

# „TERMINY METADANYCH DCMI” I MOŻLIWOŚCI ICH WYKORZYSTANIA W OPISIE RZECZOWYM<sup>1</sup>

Agnieszka Brachfogel  
Biblioteka Narodowa

*Metadane, Dublin Core, DCMI, opis rzeczowy, MARC21*

Metadane Dublin Core to jedno z najpopularniejszych zestawów elementów opisu dokumentów w repozytoriach cyfrowych lub internetowych multiwyszukiwarkach. Elementy Dublin Core (*DCMES v 1.1*<sup>2</sup>) sformułowano w połowie lat 90. w postaci zalecenia Dublin Core Metadata Initiative (DCMI) i aż do 2008 r. nie podlegały właściwie większym zmianom<sup>3</sup>. Upowszechnienie zaleceń międzynarodowych organizacji (m.in. Konsorcjum WWW) dotyczące struktur i schematów danych (m.in. sieci semantycznej, modelu RDF) spowodowało, że w 2008 r. DCMI opublikowała nowy dokument – *Terminy metadanych DCMI*, w którym uwzględniła te wytyczne. Podstawowy zestaw elementów *DCMES v 1.1* nie został wycofany, a jego zawartość, przejęta do nowego opracowania, stanowi w nim jeden z rozdziałów.

*Terminy metadanych DCMI* zawierają najnowszy wykaz wszystkich elementów metadanych utrzymywanych przez DCMI. Lista uporządkowana jest według następujących kategorii danych:

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

<sup>1</sup> Tekst jest zapisem wystąpienia przygotowanego na konferencję *Cyfrowość bibliotek i archiwów* zorganizowaną w Warszawie przez Stowarzyszenie Bibliotekarzy Polskich i Bibliotekę Narodową, 26-27 listopada 2009 r. Prezentacja dostępna jest na: <http://www.bn.org.pl/dla-bibliotekarzy/jhp-bn/aktualnosci/konferencja-cyfrowosc-bibliotek-i-archiwow>.

<sup>2</sup> *Dublin Core Metadata Element Set, Version 1.1*. [online]. [dostęp: 5.11.2010]. Dostępny w World Wide Web: <http://dublincore.org/documents/dces/>; wersja polska *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. Oprac. M. Nahotko. [online]. [dostęp: 27.10.2010]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/standard/dc.html>; patrz też: PN-ISO 15836 *Informacja i dokumentacja – Zestaw elementów metadanych Dublin Core*.

<sup>3</sup> Uzupełnieniem podstawowego 15-elementowego zestawu Dublin Core (*DCMES v 1.1*) był *Dublin Core Qualifiers*, wykaz zawierający dodatkowe elementy (tzw. kwalifikatory), które uszczegóławiały wybrane elementy zestawu *DCMES*. *Dublin Core Qualifiers* został zastąpiony przez *Terminy metadanych DCMI*, a kwalifikatory stały się integralnymi elementami nowego zestawu metadanych. Więcej patrz: <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>.

Dodatkowo, w dwóch osobnych rozdziałach, przedstawiono terminy *Słownika typów DCMI*, którymi można wskazać formę opisywanego dokumentu, oraz dwa terminy związane abstrakcyjnym modelem danych DCMI<sup>4</sup>.

Każdy termin prezentowany jest za pomocą następującego podstawowego zestawu atrybutów:

NAZWA – jednoznaczne oznaczenie terminu, niepowtarzalne wśród terminów DCMI,

ETYKIETA – nazwa terminu zrozumiała dla człowieka,

URI – jednoznaczny identyfikator terminu,

DEFINICJA – objaśnienie terminu, jego znaczenie i cechy,

RODZAJ TERMINU – rodzaj terminu zgodnie z definicją określoną w abstrakcyjnym modelu DCMI.

Większość terminów uzupełniono o dodatkowe informacje (np. komentarz będący rozszerzeniem definicji, odnośniki do innych terminów). Dzięki temu można zorientować się w strukturze metadanych, poznać zakres ich stosowania i ewentualne zależności występujące pomiędzy elementami.

