



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΕΧΝΟΛΟΓΙΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΠΟΛΙΤΙΣΜΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

Μεταδεδομένα

*Γιώργος Καρυδάκης
Τμήμα Πολιτισμικής Τεχνολογίας και
Επικοινωνίας*



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Ενότητα 3: Μεταδεδομένα

Τεχνολογίες Διαχείρισης Πολιτισμικής
Πληροφορίας

Γ. Καρυδάκης

Βασικές Έννοιες

- Αντικείμενο
 - Με «οτιδήποτε» (μαύρο κουτί) ιδιότητες συμπεριφοράς
- Πόρος
 - Χρήσιμο, αναζητήσιμο και προσδιορίσιμο αντικείμενο
- Τεκμήριο
 - Μεταφέρει πληροφορία: ήχος, βίντεο, τεκμήριο
- Κείμενο
 - Με στατική οπτική μορφή: γράμματα και εικόνες

Πόρος (Resource)

- Ένας πόρος προσδιορίζεται με ένα URI
–[απόλυτο-URI | σχετικό-URI] [“#” id-αποσπάσματος]
- Ο πόρος που προσδιορίζεται από ένα URI μπορεί να είναι αφηρημένος - π.χ. μη αναζητήσιμος δικτυακά
- Ο πόρος είναι διαφορετικός από τις οντότητες που προσδιορίζονται σε κάποια συγκεκριμένη στιγμή

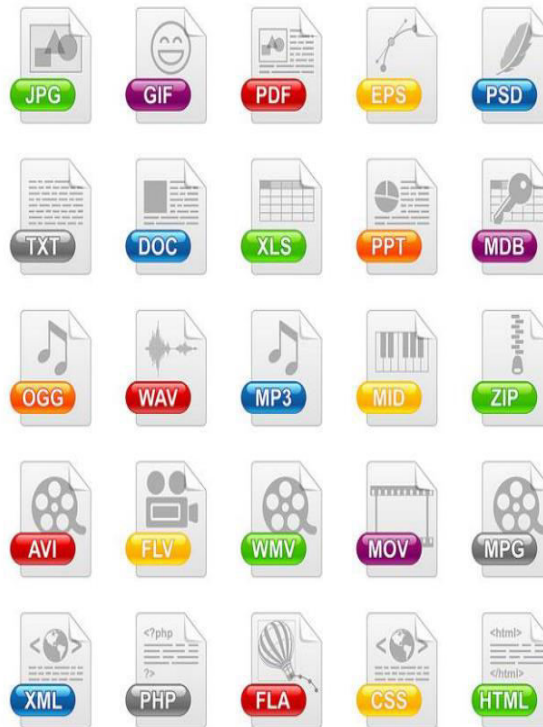
Τεκμήριο

- Οτιδήποτε προσφέρει πληροφορία στον ψηφιακό κόσμο
- Με ψηφιακή υπόσταση
 - Αρχεία διαφόρων μορφών
 - Εφαρμογές λογισμικού
 - Υπηρεσίες
- Είδη ηλεκτρονικών τεκμηρίων:
 - Μορφοποιημένα (σταθερά) - π.χ. σελίδες βιβλίων, εικόνες
 - Μορφοποιήσιμα (μεταβλητά) - π.χ. αρχεία

Ψηφιακά τεκμήρια

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

Μορφές ψηφιακών τεκμηρίων



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

Στόχοι

- Να συζητηθούν
 - Αξία
 - Αναγκαιότητα
 - Μεθοδολογία
 - Προκλήσεις των μεταδεδομένων
- Να αντιληφθούμε
 - Τάσεις
 - Εφαρμογές
 - Παραδείγματα
- Να παρουσιαστούν υπάρχοντα πρότυπα
 - Επιλογή προτύπου
 - Χρήση του προτύπου

Στόχοι

- Κατανόηση
 - Σημασίας
 - Πλαισίου
- Προβληματισμοί / Προκλήσεις
- Σωστές επιλογές
- Μεθοδολογίες
- Χρήση προτύπων

Στόχος

- Διερεύνηση των τρόπων με τους οποίους επιτυγχάνεται ... ψηφιακών πόρων που βρίσκονται στο διαδίκτυο
 - Αναζήτηση (find)
 - Ανακάλυψη (discovery)
 - Πλοήγηση (navigation)
 - Συνεύρεση (collocation)
 - Αναγνώριση (identify, DOI, URI)
 - Επιλογή (select)

