

Metadane

Agnieszka Ławrynowicz
Politechnika Poznańska

Metadane

Metadane = dane o danych?

Metadane to dane

Gdzie przeprowadzić linię podziału (dane-metadane)?

Zależy od kontekstu

Opis paczki



Metadane to opis

- opis (*description*): informacja na jakiś temat
- opisując, tworzysz zdania na temat czegoś (*subject-predicate-object*)

Metadane - definicja

Metadane = dane, które opisują inne dane

Metadane w YouTube

The screenshot shows a YouTube video player for 'Wiedźmin odcinek 4 Smok.avi'. The video is from the channel 'Kanał użytkownika MwrotekM' and has 200,881 views. The video title is 'Wiedźmin odcinek 4 Smok.avi'. The description mentions it's a serial based on Andrzej Sapkowski's stories, about Geralt's parents. The category is 'Film & Animation' and the license is 'Standard YouTube License'. The video is uploaded on Nov 19, 2011. To the right of the video player, there are several suggested videos, including 'Wiedźmin Odcinek 5/13', 'Wiedźmin Odc06 Calanthe', 'Wiedźmin 2 - najlepsze teksty', 'Wiedźmin Odc12 Falwick', 'Wiedźmin Odcinek 2/13', 'Wiedźmin Odc07 Dolina kwiatów', and 'Wiedźmin Odcinek 3/13'. Above the suggested videos, there is a section titled 'What was wrong with this ad?' with options: Inappropriate, Repetitive, and Irrelevant.

Wiedźmin odcinek 4 Smok.avi

Kanał użytkownika MwrotekM

200,881

Uploaded on Nov 19, 2011

Odcinek 4 : Smok

Serial na podstawie opowiadań Andrzej Sapkowskiego. Pewnego dnia do domu rodziców Geralta przyjeżdża stary wiedźmin. Powołując się na prawo niespodzianki zabiera chłopca do górskiej twierdzy - Kaer Morhen, gdzie Geralt ma zostać poddany mutacjom, magicznym próbom i szkoleniu, które przekształca go w wiedźmina, najemnego zabójcę potworów.

Category: Film & Animation

License: Standard YouTube License

SHOW LESS

ALL COMMENTS (433)

What was wrong with this ad?

☐ Inappropriate

☐ Repetitive

☐ Irrelevant

Google

Wiedźmin Odcinek 5/13

by Monodrag

175,454 views

Wiedźmin Odc06 Calanthe

by francuz01

33,396 views

Wiedźmin 2 - najlepsze teksty

by ProjektPierscion

26,857 views

Wiedźmin Odc12 Falwick

by francuz01

34,742 views

Wiedźmin Odcinek 2/13

by Monodrag

67,524 views

Wiedźmin Odc07 Dolina kwiatów

by francuz01

60,827 views

Wiedźmin Odcinek 3/13

by Monodrag

- metadane obiektu z portalu (statyczne),
- metadane utworzone przez użytkowników
- metadane (powiązane video) wygenerowane automatycznie przez algorytmy

