

# Πρότυπα Μεταδεδομένων

Χρήστος Παπαθεοδώρου (paratheodor@ionio.gr)

Αναπληρωτής Καθηγητής

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

και

Μονάδα Ψηφιακής Επιμέλειας,

Ινστιτούτο Πληροφοριακών Συστημάτων και Προσομοίωσης  
Ερευνητικό Κέντρο «Αθηνά»



# Περιεχόμενα

- Εισαγωγή στα μεταδεδομένα
- Dublin Core
- MODS
- Αρχειακά 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

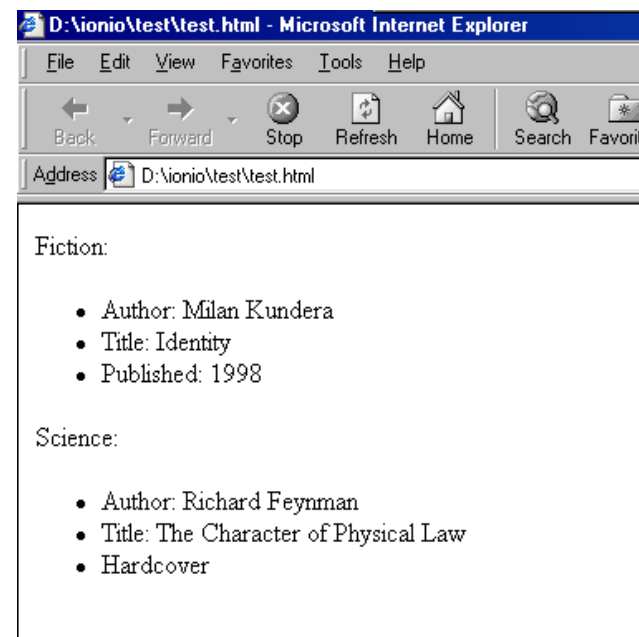
- Τα μεταδεδομένα είναι **δομημένη πληροφορία** που περιγράφει, εξηγεί, εντοπίζει ή διευκολύνει την ανάκτηση, τη χρήση ή την διαχείριση ενός ψηφιακού τεκμηρίου
- Συχνά καλούνται («**δομημένα**») «δεδομένα για άλλα δεδομένα» ή «πληροφορία για άλλη πληροφορία»
- "**Metadata** is machine understandable 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>
<UL><LI>Author: Milan Kundera</LI>
    <LI>Title: Identity</LI>
    <LI>Published: 1998</LI>
</UL>
<p> Science: </p>
<UL><LI>Author: Richard Feynman</LI>
    <LI>Title: The Character of Physical Law</LI>
    <LI>Hardcover</LI>
</UL>
</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
- eXtensible Markup Language (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)

# 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

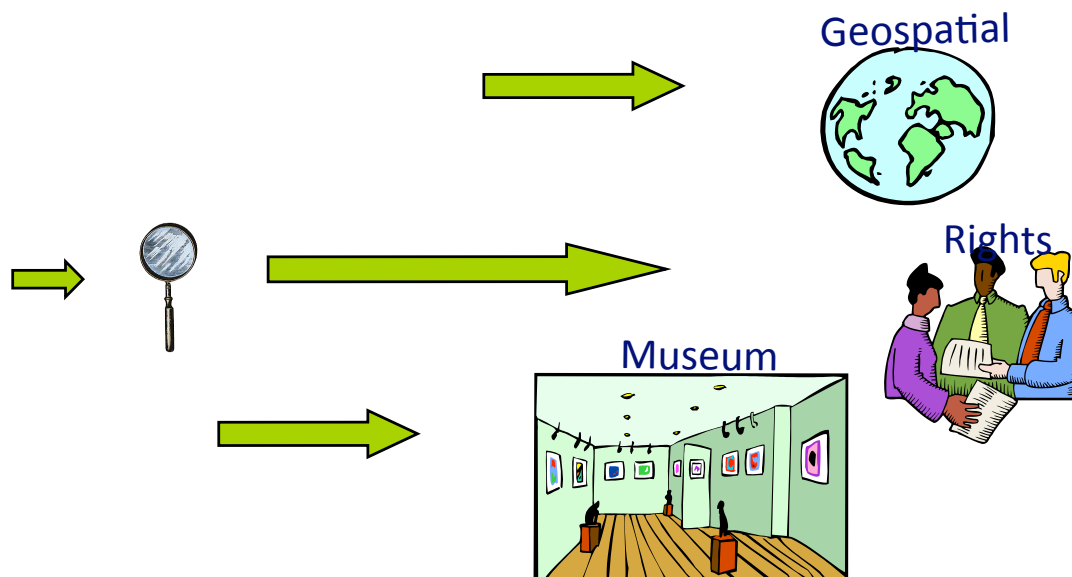
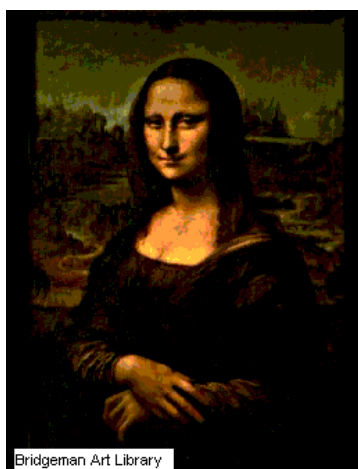
Σύμφωνα με την IFLA:

- 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

## Αλλα σχετικά με συλλογή

- Αναγνωριστικά
- Γλώσσα
- Έκδοση
- Πνευματικά δικαιώματα

▪ .....

