

Πρότυπα Μεταδεδομένων

Χρήστος Παπαθεοδώρου (paratheodor@ionio.gr)

Αναπληρωτής Καθηγητής

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

Μονάδα Ψηφιακής Επιμέλειας,

Ινστιτούτο Πληροφοριακών Συστημάτων και Προσομοίωσης
Ερευνητικό Κέντρο «Αθηνά»



Περιεχόμενα

- Εισαγωγή στα μεταδεδομένα
- Dublin Core
- MODS
- Αρχιακά μεταδεδομένα – Encoding archival description (EAD)

Ψηφιακά τεκμήρια

- Οτιδήποτε υπάρχει σε ψηφιακή μορφή και προσπελάζεται με τη βοήθεια υπολογιστή
 - κείμενα
 - εικόνες, κινούμενες εικόνες
 - ήχος, βίντεο
 - ιστοσελίδες, προγράμματα

Τύποι δεδομένων Ηλεκτρονικών τεκμηρίων

Κείμενο

.DOC

.TXT

.RTF

.PDF

Εικόνες

.BMP

.GIF

.JPEG

.TIFF

.EPS

Κινούμενες εικόνες

.ANI

.FLI

.FLC

.GIF

Ηχος

.WAV

.MID

.SND

.AUD

Βίντεο

.AVI

.MOV

.MPG

.QT

Ιστοσελίδες

.HTM

.HTML

.XML

Προγράμματα

.COM

.EXE

Πώς εντοπίζουμε και
ανακτούμε τα ψηφιακά
τεκμήρια;

Μεταδεδομένα

- Τα μεταδεδομένα είναι **δομημένη πληροφορία** που περιγράφει, εξηγεί, εντοπίζει ή διευκολύνει την ανάκτηση, τη χρήση ή την διαχείριση ενός ψηφιακού τεκμηρίου
- Συχνά καλούνται («**δομημένα**») «δεδομένα για άλλα δεδομένα» ή «πληροφορία για άλλη πληροφορία»
- "**Metadata** is machine understandable information about web resources or other things." (Tim Berners-Lee)
- Διάφορα πρότυπα μεταδεδομένων ανάλογα με το είδος του τεκμηρίου και τις πληροφοριακές ανάγκες
 - ανακάλυψη: περιγραφικά, π.χ. τίλος, συγγραφέας
 - διατήρηση: τεχνικά, π.χ. τύπος υλικού, ψηφιακής μορφή

Που Βρίσκονται τα Μεταδεδομένα

- Στην ετικέτα ενός CD (ορατά)
- Στη σελίδα τίτλου ενός τεκμηρίου (ορατά)
- Στην κορυφή μιας ιστοσελίδας (ορατά)
- Σε διαφορετική εγγραφή, π.χ. στον κατάλογο μιας βιβλιοθήκης (ορατά)
- Στην ηλεκτρονική μορφή μέσα στον πόρο (ορατά)
- Ενσωματωμένα στο ηλεκτρονικό δημοσίευμα (μέρος της κωδικοποίησης, μη ορατά)

Παράδειγμα: Ιστοσελίδα σε HTML

<HTML>

<BODY>

<p> Fiction:</p>

Author: Milan Kundera

Title: Identity

Published: 1998

<p> Science: </p>

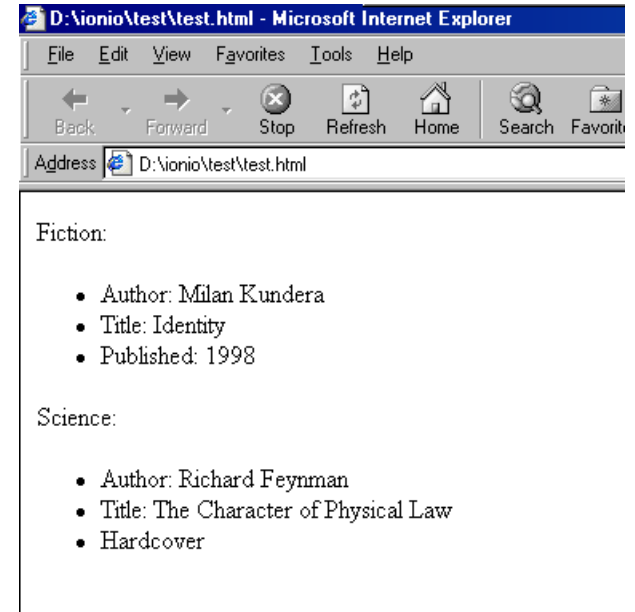
Author: Richard Feynman

Title: The Character of Physical Law

Hardcover

</BODY>

</HTML>



Πώς θα κάνουμε «ορατά» τα
μεταδεδομένα από τους
υπολογιστές (μηχαναγνώσιμη
πληροφορία);