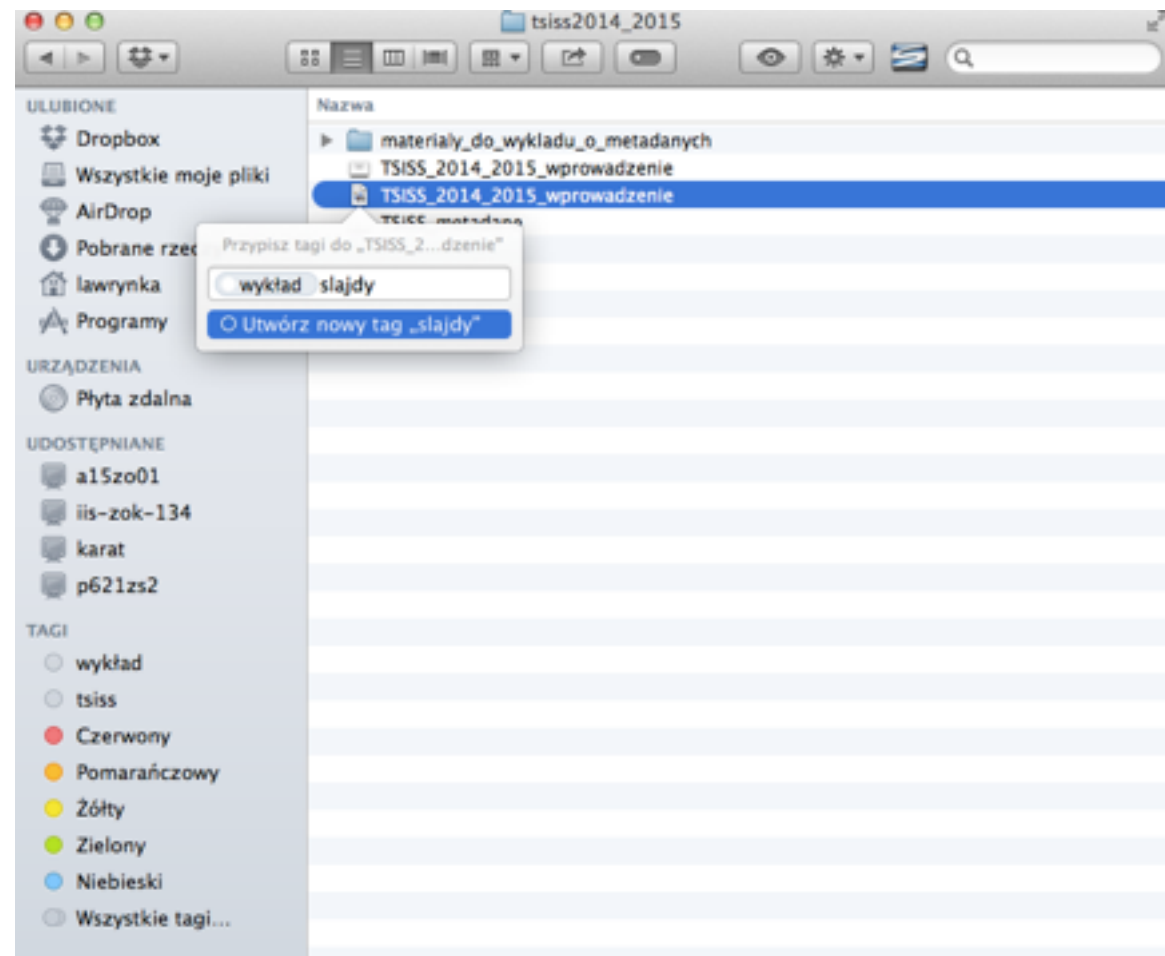
Znaczniki (tagi)

- metadane opisowe
- tagi Web 2.0, hashtagi itp.
- patrz: [https://en.wikipedia.org/wiki/Tag_\(metadata\)](https://en.wikipedia.org/wiki/Tag_(metadata))



AgnieszkaLawrynowicz @agalawrynowicz · 59 min

Dobry kurs nt. metadanych z punktu widzenia informatologii (information science) coursera.org/course/metadata #tsiss #metadata #Coursera