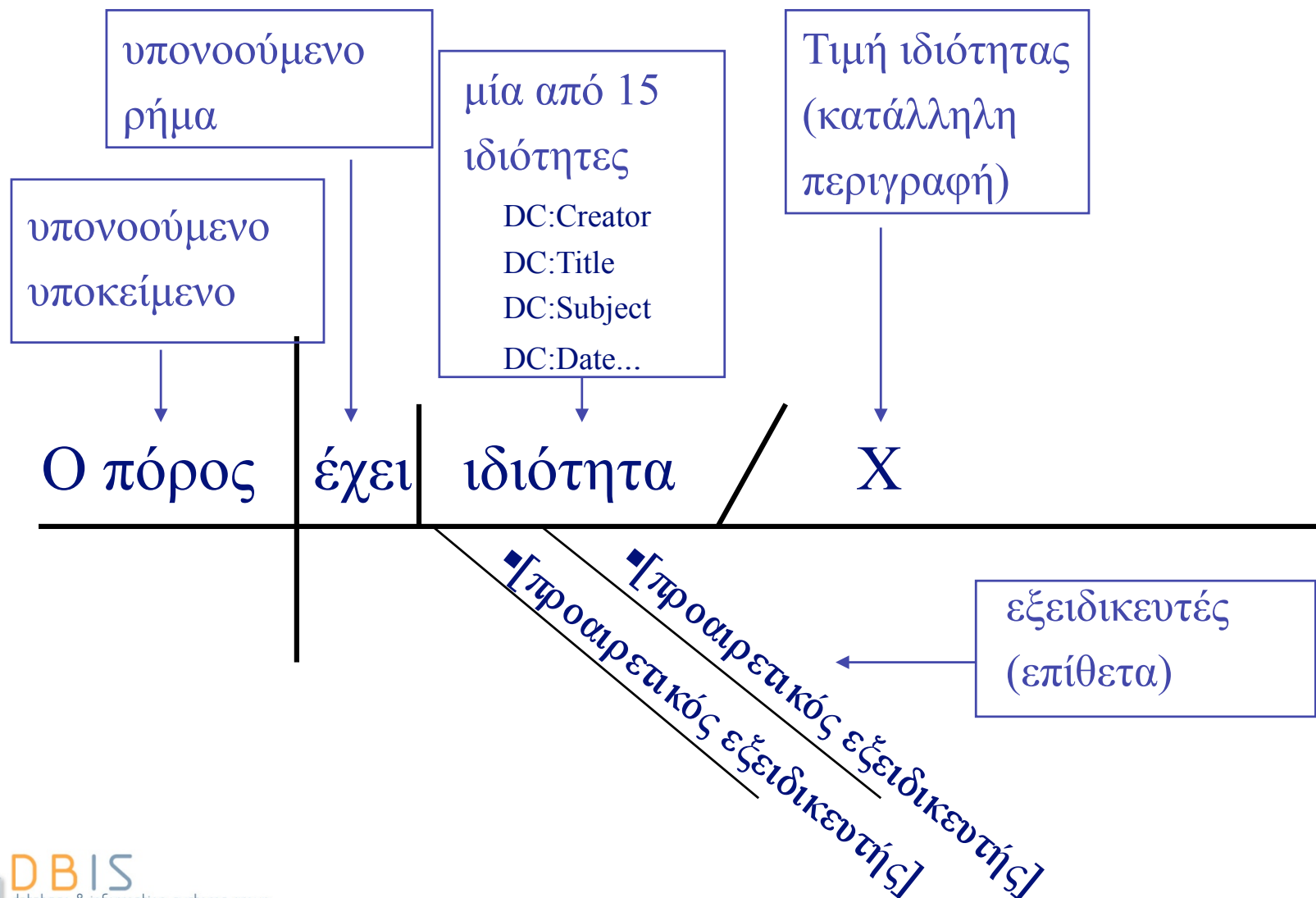
# Τα Μεταδεδομένα είναι Γλώσσα

- 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



# Dublin Core

# Το Πρότυπο «Dublin Core»

ANSI/NISO Z39.85-2001

ISSN: 1041-5653

## The Dublin Core Metadata Element Set

**Abstract:** Defines fifteen metadata elements for resource description in a cross-disciplinary information environment.

An American National Standard  
Developed by the  
National Information Standards Organization

Approved September 10, 2001  
by the  
American National Standards Institute

Published by the National Information Standards Organization  
Bethesda, Maryland



NISO Press, Bethesda, Maryland, U.S.A.

- Προτυποποίηση
  - ISO 15836-2003
  - US: NISO Z39.85-2001
  - Ευρώπη: αναγνώριση από το CEN/ISSS Workshop Agreement 13874-2000
- Συστήνεται ανταλλαγή με RDF/XML

# Χαρακτηριστικά του «Dublin Core»

- «Dublin Core» σημαίνει «Dublin, Ohio» / OCLC
- Πρωτοβουλία για να βελτιώσει την ανακάλυψη πόρων στο Διαδίκτυο
- Κοινός παρονομαστής για επικοινωνία / διαλειτουργικότητα
- Έναυσμα για σύγκλιση των προτύπων
- Επεκτάσιμο, για να καλύψει τις επιπρόσθετες ανάγκες ανακάλυψης πόρων των διαφορετικών εφαρμογών / περιοχών

# Dublin Core – Simple

- Το Dublin Core έχει 15 στοιχεία. Κάθε ένα από αυτά είναι προαιρετικό και επαναλαμβανόμενο
- Τα 15 στοιχεία 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** – (παραγωγής)
- 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
- Αναγνωριστικό (ή Προσδιοριστής ή Κωδικός Ταύτισης) / **Identifier**
  - Μοναδικό προσδιοριστικό, π.χ. URL, ISBN, ...

# Στόχοι του Dublin Core

- Απλότητα δημιουργίας και διατήρησης
  - Μη ειδικοί να δημιουργούν περιγραφικές εγγραφές για αποτελεσματική ανάκτηση σε δικτυωμένο περιβάλλον
- Κοινά κατανοητή (διαθεματική) σημασιολογία
  - Σύγκλιση κοινών, περισσότερο γενικών στοιχείων
  - Αυξημένη ορατότητα και προσβασιμότητα
  - Κατάλληλο και για τον μη ειδικό της αναζητητής
    - Τον «ψηφιακό τουρίστα»



# Χρήση του Dublin Core

- Είναι βασικός **πυρήνας** στοιχείων
- **Δεν** είναι υποκατάστατο σε πλουσιότερα περιγραφικά πρότυπα
- Παρέχει 15 «παράθυρα» ή ευρύχωρα «καλάθια» από πλουσιότερη περιγραφή πόρων
  - Φανερώνει πλούσιες περιγραφές σε απλή μορφή
  - Σημασιολογικά σταυροδρόμια, αντιστοιχίσεις σε υπάρχοντα δεδομένα

