

## Περιεχόμενο

- Τι είναι Μεταδεδομένα
- Παραδείγματα
- Δεδομένα ή Μεταδεδομένα
- Βασικές έννοιες
  - Ορολογία, ηλεκτρονικά δημοσιεύματα, πρότυπα
- Παραδόσεις Μεταδεδομένων
  - Κατηγορίες, πρότυπα

## Τι είναι Μεταδεδομένα

- “Machine understandable information about web resources or other things.”
  - Berners-Lee, 1997
- “Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics. A user might be a program or a person.”
  - Dempsey and Heery, 1998
- “Data about data”
- «structured information about information»

## Μεταδεδομένα είναι ...

- «**Δομημένα** δεδομένα για άλλα δεδομένα»
  - Που δεν είναι μέρος των αρχικών δεδομένων
- Δομημένα δεδομένα για πόρους, που μπορεί να χρησιμοποιηθούν για να υποστηρίξουν ένα μεγάλο εύρος λειτουργιών
  - Καταλογόγραφηση, Απογραφή, Μητρώο ...

## Παράδειγμα Μεταδεδομένων

- Συναντώνται καθημερινά, π.χ., σε ένα σύστημα αρχείων:
  - Όνομα αρχείου
  - Τύπος αρχείου
  - Μέγεθος αρχείου
  - Ημερομηνία αλλαγής αρχείου
  - Επιπλέον σε αρχεία εικόνων
    - Φυσικές διαστάσεις εικόνας
    - Ανάλυση, μορφή αποθήκευσης, τρόπος συμπίεσης

## Παράδειγμα Μεταδεδομένων

Για περιγραφή βιβλίων:

- ΤΙΤΛΟΣ
- ΣΥΓΓΡΑΦΕΑΣ
- ΕΚΔΟΤΗΣ
- ΗΜ/ΝΙΑ ΕΚΔΟΣΗΣ
- ISBN
- ...

Το πρότυπο MARC περιγράφει τέτοια μεταδεδομένα

## Παράδειγμα Μεταδεδομένων

Σε Βάση Δεδομένων, που περιγράφει άρθρα

- ΤΙΤΛΟΣ ΑΡΘΡΟΥ
- ΣΥΓΓΡΑΦΕΑΣ ΑΡΘΡΟΥ
- KEYWORDS
- ΤΙΤΛΟΣ ΠΕΡΙΟΔΙΚΟΥ
- ΕΚΔΟΤΗΣ ΠΕΡΙΟΔΙΚΟΥ
- ISSN
- ΤΕΥΧΟΣ, ΣΕΛΙΔΕΣ
- ...

## Παράδειγμα Μεταδεδομένων

Περιγραφής σελίδων στο WWW

- ΤΙΤΛΟΣ
- ΔΗΜΙΟΥΡΓΟΣ
- ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ
- ΠΕΡΙΓΡΑΦΗ
- ΤΥΠΟΣ – π.χ. κείμενο
- ΜΟΡΦΟΤΥΠΟ – HTML
- ...

## Παράδειγμα Μεταδεδομένων

Σε Ψηφιακή Βιβλιοθήκη Διδακτορικών

- ΤΙΤΛΟΣ
- ΔΗΜΙΟΥΡΓΟΣ (ΣΥΓΓΡΑΦΕΑΣ)
- ΠΕΡΙΓΡΑΦΗ (ΠΕΡΙΛΗΨΗ)
- ΗΜ/ΝΙΑ ΑΠΟΝΟΜΗΣ
- ΗΜ/ΝΙΑ ΕΙΣΑΓΩΓΗΣ
- COPYRIGHT NOTICE
- KEYWORDS

## Μεταδεδομένα και Δομή

- Τα μεταδεδομένα μπορεί να είναι
  - από ελάχιστα δομημένα, χαμηλής ποιότητας (π.χ. απλά σχόλια – έχοντας «τετριμμένη δομή»)
  - έως πολύ δομημένα, υψηλής ποιότητας (με πρότυπα και αυστηρούς κανόνες, π.χ. MARC+AACR, TEI, ...)
- Όσο πιο δομημένα, τόσο πιο αποτελεσματικά
  - Και πιο δύσκολα στην δημιουργία / συντήρηση
- Είτε είναι ηλεκτρονικά, είτε σε καρτέλες, είτε ...
- Θέλουμε να είναι επεξεργάσιμα από υπολογιστές
  - Για (ψηφιακή) αυτοματοποίηση

