Każdy z nich ma przypisane URI w przestrzeni nazwowej (*terms*), co jednoznacznie pozwala zidentyfikować źródło, z którego zaczerpnięte zostało hasło określające treść lub formę obiektu.

Syntaktyczne schematy kodowania (*syntax encoding schemes*)

Wśród terminów metadanych DCMI wymieniono jedenaście standardów (m.in. ISO, RFC), które są zestawieniami kodów (np. kody nazw języków, kody nazw krajów, schematy podawania dat czy współrzędnych miejsca), i – podobnie jak w przypadku słownikowych schematów kodowania – służą przede wszystkim do jednoznacznej identyfikacji źródła kodu podanego w opisie danego obiektu.

Klasy (*classes*)

Klasy są to kategorie danych, za pomocą których porządkuje się schematy metadanych w sieci semantycznej, tzn. poprzez klasy definiuje się poszczególne elementy, m.in. własności (np. własność *Twórca* należy do klasy *Agent*, stąd wiemy że chodzi o - zgodnie z definicją tej klasy – „jednostkę, która działa lub może działać”, czyli „osobę, instytucję lub agenta software'u”). Innymi słowy, klasy określają, czym są elementy (metadane), z których buduje się bazę danych. Nie można przypisać im bezpośrednio jakiejś wartości, jak to ma miejsce w przypadku własności (np. hasło *Adam Mickiewicz* może być, zależnie od sytuacji, wartością własności *Twórca* albo *Współtwórca*). To własności dopiero należą do określonej klasy (własność *Twórca* i *Współtwórca* należą do klasy *Agent*). Klasa nie jest zatem elementem opisu jakiegoś obiektu ale, leżąc w strukturze opisu, pozwala na grupowanie informacji na poziomie abstrakcyjnym (kwestia odpowiedniego przygotowania i ustawienia systemu).

Dzięki temu możliwe jest sprawowanie kontroli nad poprawnością wpisywanych wartości, przypisaniem wybranej wartości do danej własności (np. hasło *Adam Mickiewicz* w elemencie *Twórca*). Taka kontrola jest jednak możliwa tylko wówczas, gdy wartości czerpiemy ze słowników kontrolowanych. Dodatkową zaletą zastosowania klas jest to, że można się nimi posłużyć przy kształtowaniu procesu wyszukiwania, dostosowując go do indywidualnych potrzeb (np. znalezienia wszystkich danych powiązanych bezpośrednio i pośrednio z klasą *Agent*).

W przestrzeni nazw (*terms*) wymieniono dwadzieścia dwie klasy¹³. Wszystkich klas utrzymywanych przez DCMI jest więcej, bo aż trzydzieści pięć: oprócz wspomnianych dwudziestu dwóch, kolejnych dwanaście należy do przestrzeni nazw (*dcmitype*) (to są terminy wchodzące w skład *Słownika typów DCMI*¹⁴),

¹³ Patrz: Terminy metadanych DCMI, rozdział 6: Klasy.

¹⁴ Patrz: Terminy metadanych DCMI, rozdział 7: Słownik typów DCMI.

a jedna – do przestrzeni nazw (*dcam*) (chodzi o klasę *Słownikowy schemat kodowania – VocabularyEncodingScheme*¹⁵).

* * *

Upowszechnienie formatu Dublin Core oraz jego osadzenie w dynamicznie rozwijającym się Internecie wymusiło przeprowadzenie weryfikacji i, w efekcie, wprowadzenie przez DCMI odpowiednich zmian w strukturze standardu Dublin Core. Przede wszystkim uporządkowano metadane według kategorii danych: własności, klas, schematów kodowania danych, dostosowując terminologię do tej używanej w kontekście sieci semantycznej i schematu RDF. Poszczególnym terminom utrzymywanym przez DCMI nadano nowe identyfikatory URI w sposób ujednolicony i zgodny z międzynarodowymi zaleceniami. Sam wykaz uzupełniono o dodatkowe elementy, które mogą być użyteczne dla danego repozytorium.

Mimo że wykaz opublikowany jako *DCMES v1.1* nie został wycofany, to we wstępie do *Terminów metadanych DCMI* czytamy, że osoby zajmujące się implementacją metadanych Dublin Core powinny korzystać z elementów ustanowionych w nowej przestrzeni nazw (*terms*). Ze wstępnych badań wynika jednak, że w repozytoriach tworzy się struktury mieszane, budując je z różnych elementów (np. *dc: elements/1.1/* obok *dc: terms*, a także własnych, stworzonych dla potrzeb konkretnej bazy, analogicznie do pól lokalnych formatu MARC).

Pomimo rozbudowy struktura metadanych Dublin Core jest stosunkowo czytelna, choć z początku wymaga przyswojenia nowych pojęć i zrozumienia schematu, według którego funkcjonuje¹⁶. Zachowano dużą dowolność w interpretacji elementów, co można różnie oceniać: wyraźne odejście od sztywnych zasad katalogowania jest z jednej strony zaletą – daje bowiem możliwość dość swobodnego wypełniania danymi, z uwzględnieniem charakteru repozytorium i celów, którym ma służyć, ale z drugiej strony może wprowadzać użytkowników w konsternację, że te same metadane zawierają różne informacje, i odwrotnie – identyczne dane umieszczane są w różnych miejscach opisu (możemy ten fakt obserwować w istniejących repozytoriach). Ustanowienie klasy danych w pew-

¹⁵ Patrz: Terminy metadanych DCMI, rozdział 8: Terminy związane z abstrakcyjnym modelem DCMI (*DCMI Abstract Model*).

¹⁶ Patrz: *DCMI Abstract Model* <http://dublincore.org/documents/2007/06/04/abstract-model/> oraz *Resource Description Framework (RDF)* <http://www.w3.org/TR/rdf-primer/>. Praktyczną pomocą służyć może również *The Dublin Core Metadata Registry* <http://dcmi.kc.tsukuba.ac.jp/dcregistry/navigateServlet> – rejestr wszystkich obowiązujących elementów Dublin Core.

nym stopniu zaradziło temu problemowi, bo za ich pomocą zawęża się zakres użycia wybranych elementów.

Trudno określić, ile repozytoriów wprowadziło do swoich struktur elementy z nowej przestrzeni nazw, wiadomo jednak, że już się one pojawiły i zapewne będą się upowszechniać.

Summary

The author presents changes made by DCMI within the Dublin Core standard structure, resulting from its fast dissemination within the Internet.