## Własności, czyli elementy opisu

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

– własności w przestrzeni nazw (*terms*)<sup>5</sup>,

– własności w dawnej przestrzeni nazw (*elements/1.1*)<sup>6</sup>.

Własności w dawnej przestrzeni nazw */elements/1.1/* to piętnaście elementów Dublin Core, którymi dotychczas operowano. Natomiast nowością są własności w przestrzeni nazw */terms/*, które są tamtych aktualizacją i uzupełnieniem. Wykaz własności w przestrzeni nazw */terms/* składa się z pięćdziesięciu pięciu terminów<sup>7</sup>, z których:

– czterdzieści dwa to elementy znane z podstawowego i rozszerzonego zestawu Dublin Core<sup>8</sup> – zachowano nazwy i definicje, a przydano nowe URI i odpowiednio powiązano je z innymi terminami.

<sup>4</sup> Abstrakcyjny model danych DCMI (*DCMI Abstract Model* <http://dublincore.org/documents/2007/06/04/abstract-model/>) jest to opracowanie, w którym przedstawiono – w postaci diagramów i według założeń schematu RDF – zależności pomiędzy poszczególnymi elementami metadanych Dublin Core.

<sup>5</sup> Patrz: *Terminy metadanych DCMI*. Rozdział 2: Własności w przestrzeni nazw */terms/*. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <<http://www.bn.org.pl/download/document/1253606451.pdf>>.

<sup>6</sup> Patrz: *Terminy metadanych DCMI*. Rozdział 3: Własności w przestrzeni nazw */elements/1.1/*. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <<http://www.bn.org.pl/download/document/1253606451.pdf>>.

<sup>7</sup> Własności DCMI jest w sumie siedemdziesiąt jeden, z czego pięćdziesiąt pięć występuje w przestrzeni nazw */terms/*, piętnaście – w dawnej przestrzeni nazw *elements/1.1/*, a jedna – w przestrzeni nazw */dcam/* (chodzi o własność *Składnik [Member of]*).

<sup>8</sup> Wśród tych 39 elementów: 15 to elementy znane z podstawowego zestawu Dublin Core (*DCMES v.1.1*), a pozostałe 24 z wykazu *Dublin Core Qualifiers*, na który składały się elementy równorzędne wobec *DCMES* (np. *Odbiorcy*) oraz je uszczegóławiające (np. kwalifikatorem elementu *Tytuł* był *Wariant tytułu*).

Trzynastcie to nowe terminy, są to:

- prawa dostępu (*Access Rights*),
- licencja (*License*),
- metoda uzupełnienia zbioru (*Accrual Method*),
- częstotliwość uzupełniania zbioru (*Accrual Periodicity*),
- reguły uzupełniania zbioru (*Accrual Policy*),
- cytata bibliograficzna (*Bibliographic Citation*),
- zgodny z (*Conforms to*),
- data przyjęcia (*Date Accepted*),
- data copyright (*Date Copyrighted*),
- data złożenia (*Date Submitted*),
- poziom edukacyjny odbiorcy (*Education Level*),
- metoda szkolenia (*Instructional Method*),
- pośrednik (*Mediator*).

Ze wszystkich własności przestrzeni nazw */terms/* do opisu rzeczowego wykorzystać można następujące elementy:

**TEMAT** (*Subject*) – tu można podawać wszelkie informacje dotyczące treści (zawartości) opisywanego obiektu; zaleca się przy tym korzystanie ze słownictwa kontrolowanego (np. hasła przedmiotowe, symbole klasyfikacyjne).

**OPIS** (*Description*) – może zawierać „m.in.: streszczenie, spis treści, graficzne przedstawienie treści lub opis zasobu w formie swobodnego tekstu”. Jest to kolejny element dotyczący treści obiektu, ale w przeciwieństwie do *Tematu* – tu ma się większą swobodę w doborze i formie podawanych informacji.

**ABSTRAKT** (*Abstract*) – jest szczególnym przypadkiem *Opisu*, więc jego treść może równie dobrze znaleźć się w elemencie *Opisie*. Od przyjętych zasad zależy, który z tych elementów będzie wykorzystywany.