## Τι Αντικείμενα Περιγράφουν τα Μεταδεδομένα;

- Τεκμήρια σε HTML
- Ψηφιακές εικόνες
- Βάσεις δεδομένων
- Βιβλία
- Μουσειακά αντικείμενα
- Εγγραφές αρχείων
- Εγγραφές μεταδεδομένων
- ... κλπ
- Συλλογές
- Υπηρεσίες
- Φυσικούς τόπους
- Ανθρώπους
- Αφαιρετικές περιγραφές
- Έννοιες
- Γεγονότα

## (Μετά)? δεδομένα

Διαφορές στο χειρισμό:

- **Δεδομένα** – τα αντικείμενα που αναζητούμε
  - Χωρίς αυτά, τα μεταδεδομένα δεν έχουν λόγο ύπαρξης
- **Μεταδεδομένα** – στοιχεία περιγραφής των δεδομένων
  - Σε Ψηφιακή Βιβλιοθήκη:
    - Δεδομένα: video, ηλεκτρονικά βιβλία, ...
    - Μεταδεδομένα: περιγραφές με: ΤΙΤΛΟΣ, ΘΕΜΑ, ...
  - Σε OPAC:
    - Δεδομένα / Μεταδεδομένα: εγγραφές με: ΤΙΤΛΟΣ, ΘΕΜΑ, ...

## (Μετά)? Δεδομένα σε OPAC

- Τα ίδια αντικείμενα μπορεί να είναι δεδομένα ή μεταδεδομένα, ανάλογα με το σύστημα αναφοράς
- Οι εγγραφές στο OPAC με: ΤΙΤΛΟΣ, ΘΕΜΑ, ...
  - Με σύστημα αναφοράς την βιβλιοθήκη (ή την ψηφιακή βιβλιοθήκη)
    - Είναι μεταδεδομένα (τα δεδομένα είναι τα βιβλία, ...)
  - Με σύστημα αναφοράς το OPAC
    - Είναι δεδομένα (αλλά μπορεί να τα καλούμε «μεταδεδομένα»)

## Τι είναι Δεδομένα

- Τι είναι δεδομένα
  - Οτιδήποτε γνωρίζουμε ...στη γενική περίπτωση
  - Τα αντικείμενα που αναζητάμε ή αναφερόμαστε
- Τα μεταδεδομένα **είναι** δεδομένα
  - Που αφορούν άλλα δεδομένα
  - Αρα ειδική κατηγορία, πιο συγκεκριμένη ...
- Στο συμβατικό κόσμο, τα ξεχωρίζουμε:
  - Τα δεδομένα έχουν **πάντα** φυσική υπόσταση
  - Τα μεταδεδομένα έχουν φυσική υπόσταση; (καρτέλες;)
- Στον ψηφιακό κόσμο όμως;

## Είναι Δεδομένα ή Μεταδεδομένα

- Είναι Δεδομένα ή Μεταδεδομένα
  - Ο τίτλος ενός βιβλίου
  - Μια κριτική βιβλίου
  - Ο συγγραφέας ενός βιβλίου
- Είναι μεταδεδομένα
  - Όταν έχουν βοηθητικό χαρακτήρα για την αναζήτηση
- Είναι (μόνο) δεδομένα
  - Όταν τα προσδιορίζουμε, χειριζόμαστε, αναζητούμε, ...

## Σπουδαιότητα Μεταδεδομένων

- Σε μια παραδοσιακή βιβλιοθήκη
  - Παράγονται – αναγνωρίζονται (π.χ. τίτλος, συγγραφέας)
  - Συγκεντρώνονται (στον κατάλογο, ή στα ράφια ανά ομάδες)
  - Χρησιμοποιούνται στην αναζήτηση – από τον κατάλογο αλλά και από τα βιβλία
- Σε ηλεκτρονικές πηγές (π.χ. Διαδίκτυο)
  - Είναι ο τρόπος για να βρούμε την πληροφορία!
    - Εκτός από τεχνικές πάνω στο πλήρες περιεχόμενο.

## Δημοσίευμα και Μέσο

- Έντυπο
  - Σε χαρτί απαραίτητα, π.χ. «βιβλίο»
- Συμβατικό
  - Σε χαρτί ή άλλο μέσο που έχει φυσική υπόσταση
- Ψηφιακό (ή Ηλεκτρονικό)
  - Που διακινείται ηλεκτρονικά
- Π.χ. το «CDROM» τι είναι;
- Αντίστοιχα Βιβλιοθήκες συμβατικές, ψηφιακές ...

## Βασικές Έννοιες – Αντικείμενα

Από το γενικότερο προς το ειδικότερο

- 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) – κατά RFC 2396

- A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content--the entities to which it currently corresponds--changes over time, provided that the conceptual mapping is not changed in the process.

