



επιμέλεια μεταδεδομένων

Χρήστος Παπαθεοδώρου

Εργαστήριο Ηλεκτρονικής Διαχείρισης Ιστορικών Αρχείων
Τμήμα Ιστορίας και Φιλοσοφίας της Επιστήμης
Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

και

Μονάδα Ψηφιακής Επιμέλειας,
Ινστιτούτο Πληροφοριακών Συστημάτων,
Ερευνητικό Κέντρο «Αθηνά»
(paratheodor@phs.uoa.gr)



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Ταμείο
Περιφερειακής Ανάπτυξης



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΥΠΟΥΡΓΕΙΟ
ΑΝΑΠΤΥΞΗΣ ΚΑΙ ΕΠΕΝΔΥΣΕΩΝ
ΕΙΔΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΠΡΟΓΡΑΜΜΑΤΩΝ ΕΤΠΑ & ΤΣ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΕΠΑΝΕΚ

ΕΠΑΝΕΚ 2014-2020
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΑΝΤΑΓΩΝΙΣΤΙΚΟΤΗΤΑ
ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ
ΚΑΙΝΟΤΟΜΙΑ



ανάπτυξη - εργασία - αλληλεγγύη

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Τί και γιατί

- Μια εισαγωγή σε θέματα ψηφιακών συλλογών, ψηφιακών βιβλιοθηκών και συσσωρευτών ψηφιακών αντικειμένων.
- Για να γίνει αντιληπτό ότι η επιμέλεια μεταδεδομένων είναι απαραίτητη από τα πρώτα στάδια σχεδιασμού μιας ψηφιακής συλλογής, εάν θέλουμε το περιεχόμενό της να είναι 'βιώσιμο', ανοικτό και επαναχρησιμοποιήσιμο.

ερωτήματα

- Πώς θα αναζητήσω ένα πόρο;
- Πώς θα έχω πρόσβαση σε αυτόν;
- Πώς θα ξέρω πώς θα τον διαχειριστώ για να είναι βιώσιμος;
- Πώς θα τον (επανα)χρησιμοποιήσω;

μεταδεδομένα

- Τα μεταδεδομένα είναι **δομημένη πληροφορία** που περιγράφει, εξηγεί, εντοπίζει ή διευκολύνει την ανάκτηση, τη χρήση ή την διαχείριση ενός ψηφιακού τεκμηρίου (πόρου).
- Είναι πληροφορία που επεξεργάζεται και αποθηκεύεται στον υπολογιστή (**μηχαναγνώσιμη**).

