



# Διαχείριση Γνώσης (Knowledge Management)

**Ενότητα 7**: «Οργάνωση γνώσης – Μέρος Α»

Χ. Σκουρλάς, Ε. Γαλιώτου, Α. Μαρινάγη

Το μάθημα αυτό βασίζεται σε υλικό μεταπτυχιακών μαθημάτων που δίδαξαν οι Μ. Γεργατσούλης, Χ. Παπαθεοδώρου, Χ. Σκουρλάς

#### Έκτη συνάντηση

Στην έβδομη συνάντηση γίνεται παρουσίαση και συζήτηση της Οργάνωσης γνώσης στο πλαίσιο της διαχείρισης γνώσης.

## Βασικά σημεία

- Semantic Web
- Οντολογίες
- XML
- RDF
- Dublin Core

## Ο Παγκόσμιος Ιστός

- Ο παγκόσμιος ιστός είναι μια μεγάλη και επιτυχημένη προσπάθεια αν αναλογιστεί κανείς
  - την ποσότητα της διαθέσιμης πληροφορίας, και
  - το ρυθμό αύξησης των ανθρώπων που τον χρησιμοποιούν.
- Έχει αρχίσει να διαπερνά τις περισσότερες πλευρές της καθημερινής μας ζωής αλλά και της ζωής των επιχειρήσεων και οργανισμών.
- Η επιτυχία του βασίζεται στην απλότητα του.
- Διαχείριση και Ερμηνεία από τον Άνθρωπο
- Η απλότητα του HTTP και της HTML επέτρεψε στους κατασκευαστές λογισμικού, στους φορείς παροχής πληροφοριών και τους τελικούς χρήστες την εύκολη προσπέλαση στο νέο μέσο και συνέβαλε στο να αναπτύξουν σε πολύ μικρό χρονικό διάστημα μια σημαντική κρίσιμη μάζα.

# Σημασιολογικός Παγκόσμιος Ιστός (Semantic Web)

Ο Παγκόσμιος Ιστός και τα σοβαρά προβλήματα στην: ανεύρεση, εξαγωγή, αναπαράσταση, ερμηνεία, συντήρηση της πληροφορίας

- σημασιολογία της πληροφορίας η οποία είναι εύκολο να γίνει κατανοητή μηχανικά, και
- εκατομμύρια από μικρές εξειδικευμένες υπηρεσίες συλλογισμού (reasoning services) οι οποίες παρέχουν υποστήριξη στην επιτυχή επίτευξη εργασιών βασιζόμενες στην προσπελάσιμη πληροφορία

## Οι μηχανές αναζήτησης στο Σημασιολογικό Παγκόσμιο Ιστό

- αναζήτηση με βάση έννοιες αντί λέξεις κλειδιά
- Σημασιολογική εστίαση/διεύρυνση των ερωτήσεων
- ερωταπαντήσεις ως προς περισσότερα του ενός κείμενα
- χρήση τελεστών μετασχηματισμού των κειμένων
- οι οντολογίες ορίζουν την τυπική σημασιολογία της πληροφορίας διευκολύνοντας την επεξεργασία της πληροφορίας από τον Η/Υ
- οι οντολογίες ορίζουν σημασιολογία του πραγματικού κόσμου επιτρέποντας τη σύνδεση του περιεχομένου το οποίο επεξεργάζεται μηχανικά, με τη σημασία που του δίνουν οι άνθρωποι βασιζόμενοι σε κοινά αποδεκτή ορολογία.
- "μια διαμοιρασμένη και κοινή κατανόηση κάποιου τομέα, η οποία μπορεί να ανταλλαγεί μεταξύ ανθρώπων και συστημάτων εφαρμογών" Gruber

## Το πρόβλημα

Η έλλειψη κοινής αντίληψης οδηγεί σε:

- προβλήματα στην επικοινωνία μεταξύ ανθρώπων και οργανισμών.
- δυσκολίες στον προσδιορισμό των απαιτήσεων και κατά συνέπεια στην ανάπτυξη των προδιαγραφών των συστημάτων.