**SPIS TREŚCI** (*Table of Contents*) – podobnie jak *Abstrakt* jest szczególnym przypadkiem *Opisu* i od przyjętych zasad zależy, czy spis treści podany zostanie w *Opisie* czy *Spisie treści*.

**RODZAJ** (*Type*) – ogólna definicja mówi, że chodzi tu o „typ lub rodzaj zasobu”. Komentarz uzupełnia ją o wskazanie, że „najlepiej jest korzystać ze słowników kontrolowanych, jak np. *Słownik typów DCMI* (DCMI Type Vocabulary). Format pliku, obiekt fizyczny czy wymiary zasobu podajemy w elemencie *Format*”. Przykładem są określenia typu: kolekcja, oprogramowanie, obraz ruchomy lub terminy określające formę edytorsko-piśmienniczą dokumentów tekstowych np.: powieść, podręcznik, czasopismo.

**FORMAT** (*Format*) – tu podaje się cechy fizyczne opisywanego dokumentu, tzn. format pliku, nośnik lub wymiary. Informacje o rodzaju nośnika możemy podawać osobno w przeznaczonym do tego *Nośniku*.

**NOŚNIK** (*Medium*) – w tym miejscu podaje się informacje o fizycznym nośniku opisywanego obiektu, np.: płótno, papier, płyta DVD.

**ODBIORCY** (*Audience*) – wskazanie „kategorii jednostek, do których zasób jest kierowany lub dla których jest użyteczny”. Mogą to być określenia typu: kobiety, dzieci, biznesmeni.

**POZIOM EDUKACYJNY ODBIORCY** (*Education Level*) – określenie grup odbiorców poprzez wskazanie poziomu ich wykształcenia, np.: uczniowie, studenci, wykładowcy etc.

METODA SZKOLENIA (*Instructional Method*) – tu można podawać informacje o metodzie kształcenia wykorzystanej w opisywanym zasobie.

POŚREDNIK (*Mediator*) – wskazanie uczestników procesu nauczania, do których – m.in. – skierowany jest opisywany zasób (np.: nauczyciele, rodzice, trenerzy).

ZASIĘG (*Coverage*) – „umiejscawianie treści zasobu w czasie i przestrzeni, wskazywanie zakresu zastosowania lub jurysdykcji, której zasób podlega”. Preferowane są określenia słowne (nie numeryczne, a więc nie współrzędne geograficzne, ale nazwa miejsca, nie daty, ale nazwa epoki).

ZASIĘG CZASOWY (*Temporal Coverage*) – zawężenie Zasięgu zasobu do charakterystyki zasięgu chronologicznego.

ZASIĘG PRZESTRZENNY (*Spatial Coverage*) – zawężenie Zasięgu zasobu do charakterystyki zasięgu przestrzennego.

## Własności DCMI a opis przedmiotowy w formacie MARC 21

Wymienione wyżej własności DCMI przekładają się na następujące pola opisu rzeczowego w formacie MARC21:

Zestawienie własności DCMI i ich odpowiedników pól formatu MARC 21 w opisie rzeczowym

Własność DCMI	Rodzaj pola/podpola	MARC 21
TEMAT	hasło rzeczowe hasło osobowe hasło korporatywne nazwa imprezy tytuł	650 600 610 611 630
RODZAJ FORMAT NOŚNIK	forma/rodzaj określnik formalny	655 6XX \$v
ODBIORCY POŚREDNIK POZIOM EDUKACYJNY ODBIORCY METODA SZKOLENIA	określnik formalny	6XX \$v
ZASIĘG ZASIĘG PRZESTRZENNY	hasło/określnik geograficzny	651 6XX \$z
ZASIĘG ZASIĘG CZASOWY	data z hasła określnik chronologiczny	6XX \$d 6XX \$y
OPIS ABSTRAKT SPIS TREŚCI	streszczenie	520

W praktyce konwersja haseł utworzonych w formacie MARC21 na strukturę Dublin Core polega na fasetowaniu, tzn. rozdzieleniu poszczególnych elementów hasła na równorzędne jednostki. Np.:

**650 09 \$a** Przemysł tytoniowy **\$x** prawo **\$z** Polska **\$y** stan na 1925 r. **\$v**  
źródła  
(*hasło wg JHP BN*)

