Εισαγωγικές Έννοιες

Χρήστος Παπαθεοδώρου

Καθηγητής

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

(papatheodor@ionio.gr)

DBIS
database & information systems group
ionian university

Ψηφιακά τεκμήρια

- Οτιδήποτε υπάρχει σε ψηφιακή μορφή και προσπελαύνεται 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 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 τίλου 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)

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

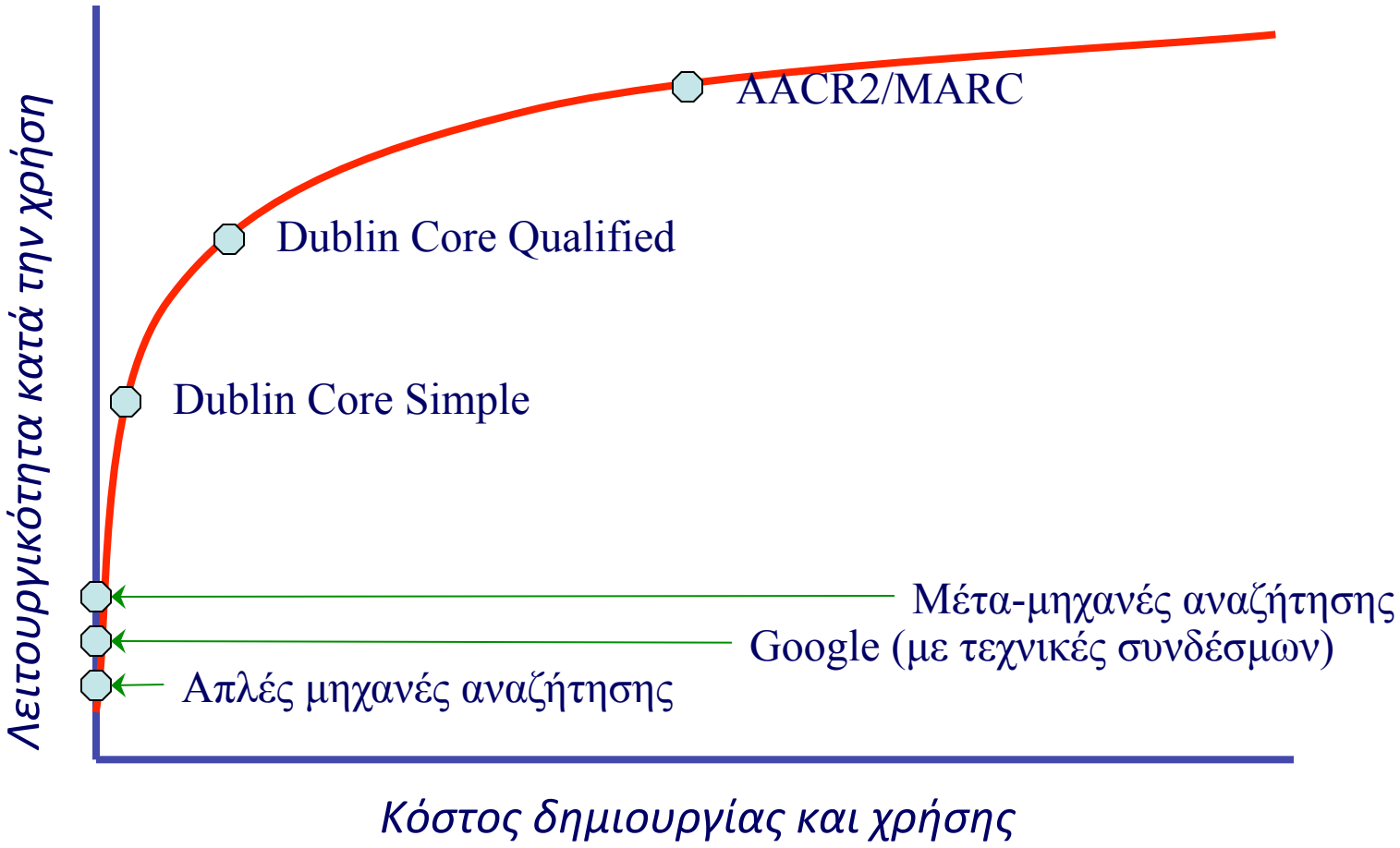
XML: κωδικοποίηση μεταδεδομένων

- Η HTML ελέγχει τον τρόπο εμφάνισης των ηλεκτρονικών τεκμηρίων, δεν ασχολείται με τη φύση τους, την περιγραφή τους, τη σημασία τους
- e**X**tensible **M**arkup **L**anguage (XML) αποτελεί μια εξαιρετικά απλή διάλεκτο της γλώσσας Standard Generalized Markup Language (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)

Στοιχεία μεταδεδομένων

Είναι μία ενότητα πληροφορίας για ένα τεκμήριο

Χαρακτηριστικά κάθε στοιχείου:

- **Ονομασία**
  - (π.χ. access_category)
- **Ορισμός**
- **Λειτουργία**
  - (access management, administration, discovery, persistent identifier, presentation, digital preservation and preservation reformatting)
- **Τύπος**
  - (Administrative, Structural, Descriptive)
- **Χρήση**
  - (frequency, requirement)
- **Επίπεδο**