# 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

# Dublin Core Qualified

# 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)
  - Σαν 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!

# Dublin Core: μη Εξειδίκευση

## Στοιχείων

- Creator
- Subject
- Publisher
- Contributor
- Type
- Identifier
- Source
- Language
- Rights

# Dublin Core: Εξειδίκευση Στοιχείων (1/2)

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 (2/2)

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



# Κωδικοποίηση Τιμών (encoding schemes)

- 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 τον ερμηνευτή (λογισμικό), η τιμή πρέπει να είναι «κατάλληλη» και χρησιμοποιήσιμη για ανακάλυψη πόρων

# Dublin Core: μη Κωδικοποίηση Τιμών

- Title
- Creator
- Description
- Publisher
- Contributor
- Rights

# Κωδικοποίηση Τιμών (1/2)

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

## Κωδικοποίηση Τιμών (2/2)

Στοιχείο	Σχήμα κωδικοποίησης
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)

# Παραδείγματα Dublin Core

# 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" lang="LANG"  
      scheme="SCHEME" content = "ELEMENT_VALUE">
```

- Για παράδειγμα:

```
<meta name = "DC.Title" content = "Το λίγο του κόσμου">
```

```
<meta name = "DC.Creator" content = "Δημουλά, Κική">
```

# 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.Date.Issued" content = "1990">
<meta name = "DC.Language" scheme = "ISO639-2" content = "gre">
</head>
<body><pre>
    ...Υπήρξα περίεργη και μελετηρή. ...
</pre></body>
</html>
```



# Dublin Core σε XML

S. Abiteboul, P. Buneman, D. Suciu “*Data on the Web: From Relations to Semistructured Data and XML*” Morgan Kaufmann Publishers, 2000.

---

```
<?xml version="1.0" encoding="UTF-8"?>
<book
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:isbn="http://www.isbn.org/def">
  <dc:title> Data on the Web: From Relations to Semistructured Data and XML</dc:title>
  <dc:creator> S. Abiteboul</dc:creator>
  <dc:creator> P. Buneman</dc:creator>
  <dc:creator> D. Suciu</dc:creator>
  <dc:publisher>Morgan Kaufmann Publishers </dc:publisher>
  <dc:date>2000</dc:date>
  <dc:type>Text </dc:type>
  <isbn:number>0-201-06673-4</isbn:number>
</book>
```

# Dublin Core – Επεκτασιμότητα

- Η αρχιτεκτονική του Dublin Core υποστηρίζει πιο εξελιγμένες λύσεις μεταδεδομένων
  - Αλλά προσοχή: προτιμάμε επέκταση του DC ή επιλογή άλλων πλουσιότερων σχημάτων;
- Έχουμε επεκτάσεις του Dublin Core σε εξειδικευμένους τομείς (domain-specific)
- Ομάδες εργασίας (του DCMI) αναπτύσσουν *Προφίλ Εφαρμογών (application profiles)* για ειδικούς τομείς

# 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, διαχείριση, πρόσβαση, ...
- Τηρώντας τα αντίστοιχα **πρότυπα** ...
- Μεγαλύτερη εξειδίκευση = μεγαλύτερη λειτουργικότητα = λιγότερη διαλειτουργικότητα!

# Τι είναι το Προφίλ Εφαρμογής

- Ένα σχήμα μεταδεδομένων που ενσωματώνει ένα **σύνολο στοιχείων** από ένα ή περισσότερα **σύνολα στοιχείων μεταδεδομένων** (ή λεξιλογίων ή χώρους ονομάτων – namespaces)
- Ένα σύνολο από **πολιτικές** που ορίζουν πώς τα στοιχεία πρέπει να εφαρμόζονται στο πεδίο της εφαρμογής
- Ένα σύνολο **οδηγιών** που ξεκαθαρίζουν τις πολιτικές που αφορούν τα στοιχεία

# 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 πολλαπλά λεξιλόγια
  - Μπορεί να προσαρμόζει (εξειδικεύει) ορισμούς για τοπικούς σκοπούς ή εξειδικευμένους τομείς (domain-specific)
  - Όλοι οι όροι πρέπει να ορίζονται στα λεξιλόγια
  - Μπορεί να περιλαμβάνει τοπικά ορισμένα λεξιλόγια

# 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
- Πρέπει να μπορεί να **δια-**λειτουργεί στο μέλλον

# Κατάλογος υπερσυνδέσμων

- [http://www.getty.edu/research/institute/standards/intrometadata/2\\_articles/index.html](http://www.getty.edu/research/institute/standards/intrometadata/2_articles/index.html)
- Metadata Demystified: A guide for publishers. NISO Press, July 2003.  
[http://www.niso.org/standards/resources/Metadata\\_Demystified.pdf](http://www.niso.org/standards/resources/Metadata_Demystified.pdf)
- Arms, Caroline R. "Some Observations on Metadata and Digital Libraries". In: *Conference on Bibliographic Control for the New Millennium*, September 2000. [http://lcweb.loc.gov/catdir/bibcontrol/arms\\_paper.html](http://lcweb.loc.gov/catdir/bibcontrol/arms_paper.html)
- 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice-Metadata Research and Application, 28 Sept. – 2 Oct. 2003, Washington USA. <http://www.ischool.washington.edu/dc2003/index.html>
- Guidance on the Structure, Content, and Application of Metadata Records for Digital Resources and Collections. Report of the IFLA Cataloguing Section Working Group on the Use of Metadata Schemas. Draft – for Worldwide Review 27 October, 2003. <http://www.ifla.org/VII/s13/guide/metaguide03.pdf>
- Dublin Core Metadata Initiative (DCMI). <http://dublincore.org/>
- The Dublin Core Metadata Element Set. ANSI/NISO Z39.85-2001.  
<http://www.niso.org/standards/resources/Z39-85.pdf>
- Metadata Encoding and Transmission Standard <http://www.loc.gov/standards/mets/>
- Metadata Object Description Schema" (MODS) <http://www.loc.gov/standards/mods/>
- MARC Standards Website. <http://lcweb.loc.gov/marc/>