παράδειγμα

Βιβλιογραφία

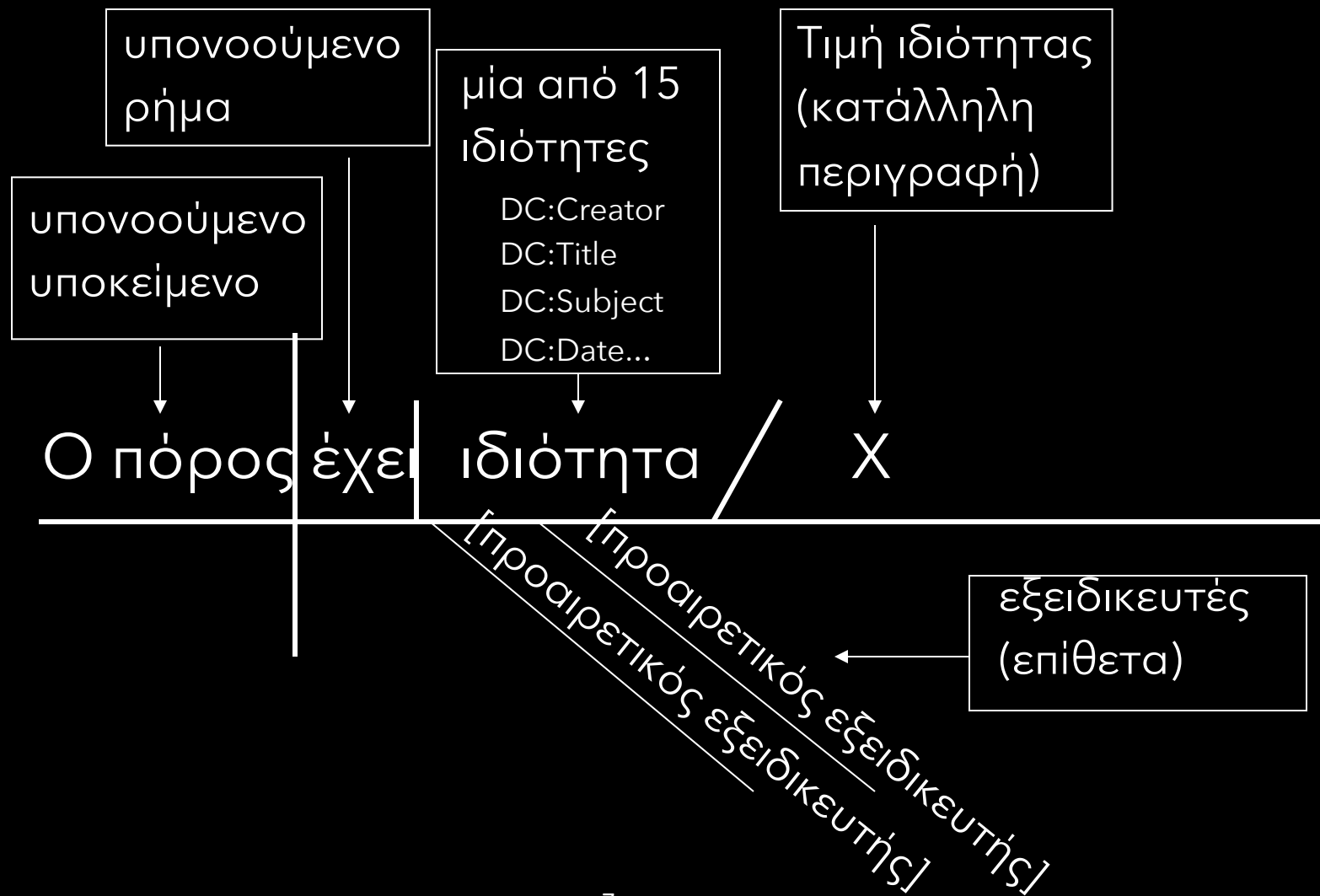
- S. Abiteboul, P. Buneman, D. Suciu *"Data on the Web: From Relations to Semistructured Data and XML"* Morgan Kaufmann Publishers, 2000.
- Norman Walsh *"A Guide to XML"* World Wide Web Journal, Vol. 2, Issue 4, 1997, pages 97-107.

```
<bibliography>
  <book>
    <author>S. Abiteboul</author>
    <author>P. Buneman</author>
    <author>D. Suciu</author>
    <title>Data on the Web: From Relations to Semistructured Data and XML</title>
    <publisher>Morgan Kaufmann Publishers</publisher>
    <year>2000</year>
  </book>
  <article>
    <author>Norman Walsh</author>
    <title>A Guide to XML</title>
    <journal>World Wide Web Journal</journal>
    <volume>2</volume>
    <issue>4</issue>
    <year>1997</year>
    <pages>97-107</pages>
  </article>
</bibliography>
```

τα μεταδεδομένα ως γλώσσα

- Τα σχήματα μεταδεδομένων είναι γλώσσες.
- Έχουν λεξιλόγια με όρους που λέγονται **στοιχεία**.
- Έχουν συντακτικό που επιτρέπει να φτιάχνουμε προτάσεις με νόημα για τους πόρους μας.
 - Τα λεξιλόγια και τα συντακτικά περιορίζουν τις δηλώσεις.
 - κάνουν, όμως, τη σημασία τους σαφή.
 - «Το βιβλίο» έχει **Title** "Gone with the Wind".
 - «Η ιστοσελίδα» έχει **Publisher** "Springer Verlag".

συντακτικό Dublin Core



πρότυπα μεταδεδομένων

- Γλώσσες κοινής αποδοχής.
- Χρειάζονται για να έχουμε **κοινή αντίληψη** των πόρων που περιγράφονται με αυτά
 - μας προσφέρουν μεγαλύτερη **δομή**
- 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.
 - ανακάλυψη: περιγραφικά, π.χ. τίλος, συγγραφέας
 - διατήρηση: τεχνικά, π.χ. τύπος υλικού, ψηφιακής μορφής

Dublin Core

- 15 στοιχεία + εξειδικευτές (qualifiers).
- Μέσο ανταλλαγής πληροφορίας και όχι πρότυπο περιγραφής πόρων.
- Απλό, με μεγάλη επιρροή σε έργα ψηφιακών βιβλιοθηκών.

