Dublin Core

Χρήστος Παπαθεοδώρου

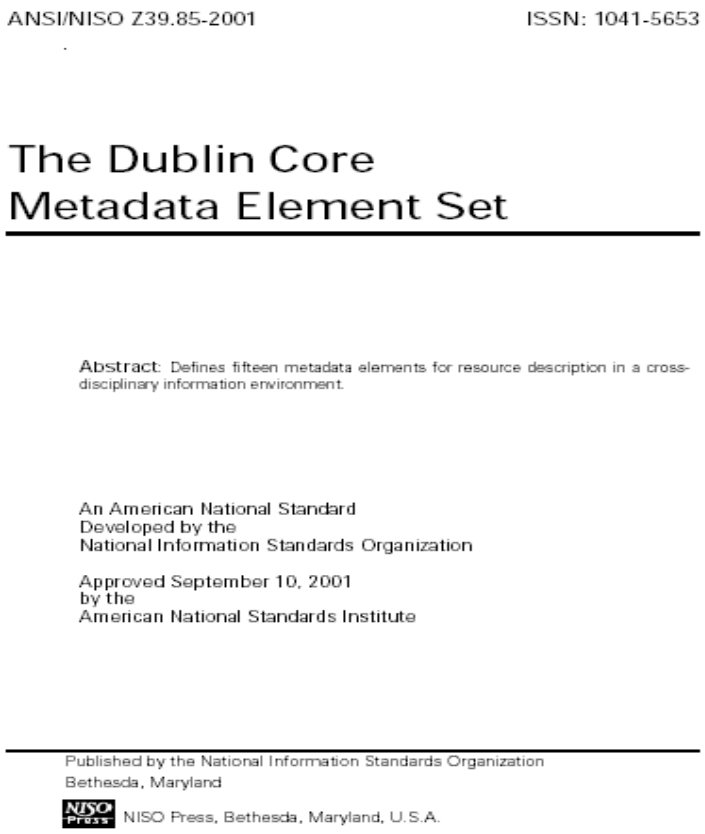
Καθηγητής

Ομάδα Βάσεων Δεδομένων και Πληροφοριακών Συστημάτων,
Τμήμα Αρχειονομίας – Βιβλιοθηκονομίας, Ιόνιο Πανεπιστήμιο

(papatheodor@ionio.gr)

DBIS
database & information systems group
ionian university

Το Πρότυπο «Dublin Core»



ANSI/NISO Z39.85-2001 ISSN: 1041-5653

The Dublin Core
Metadata Element Set

Abstract: Defines fifteen metadata elements for resource description in a cross-disciplinary information environment.

An American National Standard
Developed by the
National Information Standards Organization

Approved September 10, 2001
by the
American National Standards Institute

Published by the National Information Standards Organization
Bethesda, Maryland

NISO Press, Bethesda, Maryland, U.S.A.

- Προτυποποίηση
  - ISO 15836-2003
  - US: NISO Z39.85-2001
  - Ευρώπη: αναγνώριση από το CEN/ISSS Workshop Agreement 13874-2000
- Συστήνεται ανταλλαγή με RDF/XML

Χαρακτηριστικά του «Dublin Core»

- «Dublin Core» σημαίνει «Dublin, Ohio» / OCLC
- Πρωτοβουλία για να βελτιώσει την ανακάλυψη πόρων στο Διαδίκτυο
- Κοινός παρονομαστής για επικοινωνία / διαλειτουργικότητα (και όχι για λεπτομερή περιγραφή πόρων)
- Ένα υσμα για σύγκλιση των προτύπων
- Επεκτάσιμο, για να καλύψει τις επιπρόσθετες ανάγκες ανακάλυψης πόρων των διαφορετικών εφαρμογών / περιοχών

Dublin Core – Simple

- Το Dublin Core έχει 15 στοιχεία. Κάθε ένα από αυτά είναι προαιρετικό και επαναλαμβανόμενο
- Τα 15 στοιχεία χωρίζονται σε 3 κατηγορίες:
  - Περιεχόμενο:
    - Περιγράφουν το αντικείμενο
  - Πνευματική Ιδιοκτησία:
    - Περιγράφουν το copyright και τη δημιουργία
  - Στιγμιότυπο:
    - Περιγράφουν την εισαγωγή και διαχείριση