# Κατάλογος υπερσυνδέσμων

- OCLC, Bibliographic formats and standards. <http://www.oclc.org/bibformats/>
- UKOLN's metadata site <http://www.ukoln.ac.uk/metadata/>
- IFLA <http://ifla.inist.fr/II/metadata.htm>
- IFLA Functional Requirements for Bibliographic Records.  
[www.ifla.org/VII/s13/frbr/frbr.pdf](http://www.ifla.org/VII/s13/frbr/frbr.pdf)
- "The Value of Metadata".  
<http://www.fgdc.gov/publications/documents/metadata/metabroc.html>



# Metadata Object Description Schema (MODS)

# Dublin Core

- 15 elements + qualifiers
- Metadata container & exchange medium
- Απλό, με μεγάλη επιρροή σε έργα ψηφιακών βιβλιοθηκών
- Διπλή προσέγγιση του σημασιολογικού εύρους (fields , qualifiers)
  - Σημασιολογική ευρύτητα στοιχείων (π.χ. Creator: authors, editors, κ.α.)
  - Τα qualifiers εμποδίζουν τη διαλειτουργικότητα
- MODS: ακρίβεια και διαλειτουργικότητα

# MODS

- Metadata Object Description Schema
- Library of Congress Network Development and MARC Standards Office
- Περιγράφει τη μορφή του περιεχομένου
- 19 elements, 64 sub-elements
- Επεκτείνει το σύνολο των στοιχείων ενσωματώνοντας άλλα XML documents
- Υποσύνολο του MARC αλλά όχι με αριθμητικά πεδία
  - Όλα τα στοιχεία και τα γνωρίσματα αντιστοιχούν σε MARC 21 fields
- Crosswalk MODS - MARCXML

# MODS top level elements

titleInfo

name

typeOfresource

genre

originInfo

language

physicalDescription

abstract

tableOfContents

targetAudience

note

subject

Classification

relatedItem

location

accessCondition

**extension** (δήλωση χρήσης  
στοιχείων άλλου προτύπου)

recordInfo

identifier

# Attributes

- **ID** δίνει κωδικό στο element για αναφορές σε αυτό
- **type**
  - `<name type="personal">Christos</name>`
- **lang, xml:lang** κωδικοί από το iso639-2b, δεν υπάρχει αντίστοιχο στο MARC 21
  - `<description lang="eng">District Commissioner</description>`
  - `<titleInfo xml:lang="fr" type="translated">`
- **transliteration**
- **encoding για ημερομηνίες**
  - *w3cdtf* για ISO 8601 YYYY-MM-DD.
  - *iso8601* YYYYMMDD.
  - *marc* εφαρμόζεται μόνο στο στοιχείο `<dateIssued>`
- **Point:** για ημερομηνίες με τιμές *start* και *end*
- **xlink:** `<abstract xlink:href="http://www.ojp.usdoj.gov/bjs/abstract/cchrie98.htm"/>`

# MODS Lite (1/2)

<titleInfo>

<title>

<name>

<namePart>

<typeOfResource> text, cartographic, multimedia

<genre> εξειδίκευση του <typeOfResource>

<originInfo> προέλευση του αντικειμένου

<dateIssued>, <dateCreated>, <dateCaptured>, <dateOther>

<publisher>

<language>

<physicalDescription>

<internetMediaType> sub-element του <physicalDescription>, MIME type

<extent> sub-element του <physicalDescription>

<form>

<abstract>

<tableOfContents> τα περιεχόμενα του αντικειμένου

<note>

# MODS Lite (2/2)

<subject> όροι από taxonomy

<topic> sub-element του <subject>, θεματικός όρος

<temporal> sub-element του <subject>, χρονικά διαστήματα

<geographic> sub-element του <subject>, γεωγραφικοί όροι, γεωγραφική ιεραρχία

<cartographics> sub-element του <subject>, χωρικά δεδομένα, κλίμακες κλπ

<classification> το ταξινομικό σύστημα

<relatedItem> για τα μέρη μιας συλλογής

<identifier>

<location>

<url>

<accessCondition> περιορισμοί πρόσβασης (MARC21 506, 540)

<recordInfo> τρόπος δημιουργίας του MODS record

# titleInfo: υποχρεωτικό

- αντιστοιχεί στα DC Creator, Contributor ή MARC 21 1XX, 7XX
- Attributes
  - **type** *abbreviated* (MARC 21 field 210), *translated* (242, 246), *alternative* (246, 740), *uniform* (130, 240, 730)
  - **authority** ο τίτλος ελέγχεται από ένα record σε ένα authority file (e.g. authority="naf").
  - **displayLabel** φαίνεται πρόσθετο κείμενο (MARC 21 field 246 subfield \$i)
- Sub-elements
  - <title> χωρίς γνώρισμα αντιστοιχεί στο 245 MARC 21 field
  - <subtitle> MARC 21 fields 242, 245, 246 subfield \$b
  - <partNumber> MARC 21 fields 130, 240, 242, 243, 245, 246, 247, 730, 740 subfield \$n.
  - <partName> MARC 21 fields 130, 240, 242, 243, 245, 246, 247, 730, 740 subfield \$p
  - <nonSort> το κείμενο που βρίσκεται σε αυτές τις ετικέτες δεν λαμβάνεται υπ' όψη στην ταξινόμηση



# Name

- Attributes
  - **type** *personal, corporate, conference*
  - **authority, ID, lang, xml:lang, xlink, script, transliteration**

- Sub-elements

- **<namePart>** **type:date, family, given, termsOfAddress**
  - **<affiliation>**, **<role>**

```
<name type="personal">
  <namePart type="termsOfAddress">Dr.</namePart>
  <namePart>Brown, B. F.</namePart>
  <affiliation>Chemistry Dept., American University</affiliation>
  <role>
    <roleTerm type="text">creator</roleTerm>
    <roleTerm type="code">cre</roleTerm>
  </role>
  <role>
    <roleTerm type="text">author</roleTerm>
    <roleTerm type="code">aut</roleTerm>
  </role></name>
```