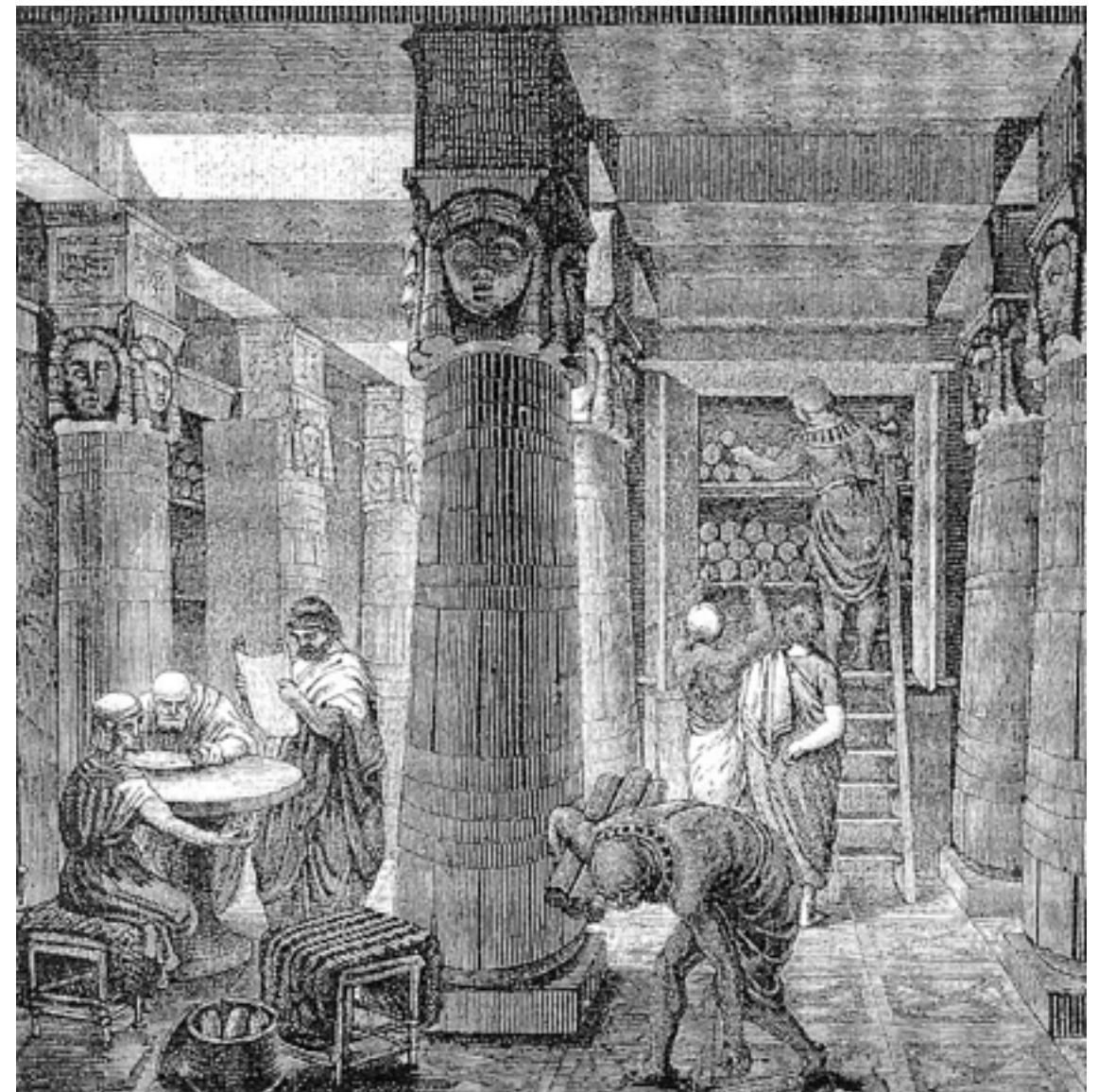
Rodzaje metadanych

- Administracyjne
- Opisowe
- Dostępowe
- Archiwizacyjne
- Techniczne/Strukturalne

Bibliotekoznawstwo

starożytna Biblioteka Aleksandryjska

- 0,5 mln zwojów
- punkty dostępowe
- tytuł, autor, temat
- prosty opis jest efektywny



Analiza tematu

- informatologia (*information science*)
- analiza tematu (*subject analysis*): ustalenie co jest tematem opisywanej rzeczy/zasobu („o czym” jest zasób)
- znaczące charakterystyki
- bez tematu: np. niektóra muzyka
- „*aboutness*” (mniej jednoznaczne niż temat)

Tezaurusy

- tezaurus ~ kontrolowane słownictwo
- lista słów + relacje pomiędzy słowami (synonimy, antonimy)

„The nice thing about standards is that you have so many to choose from”

Andrew Stuart "Andy" Tanenbaum
Computer Networks, 2nd ed., p. 254

“Standards are like toothbrushes... Everyone agrees that they’re a good idea but nobody wants to use anyone else’s.”

Murtha Baca

Library of Congress Subject Headings (LCSH)

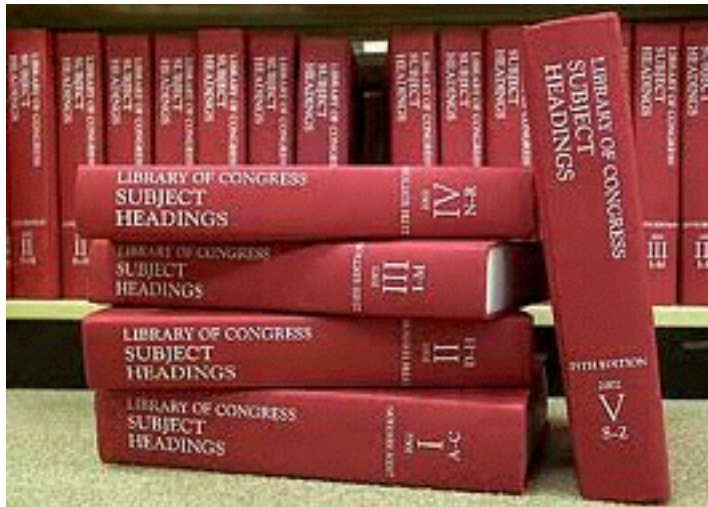
Biblioteka Kongresu Stanów Zjednoczonych