Dublin Core

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>
```

MODS

- Metadata Object Description Schema
- Library of Congress Network Development and MARC Standards Office (υποσύνολο του MARC).
- 19 στοιχεία, 64 υπο-στοιχεία

παράδειγμα - βιβλίο (1/2)

```
<?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> <publisher>Cornell University Press</publisher>
```

```
    <dateIssued encoding="marc">1999</dateIssued>
```

```
    <issuance>monographic</issuance>
```

```
  </originInfo>
```

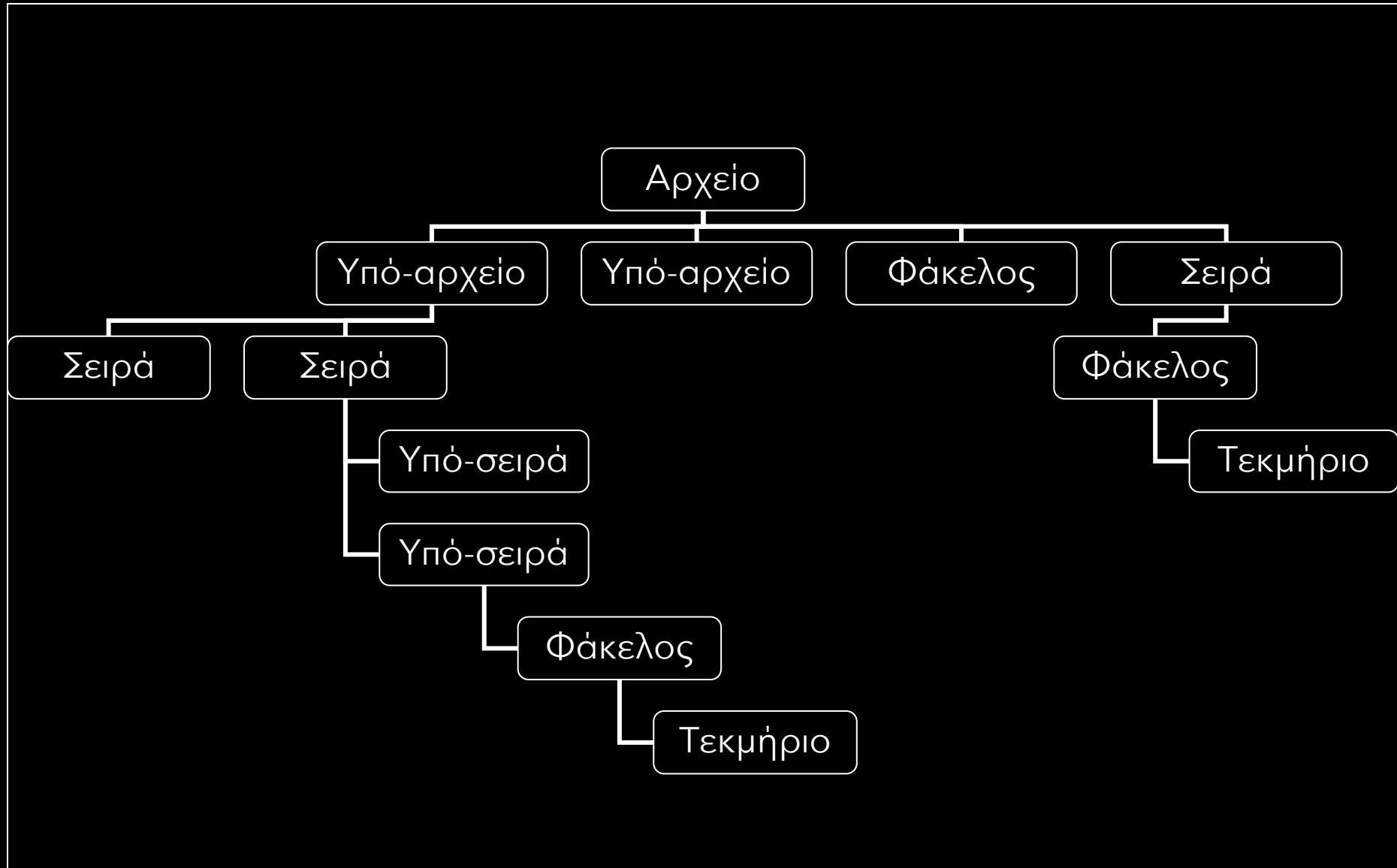
παράδειγμα - βιβλίο (2/2)

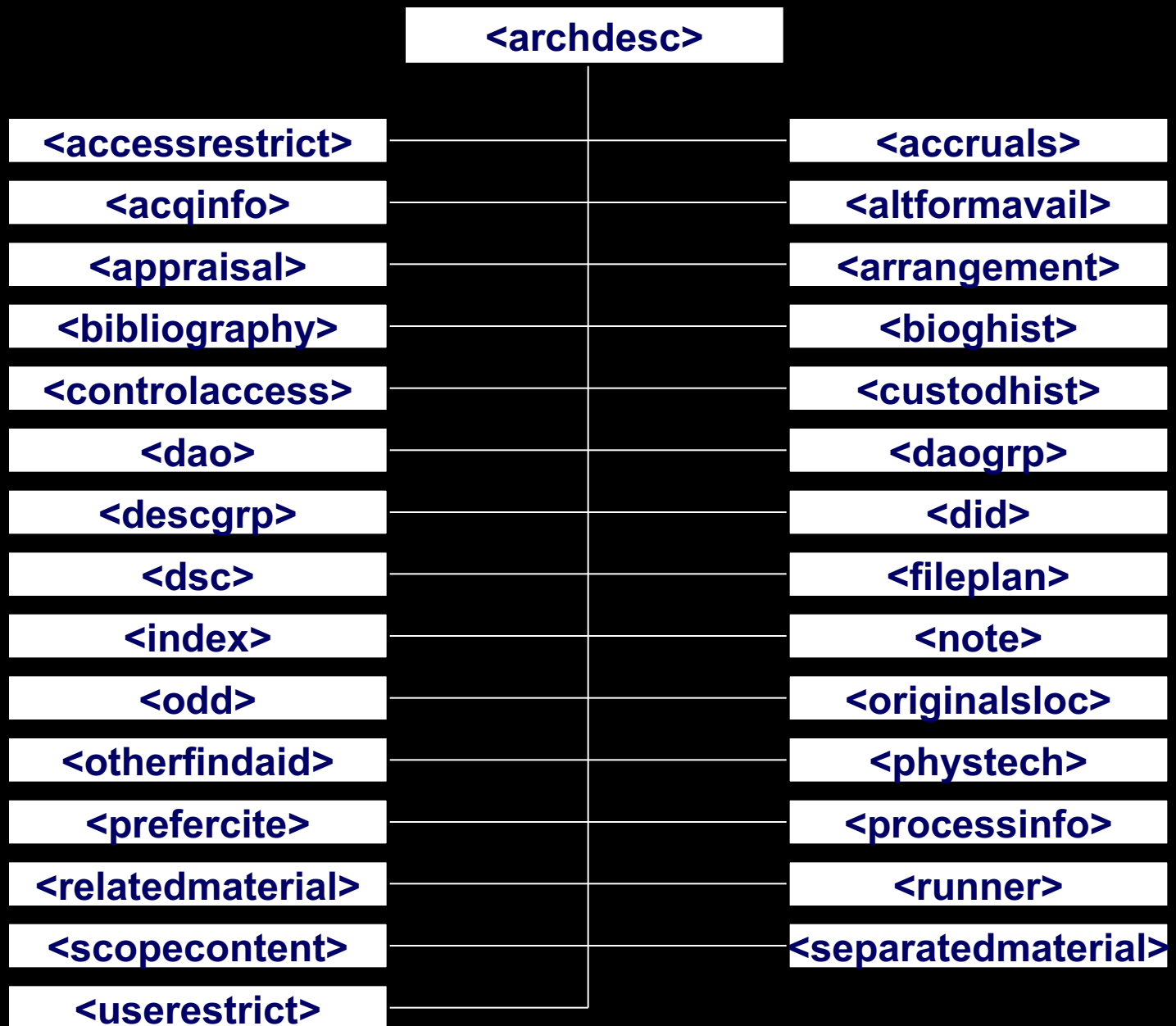
```
<language authority="iso639-2b" type="code">eng</language>
<physicalDescription>
  <form authority="marcform">print</form>
  <extent>vii, 322 p. ; 23 cm.</extent>
</physicalDescription>
<note type="statement of responsibility">Eric Alterman.</note>
<note>Includes bibliographical references (p. 291-312) and index.</note>
<subject authority="lcsch">