DC – Περιεχόμενο

- Τίτλος / **Title** – (ονομασία πηγής)
- Θέμα / **Subject**, π.χ. λέξεις-κλειδιά, ταξινομικοί κωδικοί
- Περιγραφή / **Description**
  - Π.χ. περίληψη, περιεχόμενα, περιγραφή εικόνας
- Πηγή (ή «Προέλευση) / **Source** – (παραγωγής)
- Γλώσσα / **Language** – (του περιεχομένου)
- Σχέση / **Relation** – (αναφορά σε σχετική πηγή)
  - Π.χ. έκδοση του …
- Κάλυψη / **Coverage** – (γεωγραφική ή χρονική)

DC – Πνευματική Ιδιοκτησία

- Δημιουργός / **Creator** – (πρόσωπο, οργανισμός, υπηρεσία)
- Εκδότης / **Publisher** – (πρόσωπο, οργανισμός, υπηρεσία)
- Συντελεστής (ή «Συνεργάτης» ή «Υπεύθυνος συμβολής») / **Contributor** – (πρόσωπο, οργανισμός, υπηρεσία που συμβάλλει στο περιεχόμενο)
  - Π.χ. μεταφραστής, εικονογράφος, κριτής
- Δικαιώματα / **Rights** – (κείμενο σχετικά με την πνευματική ιδιοκτησία)

DC – Στιγμιότυπο

- Ημερομηνία / **Date**
  - Π.χ. δημιουργίας, έκδοσης, μετάφρασης, πρόσκτησης, καταλογογράφησης, …
- Τύπος / **Type** – (κατηγορία, σχετικά με το περιεχόμενο)
  - Π.χ. ποίημα, λεξικό, software, home-page
- Μορφότυπο / **Format** – (φυσική ή ψηφιακή μορφή)
  - Π.χ. Macintosh-software, pdf, html, διαστάσεις, διάρκεια
- Αναγνωριστικό (ή Προσδιοριστής ή Κωδικός Ταύτισης) / **Identifier**
  - Μοναδικό προσδιοριστικό, π.χ. URL, ISBN, …

Στόχοι του Dublin Core

- Απλότητα δημιουργίας και διατήρησης
  - Μη ειδικοί να δημιουργούν περιγραφικές εγγραφές για αποτελεσματική ανάκτηση σε δικτυωμένο περιβάλλον
- Κοινά κατανοητή (διαθεματική) σημασιολογία
  - Σύγκλιση κοινών, γενικών στοιχείων
  - Αυξημένη ορατότητα και προσβασιμότητα
  - Κατάλληλο και για τον μη ειδικό της αναζήτησης
    - Τον «ψηφιακό τουρίστα»

Χρήση του Dublin Core

- Είναι βασικός **πυρήνας** στοιχείων
- ***Δεν*** είναι υποκατάστατο σε πλουσιότερα περιγραφικά πρότυπα
- Παρέχει 15 «παράθυρα» ή ευρύχωρα «καλάθια» από πλουσιότερη περιγραφή πόρων
  - Φανερώνει πλούσιες περιγραφές σε απλή μορφή
  - Σημασιολογικά σταυροδρόμια, αντιστοιχίσεις σε υπάρχοντα δεδομένα

Τα Στοιχεία Μεταδεδομένων του Dublin Core

- Διεπιστημονική ομοφωνία σε απλά σύνολα στοιχείων για ανακάλυψη πόρων
  - 15 στοιχεία (πεδία μεταδεδομένων)
  - όλα προαιρετικά
  - όλα επαναλαμβανόμενα
- Δεν προορίζεται για περιγραφή περίπλοκων πόρων
  - Η αρχική ιδέα των «απλών αντικειμένων – σαν τεκμήρια»
  - Απλότητα στη σημασιολογία, ευκολία χρήσης
- Παρέχει βασική «σημασιολογική διαλειτουργικότητα»
  - Μεταξύ επιστημονικών περιοχών, μεταξύ γλωσσών
  - Δεν παρέχει λεπτομερείς κανόνες καταλογογράφησης
- Επιτρέπει επεκτασιμότητα – σε άλλες κατηγορίες πόρων