Οι ανομοιόμορφες μέθοδοι μοντελοποίησης, οι γλώσσες και τα εργαλεία λογισμικού περιορίζουν σοβαρά:

- τη δια-λειτουργικότητα
- την επαναχρησιμοποίηση και το διαμοιρασμό εφαρμογών.

## Η λύση - Οντολογία

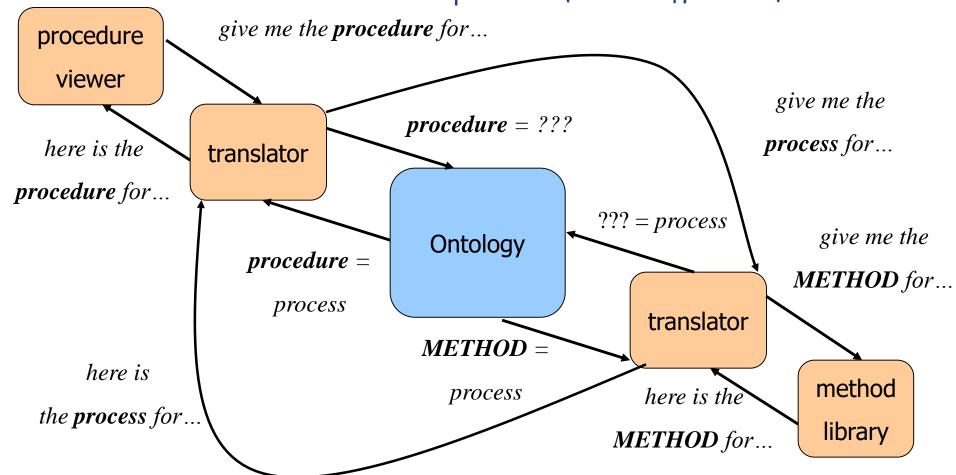
- Μια οντολογία είναι μια τυπική (formal), κατηγορηματική (explicit) προδιαγραφή μιας διαμοιρασμένης (shared) εννοιολογικής αναπαράστασης (conceptualization) Gruber
  - Ο όρος «εννοιολογική αναπαράσταση»
     (conceptualization) αναφέρεται σε ένα αφηρημένο μοντέλο φαινομένων του κόσμου στο οποίο έχουν προσδιοριστεί οι έννοιες που σχετίζονται με τα φαινόμενα αυτά.
  - Ο όρος «κατηγορηματική» (explicit) σημαίνει ότι το είδος των εννοιών που χρησιμοποιούνται, και οι περιορισμοί που αφορούν την χρήση αυτών των εννοιών είναι προσδιορισμένα με σαφήνεια.
  - Ο όρος «αυστηρή» (formal) αναφέρεται στο ότι η οντολογία πρέπει να είναι μηχανικά αναγνώσιμη.
  - Ο όρος «διαμοιρασμένη» (shared) αναφέρεται στο ότι η οντολογία πρέπει να αποτυπώνει γνώση κοινής αποδοχής στα πλαίσια μιας κοινότητας.

## Παράδειγμα Οντολογίας (σε OIL)

```
class-def animal
%plants are a class that is disjoint from animals
class-def plant subclass-of NOT animal
%it is necessary but not sufficient for a tree to be
  a plant:
class-def tree subclass-of plant
%branches are PART OF trees
class-def branch
  slot-constraint is-part-of has-value tree
%it is necessary and sufficient for a carnivore to
  be an animal:
class-def defined carnivore subclass-of animal
  slot-constraints eats value-type animal
%herbivores eat only plants OR part of plants
class-def defined herbivore subclass-of animal
  slot-constraint eats value-type plant OR
  (slot-constraint is-part-of has-value plant)
```