Που Βρίσκονται;

- Στην ετικέτα ενός CD (ορατά)
- Στη σελίδα τίτλου ενός τεκμηρίου (ορατά)
- Στην κορυφή μιας ιστοσελίδας (ορατά)
- Σε διαφορετική εγγραφή (ορατά)
 - π.χ. στον κατάλογο μιας βιβλιοθήκης
- Στην ηλεκτρονική μορφή μέσα στον πόρο (ορατά)

• Για περισσότερα στοιχεία, στο ηλεκτρονικό

Παράδειγμα

- <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"

Παραδείγματα

- Όνομα αρχείου
- Τύπος αρχείου
- Μέγεθος αρχείου
- Ημερομηνία αλλαγής αρχείου
- Επιπλέον σε αρχεία εικόνων
 - Φυσικές διαστάσεις εικόνας
 - Ανάλυση, μορφή αποθήκευσης, τρόπος συμπίεσης, ...

Ιδιότητες Μεταδεδομένων

- Μια μοναδική ονομασία για το κάθε πεδίο (tag, label, identifier, field name)
- Ένα ορισμό κάθε πεδίου
- Υποχρεωτικό ή προαιρετικό
- Επαναλαμβανόμενο ή όχι
- Οργάνωση σχέσεων μεταξύ των πεδίων (π.χ. σχέσεις ιεράρχησης)
- Περιορισμοί για τις δυνατές τιμές του πεδίου (κείμενο, αριθμητικό εύρος, ημερομηνία, ελεγχόμενο λεξιλόγιο)

Τύποι μεταδεδομένων

- Εγγενή
 - Τι περιέχει;
 - Σε τι αφορά;
 - Περιβάλλον (Context)
- Συναφή
 - Ποιος, Πού, Πώς, Πότε, Γιατί, κλπ.
- Δομή (Structure)
 - Τυπικά σύνολα

Εγγενή

- Θέμα
- Τίτλος
- Συγγραφέας
- Εκδότης
- Τόπος έκδοσης
- Ημερομηνία
- Πηγή
- Γλώσσα
- Περίληψη
- Έκδοση
- Λέξεις κλειδιά

Συναφή

- Απαιτήσεις συστήματος
- Τρόπος πρόσβασης
- Διαθεσιμότητα
- Έλεγχος (πρόσβασης)
- Κόστος
- Έκταση
- Περιγραφή κωδικοποίησης
- Περιγραφή αναθεώρησης

Μορφές

- Δεν χρειάζεται να είναι ψηφιακά
- Είναι παραπάνω από περιγραφή του αντικειμένου
- Προέρχονται από ποικιλία πηγών
- Συνεχίζουν να συσσωρεύονται
- Τα μεταδεδομένα ενός αντικειμένου μπορεί να είναι δεδομένα

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**arkup **L**anguage (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

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**

Πόσα Πρότυπα;