  <geographic>United States</geographic>
  <topic>Politics and government</topic>
  <temporal>20th century.</temporal>
</subject>
<classification authority="lcc">PN4888.P6 A48 1999</classification>
<classification edition="21" authority="ddc">071/.3</classification>
<recordInfo>
  <recordCreationDate encoding="marc">990730</recordCreationDate>
  <recordChangeDate
encoding="iso8601">20000406144503.0</recordChangeDate>
  <recordIdentifier>11761548</recordIdentifier>
</recordInfo>
</mods>
```

Encoded Archival Description (EAD)

- Εργαλεία Έρευνας (Finding Aids)
- Τελικό προϊόν της αρχειακής περιγραφής
- Πληροφοριακό εργαλείο με μεταδεδομένα που εξυπηρετούν:
 - την αναζήτηση, διαχείριση και ερμηνεία ενός αρχείου
 - εξηγούν το περιβάλλον δημιουργίας του

αρχειακή δομή





επιλογή προτύπου μεταδεδομένων

Είναι σημαντικό να καθοριστούν

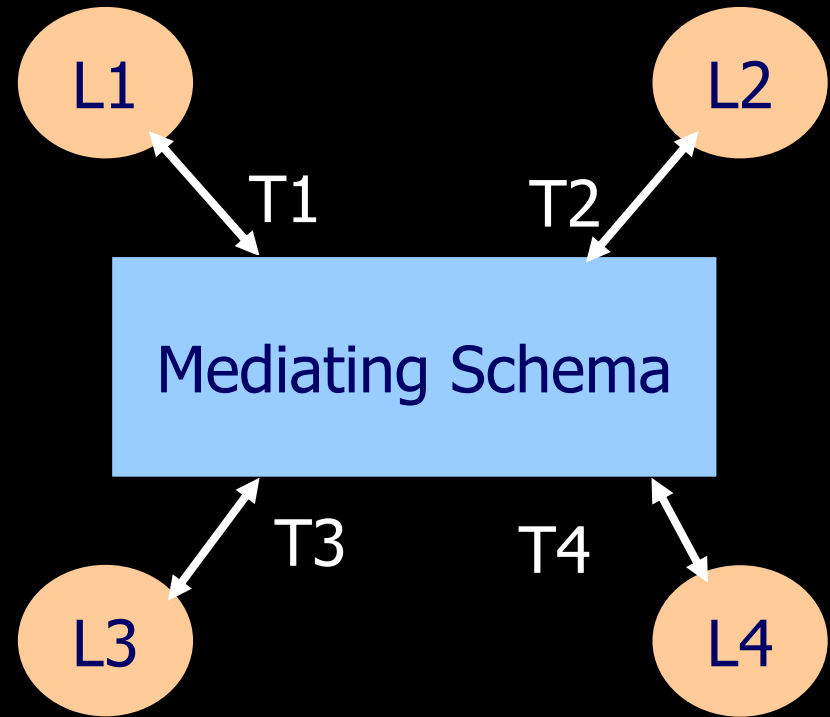
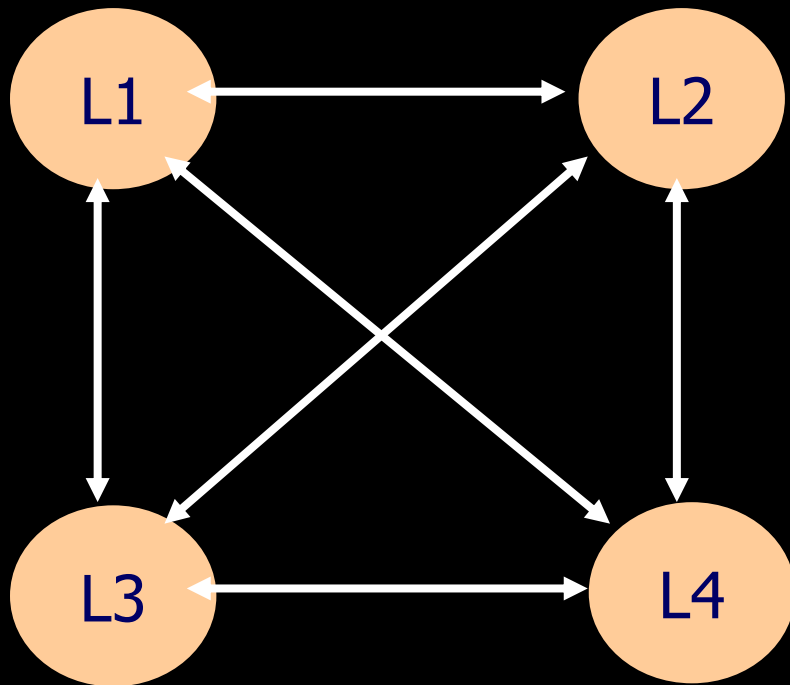
- το **πεδίο εφαρμογής**
- οι **μορφές** των ψηφιακών αντικειμένων και συλλογών