### Παράδειγμα Δια-λειτουργικότητας Uschold & Gruninger (1996)

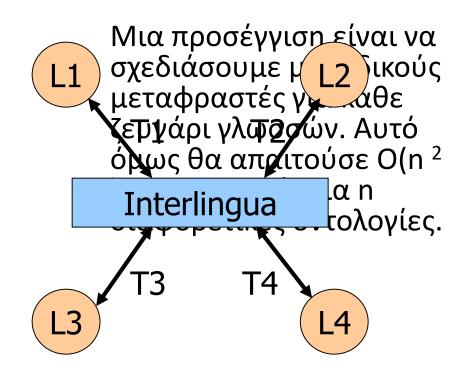
Ο όρος 'procedure' που χρησιμοποιείται από ένα εργαλείο μεταφράζεται στον όρο 'method' που χρησιμοποιείται από το άλλο με τη βοήθεια της οντολογίας, η οποία με τη σειρά της χρησιμοποιεί τον όρο 'process' για να εκφράσει την ίδια έννοια.



# Οι Οντολογίες σαν InterLingua (M. Uschold & M. Gruninger (1996) Ontologies Principles, Methods, and Applications)

Για την υποβοήθηση της δια-λειτουργικότητας οι οντολογίες μπορούν να χρησιμοποιηθούν για να υποστηρίξουν τη μετάφραση μεταξύ διαφορετικών γλωσσών και

Χρησιμοποιώντας οντολογίες σαν interlingua για την υποστήριξη της μετάφρασης, μπορούμε να μειώσουμε τον αριθμό των μεταφραστών σε O(n) για η διαφορετικές οντολογίες, αφού θα απαιτούνται μόνο μεταφραστές από την μητρική οντολογία στην οντολογία ανταλλαγής



## Η μορφή μιας οντολογίας

- Μια οντολογία οπωσδήποτε θα περιλαμβάνει ένα λεξιλόγιο όρων και κάποιες προδιαγραφές για τη σημασία τους.
- Η αναπαράσταση μιας οντολογίας μπορεί να είναι:

Ontology	Διατύπωση	
<i>Ατυπη</i> (informal)	φυσική γλώσσα	
semi-informal	δομημένο υποσύνολο φυσικής γλώσσας	
semi-formal	τεχνητή και αυστηρά ορισμένη γλώσσα	
rigorously formal	αυστηρή σημασιολογία, θεωρήματα και αποδείξεις ιδιοτήτων όπως η ορθότητα (soundness) και η πληρότητα (completeness)	

## Κατηγορίες συστατικών Οντολογίας

- Κλάσεις (classes): Έννοιες που σχετίζονται με ένα πεδίο ή κάποιες εργασίες, οι οποίες είναι συνήθως οργανωμένες σε κάποιο ταξινομικό σύστημα, π.χ., Σε μια οντολογία που αφορά το πανεπιστήμιο ο «φοιτητής» και ο «καθηγητής» αποτελούν δύο κλάσεις.
- Σχέσεις (relations): Ένας τύπος αλληλεπίδρασης μεταξύ εννοιών ενός πεδίου,
   π.χ. subclass-of, is-a
- Συναρτήσεις (functions): μια ειδική περίπτωση σχέσης στην οποία το ν-οστό στοιχείο της σχέσης προσδιορίζεται μοναδικά από τα ν-1 προηγούμενα στοιχεία, π.χ., Η τιμή-μεταχειρισμένου-αυτοκινήτου μπορεί να προσδιορίζεται σαν συνάρτηση της αρχικής τιμής του καινούριου αυτοκινήτου, του μοντέλου του αυτοκινήτου, των χαρακτηριστικών του αυτοκινήτου και των χιλιομέτρων που έχει διανύσει.
- Αξιώματα (axioms): αναπαριστούν προτάσεις που είναι πάντα αληθείς, π.χ.,
   αν ο Φ είναι δευτεροετής φοιτητής τότε μπορεί να εγγραφεί στο επιλεγόμενο μάθημα Μ.
- Στιγμιότυπα (instances): αναπαριστούν συγκεκριμένα στοιχεία, π.χ., ο φοιτητής με το όνομα Νίκος είναι ένα στιγμιότυπο της κλάσης «φοιτητής»