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

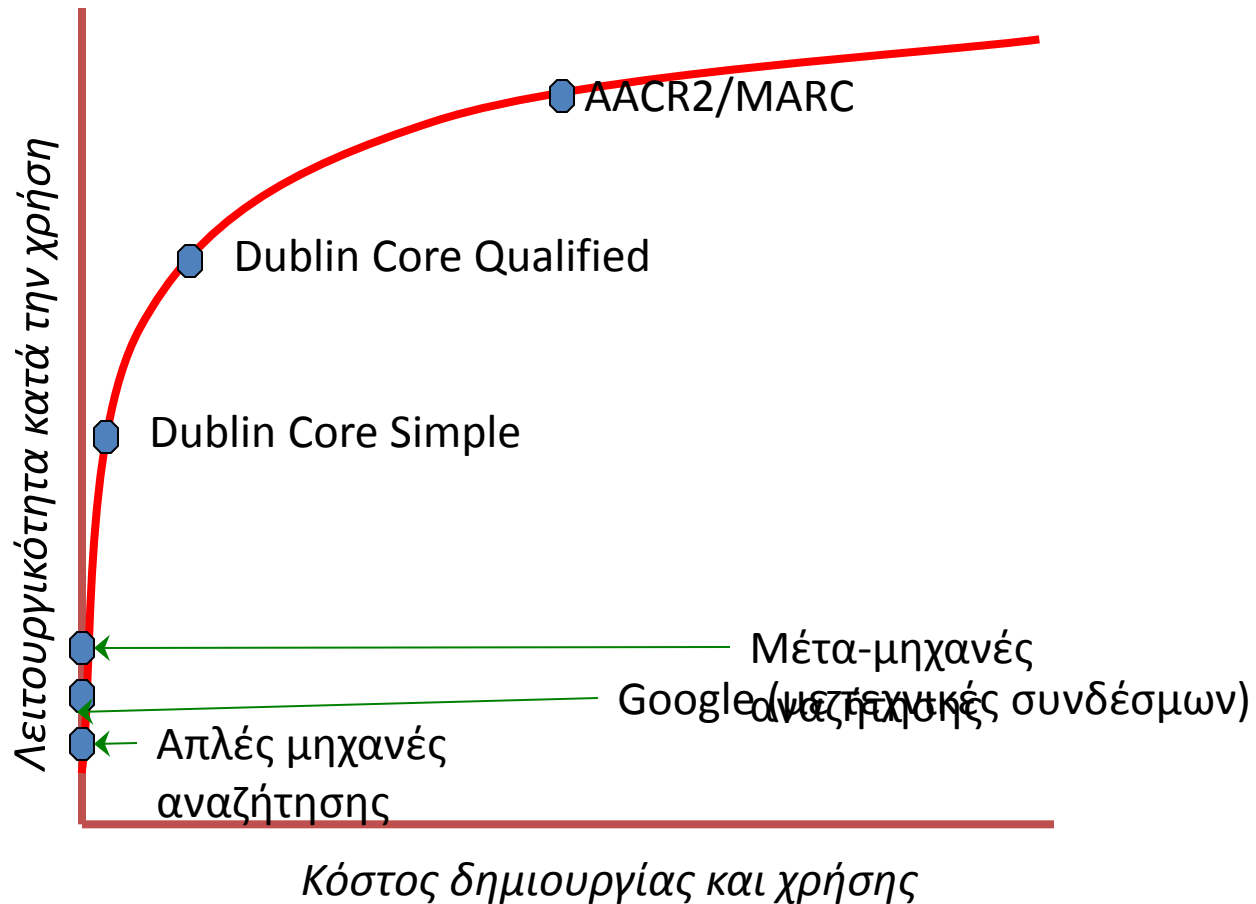
Πρότυπα Μεταδεδομένων

- 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

Πρότυπα Μεταδεδομένων

- 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)