Dublin Core Qualified

DBIS
database & information systems group
ionian university

Dublin Core – Qualified

Ως επέκταση του απλού (simple ή unqualified) Dublin Core έχουμε το εξειδικευτικό (qualified) Dublin Core, που προσφέρει:

Βελτίωση της σημασιολογικής ακρίβειας του Dublin Core ορίζοντας τους **εξειδικευτές** (qualifiers)

- Εξειδίκευση του στοιχείου (element refinement)
  - Για περισσότερη λεπτομέρεια στην περιγραφή
- Σύστημα κωδικοποίησης (encoding scheme)
  - Σαν χρήση απλών κανόνων καταλογογράφησης

Ποικιλίες Εξειδικευτών: Εξειδίκευση Στοιχείων

- Προσφέρει μεγαλύτερη **λεπτομέρεια** σε αυτόν που τη χρειάζεται
- Κάνει την σημασία ενός στοιχείου στενότερη ή πιο ειδική
  - «Date *Created*» και «Date *Modified*»
  - «*IsReplacedBy* Relation» και «*Replaces* Relation»
  - Δεν την τροποποιεί ή επεκτείνει, σε καμία περίπτωση
- Αν το λογισμικό δεν καταλαβαίνει κάποιο εξειδικευτή, μπορεί με ασφάλεια να τον αγνοήσει!

Dublin Core: μη Εξειδίκευση Στοιχείων

- Creator
- Subject
- Publisher
- Contributor
- Type
- Identifier
- Source
- Language
- Rights

Dublin Core: Εξειδίκευση Στοιχείων (1/2)

Στοιχείο	Εξειδίκευση
Title	Alternative
Description	Table Of Contents Abstract
Date	Created Valid Available Issued Modified
Format	Extent Medium
Coverage	Spatial Temporal

Dublin Core: Εξειδίκευση Στοιχείων (2/2)

Στοιχείο	Εξειδίκευση
Relation	Is Version Of Has Version Is Replaced By Replaces Is Required By Requires Is Part Of Has Part Is Referenced By References Is Format Of Has Format

Κωδικοποίηση Τιμών (encoding schemes)

- Δηλώνει ότι μια τιμή είναι
  - Ένας όρος από ελεγχόμενο λεξικό (π.χ., *Library of Congress Subject Headings*)
  - Χαρακτήρες μορφοποιημένοι με συγκεκριμένο τρόπο (π.χ., «2003-05-02» σημαίνει «2 Μαΐου», όχι «5 Φεβρουαρίου»)
- Ακόμα και αν ένα σχήμα δεν είναι γνωστό από τον ερμηνευτή (λογισμικό), η τιμή πρέπει να είναι «κατάλληλη» και χρησιμοποιήσιμη για ανακάλυψη πόρων

Dublin Core: μη Κωδικοποίηση Τιμών

- Title
- Creator
- Description
- Publisher
- Contributor
- Rights

Κωδικοποίηση Τιμών (1/2)

Στοιχείο	Σχήμα κωδικοποίησης
Date	W3C-DTF DCMI
Type	DCMI Type Vocabulary
Format	IMT
Identifier	URI
Source	URI
Language	ISO 639-2 RFC 1766
Relation	URI

Κωδικοποίηση Τιμών (2/2)

Στοιχείο	Σχήμα κωδικοποίησης
Subject	LCSH MeSH DDC UDC LCC
Coverage	DCMI Point ISO 3166 DCMI Box DCMI Period W3C-DTF

Κωδικοποίησης Τιμών του «Type»