## Κατηγορίες Οντολογιών

Κατηγορία	Επεξήγηση - παράδειγμα	
domain ontologies	γνώση γύρω από ένα συγκεκριμένο πεδίο, π.χ., ιατρική, ηλεκτρονικά	
metadata ontologies	παρέχουν ένα λεξιλόγιο για την περιγραφή του περιεχομένου ηλεκτρονικά διαθέσιμης πληροφορίας	
generic or common sense ontologies	αποτυπώνουν γενική γνώση γύρω από τον κόσμο, παρέχοντας βασικές έννοιες όπως ο χρόνος, ο χώρος, τα συμβάντα	
representational ontologies	π.χ. Frame Ontology (Gruber 1993): ορίζει έννοιες όπως frames, slots, slot constraints κ.λπ.	
method or task ontologies	παρέχουν όρους που αναφέρονται σε συγκεκριμένες εργασίες, π.χ., διάγνωση	

### Μεθοδολογία Ανάπτυξης Οντολογιών

- Δεν υπάρχουν τυποποιημένες μεθοδολογίες για τη ανάπτυξη οντολογιών παρά μόνο εμπειρικοί κανόνες.
- Φάσεις για την ανάπτυξη οντολογιών:
  - Προσδιορισμός σκοπιμότητας και πεδίου εφαρμογής
  - Κατασκευή της οντολογίας
    - Σύλληψη (capture)
    - Κωδικοποίηση (coding)
    - Ενοποίηση (integration) υπαρχουσών οντολογιών.
  - Αξιολόγηση (evaluation)
  - Τεκμηρίωση (documentation)

M. Uschold & M. Gruninger (1996) Ontologies Principles, Methods, and Applications

#### References

David Stuart Practical Ontologies for Information Professionals. London,
 UK: Facet Publishing, 2016, 184 p., ISBN: 978-1-78330-062-4.

Welcome to the Protege Ontology Library!

This page is organized into the following groupings:

- OWL ontologies
- Frame-based ontologies
- Ontologies in other formats (e.g., DAML+OIL, RDF Schema, etc.)

#### HTML versus XML

```
HTML:
<H1>Knowledge Management</H1>
     <UL> <LI>Teacher: T. Codd
          <LI>Students: 38
          <LI>Prerequisite: none
     </UL>
XML (ετικέτες που ορίζονται από τον χρήστη):
<course>
     <title>Knowledge Management</title>
     <teacher>T. Codd</teacher>
     <students>38</students>
     <req>none</req>
</course>
```

## **XML** document = δέντρο με ετικέτες

κόμβος = ετικέτα + γνωρ./τιμές + περιεχόμενα

- XML Schema: γραμματικές για περιγραφή εγκύρων δέντρων και τύπων δεδομένων
- Η XML δεν παρέχει τα απαραίτητα μέσα για την περιγραφή της σημασιολογίας

## RDF - Resource Description Framework

RDF: ένα μοντέλο δεδομένων ανεξάρτητο από το πεδίο εφαρμογής, και ανεξάρτητο από συγκεκριμένη εφαρμογή.

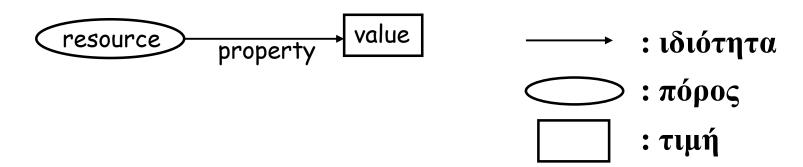
Τα RDF δεδομένα είναι πιθανόν να εμφανίζονται ή να μην εμφανίζονται σε μορφή XML,

Το μοντέλο της RDF αποτελείται από ένα σύνολο δηλώσεων (statements).