# Παράδειγμα – βιβλίο (1/3)

```
<?xml version="1.0" encoding="UTF-8" ?>
<mods
  <titleInfo>
    <title>Sound and fury :</title>
    <subTitle>the making of the punditocracy /</subTitle>
  </titleInfo>
  <name type="personal">
    <namePart>Alterman, Eric</namePart>
    <role>
      <roleTerm type="text">creator</roleTerm>
    </role>
  </name>
  <typeOfResource>text</typeOfResource>
  <genre authority="marc">bibliography</genre>
  <originInfo>
    <place>
      <code authority="marc" nyu</code>
      <text >Ithaca, N.Y.</text>
    </place>
```

# Παράδειγμα – βιβλίο (2/3)

```
<publisher>Cornell University Press</publisher>
<dateIssued>c1999</dateIssued>
<dateIssued encoding="marc">1999</dateIssued>
<issuance>monographic</issuance>
</originInfo>
<language authority="iso639-2b" type="code">eng</language>
<physicalDescription>
  <form authority="marcform">print</form>
  <extent>vii, 322 p. ; 23 cm.</extent>
</physicalDescription>
<note type="statement of responsibility">Eric Alterman.</note>
<note>Includes bibliographical references (p. 291-312) and index.</note>
<subject authority="lcsht">
  <geographic>United States</geographic>
  <topic>Politics and government</topic>
  <temporal>20th century.</temporal>
</subject>
```

# Παράδειγμα – βιβλίο (3/3)

<classification authority="lcc">PN4888.P6 A48 1999</classification>

<classification edition="21" authority="ddc">071/.3</classification>

<recordInfo>

<recordContentSource>DLC</recordContentSource>

<recordCreationDate encoding="marc">990730</recordCreationDate>

<recordChangeDate encoding="iso8601">20000406144503.0</recordChangeDate>

<recordIdentifier>11761548</recordIdentifier>

</recordInfo>

</mods>

# Παράδειγμα – χάρτης (1/3)

```
<?xml version="1.0" encoding="UTF-8" ?>
<mods>
  <titleInfo>
    <title>Campbell County, Wyoming /</title>
  </titleInfo>
  <name type="corporate">
    <namePart>Campbell County Chamber of Commerce</namePart>
  </name>
  <typeOfResource>cartographic</typeOfResource>
  <genre authority="marc">map</genre>
  <originInfo>
    <place>
      <placeTerm authority="marccountry" type="code">wyu</placeTerm>
    </place>
    <place>
      <placeTerm type="text">[Gillette, Wyo.]</placeTerm>
    </place>
    <publisher>Campbell County Chamber of Commerce</publisher>
```

# Παράδειγμα – χάρτης (2/3)

```
<dateIssued>[1982?]</dateIssued>
<dateIssued encoding="marc">1982</dateIssued>
<issuance>monographic</issuance>
</originInfo>
<language>
  <languageTerm authority="iso639-2b" type="code">eng</languageTerm>
</language>-
<physicalDescription>
  <extent>1 map ; 33 x 15 cm.</extent>
</physicalDescription>
<note type="statement of responsibility">this map reproduced by Campbell County Chamber
  of Commerce.</note>
<note>In lower right corner: Kintzels-Casper.</note>
<subject authority="lcsb">
  <geographic>Campbell County (Wyo.)</geographic>
  <topic>Maps</topic>
</subject>
<classification authority="lcc">G4263.C3 1982 .C3</classification>
```

# Παράδειγμα – χάρτης (2/3)

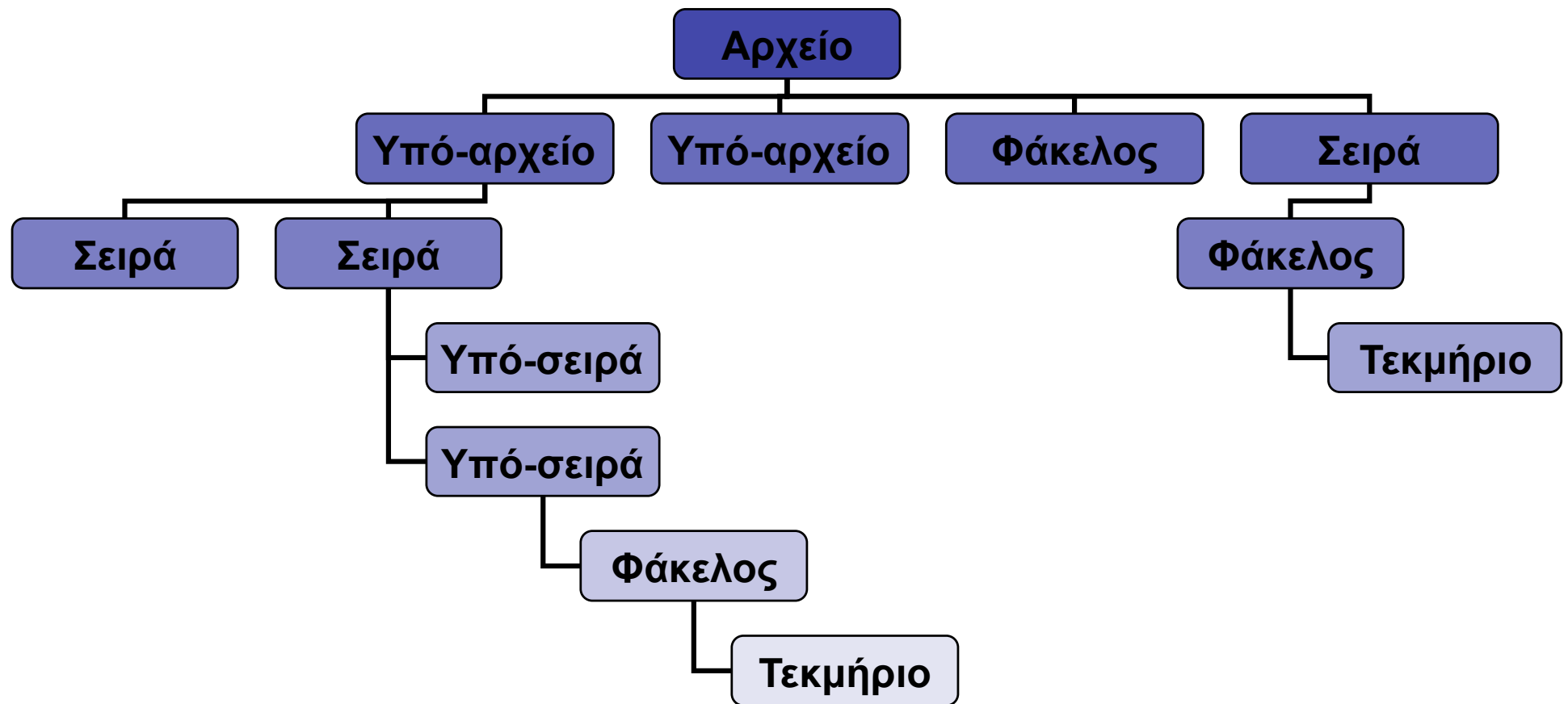
```
<recordInfo>  
  <recordContentSource>DLC</recordContentSource>  
  <recordCreationDate encoding="marc">830222</recordCreationDate>  
  <recordChangeDate encoding="iso8601">19830426000000.0</recordChangeDate>  
  <recordIdentifier>5466714</recordIdentifier>  
</recordInfo>  
</mods>
```