## 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

## «Ηλεκτρονικό» Δημοσίευμα (Τεκμήριο)

- Οτιδήποτε προσφέρει πληροφορία στον ψηφιακό κόσμο
- Με ψηφιακή υπόσταση
    - Αρχεία διάφορων μορφών
    - Εφαρμογές λογισμικού
  - Υπηρεσίες
- Είδη ηλεκτρονικών τεκμηρίων:
- Μορφοποιημένα (σταθερά) – π.χ. σελίδες βιβλίων, εικόνες
  - Μορφοποιήσιμα (μεταβλητά) – π.χ. αρχεία επεξεργαστή κειμένου, πίνακες αριθμητικών δεδομένων
  - Δυναμικά, δημιουργούμενα κατά την ανάκτηση τους (από προγράμματα)

## Μορφές Ηλεκτρονικών Τεκμηρίων

- Κείμενο
- Εικόνες
- Κινούμενες εικόνες
- Ήχος
- Βίντεο
- Ιστοσελίδες
- Προγράμματα
- ...

## Παραδείγματα Τύπων Αρχείων Ηλεκτρονικών Τεκμηρίων

- Κείμενο
  - .DOC
  - .TXT
  - .RTF
  - .PDF
  - .PS
  - .TEX
- Ιστοσελίδες
  - .HTM
  - .HTML
  - .XML
  - .HTMLS
  - .DHTML
- Προγράμματα
  - .COM
  - .EXE
  - .PL
- Άλλα δεδομένα
  - .XLS .PPT .MDB
  - .ZIP

## Παραδείγματα Τύπων Αρχείων Ηλεκτρονικών Τεκμηρίων

- Εικόνες
  - .BMP
  - .GIF
  - .JPEG
  - .TIF
  - .EPS
- Κινούμενες εικόνες
  - .ANI
  - .FLI
  - .FLC
- Ήχος
  - .WAV
  - .MID
  - .SND
  - .AUD
- Βίντεο
  - .AVI
  - .MOV
  - .MPG
  - .QT

## Διαχείριση Τεκμηρίων

- Παραγωγή
- Προσδιορισμός
- Αποθήκευση
- Αναζήτηση
- Ανάκτηση
- Έλεγχος έκδοσης
- Διαχείριση ροής εργασιών
- Παρουσίαση
- Διάδοση

## Ψηφιακό Περιεχόμενο και Βιβλιοθήκη

- **Ψηφιακό Περιεχόμενο:** Πληροφορία (επεξεργασμένη ή μη) σε οποιοδήποτε υπολογιστικό σύστημα ή διαδικτυακή πλατφόρμα
- **Ψηφιακή Βιβλιοθήκη:** Συγκέντρωση (ή εξασφάλιση πρόσβασης) πληροφορίας (περιεχόμενο) σε ψηφιακά μέσα, οργανωμένο για εύκολη πρόσβαση
  - Δεν είναι η «Αυτοματοποιημένη βιβλιοθήκη»

## Ανταλλαγή Δεδομένων

- Τα δεδομένα (και τα μεταδεδομένα) τα
  - Χρησιμοποιούμε (εσωτερικά στην εφαρμογή)
  - Αποθηκεύουμε (σε αρχεία)
  - Μετατρέπουμε, σε άλλες μορφές [εύκολα;]
  - **Ανταλλάσσουμε**
    - Με αποθήκευση σε αρχεία (και φύλαξη για μετά)
    - Μέσω πρωτοκόλλων (μεταξύ εφαρμογών)

## Πρότυπα για Ανταλλαγή

- Για εικόνα, ήχο, βίντεο, κλπ
- Για γραφήματα, πίνακες
- Για μορφοποιημένα κείμενα
- Για κωδικοποίηση γραμμάτων / γραμματοσειρές
- Για βάσεις δεδομένων
- Για μεταδεδομένα
- Για πρωτόκολλα επικοινωνίας
- ...

## Πρότυπα

- Χρειάζονται για να έχουμε κοινή αντίληψη των δεδομένων που περιγράφονται με αυτά
- Μας περιορίζουν στην **ευελιξία**
- Προσφέροντας μεγαλύτερη **δομή**
- ⇒ Επεξεργάζονται ευκολότερα **μηχανικά**
- Αναγκαιότερα σε **ψηφιακά αντικείμενα**
  - Αφού στα συμβατικά έχουμε αισθητήρια αντίληψη
- Πρώιμη έκφρασή τους, για κριτική, με τα RFC (Request For Comment)

## Σχήμα (Schema)

Σχήμα έχουν:

- Βάσεις δεδομένων, XML, RDF, μεταδεδομένα

Σχήμα είναι ο ορισμός της δομής των δεδομένων που περιγράφουν, δηλαδή

- Των «πεδίων» που χρησιμοποιούνται
- Των τύπων των τιμών τους
- Των κανόνων και περιορισμών

## Παραδόσεις Μεταδεδομένων

- Τα Μεταδεδομένα είναι δομημένα δεδομένα για άλλα δεδομένα
- Τρεις παραδόσεις:
  - Συστήματα διαχείρισης βάσεων δεδομένων
    - Σχήματα των σχεσιακών βάσεων δεδομένων
  - Παραδοσιακοί κατάλογοι βιβλιοθήκης
  - Μεταδεδομένα Ιστού

## Καταλογογράφηση Βιβλιοθήκης

- 19ος αιώνας: βιβλιογραφικές εγγραφές (καρτέλες) και κανόνες καταλογογράφησης
- 1960+: Machine-Readable Card Catalogs (παραλλαγές MARC)
- 1970-1980: on-line δίκτυα βιβλιοθηκών και επέκταση κανόνων καταλογογράφησης (AACR2, RAK)

## Συστήματα Οργάνωσης Γνώσης στο Ψηφιακό Περιβάλλον

- Σχήματα οργάνωσης πληροφορίας και διευκόλυνσης διαχείρισης γνώσης
- Λειτουργούν ως γέφυρες μεταξύ των αναγκών πληροφόρησης των χρηστών και του υλικού μιας συλλογής.
  - λίστες όρων (λεξικά)
  - σχήματα ταξινόμησης (LCSH)
  - λίστες σχέσεων όρων (θησαυροί)
- Ποικίλουν ως προς την περιπλοκότητα, τη δομή και τη λειτουργία
- Μπορούν να βελτιώσουν την οργάνωση των ψηφιακών βιβλιοθηκών και να διευκολύνουν την πρόσβαση στο περιεχόμενό τους.

## Μεταδεδομένα Παραδοσιακής Βιβλιοθήκης

- Ευρετήρια (κατάλογοι)
- Σχήματα ταξινόμησης (LCCS, DEWEY)
- Ταξιθέτηση (οργάνωση στο χώρο)

## Μεταδεδομένα Ψηφιακής Βιβλιοθήκης

- Περιγραφικά: *Συγγραφέας, τίτλος, λέξεις-κλειδιά, έκδοση, θέματα*
- Δομικά: *Συστατικά μέρη, σχέσεις συστατικών, πλοήγηση, είδη αντικειμένων, κλπ.*
- Διοικητικά: *Ημερομηνία δημιουργίας, τελευταία τροποποίηση, υπεύθυνος, οδηγίες χειρισμού*
- Όροι & Προϋποθέσεις: *Δικαιώματα χρήσης, πρόσβασης, εκτύπωσης, αντιγραφής, επαναχρησιμοποίησης, αναδημοσίευσης; κόστος*
- Αναζήτηση-εύρεση
- Αποθήκευση-παρουσίαση

## Υποστηριζόμενες Λειτουργίες Ψηφιακής Βιβλιοθήκης

- Εύρεση
- Έλεγχος πρόσβασης
- Παρουσίαση
- Διαχείριση
- Ψηφιακή συντήρηση
- Αναδιάρταξη

## Ιδιότητες Μεταδεδομένων

- Μια μοναδική ονομασία για το κάθε πεδίο (tag, label, identifier, field name)
- Ένα ορισμό κάθε πεδίου
- Το εάν ή όχι ένα πεδίο είναι υποχρεωτικό, ή προαιρετικό
- Το εάν είναι ή όχι επαναλαμβανόμενο
- Την οργάνωση σχέσεων μεταξύ των πεδίων (π.χ. σχέσεις ιεράρχησης)
- Περιορισμοί για τις δυνατές τιμές του πεδίου (κείμενο, αριθμητικό εύρος, ημερομηνία, ελεγχόμενο λεξιλόγιο)
- Προαιρετική υποστήριξη για στοιχεία τοπικού ενδιαφέροντος

## MARC21 (USMARC)

0XX = Control information, numbers, codes  
1XX = Main entry  
2XX = Titles, edition, imprint  
3XX = Physical description, etc.  
4XX = Series statements  
5XX = Notes  
6XX = Subject access fields  
7XX = Name, etc. added entries or series; linking  
8XX = Series added entries; holdings and locations  
9XX = Reserved for local implementation

## USMARC

Περιγράφεται τυπικά από πολλές προδιαγραφές:

- *USMARC Format for Bibliographic Data*
- *USMARC Format for Holdings Data*
- *USMARC Format for Authority Data*
- *USMARC Format for Classification Data*
- *USMARC Format for Community Information*