Temat: *Przemysł tytoniowy – prawo*  
Zasięg przestrzenny: *Polska*  
Zasięg czasowy: *1925 r.*  
Rodzaj: *źródła (ustawa)*

**650 0 \$a** Rośliny aromatyczne **\$x** uprawy **\$z** Polska **\$v** czasopisma  
(*hasło wg KABA*)

Temat: *Rośliny aromatyczne – uprawy*  
Zasięg przestrzenny: *Polska*  
Rodzaj: *czasopismo*

Problemy przy konwersji haseł na strukturę Dublin Core dotyczą przede wszystkim odpowiedniego wyboru i przygotowania haseł, prawidłowego zestawienia poszczególnych formatów (tzw. mapowania), a także ujednoczenia form niektórych wartości (np. liczbę mnogą *czasopisma*, *źródła* należałoby zastąpić liczbą pojedynczą *czasopismo*, *źródło*; zamiast ogólnego terminu *źródło* należałoby podawać formę opisywanego obiektu, np. *ustawa*, *rozporządzenie*, *statut*).

### **Przykłady zastosowania własności z przestrzeni nazw /terms/**

W istniejących bazach najczęściej stosowane są struktury mieszane, tzn. podstawę stanowią elementy Dublin Core w dawnej przestrzeni nazw */elements/1.1/*, ale niemal zawsze uzupełniane są elementami własnymi, stworzonymi dla potrzeb danej bazy. Marginalnie można spotkać elementy z nowej przestrzeni nazw */terms/*.

Elementy Dublin Core z nowej przestrzeni nazw */terms/* wykorzystano na przykład w następujących projektach:

- CONTENTdm
- MICHAEL Project
- XMETADIS.

**CONTENTdm** to oprogramowanie służące do przesyłania, opisywania i udostępniania kolekcji cyfrowych umożliwiające przeszukiwanie zawartości kolekcji cyfrowych (jest to rodzaj multiwyszukiwarki i repozytorium w jednym). Stworzone zostało przez specjalistów z University of Washington, a obecnie jest produktem oferowanym przez OCLC. Znalazły się tu następujące nowe własności Dublin Core (wyróżniony element należy do opisu rzeczowego):

Date Created <dcterms:created>  
Date Published <dcterms:issued>  
Collection Identifier <dcterms:isPartOf>

### **Subject (Place) <dcterms:spatial>**

Title (Alternative) <dcterms:alternative>

Metacollection Identifier <dcterms:isPartOf>

**MICHAEL Project** to europejski wielojęzyczny serwis umożliwiający przeszukiwanie cyfrowych kolekcji muzeów, archiwów, bibliotek i innych instytucji kultury. W strukturze tej bazy znajdują się następujące nowe własności Dublin Core (wyróżnione elementy należą do opisu rzeczowego):

Size [dcterms:extent]

Standard [dcterms:conformsTo]

**Spatial Coverage [dcterms:spatial]**

**Period [dcterms:temporal]**

Super-Collection [dcterms:isPartOf]

Sub-Collection [dcterms:hasPart]

**XMETADIS** to format zestawu metadanych Niemieckiej Biblioteki Narodowej sporządzony dla dysertacji i prac postdoktorskich udostępnianych online. W jego strukturze znalazły się następujące nowe własności Dublin Core (wyróżniono elementy należące do opisu rzeczowego):

<dcterms:alternative>

**<dcterms:abstract>**

<dcterms:issued> lub <dcterms:dateAccepted>

**<dcterms:medium>**

## **Kontrola słownictwa**

Opracowanie rzeczowe udostępnianych dokumentów powinno opierać się na słownictwie kontrolowanym. Dlatego wśród terminów metadanych DCMI wymienia się słowniki oraz wykazy, które stanowią źródło formalnego słownictwa. Podzielono je na dwie kategorie:

- słownikowe schematy kodowania (*vocabulary encoding schemes*),
- syntaktyczne schematy kodowania (*syntax encoding schemes*).

Do **słownikowych schematów kodowania**<sup>9</sup> należą:

- słowniki haseł i terminów przedmiotowych (LSCH, MeSH),
- klasyfikacje (DDC, LCC, NLM, UKD),
- inne (IMT, Słownik typów DCMI, TGN).