największa biblioteka świata
ponad 142 mln dokumentów
rok założenia: 1800

- LCSH dostarcza tezaurusa - kontrolowanego słownictwa
- LCSH dostarcza struktury do rozszerzania dostępnego zbioru terminów



LCSH



- LCHS: kompletny tezaurus
- uwaga: słowniki są przeważnie wycinkiem języka (niekompletne) i NIE są słownictwem kontrolowanym
- *subject heading, index term, descriptor*

Personal name [Sapkowski, Andrzej, author.](#)
Uniform title [Chrzest ognia, English](#)
Main title Baptism of fire / Andrzej Sapkowski ; translated by David French.
Edition First U.S. edition.
Published/Produced New York, NY : Orbit, 2014.

Description
 389 pages ; 21 cm.

ISBN
 9780316219181 (pbk.)
 0316219185 (pbk.)

LC classification (full)
 PG7178.A65 C4813 2014

LC classification (partial)
[PG7178.A65](#)

Related names
[French, David, translator.](#)

Summary
 The Wizards Guild has been shattered by a coup and, in the uproar, Geralt was seriously injured. The Witcher is supposed to be a guardian of the innocent, a protector of those in need, a defender against powerful and dangerous monsters that prey on men in dark times. But now that dark times have fallen upon the world, Geralt is helpless until he has recovered from his injuries. While war rages across all of the lands, the future of magic is under threat and those sorcerers who survive are determined to protect it. It's an impossible situation in which to find one girl--Ciri, the heiress to the throne of Cintra, has vanished--until a rumor places her in the Nifigaard court, preparing to marry the Emperor. Injured or not, Geralt has a rescue mission on his hands.

Subjects
[Monsters--Fiction.](#)
[Magic--Fiction.](#)
[Heiresses--Fiction.](#)

Form/Genre
[Fantasy fiction.](#)

Series
 The Witcher ; 3

Dublin Core



- DCMI: międzynarodowa społeczność zainteresowana interoperacyjnymi metadanymi: bibliotekoznawcy, informatycy, muzea, instytucje naukowe itd.
- „Dublin”: odnosi się do Dublina w Ohio (USA): początki 1995 OCLC (Online Computer Library Center)/NCSA (National Center for Supercomputing Applications) Metadata Workshop
- „Core”: odnosi się do terminów metadanych ”broad and generic being usable for describing a wide range of resources”



Dublin Core

- wszystkie element opcjonalne
- wszystkie elementy powtarzalne
- elementy mogą być wyświetlane w dowolnej kolejności
- rozszerzalne

15 prostych elementów

Creator	Title (+wariant tytułu)	Subject
Contributor	Date (utworz., wyd. mod.)	Description (np. spis treści)
Publisher	Type	Format (rozmiar, nośnik)
Coverage (czas, miejsce)	Rights	Relation (np. ma wersję)
Source	Language	Identifier

Przykład



The screenshot displays the website of the Wielkopolska Biblioteka Cyfrowa (Greater Poland Digital Library). The header features the library's logo and name, along with a navigation bar containing links to the home page, collections, new account, login, and contact. A sidebar on the left provides options for viewing the publication, including description, information, content, and download. The main content area shows the details of a specific publication, 'Heraldyka polska wieków średnich' by Franciszek Piekosiński. The details include the title, author, subject, and keywords, all in Polish. A 'Pokaż treść!' (Show content!) button is prominently displayed next to the title. The bottom of the page includes a section for exporting metadata and a list of favorite positions.

Wielkopolska Biblioteka Cyfrowa

obecnie czytających: 380 STRONA GŁÓWNA KOLEKCJE NOWE KONTO LOGOWANIE KONTAKT

Wydanie

- Opis
- Informacje
- Treść
- Treść (nowe okno)
- Pobierz
- Podobne wydania
- Opcje wyświetlania

Zobacz także

- Rocznik Towarzystw...
- Heraldyka polska w...
- Heraldyka polska w...
- Rocznik Polskiego ...
- Insignia seu Cleno...

