

## Μεταδεδομένα

### Μια εισαγωγή στα μεταδεδομένα

Παπαδάκης Γιάννης

Δευτέρα, 12/11/2007

## Ηλεκτρονικό τεκμήριο

Τεκμήριο που υπάρχει σε ηλεκτρονική μορφή και στο οποίο η πρόσβαση επιτυγχάνεται με τη βοήθεια της τεχνολογίας υπολογιστών

## Ηλεκτρονικά τεκμήρια

- data files
  - computer files
    - electronic resources
      - things - objects
        - resource collections

ψηφιακοί πόροι – ψηφιακές συλλογές

## Καταλογογράφηση/Περιγραφή Ηλεκτρονικών τεκμηρίων

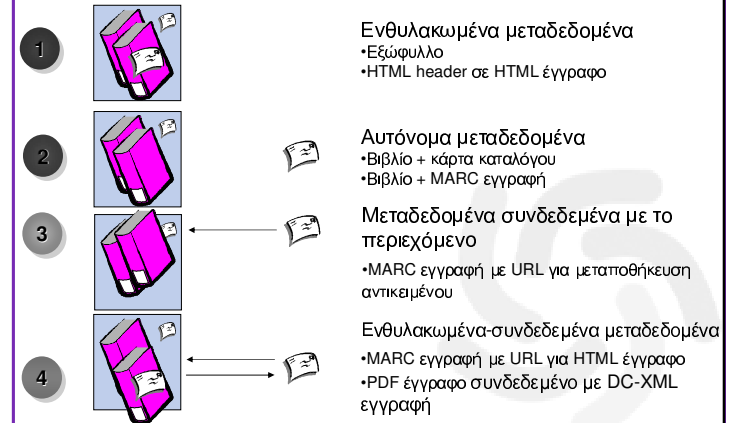
- Αγγλοαμερικανικοί κανόνες καταλογογράφησης
  - AACR2
- Λειτουργικές απαιτήσεις για βιβλιογραφικές εγγραφές
  - Functional Requirements for Bibliographic Records-FRBR
- Διατάξεις δεδομένων - formats
  - MARC, XML
- Πρότυπα μεταδεδομένων (metadata)
  - Dublin Core, EAD, FGDC κ.ά.
- Βιβλιογραφική περιγραφή (description)
  - ISBDs π.χ. ISBD(ER)
- Βιβλιογραφικές αναφορές (references)
  - ISO 690-2, ΕΛΟΤ 560-2

## Περιγραφή Ηλεκτρονικών τεκμηρίων

- "One recognisable form of *metadata* is the card index catalogue in a library; the information on that card is metadata about a book. Perhaps without knowing it, you use metadata in your work every day, ..." (Paul Miller)
- "*Metadata* really is nothing more than data about data; a catalog record is metadata; so is a TEI header, or any other form of description. We could call it cataloging, but for some people that term carries excess baggage, like Anglo-American Cataloging Rules and USMARC. So to some extent this is a "you call it corn, we call it maize" situation, but metadata is a good neutral term that covers all the bases." (Priscilla Caplan)
- "*Metadata* is machine understandable information about web resources or other things." (Tim Berners-Lee)

Πηγή: Gradmann, S. Cataloguing vs. Metadata: old wine in new bottles? 64th IFLA General Conference August 16 - August 21, 1998

## Μεταδεδομένα και περιεχόμενο



## Τύποι μεταδεδομένων

### Διαχειριστικά μεταδεδομένα:

- Δεδομένα για τα μεταδεδομένα
- (π.χ. Αριθμός εγγραφής)