- ο **βαθμός λεπτομέρειας** στην περιγραφή και παρουσίαση

προκειμένου να αξιολογηθούν τα πρότυπα μεταδεδομένων και επιλεγεί το κατάλληλο.

διαλειτουργικότητα μεταδεδομένων

- Πρόβλημα:
 - Τι κάνουμε στην περίπτωση που θέλουμε να συνδέσουμε τη συλλογή μας με άλλες συλλογές με διαφορετικά σχήματα στο περιβάλλον του Παγκόσμιου Ιστού;
- Μετασχηματισμός των μεταδεδομένων της συλλογής μας στο σχήμα κάποιας άλλης συλλογής.
 - Απαιτείται πίνακας αντιστοίχισης ενός σχήματος μεταδεδομένων σε ένα άλλο σχήμα.

διαλειτουργικότητα μεταδεδομένων



συσσωρευτές

- Θεματικοί ή εθνικοί μηχανισμοί αυτοματοποιημένης συγκέντρωσης και επιμέλειας μεταδεδομένων ψηφιακών συλλογών μέσω πρωτοκόλλων συγκομιδής μεταδεδομένων.
- Διευκολύνουν την έκθεση των τοπικών συλλογών, την **ενοποίηση** ετερογενών πηγών, την **ενιαία** αναζήτηση από τον τελικό χρήστη.
- **Διαλειτουργικότητα**: μετασχηματισμός των μεταδεδομένων των συλλογών στο σχήμα του συσσωρευτή.

Europeana



europæana data model