Κόστος και Λειτουργικότητα Αναζήτησης και Μεταδεδομένων



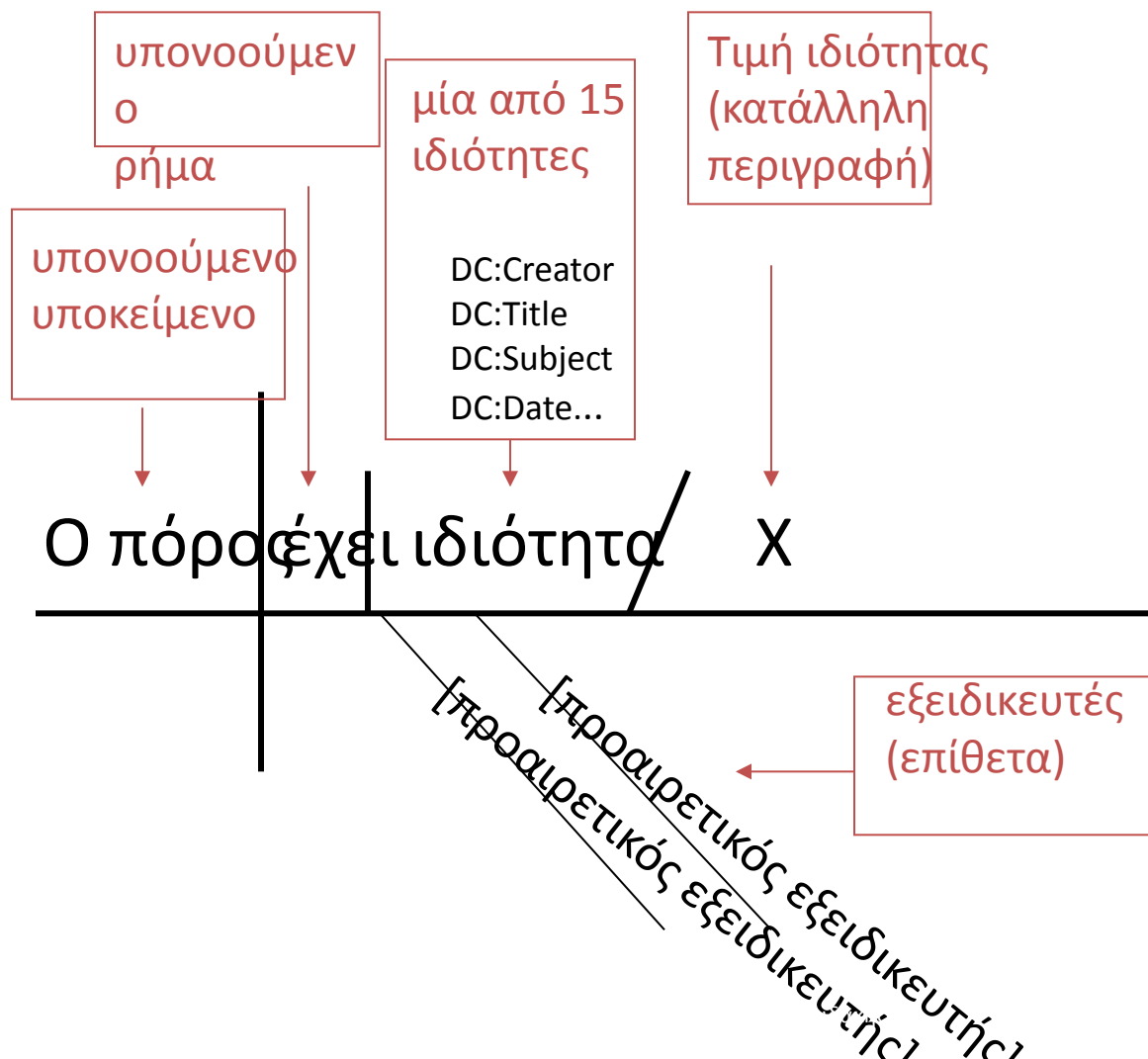
Τα Μεταδεδομένα είναι Γλώσσα

- 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)