# Encoded Archival Description (EAD)



# Αρχειακά μεταδεδομένα

Αρχείο είναι το σύνολο των τεκμηρίων ανεξαρτήτως χρονολογίας, σχήματος και ύλης που έχει δεχθεί ή παραγάγει οποιοδήποτε φυσικό ή νομικό πρόσωπο, οποιοσδήποτε οργανισμός, δημόσιος ή ιδιωτικός, στα πλαίσια των δραστηριοτήτων του.



# Εργαλεία έρευνας (Finding aids)

- Τελικό προϊόν της αρχειακής περιγραφής
- Πληροφοριακό εργαλείο το οποίο περιλαμβάνει μεταδεδομένα που εξυπηρετούν την αναζήτηση, διαχείριση και ερμηνεία ενός αρχείου και εξηγούν το περιβάλλον δημιουργίας του

# ΞΕΚΙΝΩΝΤΑΣ...

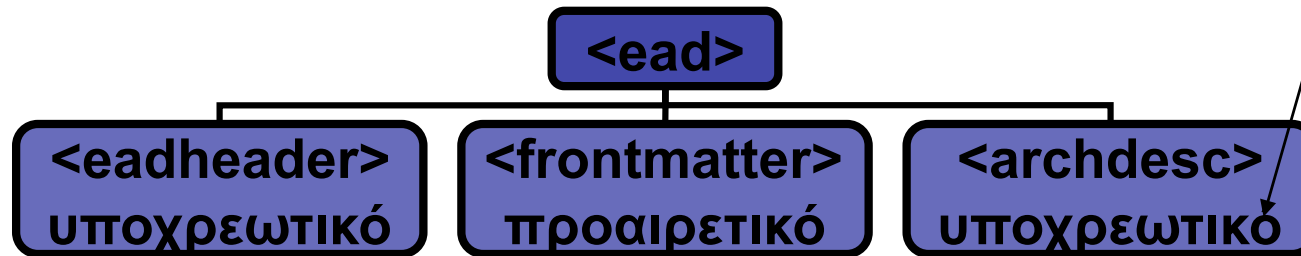
- Απαραίτητες πληροφορίες
  - Τίτλος
  - Όνομα δημιουργού και σκοπός δημιουργίας
  - Ημερομηνίες
  - Μοναδικός αριθμός ταυτοποίησης
  - Διαδικασία πρόσκτησης
  - Περιορισμοί πρόσβασης και / ή αναπαραγωγής
  - Υλικό που έχει αποκοπεί από το κύριο μέρος

# EAD

- Περιγραφικά μεταδεδομένα
- Έμπνευση: Text Encoding Initiative (TEI, 1987)
- Πρώτη προσπάθεια: Πανεπιστήμιο Μπέρκλεϋ της Καλιφόρνια (1993)
- SGML/XML DTD
- EAD version 1.0: 1998
- EAD version 2002: η νέα έκδοση
- Η ανάπτυξη του EAD ξεκίνησε από τις ΗΠΑ
- Γρήγορα υπήρξε ενδιαφέρον από τη παγκόσμια αρχειακή κοινότητα
- Society of American Archivists (<http://www.archivists.org/>) και Library of Congress (<http://www.loc.gov/ead/>)

Περιλαμβάνει επισκόπηση του αρχείου (<did>), διαχειριστικές και συμπληρωματικές πληροφορίες (π.χ. <bioghist>, <accessrestrict> κτλ), περιγραφή συστατικών μερών (<dsc>)

## <ead>



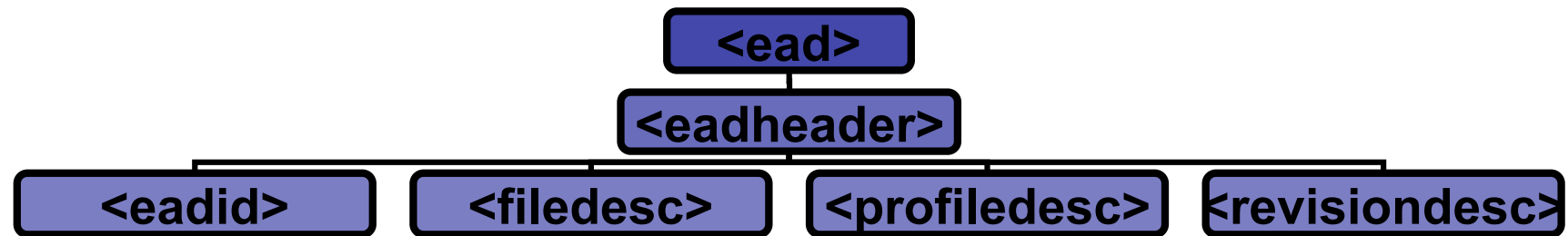
Πληροφορίες για το εργαλείο έρευνας **και όχι για το αρχείο**. Μεταδεδομένα ηλεκτρονικού εργαλείου έρευνας

Χρησιμοποιείται για την παροχή πληροφοριών έκδοσης, όπως της σελίδας τίτλου του έντυπου εργαλείου έρευνας ή άλλων προκαταρκτικών πληροφοριών