edm:providedCHO

dc:title=The history of the most renowned Don Quixote of Mancha [Texto impreso]:
and his trusty squire Sancho Pancha [sic] / now made English according ... by J.P.
dcterms:extent= [16], 616, [3] p., [7] h. de lám.; Fol.
dcterms:issued= 1687
dc:language=eng
dc:contributor=Phillips, John
dc:contributor=Hodgkin, Thomas
dc:contributor= Whitwood, William

edm:aggregatedCHO

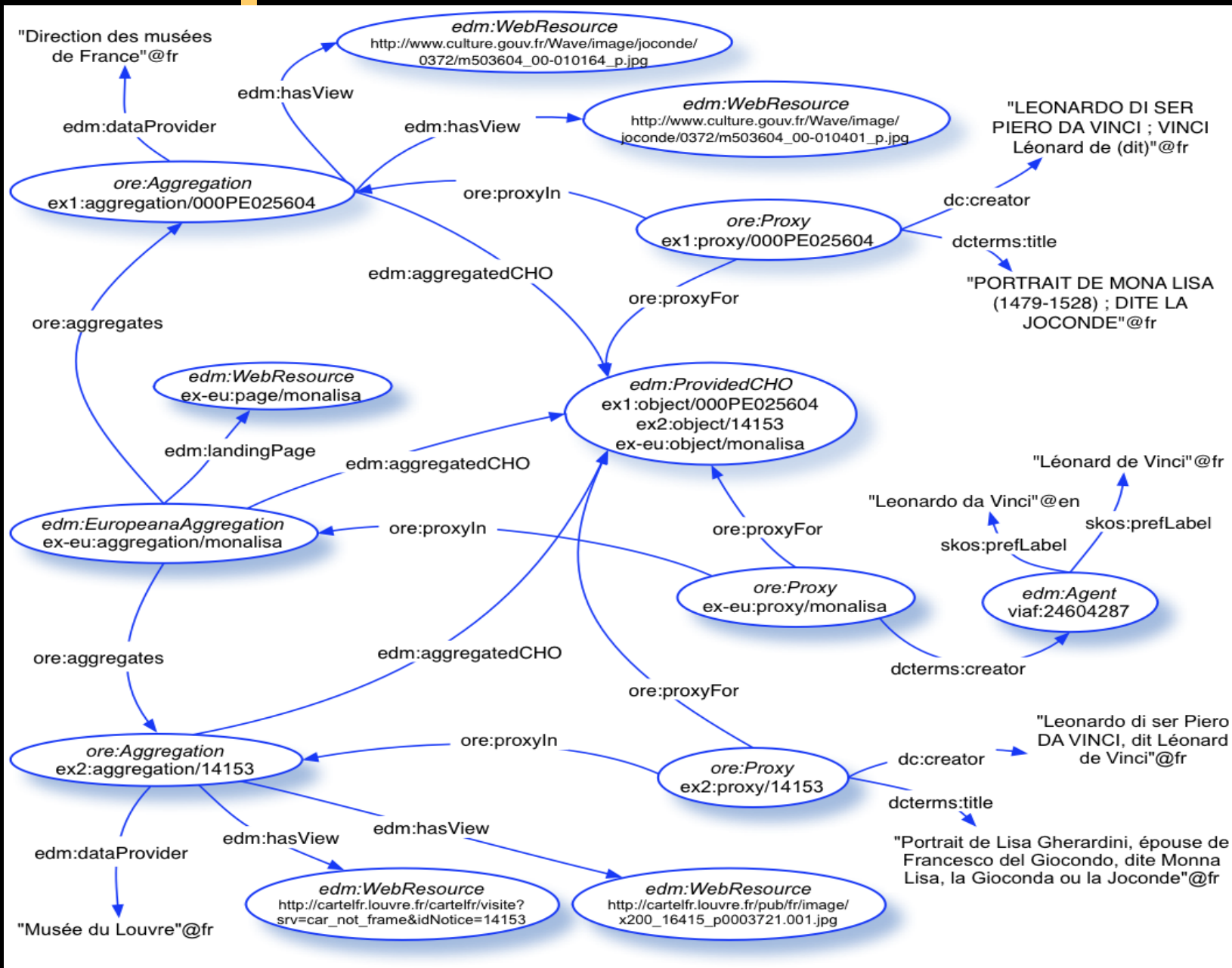
edm:webResource

http://bibliotecadigitalhispanica.bne.es:80/webclient/DeliveryManager?pid=1811612&custom_att_2=simple_viewer

edm:hasView

ore:Aggregation

europaana data model



οντολογίες και διαλειτουργικότητα μεταδεδομένων

- Σαφής τυπικός φορμαλισμός που ορίζει τις οντότητες ενός θεματικού πεδίου και τις μεταξύ τους σχέσεις.
- Αναπαράσταση γνώσης - συνεπάγεται την εξαγωγή συμπερασμάτων.
- Ανάπτυξη οντολογιών για ένα θεματικό χώρο (domain ontologies), CIDOC CRM για πολιτιστικό περιεχόμενο
- Οντολογία ως σχήμα διαμεσολάβησης

σημασιολογική διαλειτουργικότητα