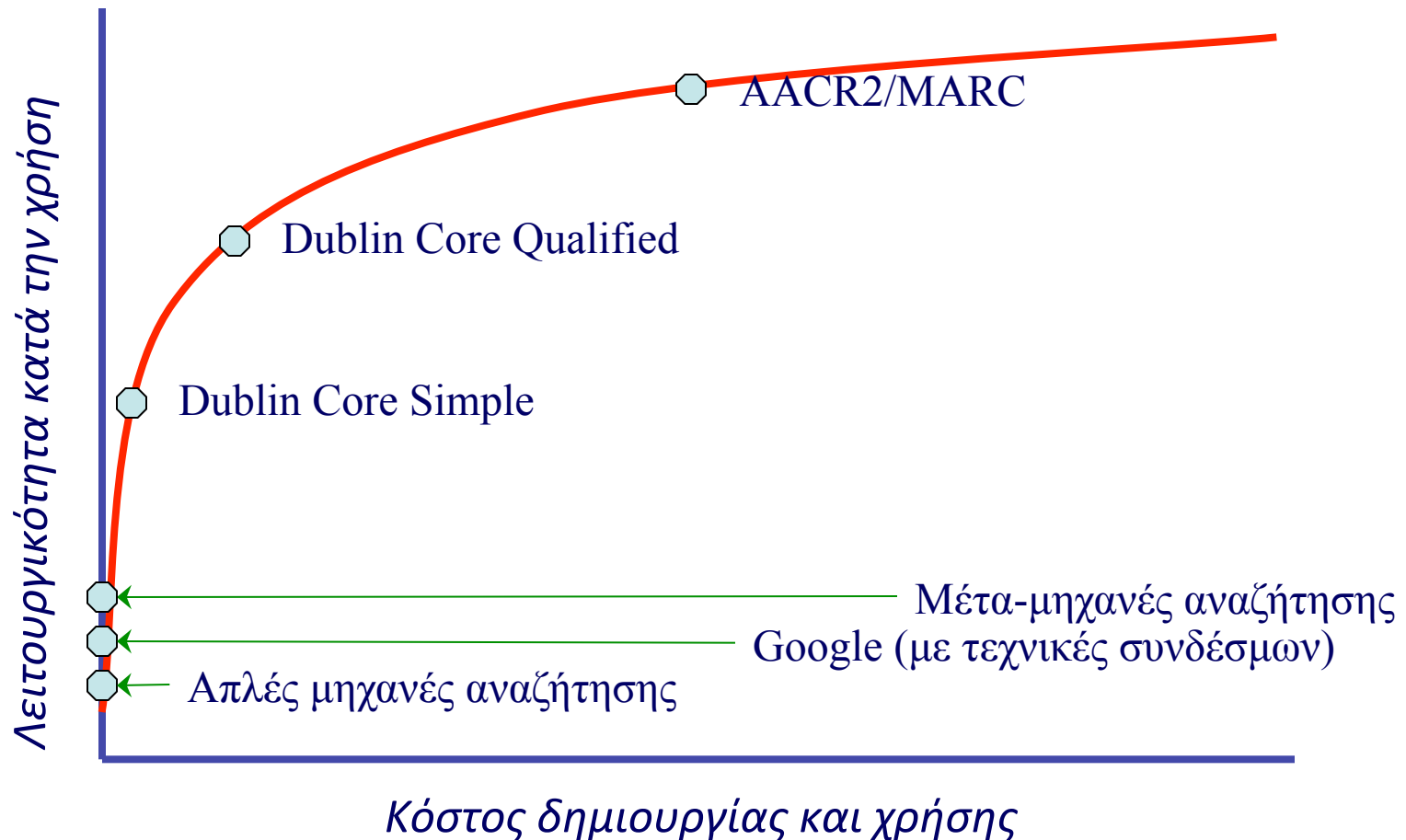
XML: κωδικοποίηση μεταδεδομένων

- Η HTML ελέγχει τον τρόπο εμφάνισης των ηλεκτρονικών τεκμηρίων, δεν ασχολείται με τη φύση τους, την περιγραφή τους, τη σημασία τους
- **e**Xtensible **M**arku**p** **L**angu**a**ge (XML) αποτελεί μια εξαιρετικά απλή διάλεκτο της γλώσσας **S**tandard **G**eneralized **M**arkup **L**anguage (SGML), η οποία αναπτύχθηκε με στόχο να διευκολύνει το χειρισμό, επεξεργασία, διακίνηση και αποθήκευση τεκμηρίων στον Παγκόσμιο Ιστό (web)