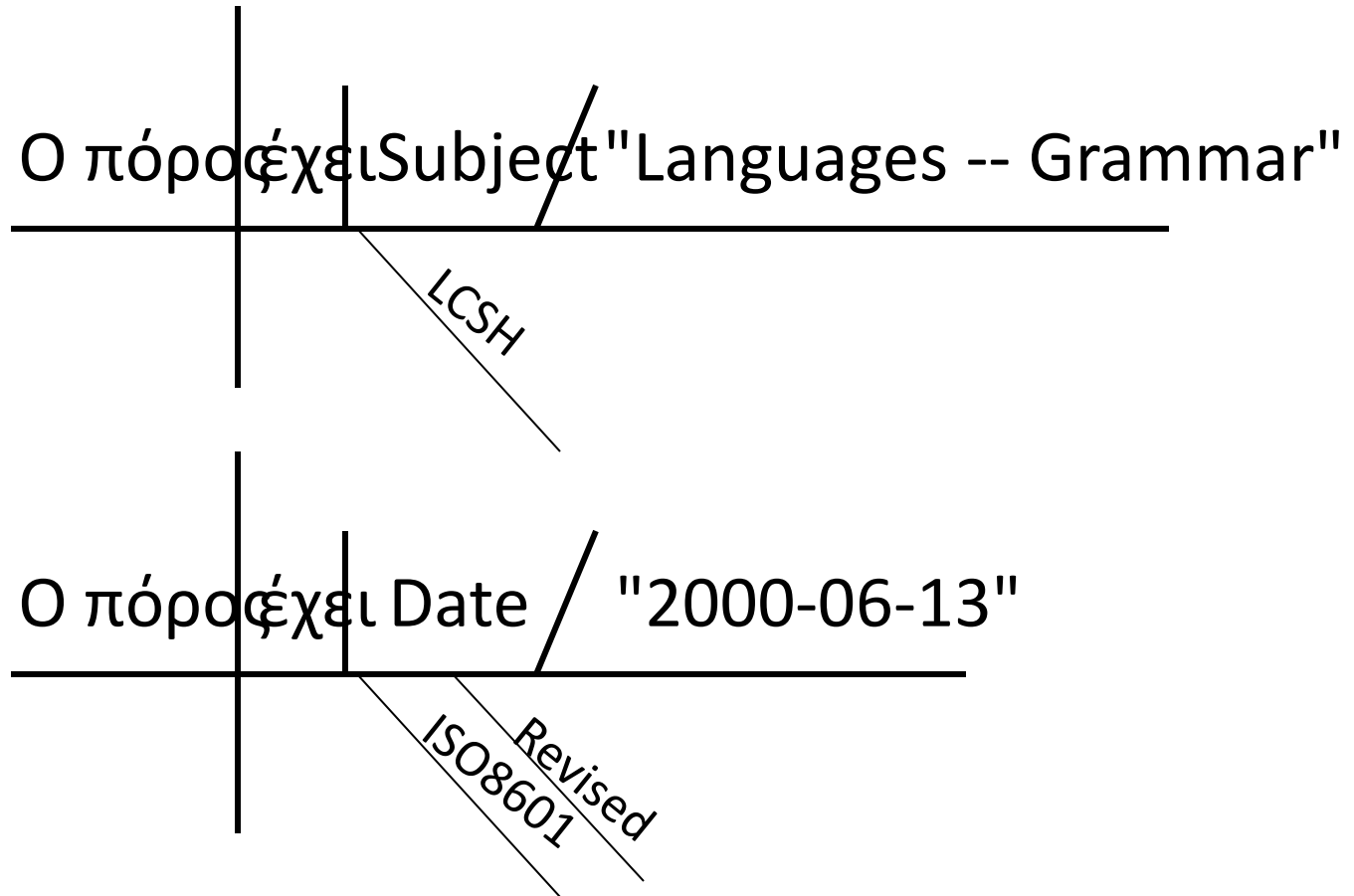
Τα Μεταδεδομένα ως Γλώσσα

- Διαφορετικές κοινότητες (περιγραφής πόρων) μιλούν διαφορετικές γλώσσες
 - 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

Κανόνες Γραμματικής DC



Παραδείγματα Γραμματικής DC



Οι Γλώσσες Μεταδεδομένων είναι Πολυγλωσσικές

- Τα μεταδεδομένα δεν είναι ομιλούμενη γλώσσα
- Τα ουσιαστικά των μεταδεδομένων – τα «στοιχεία» - είναι σύμβολα που αντιπροσωπεύουν έννοιες εκφράσιμες σε πολλαπλές φυσικές γλώσσες
- Τα πρότυπα μπορεί να έχουν δεκάδες

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**

Σύνταξη Μεταδεδομένων

Κάθε γλώσσα έχει ένα (ή περισσότερους) τρόπους **γραφής** της γλώσσας, για ανταλλαγή (ή αποθήκευση) πληροφοριών

- ISO2709, για MARC
- HTML – “meta”
- RDF, βασιζόμενο σε XML
- DC, MODS

Διαλειτουργικότητα με Μεταδεδομένα

Κύριες προσεγγίσεις
διαλειτουργικότητας

- Αναζήτηση σε διαφορετικά συστήματα
 - Δεν έχουν τα ίδια μεταδεδομένα, αλλά αντιστοιχούν τις δυνατότητες αναζήτησης σε ένα κοινό σύνολο
 - Π.χ. Z39.50

Προκλήσεις και Τάσεις στα Μεταδεδομένα

- «Απλού και ευρύ» και «πλούσιου αλλά συγκεκριμένου»
 - απλότητα ↔ λειτουργικότητα
- «Γενικού» και «προσαρμοσμένου»
 - διαλειτουργικότητα ↔ επεκτασιμότητα
- «Αναγνώσιμου από ανθρώπους» και «επεξεργάσιμου από μηχανές»
 - σε δημιουργία και χρήση

Παρανοήσεις για Μεταδεδομένα

- *Η πιο λεπτομερής περιγραφή μεταδεδομένων είναι πάντα καλύτερη;*
- Συνήθως ισχύει για συμβατικά τεκμήρια
 - Που δεν διαθέτουμε **άλλη πληροφορία** για αυτά
- Αλλά θα γίνουν και περισσότερα