# <eadheader>

- Υποχρεωτικό στοιχείο
- Βασισμένο στο TEI header
- Πληροφορίες για το εργαλείο έρευνας **και όχι για το αρχείο**
- Μεταδεδομένα ηλεκτρονικού εργαλείου έρευνας
- «Ηλεκτρονική σελίδα τίτλου»

# <eadheader>





# <eadheader>

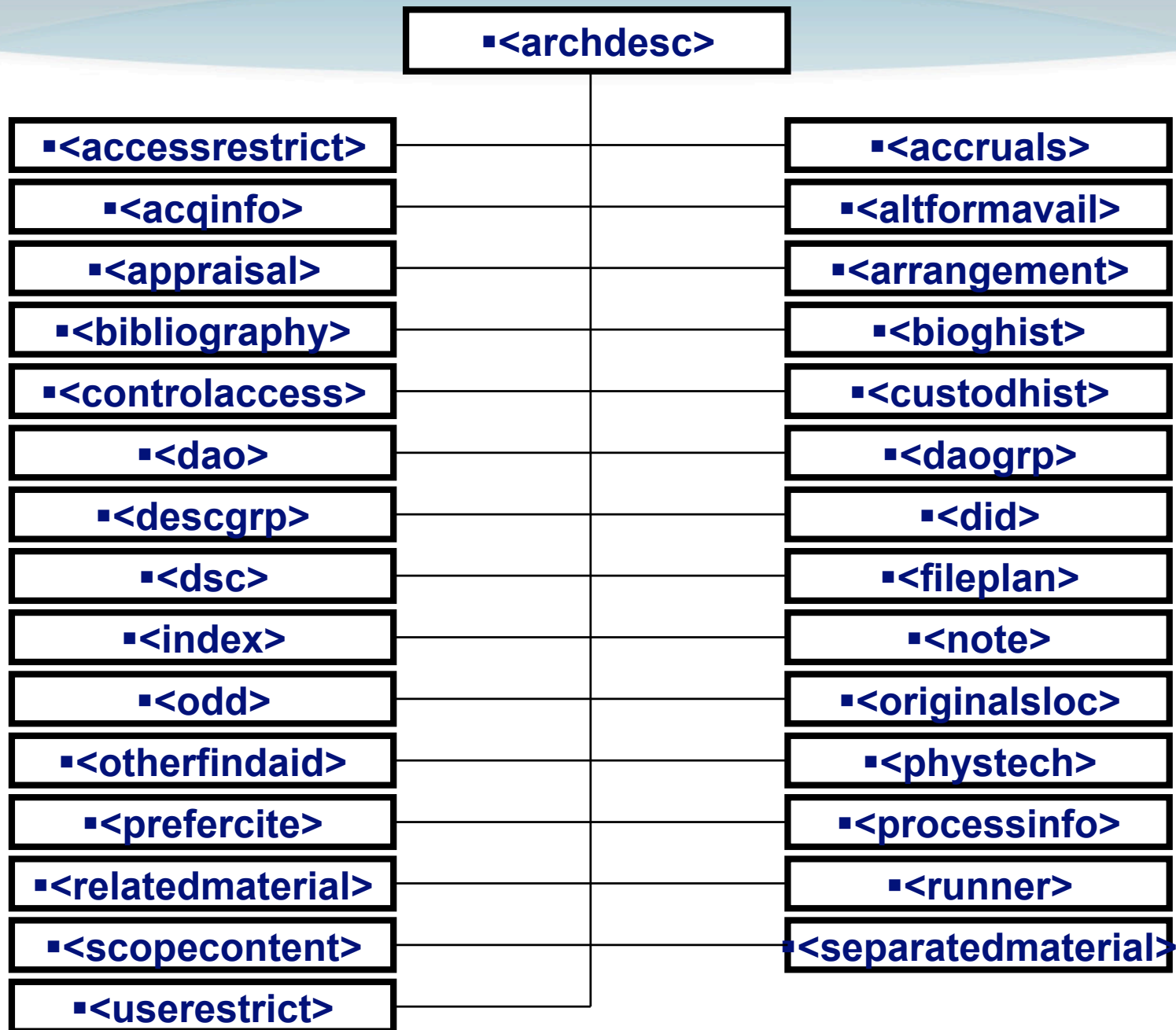
- <eadid>
  - Υποχρεωτικό στοιχείο
  - Κωδικός ταύτισης του εργαλείου έρευνας
- <filedesc>
  - Υποχρεωτικό στοιχείο
  - Βασικές βιβλιογραφικές πληροφορίες για το εργαλείο έρευνας
  - <titlestmt>, <editionstmt>, <publicationstmt>, <seriesstmt>, <notestmt>

# <eadheader>

- <profiledesc>
  - Πληροφορίες για την κωδικοποίηση του εργαλείου έρευνας
  - <creation>, <language>, <desrules>
- <revisiondesc>
  - Πληροφορίες για αλλαγές στο εργαλείο έρευνας
  - <list>, <change>

# <archdesc>

- Υποχρεωτικό στοιχείο
- Εργαλείο έρευνας
- Περιλαμβάνει
  - Επισκόπηση του αρχείου (<did>)
  - Διαχειριστικές και συμπληρωματικές πληροφορίες (π.χ. <bioghist>, <accessrestrict> κτλ)
  - Περιγραφή συστατικών μερών (<dsc>)



# <archdesc>

- <did>
  - Υποχρεωτικό στοιχείο
  - Εμφανίζεται και μέσα στην περιγραφή των συστατικών μερών (<c> και <c01> - <c12>)
  - Βασικές πληροφορίες για το σύνολο του αρχείου (υψηλότερο επίπεδο)
  - <unittitle>, <origination>, <langmaterial>, <physdesc>, <unitdate>, <unitid>, <abstract>, <physloc>, <repository> ...

# <archdesc>

- <accessrestrict>
  - Διαθεσιμότητα του αρχείου
- <acqinfo>
  - Πηγή πρόσκτησης και συνθήκες πρόσκτησης
- <bioghist>
  - Βιογραφία ή διοικητική ιστορία δημιουργού
- <controlaccess>
  - Σημεία πρόσβασης
  - <corpname>, <famname>, <geogname>, <persname>, <subject>, <title>, <genreform> ...