Μια δήλωση RDF (RDF statement) συνίσταται από:

Πόρους (resources) (= κόμβοι) οι οποίοι έχουν ιδιότητες (properties) οι οποίες έχουν τιμές (values) (=κόμβοι,strings)

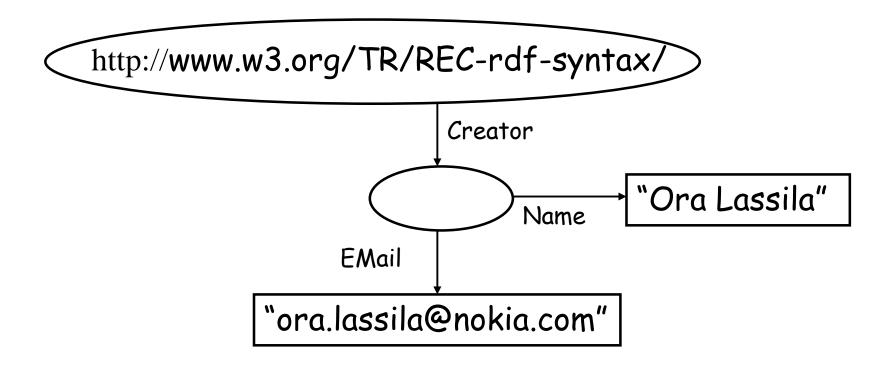
υποκείμενο (subject) κατηγόρημα (predicate) αντικείμενο (object



#### "http://www.w3.org/TR/REC-rdf-syntax/ has creator Ora Lassila"

http://www.w3.org/TR/REC-rdf-syntax/
creator
"Ora Lassila"

# "The individual whose name is Ora Lassila and whose email is ora.lassila@nokia.com is the creator of http://www.w3.org/TR/REC-rdf-syntax/"



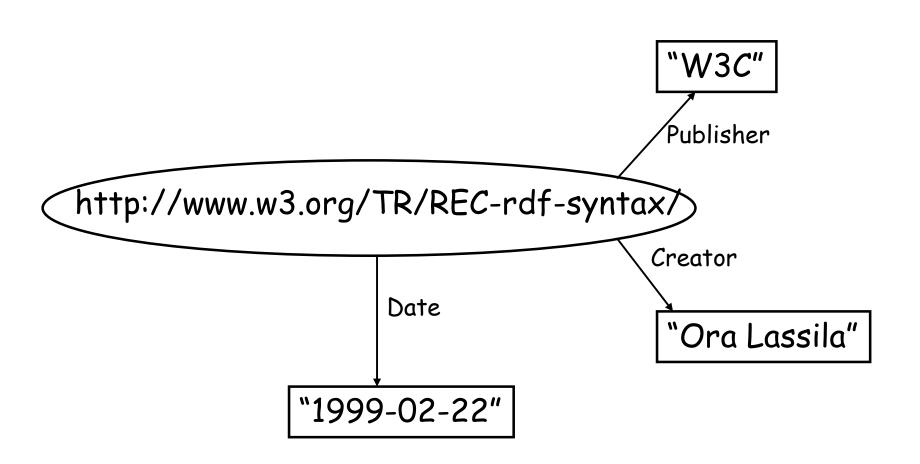
```
Αντίστοιχες τριάδες:

{Creator, [http://www.w3.org/TR/REC-rdf-syntax/], [X]}

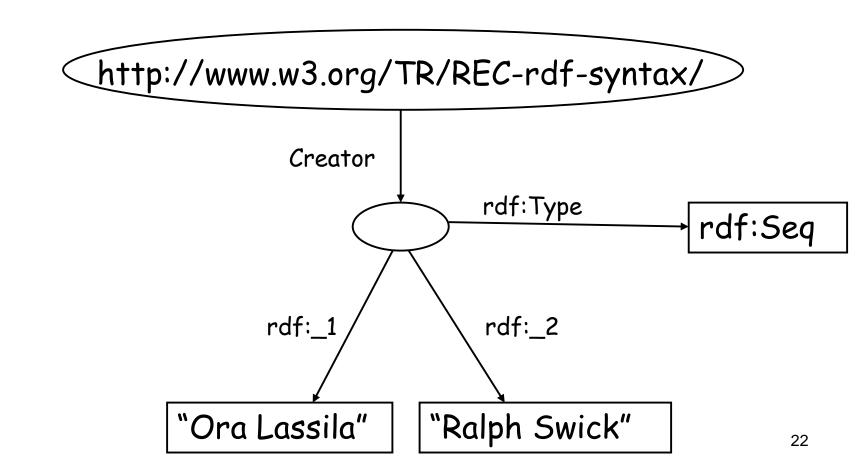
{Name, [X], "Ora Lassila" }

{EMail, [X], "ora.lassila@nokia.com" }
```