## Ένα Μεγάλο Σχήμα

- Η εγγραφή MARC / AACR2 συνδυάζει:
  - Περιγραφή (των διαφορετικών τύπων αντικειμένων, όπως τα βιβλία και οι ταινίες)
  - Ανάλυση θέματος (τι αφορά ένα βιβλίο ή μια ταινία, όπως οι "εμπορικές ενόσεις - Αυστραλία")
  - Πληροφορίες holdings: θέσεις που βρίσκονται
  - Εγγραφές που δημιουργούνται από ιδιαίτερα εκπαιδευμένους επαγγελματίες
- Φιλοσοφία: ένα μεγάλο σχήμα («ο κατάλογος καρτών») μπορεί να ικανοποιήσει τις ανάγκες των χρηστών για όλους τους τύπους εργασιών

## Καθορισμός Μεταδεδομένων

- Δεν υπάρχει μία λύση για όλους, "one size fits all"
- Εξαρτώνται από τις λειτουργίες που υποστηρίζουν
- Πρέπει να καθορίζονται με βάση **πρότυπα**:
  - Αξιοποιούν την εμπειρία και επιτηδειότητα των άλλων
  - Παρέχουν μια βάση κοινής πρακτικής
  - Αντικατοπτρίζουν συναίνεση, διευκολύνοντας την ανταλλαγή, πρόσβαση και διαλειτουργικότητα
  - Μπορεί να υποστηρίζονται σε εργαλεία λογισμικού
- «Διαχειριστικά» μεταδεδομένα – διαφορετικά σε κάθε σύστημα

## Ο Μύθος της Κατηγοριοποίησης

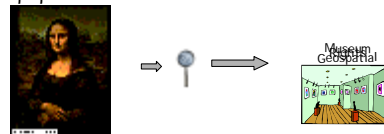
...there is no classification of the universe that is not fictional and conjectural.

...δεν υπάρχει κατηγοριοποίηση του σύμπαντος που να μην είναι φανταστική και αυθαίρετη

– Jorge Luis Borges

## Φακοί και Όψεις

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



## Πόσα Πρότυπα Μεταδεδομένων

- Δεν υπάρχει ένα μοναδικό διεθνές πρότυπο για Μεταδεδομένα, γιατί:
  - Διαφορετικά επίπεδα πολυπλοκότητας, από πλούσιες μέχρι απλές μορφές
  - Κάμποσα σχήματα μεταδεδομένων έχουν προταθεί για διαφορετικά επίπεδα και απαιτήσεις
  - Γιατί επεκτείνουμε τα υπάρχοντα πρότυπα
- Πότε πρέπει να αναζητούμε άλλο πρότυπο ή άλλη σύσταση (RFC – Request For Comment);

## Πρότυπα ... το Πρόβλημα

- Πως μπορεί ένα project να κάνει καλή χρήση των συστάσεων και προτύπων μεταδεδομένων για να αντιμετωπίσει τις ειδικές απαιτήσεις των εφαρμογών του;
  - Υπάρχουν πολλές συστάσεις και λίγες οδηγίες χρήσης
  - Πώς βρίσκουμε πώς τα άλλα projects χρησιμοποιούν τις συστάσεις και τα πρότυπα;
  - Στην πράξη, γενικά πρότυπα (π.χ., Dublin Core) πρέπει να συνδυαστούν με πρότυπα που εξαρτώνται από τις εφαρμογές (π.χ., MPEG-7)

## Συνδυάζοντας Πολλαπλά Πρότυπα στην Πράξη

- Εξειδικευμένες εφαρμογές [κυβερνητικής πληροφόρησης, εκπαίδευσης, μαθηματικών, πολυμέσων, κλπ] μπορεί να χρειάζονται να:
  - Χρησιμοποιούν στοιχεία γενικής χρήσης, π.χ. Dublin Core
  - Χρησιμοποιούν στοιχεία από άλλα, πιο εξειδικευμένα στο πεδίο, πρότυπα
  - Να ακολουθούν άλλα projects στις πρακτικές των σχημάτων τους
  - Να επινοούν τοπικά στοιχεία εκτός της εμβέλειας των υπαρχόντων προτύπων

## Μεταδεδομένα και Πρότυπα

- Για να είναι τα μεταδεδομένα χρήσιμα και επωφελή είναι ουσιαστικό
  - Η Δομή, Σημασιολογία και Σύνταξη να συμμορφώνονται με τα πρότυπα
  - Η απεικόνιση της ουσίας των πηγών
  - Ένα καταναμημένο μοντέλο μεταδεδομένων