

Część 2

Wszystkie nowoczesne standardy zapisu informacji używają jednego języka zapisu, XML. Jest to język uniwersalny, prosty i łatwy do opanowania, a jednocześnie ma ogromną moc ekspresji. Adres Instytutu możemy w XML zapisać płasko:

```
<adres>180 Second Avenue, New York, NY</adres>
```

albo hierarchicznie:

```
<galaktyka nazwa="Droga Mleczna">
  <gwiazda nazwa="Sol">
    <planeta nazwa="Mars"/>
    <planeta nazwa="Ziemia">
      <kontynent nazwa="Ameryka Północna">
        <panstwo nazwa="USA">
          <stan nazwa="Nowy Jork"> [...] itp.
        <stan>
      <panstwo>
    <kontynent>
  <planeta>
</gwiazda>
</galaktyka>
```

EAD jest standardem (wyrażanym w XML) opracowanym dla archiwów i jest bardzo typowym przykładem opisu hierarchicznego. Jest odbiciem typowej organizacji archiwum, gdzie kolekcja (zespół archiwalny, fonds) może być podzielona na pod-zespoły (subfonds), te z kolei na serie, podserie, grupy, podgrupy itp. Często organizacja taka nie jest sprawą wyboru, gdy na przykład oryginalny twórca danej kolekcji tak ją właśnie uporządkował. Zasada szacunku dla oryginalnego twórcy kolekcji (respect de fonds) wymaga pozostawienia w miarę możliwości oryginalnej organizacji.

EAD, który definiuje zespół archiwalny jako najwyższy element w hierarchii danego repozytorium, pozwala na zdefiniowanie aż do 10 poziomów zależnych, a więc odzwierciedlenie

naprawdę drzewiastej a nawet krzaczastej hierarchii. Archiwista opisuje wtedy osobno każdy z poziomów, starając się zbalansować intencję twórcy z wymogami zdrowego rozsądku. Wielopoziomowość ułatwia 'ogarnięcie' myślą zasobu, ale wcale nie ułatwia a często wręcz utrudnia znalezienie poszukiwanej informacji, która może się znaleźć na wielu różnych gałęziach drzewa, albo na różnych poziomach opisu.

Dublin Core (DC) jest również standardem zapisu danych wyrażanym w XML. W odróżnieniu od EAD, jest systemem płaskim - nie definiuje żadnej hierarchii. O ile EAD ma około 150 etykiet, DC ma 15 podstawowych etykiet takich jak tytuł, opis, data, itp. co bardzo upraszcza opis dowolnego zasobu. Praca nad standardem Dublin Core rozpoczęła się w roku 1995, jest to 'stary' i powszechnie używany standard. Z czasem okazało się, że często można uściślić znaczenie jakiejś etykiety, poprzez wprowadzenie dodatkowego zawężenia (kwalifikatora). Przykładowo etykieta 'coverage' - obszar pokrycia, może być (w DC) uściślona przez użycie kwalifikatora 'spatial' - przestrzenny, i 'temporal' - czasowy. Łatwo wtedy przypisać danemu dokumentowi miejsca i daty używając tych dwóch kwalifikatorów. Mimo tego DC jest systemem płaskim - można przypisać jedno lub więcej miejsc, np. spatial = "180 Second Avenue, New York, NY", ale nie hierarchiczną strukturę informacji o takim miejscu.

Postęp w wyszukiwarkach czy w ogólności w dziale AI (sztucznej inteligencji, choć teraz rzadko używa się tego terminu) pozwala na znalezienie informacji nawet wtedy, kiedy nie jest ona rygorystycznie opisana. Każdy z nas używa wyszukiwarki Google i innych, choć może nie każdy zauważył niesamowite zwiększenie trafialności poszukiwania. Google używa wskazówek z kontekstu (co niektórzy mają tej firmie za złe) aby w dramatyczny sposób zwiększyć dokładność trafienia. Użycie prostych etykiet daje przeszukiwarkom dodatkową informację, zwiększając szanse na trafienie na poszukiwaną informację.

Którego standardu używamy w archiwum Instytutu? Zarówno EAD jak i DC znalazły swoje miejsce. Hierarchiczny EAD jest używany do autorytatywnego (w miarę możliwości) opisu zasobu archiwalnego. Na naszej stronie możecie zobaczyć [dwa poziomy](#) takiego opisu (źródła EAD nie są jeszcze dostępne, właśnie pracujemy nad konwersją). Przy digitalizacji zasobów używamy standardu Dublin Core. Najważniejszym naszym celem jest tu znajdywalność informacji, i użycie płaskiego standardu DC ułatwia bardzo zarówno pracę nad archiwami jak i ich prezentację. Zastosowanie etykiety

Przestrzenny obszar pokrycia

można wypróbować w naszej

[wystawie online](#)

, gdzie udostępnione są dokumenty z ośmiu zespołów archiwalnych Instytutu.

Standardy metadanych dla archiwów: płaskie czy hierarchiczne? (Cz. 2)

Wpisany przez Marek Zieliński

poniedziałek, 23 lutego 2015 00:00 - Poprawiony poniedziałek, 02 marca 2015 16:10

Marek Zieliński

Artykuł ukazał się 22 sierpnia 2012 w *Blogu archiwistów i bibliotekarzy Instytutu Piłsudskiego*

Może Cię też zainteresować:

- [Wstęp do standardów metadanych](#)
- [Standardy metadanych: Dublin Core](#)
- [Standardy metadanych: EAD](#)
- [Oprogramowanie dla archiwistów](#)