Αναγκαιότητα Προτύπων Μεταδεδομένων

- Χρειάζονται για να έχουμε **κοινή αντίληψη** των δεδομένων που περιγράφονται με αυτά
- Μας προσφέρουν μεγαλύτερη **δομή**
 - Μας περιορίζουν στην **ευελιξία**
 - Επεξεργάζονται ευκολότερα **μηχανικά**
- Είναι αναγκαιότερα σε **ψηφιακά** αντικείμενα
 - Αφού στα συμβατικά έχουμε αισθητήρια αντίληψη για αναγνώριση, θέση, θέματος, ...

Κόστος και Λειτουργικότητα Αναζήτησης και Μεταδεδομένων



Πόσα Πρότυπα Μεταδεδομένων

- Δεν υπάρχει **ένα** μοναδικό διεθνές πρότυπο για μεταδεδομένα, γιατί:
 - Χρειαζόμαστε διαφορετικά **επίπεδα πολυπλοκότητας**, από πλούσιες μέχρι απλές περιγραφές
 - **Υπάρχουν** κάμποσα σχήματα μεταδεδομένων, για διαφορετικά επίπεδα και απαιτήσεις
 - **Επεκτείνουμε** τα υπάρχοντα πρότυπα

Πρότυπα Μεταδεδομένων

- AACR2
- MARC
- Text Encoding Initiative - TEI Header (1990)
- Dublin Core - DC (1995)
- Encoded Archival Description – EAD (1996)
- Open Archives Initiative - OAI
- Visual Resources Association Core– VRA (1997)
- Federal Geographic Data Committee for Digital Geospatial Metadata – FGDC
- Data Documentation Initiative – DDI (1997)
- Gateway to Educational Materials – GEM (1999)
- Government (Global) Information Locator Service - GILS
- Metadata Encoding and Transmission Standard – METS
- Metadata Object Description Schema – MODS
- Computer Interchange of Museum Information – CIMI
- Interoperability of Data in E-Commerce Systems – INDECS
- Online Information Exchange – ONIX (2000)
- Extended Markup Language – XML, MARCXML
- Australian Recordkeeping Metadata Schema (RKMS)

Κατηγορίες Μεταδεδομένων

■ Διαχειριστικά (administrative)

Με πληροφορίες για τη διαχείριση και διατήρηση της εγγραφής, όπως δημιουργία, μετατροπή και σχέσεις με άλλες εγγραφές, π.χ. κάτοχος, ημερομηνίες δημιουργίας ή/και μεταβολής, γλώσσα εγγραφής, διαχείρισης δικαιωμάτων κλπ.

Δομικά (structural)

Με πληροφορίες για την αποθήκευση και παρουσίαση

Περιγραφικά (descriptive)

Με πληροφορίες που περιγράφουν ιδιότητες και περιεχόμενο του αντικειμένου στο οποίο αναφέρεται η εγγραφή και συμβάλλουν στην ανάκτηση (ανακάλυψη), π.χ. τίτλος, συγγραφέας, θεματικοί όροι, περίληψη