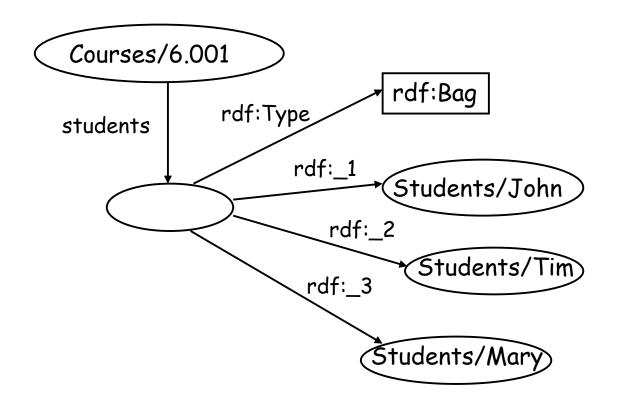
"http://www.w3.org/TR/REC-rdf-syntax/ was created at 1999-02-22, its creator is Ora Lassila and its publisher is W3C"



Τα περιβλήματα (Containers) χρησιμοποιούνται για την αναπαράσταση συλλογών από πόρους και επιτρέπουν ομαδοποίηση των πόρων (ή των τιμών). Τύποι: bag — συλλογή χωρίς διάταξη, seq — διατεταγμένη συλλογή ("sequence"), alt — αναπαριστά εναλλακτικές τιμές. "http://www.w3.org/TR/REC-rdf-syntax/ has as creator(s) Ora Lassila and Ralph Swick"

