Przy opisie jakiegoś zasobu - książki, kolekcji medali, dokumentu, obrazu - mamy w zasadzie dwie możliwości. Dokonać opisu w postaci narracji (zwykle sięgamy do tego, co jest pod ręką, czyli w dobie komputerów po procesor tekstu). Albo dokonać opisu w postaci struktury - na przykład w tabeli arkusza rozliczeniowego. Opis w postaci narracji pozwala na pełna ekspresję intencji badacza, i jednocześnie prawie uniemożliwia dalsza automatyczną przeróbkę danych. W Instytucie Piłsudskiego mamy doskonale zrobiony przez doświadczonego fachowca opis kolekcji falerystycznej, który ma postac narracji. Poszczególne elementy opisu (nazwa odznaki, jej twórca, miejsce stworzenia, daty itp.) są graficznie uwydatnione- np przez użycie czcionki wytłuszczonej, kursywy, przez oddzielanie elementów przecinkami, średnikami itp. Jednocześnie, jeśli jakiejś informacji brak, po przecinku znajdzie się już inny element opisu. Setki stron takiego tekstu wymagają wielu dni a nawet tygodni pracy aby zrobić prostą tabelkę którą można wyświetlić na stronie internetowej, gdyż zautomatyzowanie konwersji jest prawie niemożliwe.

Nawet prosta tabela w arkuszu rozliczeniowym daje strukturę - tytuł bedzie zawsze np. w kolumnie trzciej a data w siódmej itp. Jeśli więc użyliśmy jakiejś struktury, i nie mieszalismy np. miejsca z datą, mamy podstawy do użycia danych w różny sposób, taki, jakiego w danym momencie potrzebuje projektant wystawy, witryny internetowej czy inwentarza. Przy użyciu standardów metadanych najważniejszą decyzją jest użycie struktury adekwatnej do opisywanego zasobu. Przetłumaczenie tej struktury na taki czy inny standard metadanych jest wtedy zajęciem trywialnym. Mówiąc trywialnym mam na myśli to, że da się zautomatyzować - kiedy raz stworzymy algorytm konwersji, przeróbka 100 czy 100 tysięcy rekordów to tylko sprawa zapuszczenia komputera na sekundy albo godziny pracy.

Standardy metadanych mają strukturę, ale maja jeszcze cos innego - składnię która rozumieją wszyscy (albo różne programy). Tak, jak np. możemy komuś posłać dokument w formacie pdf i mamy dużą szanse że dokument zostanie otwarty, tak samo możemy posłać dane w określonym standarcie metadanych, i odbiorca będzie mógł te dane odczytać i użyć. A więc zamiast używać artbitralnych nagłówków kolumn w tabelce, będziemy używali standardowych “nagłówków” których znaczenia nie trzeba będzie tłumaczyć, bo jest powszechnie znane.

Dublin Core (DC) jest najprostszym i najbardziej podstawowym standardem metadanych. Powstał w roku 1995, i powinien byc uczony w szkołach jako część podstawowej wiedzy o

Internecie, gdyby tylko rozwój Internetu był nieco wolniejszy. Pozostaje nam teraz alternatywa, w postaci kursów dokształcających (co dotyczy zresztą wszystkich elementow wiedzy o Internecie). Dublin Core jest standardem, który najlepiej pasuje do metafory tabelki. W swojej podstawowej wersji posiada 15 elementów najczęściej używanych przy opisie zasobu.

Tabela 1. Wszystkie elementy (elements) Dublin Core (wersja 1.1). Namespace - ***dc***

Element	***Etykieta***	***Opis***
title	Tytuł	Nazwa zasobu.
creator	Twórca	Jednostka w głównej
contributor	Współtwórca	Jednostka odpowiedzi
subject	Temat	Tematyka opis
description	Opis	Opis zasobu.
publisher	Wydawca	Jednostka odpowiedzi
date	Data	Pojedyncza da
type	Rodzaj	Istota lub gatu
format	Format	Format pliku, fi
identifier	Identyfikator	Najlepiej jest id
source	Źródło	Zasób, z któreg
language	Język	Język zasobu.

relation	Powiązanie	Powiązany	zas
coverage	Zasięg	Umiejscowienie	
rights	Prawa	Informacje	o pr

Powyższa lista pochodzi z pierwszej wersji standardu (wersja 1.1), i jest szeroko stosowana. Jest to lista podstawowych haseł pozwalających na opisanie jakiegoś zasobu. Widzimy tu wpływ starszego systemu katalogowania książek (mamy osobne pole na twórcę i współtwórcę zasobu, choć często rozdzielenie tych dwóch elementów nie jest jednoznaczne). Opis może być zawsze zrobiony w sposób bardziej szczegółowy, niż pozwala na to DC, ale użycie DC pozwala na wyodrębnienie podstawowych elementów opisu w sposób pozwalający na łatwą wymiane danych.