### Περιγραφικά μεταδεδομένα:

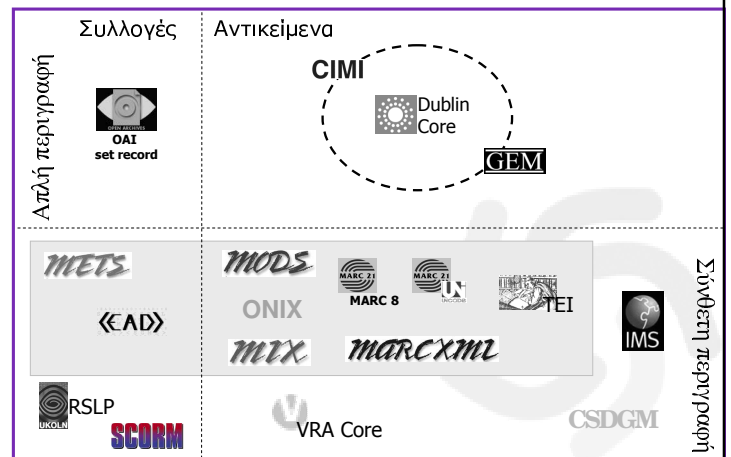
- Περιγραφή του αντικειμένου για ανάκτηση και εντοπισμό
- (π.χ. τίτλος)



### Τεχνικά μεταδεδομένα:

- Τεχνικά χαρακτηριστικά του αντικειμένου
- (π.χ. Μέγεθος αρχείου)

## Η συνολική εικόνα



## [ Dublin Core ]

- Πρωτοβουλία για να βελτιώσει την ανακάλυψη πόρων στο Διαδίκτυο
- Κοινός παρονομαστής για επικοινωνία και διαλειτουργικότητα
- Έναυσμα για σύγκλιση των προτύπων

## [ Σύνομη ιστορία του DCMI ]

- 1994: 2ο WWW συνέδριο, Chicago
  - Ανεπίσημες συζητήσεις
- 1995: NCSA/OCLC Ημερίδα μεταδεδομένων, Dublin, Ohio
  - Το 13-στοιχείων μεταδεδομένων "Dublin Core" για πόρους Διαδικτύου
- 1996: DC-2 Warwick Ημερίδα
  - "Warwick Framework": τα μεταδεδομένα είναι «modular»
- 1996: DC-3 Dublin, Ohio Ημερίδα
  - Επέκταση των 13 στοιχείων σε 15 (λιγότερο κειμενο-κεντρικά)
- 1997-: RDF: επίσημη έκφραση του **Warwick Framework**
- 2000: Το DCMI συστήνει προσδιοριστές (qualifiers)

## [ 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
- Ένα σύνολο από 15 ευρύχωρα «καλάθια» ...

## [ Εισαγωγή στο Dublin Core ]

- Η διαλειτουργικότητα απαιτεί
  - Χρήση κανόνων περιεχομένου και προτύπων
  - Σαφήνεια για τους πόρους που περιγράφονται
- Οι πραγματικοί πόροι είναι πιο περίπλοκοι από τα (μόνιμα) αντικείμενα τύπου τεκμηρίων
  - Τα χαρακτηριστικά των πόρων αλλάζουν με το χρόνο
  - Οι ενέργειες των πρακτόρων (agents) προκαλούν αλλαγές

## Πρότυπα και Dublin Core

- Προτυποποίηση
  - Ευρώπη: αναγνώριση από το CEN/ISSS Workshop Agreement 13874 (2000)
  - US: επικύρωση από το NISO Z39.85 (2001)
  - Αποστολή για ISO
- **Δεν** είναι υποκατάστατο σε πλουσιότερα περιγραφικά πρότυπα
- Παρέχει 15 «παράθυρα» από πλουσιότερη περιγραφή πόρων
  - Φανερώνει πλούσιες περιγραφές σε απλή μορφή
  - Σημασιολογικά σταυροδρόμια, αντιστοιχίσεις σε υπάρχοντα δεδομένα
  - Είναι περισσότερο αντικείμενο εξαγωγής – παρά δημιουργίας

## Dublin Core: Τα 15 στοιχεία

- Τίτλος / **Title**
- Θέμα / **Subject**
- Περιγραφή / **Description**
- Πηγή / **Source**
- Γλώσσα / **Language**
- Σχέση / **Relation**
- Κάλυψη / **Coverage**
- Δημιουργός / **Creator**
- Εκδότης / **Publisher**
- Συντελεστής / **Contributor**
- Δικαιώματα / **Rights**
- Ημερομηνία / **Date**
- Τύπος / **Type**
- Μορφότυπο / **Format**
- Κωδικός / **Identifier**