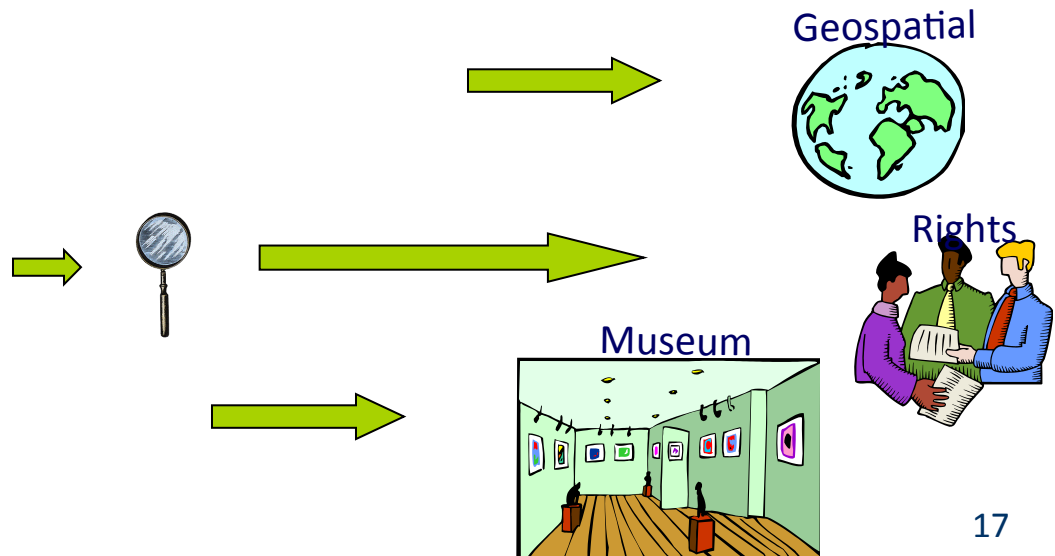
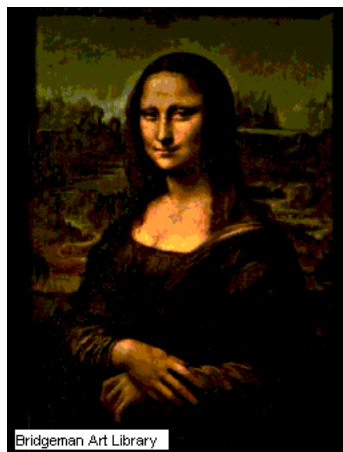
Κατηγορίες Μεταδεδομένων

Σύμφωνα με την IFLA:

- Administrative
- Descriptive
- Analytical (subject)
- Rights management
- Technical
- Other, as determined

Φακοί και Όψεις

- Όλες οι κατηγοριοποιήσεις παρέχουν και **πρέπει να παρέχουν** μια μεροληπτική όψη της πραγματικότητας
- Κάθε όψη τονίζει συγκεκριμένα χαρακτηριστικά και κρύβει άλλα



Επιλογή μεταδεδομένων

Είναι σημαντικό να καθοριστούν

- το πεδίο εφαρμογής
- τα κριτήρια
- οι μορφές των ψηφιακών αντικειμένων και συλλογών
- ο βαθμός λεπτομέρειας στην περιγραφή και παρουσίαση

προκειμένου να αξιολογηθούν τα πρότυπα μεταδεδομένων και επιλεγεί το κατάλληλο.

Ερώτηση: Ποια μπορεί να είναι τα βασικά στοιχεία που θα περιγράφουν μια δεδομένη ψηφιακή συλλογή;

Απάντηση (παράδειγμα)