  - (Συλλογή ή κατηγορία αντικειμένων (π.χ. εικόνες), πρωτογενές αντικείμενο, ή τμήμα ενός αντικειμένου ή ψηφιακό αρχείο)

Στοιχεία μεταδεδομένων–IFLA

- Subject
- Date
- Conditions of use
- Publisher
- Name assigned to the resource
- Language/mode of expression
- Resource identifier
- Resource type (i.e., what the resource ***is***, rather than what it is *about*)
- Author/creator
- Version

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

Αλλα σχετικά με συλλογή

- Αναγνωριστικά
- Γλώσσα
- Έκδοση
- Πνευματικά δικαιώματα
- ……

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

Στοιχεία δεδομένων – IFLA

Descriptive

- o Title (also alternative and parallel titles; subtitles; short titles; etc.)
- o Creator (author; composer; cartographer; artist; etc.)
- o Date
- o Publisher
- o Unique identifiers (ISBN; ISSN; etc.)
- o Dynamic links (URI; URL; etc.)
- o Summary; descriptive note; review; etc.
- o Audience level
- o Physical media; format; etc.
- o Language of the item or object
- o Version

Στοιχεία δεδομένων – IFLA

Analytical

- o Controlled subject terms, e.g., subject headings, descriptors
- o Subject/topic keywords
- o Abstract; Table of Contents (TOC)
- o Codes derived from classification systems or categorization schemes
- o Other elements of local importance, e.g., department affiliation; links to other related e-content