Dane dostarcza:

FEDERACJA BIBLIOTEK CYFROWYCH

Powered by Europeana

Eksport metadanych

OAI-PMH RDF

RIS BIBTEX

Ulubione pozycje

Opis wydania

Heraldyka polska wieków średnich

Tytuł:
Heraldyka polska wieków średnich

Twórca:
Piekosiński Franciszek (1844-1906)

Temat i słowa kluczowe:
heraldyka średniowiecze ; heraldyka polska 16w. ; herby średniowieczne ; heraldyka polska 15w. ; heraldyka polska 14w.

Opis:
Estr. III, 370

Wydawca:
Akademia Umiejętności

Miejsce wydania:
Kraków

Typ obiektu:
monografie heraldyczne

Identyfikator:
oai:www.wbc.poznan.pl:1355

Język:
pol

Prawa do dysponowania publikacją:
Biblioteka Kórnicka PAN

Data wydania:
1889

Właściciel praw:
PAN Biblioteka Kórnicka

Pokaż treść!

Dodatkowe kwalifikatory

- opierają się na 15 elementach Dublin Core i dokładają dalsze poziomy szczegółośc
- dwa typy:
 - uszczegółowienie elementu
 - schemat kodowania

Uszczegółowienie elementu

Zawęża znaczenie elementu DC

- np. *medium* uszczegóławia *format*
- np. *bibliographicCitation* uszczegóławia *identifier*

Rozszerzenia są własnościami jak inne własności

- `<dateCopyrighted>2014-08-12</dateCopyrighted>`
a *NIE*
- ~~`<date.dateCopyrighted>2014-08-12</date.dateCopyrighted>`~~

Schemat kodowania

- słownictwo
 - wskazanie, że wartość pochodzi z kontrolowanego słownika (np. LCSH)
- składnia
 - łańcuchy znakowe formatowane w standardowy sposób (np. format daty zgodny z ISO 8601)

Zasady

- *Dumb-Down*
- *One-to-One*

Zasada *dumb-down*

- Prosty DC nie używa uszczegółowień elementów ani schematów kodowania; wyrażenia zawierają jedynie łańcuchy znakowe
- Kwalifikowany DC używa ich i innych cech abstrakcyjnego modelu DCMI
- Translacja złożonych (kwalifikowanych) opisów (własności i wartości) do prostego DC