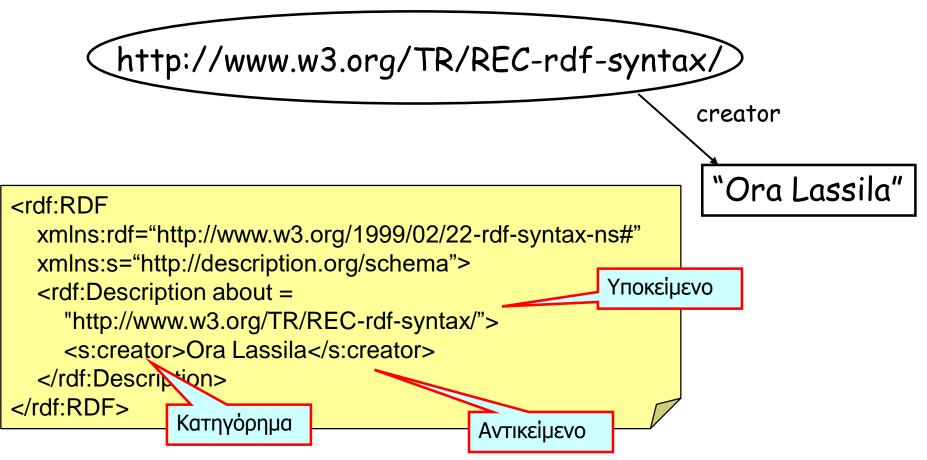


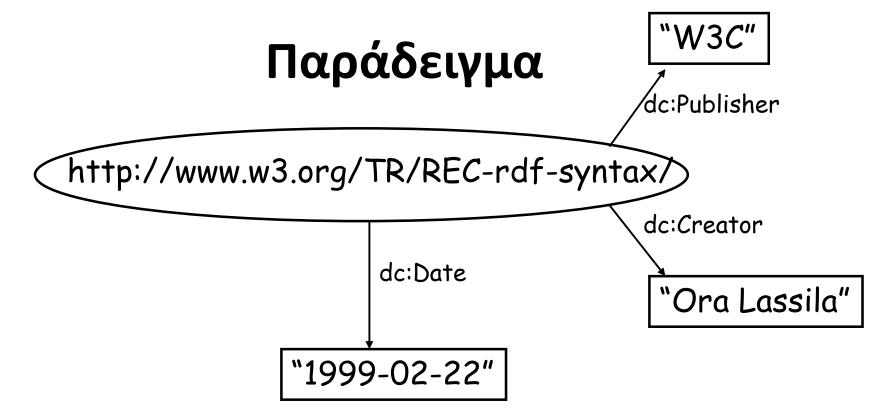
#### "The students in course 6.001 are John, Tim and Mary"



## Βασική σειριακή σύνταξη σε RDF XML

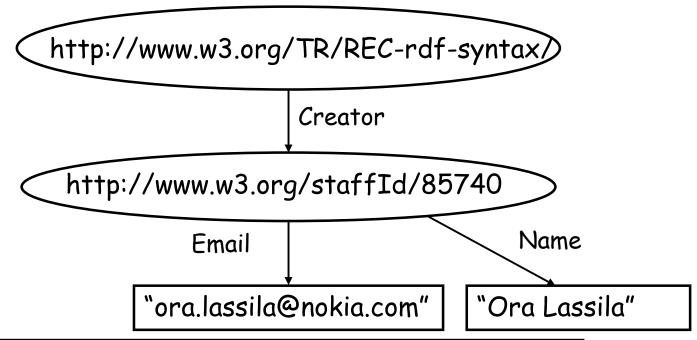
Η σύνταξη ομαδοποιεί πολλαπλές δηλώσεις για τον ίδιο πόρο σε ένα στοιχείο με ετικέτα *Description* 





```
<rdf:RDF
   xmIns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmIns:dc="http://purl.org/dc/elements/1.1/">
   <rdf:Description about = "http://www.w3.org/TR/REC-rdf-syntax/">
   <dc:publisher>W3C</dc:publisher>
   <dc:creator>Ora Lassila</dc:creator>
   <dc:date>1999-02-22</dc:date>
   </rdf:Description>
</rdf:RDF>
```

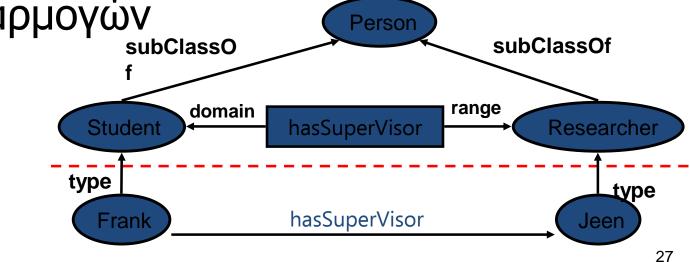
## Παράδειγμα με αναφορές



```
<rdf:RDF xmlns:rdf="..." xmlns:s="...">
    <rdf:Description about = "http://www.w3.org/TR/REC-rdf-syntax/">
        <s:Creator rdf:resource="http://www.w3.org/staffld/85740" />
        </rdf:Description>
        <s:Name>Ora Lassila</s:name>
        <s:Email>ora.lassila@nokia.com</s:Email>
        </rdf:Description>
        </rdf:RDF>
```

#### RDF Schema

- Ορίζει μικρά **λεξιλόγια** για την RDF:
  - Class, subClassOf, type
  - Property, subPropertyOf
  - domain, range
- Τα λεξιλόγια μπορεί να χρησιμοποιηθούν για να οριστούν άλλα λεξιλόγια για συγκεκριμένα πεδία εφαρμογών