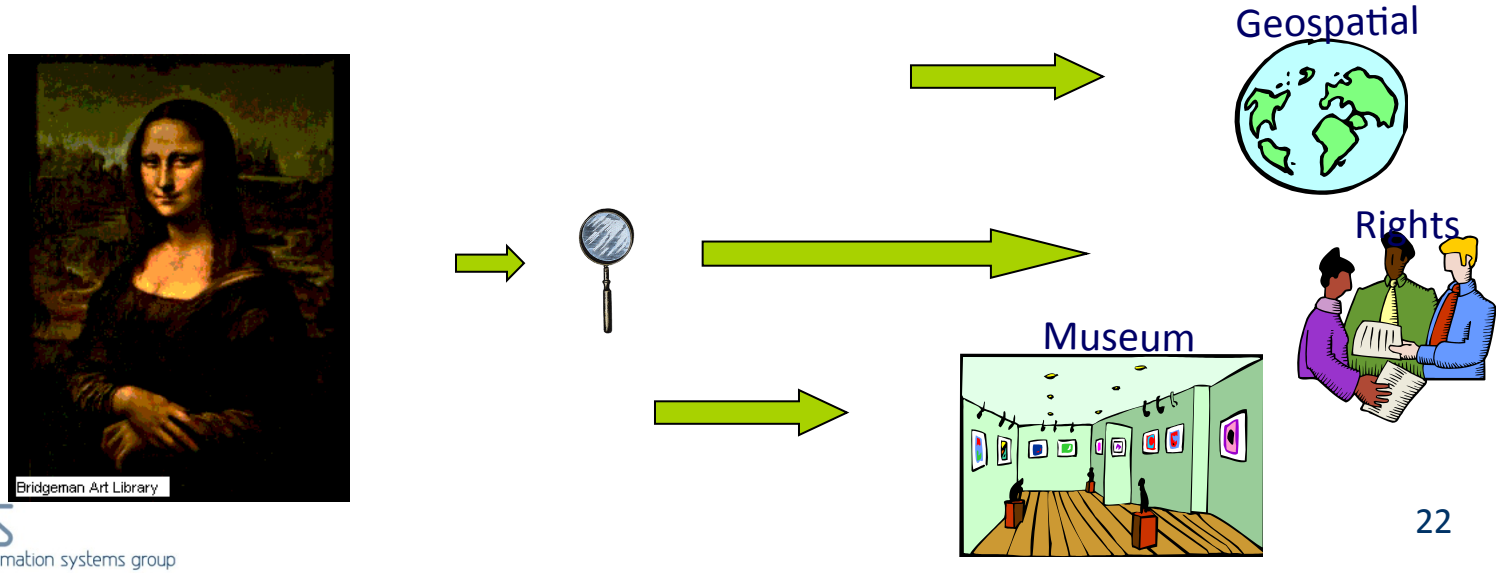
Στοιχεία δεδομένων - IFLA (5)

Technical

- o Digitizing equipment specifications
- o Camera positions
- o Shooting conditions
- o Coding parameters
- o Voice recognition and/or read-back hardware and software
- o Optical scanner specifications
- o Image rendering equipment
- o Type of file and conversion software requirements