Υποχρεωτικά

- Τίτλος
- Συγγραφέας
- Θέμα
- Ηλεκτρονική διεύθυνση

Επιθυμητά

- Εκδότης
- Ημερομηνία
- Κατηγορία
- Σχόλια

Άλλα σχετικά με συλλογή

- Αναγνωριστικά
- Γλώσσα
- Έκδοση
- Πνευματικά δικαιώματα
-

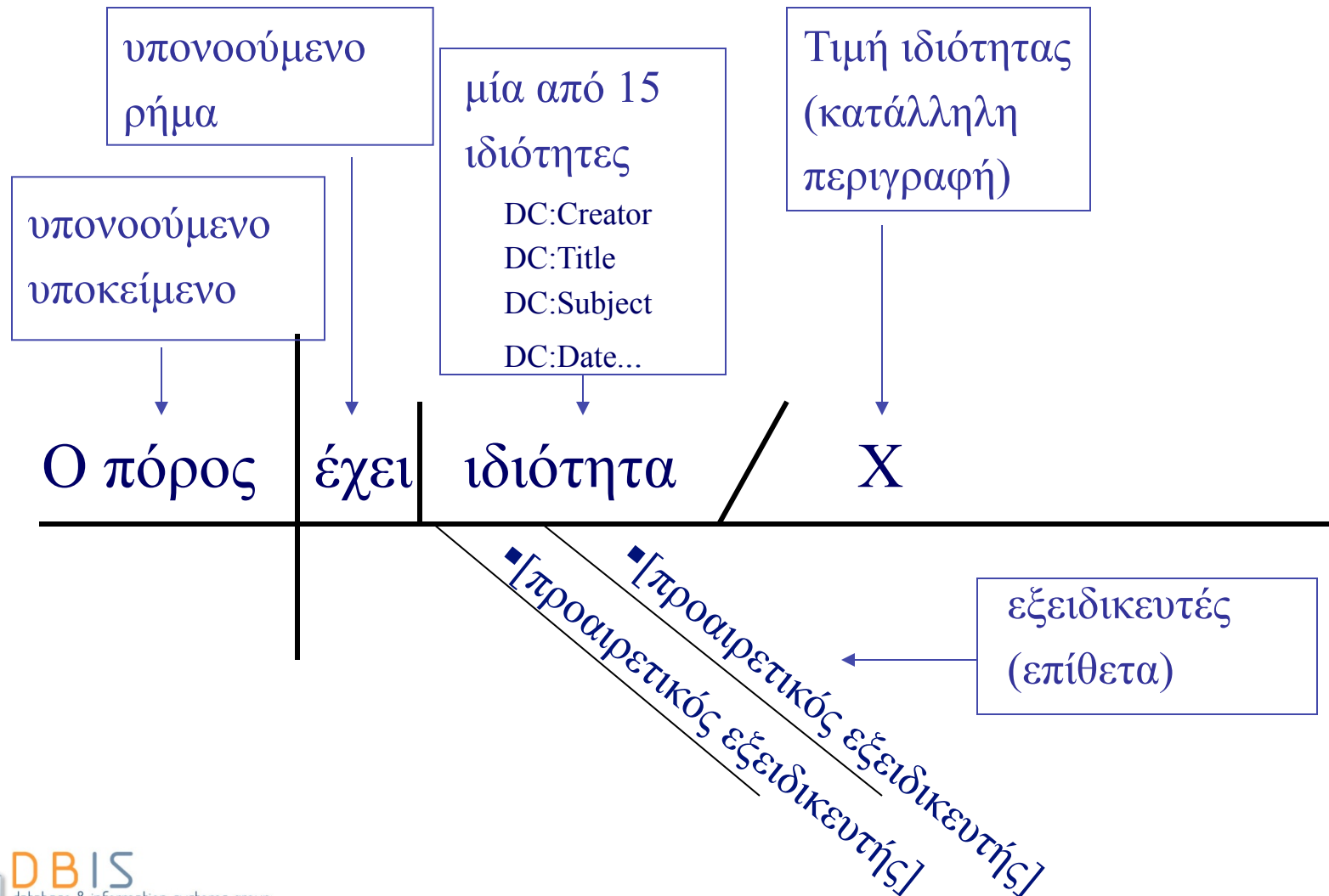
Τα Μεταδεδομένα είναι Γλώσσα

- Τα σχήματα μεταδεδομένων είναι γλώσσες για να γίνονται δηλώσεις για τους πόρους:
 - «Το βιβλίο» έχει **Title** "Gone with the Wind"
 - «Η ιστοσελίδα» έχει **Publisher** "Springer Verlag"
- Οι όροι των λεξιλογίων (στοιχεία) ορίζονται με πρότυπα – όπως το Dublin Core
- Οι γραμματικές των μεταδεδομένων περιορίζουν τις δηλώσεις και τα μοντέλα δεδομένων που μπορεί να χρησιμοποιηθούν – κάνοντας όμως τη σημασία τους πιο έκδηλη και σαφή

Τα Μεταδεδομένα ως Γλώσσα

- Διαφορετικές κοινωνίες (περιγραφής πόρων) μιλούν διαφορετικές γλώσσες
 - Π.χ. Τι γλώσσες περιγράφουν Ιατρική σε Ελλάδα και Αγγλία; Η Ιατρική ορολογία είναι κοινή.
 - Και οι φυσικοί ομιλητές των διαφορετικών γλωσσών πρέπει να επικοινωνήσουν ...
 - Και οι τουρίστες (και οι ομιλητές των διαφορετικών γλωσσών) μαθαίνουν να «μιλούν πρωτόγονα» για να επικοινωνήσουν

Κανόνες Γραμματικής DC



Παραδείγματα Γραμματικής DC

Ο πόρος	έχει	Subject	"Languages -- Grammar"
		<i>LCSH</i>	
Ο πόρος	έχει	Date	"2000-06-13"
		<i>ISO8601 Revised</i>	

Οι Γλώσσες Εξελίσσονται με τη Χρήση

- Αναπόφευκτα, οι γλώσσες **αντιστέκονται** στην **σταθερότητα**
- Οι άνθρωποι **«επεκτείνουν»** τους αυθεντικούς ορισμούς
- Οι υλοποιητές **παρεξηγούν** την επιδιωκόμενη σημασία ή χρήση των στοιχείων
- Οι υλοποιητές χρησιμοποιούν **τοπικούς όρους** και **εκφράσεις**
- Αν η εφαρμογή δεν ταιριάζει με το πρότυπο, το πρότυπο συχνά **«προσαρμόζεται»** να ταιριάζει στην εφαρμογή