- Να συσχετίσουμε εννοιολογικά τα λεξιλόγια των μεταδεδομένων (πληροφορία) που έχουμε για τους πόρους και τις συλλογές μας.
- Να επαυξήσουμε την πληροφορία αναδεικνύοντας νέες σχέσεις μεταξύ των μεταδεδομένων.

Example 2: Document

“The following declaration has been approved:

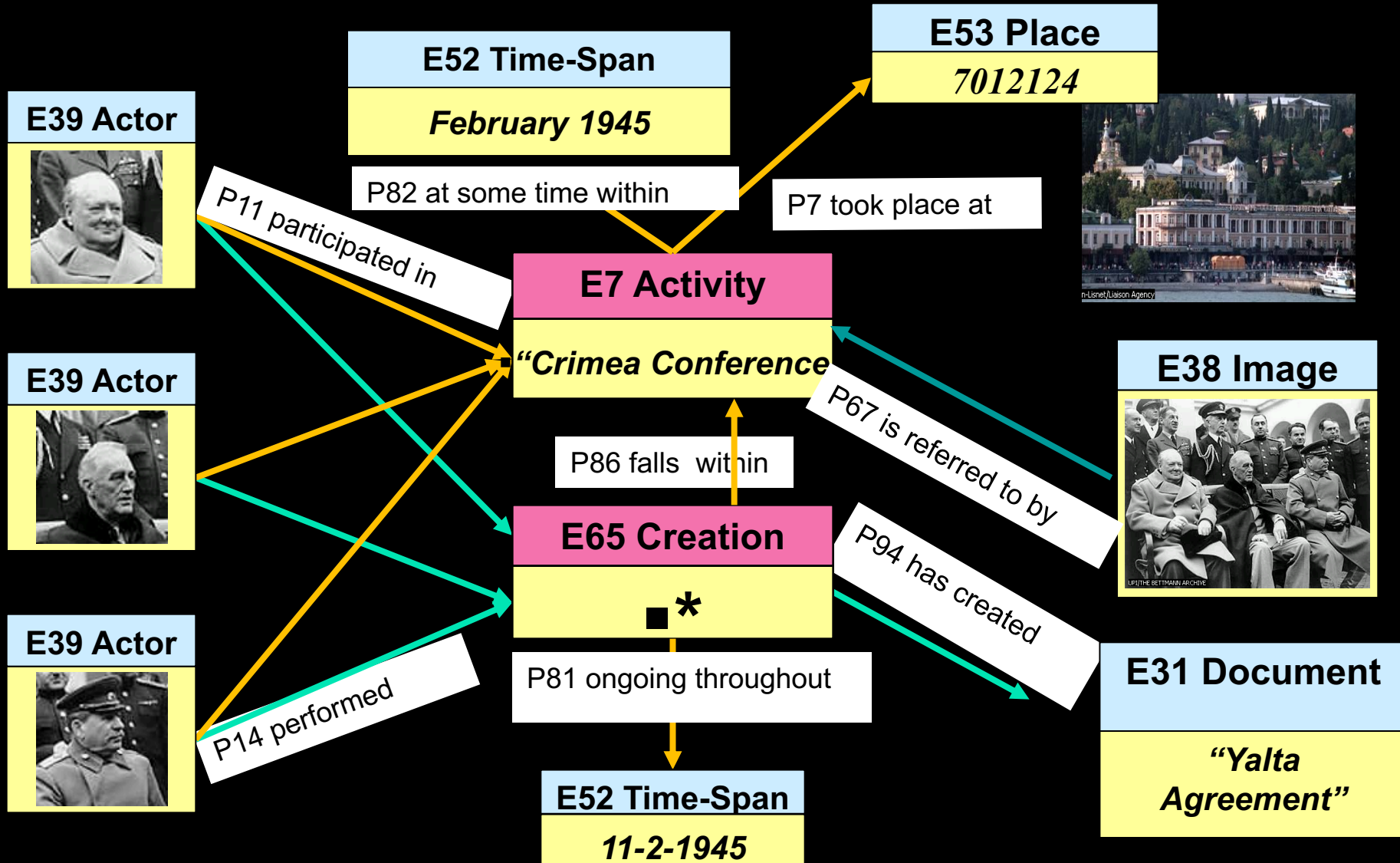
The Premier of the Union of Soviet Socialist Republics, the Prime Minister of the United Kingdom and the President of the United States of America have consulted with each other in the common interests of the people of their countries and those of liberated Europe. They jointly declare their mutual agreement to concert... and to ensure that Germany will never again be able to disturb the peace of the world...”

Type: Text
Title: Protocol of Proceedings of Crimea Conference
Title.Subtitle: II. Declaration of Liberated Europe
Date: February 11, 1945.
Creator: Premier of the Union of Soviet Socialist Republics
Prime Minister of the United Kingdom