- Text (βιβλία, άρθρα, email, fax, …)
- Image (εικόνες, κινούμενες εικόνες, διαγράμματα, …)
- Sound (μουσικό CD, αρχείο μουσικής, ομιλία, …)
- Interactive Resource (υπηρεσίες συζήτησης, …)
- Service (τραπεζικές υπηρεσίες, υπηρεσία Z3950, …)
- Software (λογισμικό υπολογιστή)
- Dataset (πίνακες, βάσεις δεδομένων, …)
- Event (συνάντηση, έκθεση, συνέδριο, δίκη, γιορτή, …)
- Collection (συλλογή τεκμηρίων)

Παραδείγματα Dublin Core

DBIS
database & information systems group
ionian university

Κωδικοποίηση Dublin Core σε HTML

- Οι κανόνες σύνταξης των μεταδεδομένων υπαγορεύονται από τις γλώσσες κωδικοποίησης: HTML, XML, …
- Το πρόθεμα DC πριν από ένα στοιχείο δεδομένων προσδιορίζει ότι αυτό ορίζεται με το πρότυπο Dublin Core
- Γενικά, η σύνταξη είναι ως εξής:

<meta name = “PREFIX.ELEMENT_NAME” lang=“LANG”
scheme=“SCHEME” content = “ELEMENT_VALUE”>

- Για παράδειγμα:

<meta name = “DC.Title” content = “Το λίγο του κόσμου”>
<meta name = “DC.Creator” content = “Δημουλά, Κική”>

Παράδειγμα, HTML με DC

```
<html>
<head>
<title>Το λίγο του Κόσμου </title>
<meta name = “DC.Title”  content = “Το λίγο του κόσμου”>
<meta name = “DC.Creator” content = “Δημουλά, Κική”>
<meta name = “DC.Publisher” content = “Στιγμή”>
<meta name = “DC.Date.Issued” content = “1990”>
<meta name = “DC.Language” scheme = “ISO639-2” content = “gre”>
</head>
<body><pre>
    …Υπήρξα περίεργη και μελετηρή. …
</pre></body>
</html>
```

Seated woman - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://www.tased.edu.au/tasimg/allport/12406045.htm What's Related

STATE LIBRARY of tasmania

Images from the Heritage Collections

Click on the image to see a larger version
Search the Heritage Collections image database

Creator : Bock, Thomas, 1793-1855.
Title : Seated woman
Location : Allport Library and Museum of Fine Arts
Description : 1 drawing : pencil, pen and ink on paper ; 13 x 12 cm. (irregular).
Notes : Attributed to Thomas Bock. Unsigned and undated. Title inscribed in pencil lower right: Seated woman Four sketches one woman and three men. Recto: Seated woman ; verso: Four sketches one woman and three men. Exact measurements 123 x 116 mm. (irregular). Condition on accession: Dirt, stained. Two corners creased, one corner torn.
Publisher/Date :
Subjects :
Women -- Portraits.
Men -- Portraits.

Reading file...

- <HTML><HEAD><TITLE>Seated woman</TITLE>
- <META NAME="description" CONTENT="Seated woman by Bock, Thomas, 1793-1855.">
- <META NAME="keywords" CONTENT="Women -- Portraits. Men -- Portraits.">
- <META NAME="DC.Creator" SCHEME="LCNA" CONTENT="Bock, Thomas, 1793-1855.">
- <META NAME="DC.Title" CONTENT="Seated woman">
- <META NAME="DC.Subject" SCHEME="LCSH" CONTENT="Women -- Portraits.">
- <META NAME="DC.Subject" SCHEME="LCSH" CONTENT="Men -- Portraits.">
- <META NAME="DC.Description" CONTENT="Image of item held in Allport Library and Museum of Fine Arts, State Library of Tasmania. 1 drawing : pencil, pen and ink on paper ; 13 x 12 cm. (irregular).">
- <META NAME="DC.Publisher" CONTENT="State Library of Tasmania">
- <META NAME="DC.Type" CONTENT="text/image">
- <META NAME="DC.Format" CONTENT="text/html">
- <META NAME="DC.Format" CONTENT="image/jpeg">
- <META NAME="DC.Date" SCHEME="ISO8601" CONTENT="2000-03-08">
- <META NAME="DC.Identifier" SCHEME="URL" CONTENT="http://www.tased.edu.au/tasimg/allport/12406045.htm">
- <META NAME="DC.Language" SCHEME="RFC 1766" CONTENT="en">
- <META NAME="DC.Relation" CONTENT="IsFormatOf Item held in Allport Library and Museum of Fine Arts, State Library of Tasmania">
- <META NAME="DC.Coverage" CONTENT="Tasmania">
- <META NAME="DC.Rights" CONTENT="http://www.tased.edu.au/tasimg/copyrt.htm">
- <META NAME="DC.identifier.URL.thumbnail" CONTENT="http://www.tased.edu.au/tasimg/allport/tn150/12406045.jpg">