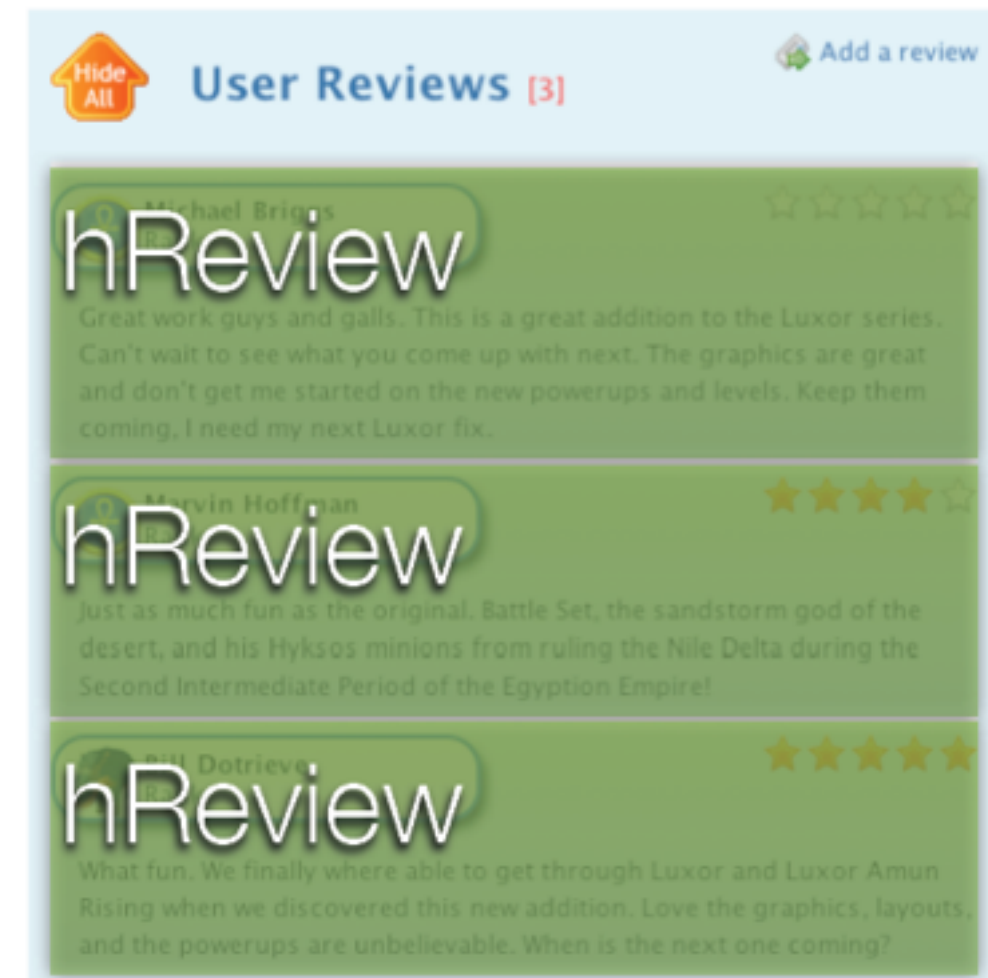
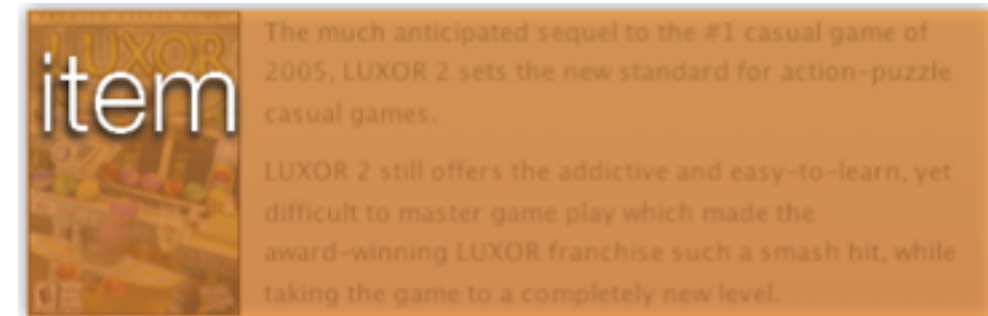
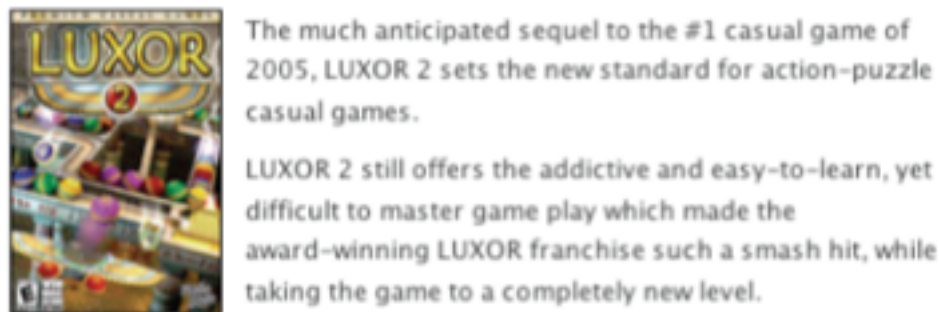
Zasada One-to-One

- Twórz jeden opis (*description*) metadanych dla jednego i tylko jednego zasobu
 - nie opisuj zdigitalizowanego obrazu tak jakby to był oryginalny obraz
 - nie opisuj obydwu obrazu i malarza w tym samym opisie
 - opisz kompozytora i jego pracę w dwóch osobnych opisach
- Zgrupuj powiązane opisy w zbiór opisów, rekord (*description set*)

Co to są mikroformaty?