# <archdesc>

- <scopecontent>
  - Επισκόπηση περιεχομένου του αρχείου (σε μορφή κειμένου)
- <separatedmaterial>
  - Υλικό από την ίδια πηγή που έχει αποκοπεί φυσικά ή εκκαθαριστεί από το αρχείο
- <userrestrict>
  - Πληροφορίες για τις συνθήκες που επηρεάζουν τη χρήση του αρχείου εφόσον έχει επιτραπεί η πρόσβαση

# <archdesc>

- <dsc>
  - Προαιρετικό στοιχείο
  - Αναλυτική περιγραφή του αρχείου
  - Περιγραφή από το γενικό στο ειδικό
  - Συστατικά μέρη: <c> ή <c01> - <c12>



# <archdesc>

- Για τη σύνδεση με ένα ή περισσότερα ψηφιακά αντικείμενα

- <daogrp> Digital Archival Object Group

- <daodesc> Digital Archival Object Description

- <daoloc> Digital Archival Object Location

<daogrp>

<daodesc>

<head>Image Sampler</head>

<p>Explanatory paragraph</p>

</daodesc>

<daoloc href="//images/Baker109.jpeg">

<daodesc>

<p>Ella Baker, head-and-shoulders portrait</p>

</daodesc>

</daoloc>

</daogrp>

# EAD – Γνωρίσματα

- Πρόσθετες πληροφορίες για τα στοιχεία
- Μεταδεδομένα για τα στοιχεία
- Επεξεργασία από λογισμικό
- «Κωδικοποιημένα δεδομένα»
- Τιμές
  - Σταθερές
  - Προτεινόμενες
  - Ανεξάρτητες

# Γνωρίσματα

- level
  - Ιεραρχικό επίπεδο του τεκμηρίου που περιγράφεται
  - <archdesc> (Υποχρεωτικό), <c> και <c01> - <c12>
  - Τιμές: collection, fonds, class, recordgrp, series, subfonds, subgrp, subseries, file, item, otherlevel
  - π.χ. <archdesc level="fonds">  
    <c level="item">  
    <c02 level="subseries">

# Γνωρίσματα

- type
  - Διαθέσιμο για αρκετά στοιχεία
    - <unitdate> (bulk, inclusive, other)
    - <archdesc> (inventory, register, other)
    - <dsc> (analyticover, combined, in-depth, othertype)

# Γνωρίσματα

- audience (external, internal)
- countrycode (π.χ. GR)
- datechar (π.χ. creation, accumulation, ή modification)
- findaidstatus (π.χ. “edited-full-draft” )
- relatedencoding (π.χ. MARC 21)
- encodinganalog (π.χ. 500)
- langcode (π.χ. gre)
- mainagencycode (π.χ. ELIA)
- normal (π.χ. 1756/1868)

# Παράδειγμα

```
<archdesc level="fonds">
```

```
<did>
```

```
<unittitle>Αρχείο Γραμματείας / Υπουργείου επί των Οικονομικών </unittitle>
```

```
<unitid countrycode="GR" repositorycode="GAK">256.12</unitid>
```

```
<unitdate normal="1833/1862">1833-1862</unitdate>
```

```
<langmaterial>
```

```
<language langcode="fre">Κυρίως γαλλικά</language>
```

```
<language langcode="gre">και δευτερευόντως
```

```
</langmaterial>
```

```
<origination>
```

```
<corpname>Ανάκτορα, Γραμματεία / Υπουργείο επί των
```

```
</origination>
```

```
</did>
```

Ελλη

Οικον

<accessrestrict><p>Δεν υπάρχουν περιορισμοί πρόσβασης </p> </accessrestrict>

<acqinfo><p>Η πρόσκτηση από τα Γ.Α.Κ. πολλών αρχείων του δημόσιου τομέα του 19ου αιώνα ήταν ιδιαίτερα δυσχερής... </p> </acqinfo>

<bioghist><p>Στις 25 Ιανουαρίου / 1 Φεβρουαρίου 1833 καθιερώθηκε, με διάταγμα της Αντιβασιλείας, οι διευθύνοντες στα Υπουργεία να φέρουν τον τίτλο του «Γραμματέως της Επικρατείας»... </p> </bioghist>

<controlaccess>

<controlaccess>

<subject>Ελληνική Επανάσταση</subject>

<subject>Οικονομικά</subject>

</controlaccess>

</controlaccess>

<scopecontent><p>Το αρχείο της Γραμματείας διαρθρώνεται ως εξής:...</p> </

scopecontent>

```
<dsc>
  <c01 level="subfonds">
    <did>
      <unittitle>Υπο-αρχείο Ανακτόρων</unittitle>
      <unitdate>1833-1844</unitdate>
      <origination>
        <corpname>Ανάκτορα</corpname>
      </origination>
    </did>
  <c02 level="series">
    <did>
      <unittitle>Δάση</unittitle>
      <unitdate>1833-1844</unitdate>
    </did>
    <c03 level="file">...</c03>
  </c02>
</c01>
</dsc>
```



# Χρήσιμες ηλεκτρονικές πηγές

- Encoded Archival Description (EAD) - Official EAD Version 2002 Web Site. (<http://www.loc.gov/ead/>)
- EAD Help Pages (<http://jefferson.village.virginia.edu/ead/>)
- Encoded Archival Description - An Introduction and Overview (<http://www.dlib.org/dlib/november99/11pitti.html>)
- THE EAD COOKBOOK 2002 (<http://www.iath.virginia.edu/ead/ead2002cookbookhelp.html>)
- Recommended Best Practices for Encoded Archival Description Finding Aids at the Library of Congress (<http://www.loc.gov/ead/practices/lcp2002.html>)

# Χρήσιμες ηλεκτρονικές πηγές

- RLG Best Practice Guidelines for Encoded Archival Description (<http://www.rlg.org/en/pdfs/bpg.pdf>)
- RLG EAD Report Card (<http://www.rlg.org/ead-report-card/>)
- Encoded Archival Context (<http://www.iath.virginia.edu/eac/>)
- ISAD (G) ([http://www.ica.org/biblio/cds/isad\\_g\\_2e.pdf](http://www.ica.org/biblio/cds/isad_g_2e.pdf))
- ΔΙΠΑΠ (Γ) (<http://www.eae.org.gr/dipap.pdf>)