Dublin Core σε XML

S. Abiteboul, P. Buneman, D. Suciu "*Data on the Web: From Relations to Semistructured Data and XML*" Morgan Kaufmann Publishers, 2000.

```
<?xml version="1.0" encoding="UTF-8"?>
<book
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:isbn ="http://www.isbn.org/def">
        <dc:title> Data on the Web: From Relations to Semistructured Data and XML</dc:title>
        <dc:creator> S. Abiteboul</dc:creator>
        <dc:creator> P. Buneman</dc:creator>
        <dc:creator> D. Suciu</dc:creator>
        <dc:publisher>Morgan Kaufmann Publishers <dc:publisher>
        <dc:date>2000</dc:date>
        <dc:type>Text </dc:type>
        <isbn:number>0-201-06673-4</isbn:number>
</book>
```

Dublin Core – Επεκτασιμότητα

- Η αρχιτεκτονική του Dublin Core υποστηρίζει πιο εξελιγμένες λύσεις μεταδεδομένων
  - Αλλά προσοχή: προτιμάμε επέκταση του DC ή επιλογή άλλων πλουσιοτέρων σχημάτων;
- Έχουμε επεκτάσεις του Dublin Core σε εξειδικευμένους τομείς (domain-specific)
- Ομάδες εργασίας (του DCMI) αναπτύσσουν *Προφίλ Εφαρμογών* (*application profiles*) για ειδικούς τομείς

Τι είναι το Προφίλ Εφαρμογής

- Ένα σχήμα μεταδεδομένων που ενσωματώνει ένα **σύνολο στοιχείων** από ένα ή περισσότερα **σύνολα στοιχείων μεταδεδομένων** (ή λεξιλογίων ή χώρους ονομάτων – namespaces)
- Ένα σύνολο από **πολιτικές** που ορίζουν πώς τα στοιχεία πρέπει να εφαρμόζονται στο πεδίο της εφαρμογής
- Ένα σύνολο **οδηγιών** που ξεκαθαρίζουν τις πολιτικές που αφορούν τα στοιχεία

Λεξιλόγια και Προφίλ Εφαρμογών

- Το **Λεξιλόγιο δηλώνει** όρους και ορισμούς
  - Λεξιλόγιο Dublin Core = Πρότυπο Dublin Core
- Το **Προφίλ Εφαρμογής επανα-χρησιμοποιεί** όρους από ένα ή περισσότερα λεξιλόγια
  - Μπορεί να συνδυάζει όρους από πολλαπλά λεξιλόγια
  - Μπορεί να προσαρμόζει (εξειδικεύει) ορισμούς για τοπικούς σκοπούς ή εξειδικευμένους τομείς (domain-specific)
  - Όλοι οι όροι πρέπει να ορίζονται στα λεξιλόγια
  - Μπορεί να περιλαμβάνει τοπικά ορισμένα λεξιλόγια

Συμπεράσματα

- Χρήση προτύπων και καθιερωμένων κανόνων
- Όχι αυτοσχεδιασμοί στην περιγραφή των δεδομένων
- Έστω και να φαινομενικά επαρκούν για την εφαρμογή
- Δεν φτάνει να δουλεύει μια εφαρμογή
- Πρέπει να μπορεί να **δια**-λειτουργεί στο μέλλον