## Dublin Core – Simple

- Συστήνεται ανταλλαγή με RDF/XML
- 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**
- Γλώσσα / **Language** – (του περιεχομένου)
- Σχέση / **Relation** – (αναφορά σε σχετική πηγή)
  - Π.χ. έκδοση του ...
- Κάλυψη / **Coverage** – (γεωγραφική ή χρονική)

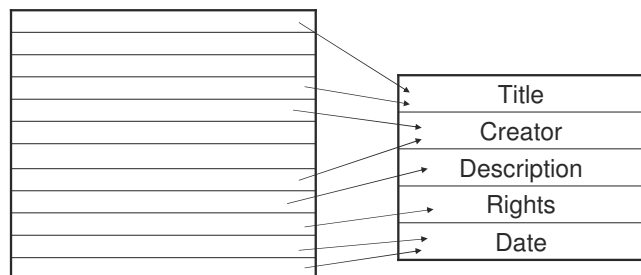
## [ DC – Πνευματική Ιδιοκτησία ]

- 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
- Κωδικός (Ταύτισης ή «Προσδιοριστής») / **Identifier**
  - Μοναδικό προσδιοριστικό, π.χ. URL, ISBN, ...

## [ Χρήση του Dublin Core ]



Πλούσια περιγραφή

Περιγραφή σε DC simple

## [ 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)
    - Σα 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!

## Εξειδίκευση σε Dublin Core

- Νέοι προσδιορισμοί / επίπεδα
  - π.χ. Degree = τύπος πτυχίου
- Δομημένο αλλά μη περιορισμένο σύνολο!
- Κανόνας «γενικότερου πεδίου», π.χ:
  - TITLE
  - TITLE.ALTERNATIVE
  - TITLE.TRANSLATED
  - TITLE.ALTERNATIVE.TRANSLATED
  - TITLE.FULL

## Dublin Core: Εξειδίκευση Στοιχείων

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

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

## Ποικιλίες Εξειδικευτών: Σύστημα Κωδικοποίησης Τιμών

- 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 το λογισμικό, η τιμή πρέπει να είναι «κατάλληλη» και χρησιμοποιήσιμη για ανακάλυψη πόρων

## Dublin Core: Κωδικοποίηση τιμών

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

## Dublin Core: Κωδικοποίηση τιμών

Στοιχείο	Εξειδίκευση
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)

## 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" content = "ELEMENT_VALUE">`
  - Για παράδειγμα :  
`<meta name = "DC.Title" content = "Παρουσίαση DC">`  
`<meta name = "DC.Creator" content = "Κατιδάκης, Σαράντος">`

## Κωδικοποίηση σε HTML 4.0

- Η HTML 4.0, καθιέρωσε τους προσδιοριστές (qualifiers) και τις ιδιότητες στις ετικέτες META για τη γλώσσα των μεταδεδομένων (LANG) και το πρότυπο κωδικοποίησης ή περιχομένου (SCHEME)
- Η κωδικοποίηση μεταδεδομένων (π.χ. σε Dublin Core) τοποθετείται στο τμήμα του αρχείου HTML ανάμεσα στις θέσεις <HEAD> και </HEAD>.
- Η HTML αδυνατεί να περιγράψει σύνθετες περιπτώσεις – τότε χρησιμοποιείται το πρότυπο RDF
- Για παράδειγμα, ένα αρχείο HTML με ενσωματωμένη DC κωδικοποίηση, θα έχει την παρακάτω μορφή:

## 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.DateIssued" content = "1990">`
- `<meta name = "DC.Language" scheme = "ISO639-2" content = "gre">`
- `</head>`
- `<body> mpla - mpla </body>`
- `</html>`

## Τέλος παρουσίασης

- Ευχαριστώ για την προσοχή σας!