Każde z nich ma przypisane URI w przestrzeni nazwowej */terms/*, co jednoznacznie pozwala zidentyfikować źródło, z którego zaczerpnięte zostało hasło określające treść lub formę zasobu.

<sup>9</sup> Patrz: *Terminy metadanych DCMI*. Rozdział 4: Słownikowe schematy kodowania. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <<http://www.bn.org.pl/download/document/1253606451.pdf>>.

Do **syntaktycznych schematów kodowania**<sup>10</sup> należą standardy (m.in. ISO, RFC), które są zestawieniami różnego typu kodów (np. nazw języków, 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 zasobu.

Poza schematami kodowania danych osobną kategorię stanowią **klasy**<sup>11</sup>, za pomocą których porządkuje się metadane w sieci semantycznej. Klasy definiują poszczególne elementy, m.in. własności (np. własność *Zasięg czasowy* należy do klasy *Okres*, stąd wiadomo że chodzi o – zgodnie z definicją tej klasy – „nazwę przedziału czasu lub daty jego początku i końca”). Klasy określają, czym są elementy, z których buduje się bazę danych. Klasa nie jest zatem elementem opisu jakiegoś obiektu, ale pozwala na odpowiednie grupowanie informacji (kwestia odpowiedniego przygotowania i ustawienia systemu).

Dzięki temu możliwa jest kontrola poprawności przypisywania wartości danej własności (np. hasło *Adam Mickiewicz* do elementu *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 daną klasą).

## Podsumowanie

Podstawowe zmiany, które wprowadzono do formatu Dublin Core, polegają na:

- przemianowaniu i pogrupowaniu poszczególnych elementów zgodnie ze schematem RDF,
- osadzeniu elementów opisu w nowej przestrzeni nazw */terms/* i określeniu relacji pomiędzy własnościami z nowej przestrzeni nazw */terms/* i dawnej */elements/1.1/*,
- uzupełnieniu zestawu o nowe elementy.

Z pięćdziesięciu pięciu własności nowej przestrzeni nazw */terms/* czternastoma można posłużyć się do opisu treści (zawartości) lub formy zasobu, trzy, tj.: Poziom edukacyjny odbiorcy (*Education Level*), Metoda szkolenia (*Instructional Method*) oraz Pośrednik (*Mediator*) są całkowicie nowymi elementami.

Zwiększenie liczby elementów Dublin Core pozwala tworzyć precyzyjniejsze opisy obiektów udostępnianych w repozytoriach. W istniejących bazach już

<sup>10</sup> Patrz: *Terminy metadanych DCMI*. Rozdział 5: Syntaktyczne schematy kodowania. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <<http://www.bn.org.pl/download/document/1253606451.pdf>>.

<sup>11</sup> Patrz: *Terminy metadanych DCMI*. Rozdział 6: Klasy. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <<http://www.bn.org.pl/download/document/1253606451.pdf>>. W przestrzeni nazw */terms/* wymieniono dwadzieścia dwie klasy. 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*), a jedna – do przestrzeni nazw */dcam/* (chodzi o klasę *Słownikowy schemat kodowania [VocabularyEncodingScheme]*).

korzysta się z tej możliwości, zastępując ogólne terminy (np. *Zasięg*) elementami o węższym zakresie (np. *Zasięg przestrzenny*, *Zasięg czasowy*). Wydaje się jednak, że w większości repozytoriów zmiany w strukturach metadanych będą marginalne, i jak to widać na przedstawionych wyżej przykładach, będą mieszanką elementów starych i nowych oraz własnych, wykorzystywanych lokalnie w danej bazie.

## Summary

DCMI Metadata Terms is an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative. Included are the fifteen terms of the Dublin Core Metadata Element Set (v. 1.1). The hierarchy and appropriate linking between the terms has been created. Number of the subject elements has increased. Partly it is the refinements of the basic elements but some of them are completely new. Although the conversion into and from MARC fields is already possible, some difficulties are encountered and need to be solved. Usage of the DCMI Metadata Terms has not become widespread yet, however the elements enter the structures of the catalogues.