President of the United States of America
Publisher: State Department

Example 3: Getty TGN record

TGN ID: 7012124
Names: Yalta (C,V), Jalta (C,V)
Types: inhabited place(C), city (C)
Position: Lat: 44 30 N, Long: 034 10 E
Hierarchy: Europe (continent) ← Ukrayina (nation) ← Krym (autonomous republic)
Note: ...Site of conference between Allied powers in WWII in 1945...
Source: TGN, Thesaurus of Geographic Names

σημασιολογική ολοκλήρωση



συμπεράσματα

- Οι ψηφιακές συλλογές έχουν τη δυνατότητα να απευθύνονται σε ένα μεγάλο μέρος του κοινού και να συνδέονται με άλλου σκοπού δραστηριότητες.
- Οφείλουν λαμβάνουν υπ' όψιν τις απαιτήσεις της κοινότητας, το εύρος των οποίων μπορεί να είναι πολύ μεγάλο και να ποικίλει.
- Η ψηφιακή επιμέλεια οφείλει να βλέπει τις δυνητικές χρήσεις του περιεχομένου.
- Ο ρόλος των επαγγελματιών της πληροφόρησης είναι σημαντικός για την επιμελή διαχείριση των μεταδεδομένων, αλλά και για την υποστήριξη, την κατάρτιση και τη διασύνδεση των κοινοτήτων.