Dublin Core jest standardem płaskim, to znaczy że przy opisie zasobu (np. książki) możemy wyodrębnić np. wydawcę, ale nie mamy struktury na szczegółowy opis tego wydawcy (np. nazwa, adres itp). W miarę rozszerzania użycia DC wiązało to ręce użytkowników, którzy mieli potrzebę lepszego sprecyzowania (uszczegółowienia) danego elementu. Drugim impulsem do rozbudowy DC było powstawanie standardów Symantic Web, co pozwalało na uściślenie tego, co można umieszczać w danych polach. Np. pole publisher (Wydawca) może odnosić sie do organizacji Wydawcy jako link, albo zawierać tylko tekst z nazwą Wydawcy. Aby sprostac tym dwóm wymaganiom, w roku 2008 opublikowana została nowa wersja standardu. Nowa wersja posiada 55 etykiet, i aby odróżnić je od oryginanych 15 elementów (elements) stosuje sie nazwę terminy (terms). Te 55 terminów zawiera w sobie oryginalne 15 elementów, choć zasady ich użycia mogą być bardziej uściślone. Terminy definiują nowe własności, ale w wielu przypadkach uszczegóławiają znaczenie elementów z oryginalnego zestawu. Kolejna tabelka przedstawia wybrane terminy które uściślają znaczenie elementów DC.

Tabela 2. Wybrane terminy (terms) Dublin Core. Namespace - ***dcterms***

Element	***Termin (uszczegó- łowienie)***

Etykieta

Opis

contributor	**creator**	Twórca
title	**alternative**	Wariant tytułu
coverage	**spatial**	Zasięg przestrzenny
temporal	Zasięg czasowy	Charakterystyka zasię
date	**available**	Data udostępnienia
created	Data utworzenia	Data utworzenia zasob
dateAccepted	Data przyjęcia	Data zatwierdzenia za
dateCopyrighted	Data copyright	Data copyright
dateSubmitted	Data złożenia	Data przedłożenia zas
issued	Data wydania	Data oficjalnego wyda
modified	Data modyfikacji	Data zmiany zasobu.
valid	Data ważności	Data (często zakres da
description	**tableOfContents**	Spis treści
abstract	Abstrakt	Streszczenie zasobu.
format	**medium**	Nośnik
identifier	**bibliographicCitation**	Cytata bibliograficzna

relation	**conformsTo**	Zgodny z
	hasFormat	Ma format	Zasób o tej samej zaw
	hasPart	Ma część	Zasób, który stanowi fi
	isPartOf	Jest częścią	Zasób, którego część f

[...]

Wszystkie tłumaczenia polskie w obu tabelkach: Przekład : Agnieszka Brachfogel (Biblioteka Narodowa), Współpraca : Dariusz Paradowski (Biblioteka Narodowa), Data wydania: 2009-09-08.

Terminy, tak jak i elementy, są unikalne w przestrzeni nazw (namespace) Dublin Core. Można więc używać *created* aby okreslic datę utworzenia zasobu, ale można także zaznaczyć, że jest to uściślenie pola *date* (np. date.created). Jak widzimy z tabelki, daty i relacje uzyskały w nowej wersji szczególnie wiele uściśleń, co odpowiada potrzebom również archiwistów.

Dublin Core poza 55 terminami posiada także definicjie klas i schematów kodowania. Pola mogą mieć zawartość tekstową (np. *abstract*) Inne moga być ograniczone do słowa ze słownika “haseł wzorcowych”, albo zdefiniowane według określonych, jednoznacznych reguł (np. daty). Oprogramowanie stosowane do opisywania zasobów posiada często odpowiednie narzędzia które ułatwiaja stosowanie tych reguł.

Początkowo Dublin Core (wersja 1.1) był bardzo prostym standardem, i zawierał prawie wyłacznie definicje elementów (sementyke) z minimalnymi wzmiankami o składni. Nowsza wersja zawiera już więcej elementow składni - podobnie jak języki naturalne staje się bardziej złożona. Aby mówić pełnym językiem Dublin Core należy nauczyć się nie tylko terminów, ale także klas i schematów, oraz szczegółów składni. Dlatego też wersja 1.1 jest ciągle dużo bardziej popularna niż nowa, i prawdopodobnie przez długi czas tak zostanie.

Jako interesujące ćwiczenie, polecałbym wzięcie najbliżej leżącej na biurku książki, i wypisanie (minimum) 10 metadanych które opisują tę książkę, i odpowiadają elementom i/lub terminom Dublin Core. W uwagach można podać, czy się to udało.

Więcej o Dublin Core

- Elementy Dublin Core 1.1
- Terminy Dublin Core
- Przewodnik po Dublin Core

Marek Zieliński

Artykuł ukazał się 20 maja 2013 w *Blogu archiwistów i bibliotekarzy Instytutu Piłsudskiego*

Może Cię też zainteresować:

- Wstęp do standardów metadanych