- sposób nadania znaczenia elementom HTML i jawnego pokazania struktur danych na stronach HTML
- zaprojektowane dla ludzi w pierwszej kolejności, w drugiej dla maszyn
- zbiór prostych, otwartych formatów danych, zbudowanych w oparciu o **istniejące i szeroko zaadaptowane standardy** (np. (X)HTML)
- rozwiązują pojedynczy, specyficzny problem (np. reprezentację informacji geograficznej, kalendarzowej)

Ilustracja mikroformatów



Rodzaje mikroformatów

- **Elementarne mikroformaty** (jeden znacznik)
 - Rel-home (strona domowa) `<link href="http://technorati.com" rel="home" />`
 - Rel-License (licencja) `<a href="http://creativecommons.org/licenses/by/2.0/" rel="license">cc by2.0</a>`
 - Inne: rel-tag, rel-enclose, xfn-tags
- **Złożone mikroformaty**
 - Często oparte na istniejącym standardzie
 - np. hCard, hCalendar, hEvent, hReview

RDFa

RDFa = RDF w atrybutach

- rekomendacja W3C
- zbiór nowych atrybutów (X)HTML do wyrażenia metadanych wewnątrz (X)HTML
- format serializacji RDF, gdzie trójki RDF są "osadzone" w (X)HTML
- niezależne od dziedziny (w przeciwieństwie do dedykowanych dla danej dziedziny mikroformatów)

Składnia: wykorzystanie RDFa w XHTML

- **Odpowienie atrybuty XHTML:** @rel, @rev, @content, @href, @src
- **Nowe atrybuty, specyficzne dla RDFa:** @about, @property, @resource, @datatype, @typeof

Opis podstawowych atrybutów RDFa

- **@about**, **@src** – URI, które specyfikuje zasób opisywany przez metadane
- **@rel**, **@rev** – określają relację (relację odwrotną) z innym zasobem
- **@href**, **@resource** – określają stowarzyszony zasób
- **@property** – określa własność dla zawartości ('content') elementu
- **@content** – opcjonalny atrybut, który nadpisuje zawartość elementu, używając atrybutu 'property'
- **@datatype** – opcjonalny atrybut, który określa typ danych tekstu do wykorzystania z atrybutem 'property'
- **@typeof** – opcjonalny atrybut, który określa typ(y) RDF podmiotu (zasobu opisywanego przez metadane)

Mikrodane

- Mikrodane = specyfikacja grupy WHATWG, wykorzystywana do zagnieżdżania metadanych w istniejącej treści stron Web
- wyszukiwarki internetowe, crawlery, przeglądarki mogą ekstrahować i przetwarzać mikrodane żeby wzbogacić usługi przeglądarkowe

Model mikrodanych

- Model Mikrodanych składa się z grup par nazwa-wartość
- *Item* - każda z poszczególnych grup
- Każdy *item* może posiadać typ, globalny identyfikator oraz listę par nazwa-wartość
- Każda nazwa na tej liście zwana jest właściwością (ang. property) i posiada jedną lub więcej wartości
- Każda wartość jest łańcuchem znaków lub grupą par nazwa-wartość (*item*)