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)

Κατηγοριοποίηση

- Administrative
- Descriptive
- Preservation
- Technical
- Use

Table 1. Different Types of Metadata and Their Functions

Type	Definition	Examples
Administrative ⁷	Metadata used in managing and administering information resources	<ul style="list-style-type: none"> - Acquisition information - Rights and reproduction tracking - Documentation of legal access requirements - Location information - Selection criteria for digitization - Version control and differentiation between similar information objects - Audit trails created by recordkeeping systems
Descriptive	Metadata used to describe or identify information resources	<ul style="list-style-type: none"> - Cataloging records - Finding aids - Specialized indexes - Hyperlinked relationships between resources - Annotations by users - Metadata for recordkeeping systems generated by records creators
Preservation	Metadata related to the preservation management of information resources	<ul style="list-style-type: none"> - Documentation of physical condition of resources - Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration

Technical	Metadata related to how a system functions or metadata behave	<ul style="list-style-type: none"> - Hardware and software documentation - Digitization information, e.g., formats, compression ratios, scaling routines - Tracking of system response times - Authentication and security data, e.g., encryption keys, passwords
Use	Metadata related to the level and type of use of information resources	<ul style="list-style-type: none"> - Exhibit records - Use and user tracking - Content re-use and multi-versioning information

Όψεις

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

Στοιχεία δεδομένων

- Στοιχείο δεδομένων (data element)
 - ενότητα δεδομένων για την οποία ο ορισμός (definition), η αναγνώριση (identification), η αναπαράσταση (representation) και οι επιτρεπτές τιμές (values) καθορίζονται μέσω ενός συνόλου ιδιοτήτων (attributes).
- Τα στοιχεία δεδομένων εφαρμόζονται στα πληροφοριακά μοντέλα επιχειρήσεων, στα μοντέλα δεδομένων, στα διαγράμματα ροής δεδομένων, στο σχεδιασμό βάσεων δεδομένων, στις προδιαγραφές διεπαφών και

Επίπεδα Μεταδεδομένων

- Set
 - ψηφιακή συλλογή αντικειμένων
- Aggregate
 - κατηγορία ψηφιακών αντικειμένων
 - π.χ. εικόνες
- Primary Object
 - πρωτογενές αντικείμενο, ισοδύναμο του φυσικού, π.χ. βιβλίο
- Intermediate Object
 - άποψη ή συστατικό του πρωτογενούς αντικειμένου
 - π.χ. παρουσίαση των σελίδων του βιβλίου ως εικόνες ή ως κείμενο
- Terminal Object

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

Παράδειγμα

Υποχρεωτικά

- Τίτλος
- Συγγραφέας
- Θέμα
- Ηλεκτρονική διεύθυνση

Παράδειγμα

Επιθυμητά

- Εκδότης
- Ημερομηνία
- Κατηγορία
- Σχόλια

Παράδειγμα

Άλλα σχετικά με συλλογή

- Αναγνωριστικά
- Γλώσσα
- Έκδοση
- Πνευματικά δικαιώματα
- ...

Προσαρμογή Μεταδεδομένων

- Τα μεταδεδομένα **προσαρμόζονται** σε πολλαπλές παραλλαγές / ποικιλίες
 - Αποτυπώνοντας την **ανομοιομορφία** των δημιουργών και διατηρητών μεταδεδομένων
 - Προσφέροντας για κάθε **κοινότητα** (συγκεκριμένο χώρο) εξειδικευμένη λειτουργικότητα, δημιουργία.

Προφίλ Εφαρμογής

- Ένα σχήμα μεταδεδομένων που ενσωματώνει ένα **σύνολο στοιχείων** από ένα ή περισσότερα **σύνολα στοιχείων μεταδεδομένων**
 - λεξιλογίων
 - χώρους ονομάτων – namespaces
- Ένα σύνολο από **πολιτικές** που ορίζουν πώς τα στοιχεία πρέπει να

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

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

Εργασία

- Γράψτε μια παράγραφο για κάθε ένα από τα πρότυπα που αναφέρονται στις διαφάνειες 22-23
- Βρείτε ενδεικτικά αρχεία για τρία από αυτά