Φακοί και Όψεις

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



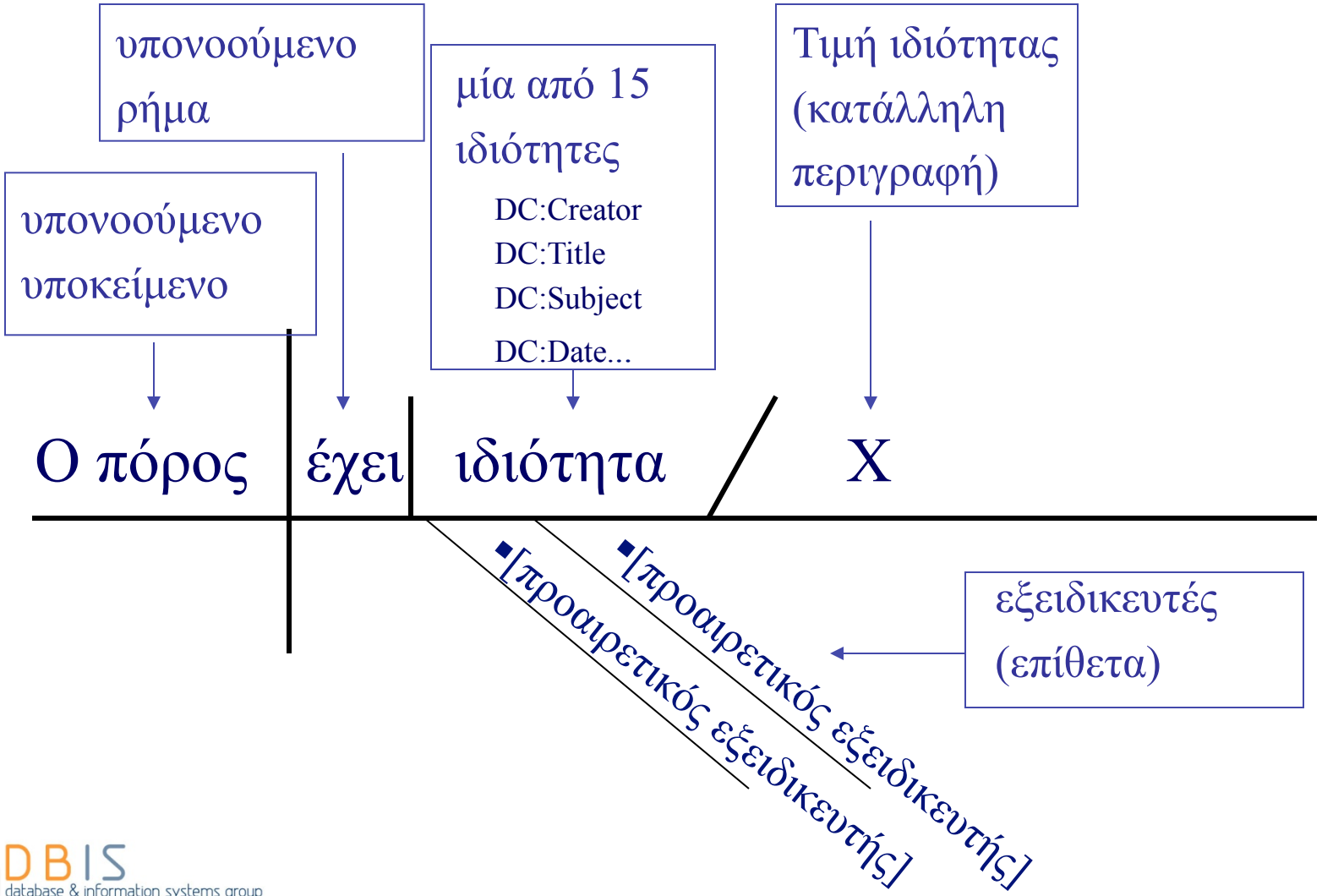
Τα Μεταδεδομένα είναι Γλώσσα

- 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
- Το συντακτικό των μεταδεδομένων ορίζεται από τα πρότυπα και θέτει περιορισμούς στις δηλώσεις που μπορεί να εκφραστούν – κάνοντας όμως τη σημασία τους πιο έκδηλη και σαφή