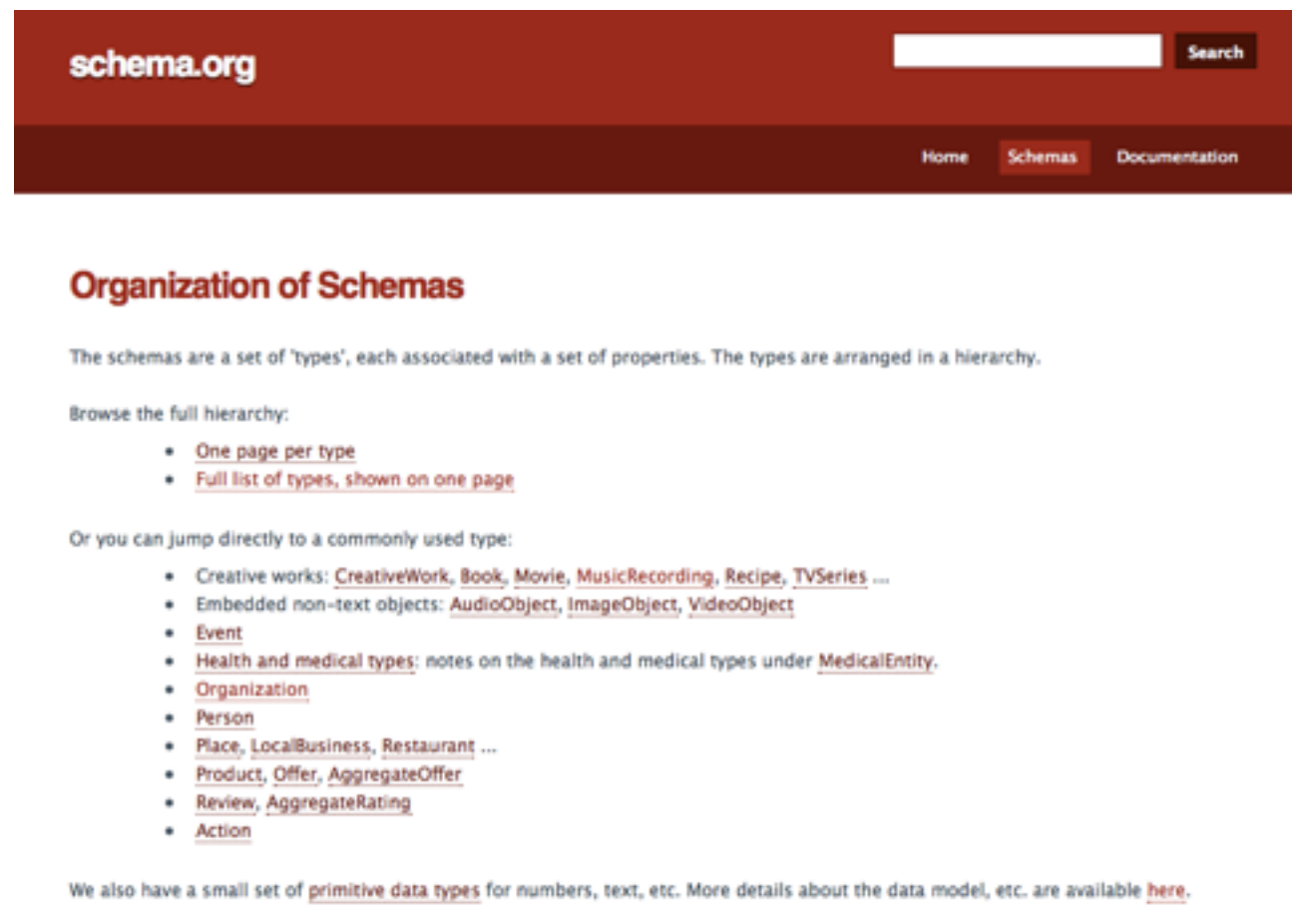
Atrybuty globalne

- **itemscope** – tworzy *Item* i wskazuje, że elementy podrzędne tego elementu, zawierają informację o nim
- **itemtype** – na wartość składa się nieuporządkowany zbiór unikalnych tokenów reprezentujących poprawny URL słownika, który opisuje element i jego własności
- **itemid** – unikalny identyfikator obiektu
- **itemprop** – wskazuje, że znacznik zawiera wartość danej własności; nazwa własności oraz kontekst wartości opisane są w słowniku *itemu*
- **itemref** – zawiera listę identyfikatorów elementów HTML (a nie obiektów *item*), które nie są potomkami danego obiektu item, by wskazać dodatkowe elementy do przeszukania w celu znalezienia par nazwa-wartość obiektu item

Słowniki

- Semantyka dostarczana za pomocą słowników Mikrodanych
- Możliwe zarówno stworzenie własnego słownika, jak i użycie gotowego (dostępnego w sieci)
- Najczęściej używane typy słownikowe udostępnia [Schema.org](https://schema.org) (m.in. Person, Event, Organization, Product, Review)

schema.org



- inicjatywa (od 2011) Bing, Google i Yahoo! (operatorów największych wyszukiwarek internetowych)
- cel: tworzenie i wspieranie zbioru schematów do oznaczania danych strukturalnych na stronach Web
- proponuje używać ontologii ze schema.org i mikrodanych w HTML5 aby oznaczyć metadanymi treść strony

Przykład

ze strony schema.org:

```
<div itemscope itemtype ="http://schema.org/Movie">  
  <h1 itemprop="name"&g;Avatar</h1>  
  <div itemprop="director" itemscope itemtype="http://  
schema.org/Person">  
    Director: <span itemprop="name">James Cameron</  
span> (born <span itemprop="birthDate">August 16,  
1954)</span>  
  </div>  
  <span itemprop="genre">Science fiction</span>  
  <a href=" ../movies/avatar-theatrical-trailer.html"  
itemprop="trailer">Trailer</a>  
</div>
```

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)



Cross-walks

- „mapowania” z jednego schematu na drugi
- Przykład (z Wikipedii):

MARC field		Dublin Core element
260\$c (Date of publication, distribution, etc.)	→	Date.Created
522 (Geographic Coverage Note)	→	Coverage.Spatial
300\$a (Physical Description)	→	Format.Extent