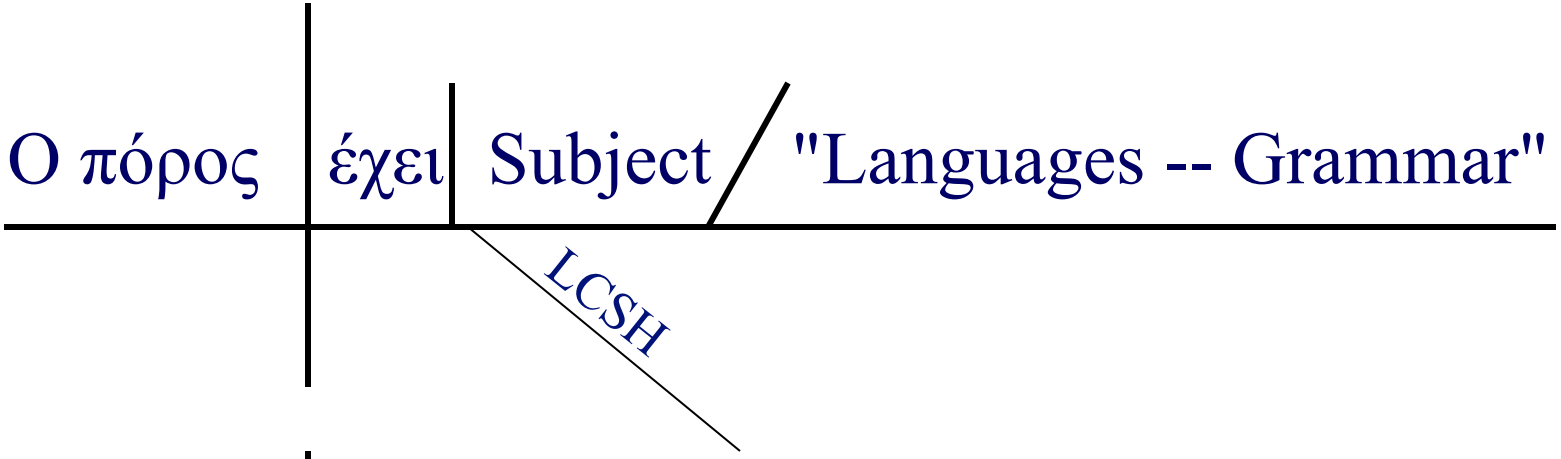
Τα Μεταδεδομένα ως Γλώσσα

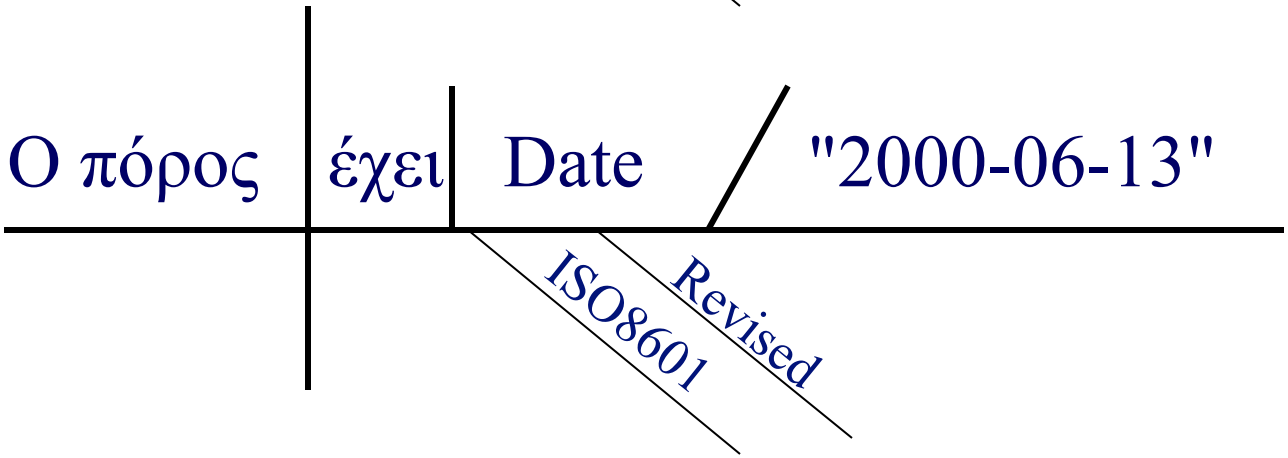
- Ορίζουν μια κοινή ορολογία – λεξιλόγιο για την περιγραφή πόρων από μια κοινότητα
- Ορίζουν ένα πεδίο κοινής επικοινωνίας – ανταλλαγής πληροφορίας εντός μιας κοινότητας και μεταξύ κοινοτήτων
- Τα ουσιαστικά των μεταδεδομένων – τα «στοιχεία» - είναι σύμβολα που αντιπροσωπεύουν έννοιες εκφράσιμες σε πολλαπλές φυσικές γλώσσες

Κανόνες Γραμματικής DC



Παραδείγματα Γραμματικής DC





Οι Γλώσσες Εξελίσσονται με τη Χρήση

- Αναπόφευκτα, οι γλώσσες **αντιστέκονται** στην **σταθερότητα**
- Οι ορισμοί των όρων των λεξιλογίων «**επεκτείνονται**» ή τροποποιούνται
- Τα πρότυπα μπορεί να έχουν δεκάδες μεταφράσεις
- Οι χρήστες των προτύπων μεταδεδομένων **παρεξηγούν/ παρερμηνεύουν** τη σημασία ή χρήση των στοιχείων
- Οι σχεδιαστές σχημάτων μεταδεδομένων χρησιμοποιούν **τοπικούς όρους** και **εκφράσεις**
- Αν η εφαρμογή δεν ταιριάζει με το πρότυπο, το πρότυπο συχνά **«προσαρμόζεται»** να ταιριάζει στην εφαρμογή