## RDF Schema (Παράδειγμα)

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
       xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<rdfs:Class rdf:ID="Person">
 <rdfs:comment>Person Class</rdfs:comment>
 <rdfs:subClassOf rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Resource"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Student">
 <rdfs:comment>Student Class</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Person"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Teacher">
 <rdfs:comment>Teacher Class</rdfs:comment>
 <rdfs:subClassOf rdf:resource="#Person"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Course">
 <rdfs:comment>Course Class</rdfs:comment>
 <rdfs:subClassOf rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Resource"/>
</rdfs:Class>
```

```
<rdf:Property rdf:ID="teacherOf">
 <rdfs:comment>Teacher of a course</rdfs:comment>
 <rdfs:domain rdf:resource="#Course"/>
 <rdfs:range rdf:resource="#Teacher"/>
</rdf:Property>
<rdf:Property rdf:ID="studentsOf">
 <rdfs:comment>List of Students of a course in alphabetical order</rdfs:comment>
 <rdfs:domain rdf:resource="#Course"/>
 <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Seq"/>
</rdf:Property>
<rdf:Property rdf:ID="nameOf">
 <rdfs:comment>Name of a Person or Course</rdfs:comment>
 <rdfs:domain rdf:resource="#Person"/>
 <rdfs:domain rdf:resource="#Course"/>
 <rdfs:range rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Literal"/>
</rdf:Property>
</rdf:RDF>
```

### Μεταδεδομένα

- Τα μεταδεδομένα είναι δομημένη πληροφορία που περιγράφει, εξηγεί, εντοπίζει ή διευκολύνει την ανάκτηση, τη χρήση ή τη 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. τίτλος, συγγραφέας
  - διατήρηση: τεχνικά, π.χ. τύπος υλικού, ψηφιακής μορφή

### Που Βρίσκονται τα Μεταδεδομένα

- Στην ετικέτα του CD
- Στη σελίδα τίτλου
- Στην κορυφή της ιστοσελίδας
- Στον κατάλογο της βιβλιοθήκης
- Στην ηλεκτρονική μορφή μέσα στον πόρο
- Μέρος της κωδικοποίησης στο ηλεκτρονικό δημοσίευμα (μη ορατά)

## Παράδειγμα: Ιστοσελίδα σε HTML

```
<HTML>
<BODY>
 Fiction:
<UL><LI>Author: Milan Kundera</LI>
     <LI>Title: Identity</LI>
     <LI>Published: 1998</LI>
</UL>
 Science: 
<UL><LI>Author: Richard Feynman</LI>
    <LI>Title: The Character of Physical Law</LI>
     <LI>Hardcover</LI>
</UL>
</BODY>
</HTML>
```

Η eXtensible Markup Language XML αναπτύχθηκε ως απλή διάλεκτος της γλώσσας SGML για να διευκολύνει τη διαχείριση, επεξεργασία, διακίνηση 32 και αποθήκευση τεκμηρίων στον Παγκόσμιο Ιστό.

#### **Standards | Librarians and Archivists | Library of Congress**

https://www.loc.gov/librarians/standards

#### **Resource Description Formats**

- BIBFRAME (Bibliographic Framework Initiative)
   Linked data model, vocabulary, and tools for expressing bibliographic data
- <u>EAD (Encoded Archival Description)</u>
   XML markup designed for encoding archival finding aids
- MADS (Metadata Authority Description Standard)
   XML markup for authority data from MARC 21 records and original authority data
- MARC 21 formats
   Representation and communication of descriptive metadata about library items
- MARCXML
   XML representation of MARC 21 data
- MODS (Metadata Object Description Standard)
   XML markup for metadata from existing MARC 21 records and original resource description
- VRA Core
   XML schema and data format description of visual culture and images that document them

#### **Digital Library Standards**

- <u>ALTO</u> Technical metadata for Optical Character Recognition (OCR)
- <u>AudioMD and VideoMD</u> XML schemas for technical metadata on audio- and videobased digital objects
- METS (Metadata Encoding & Transmission Standard) Structure for encoding descriptive, administrative, and structural metadata
- <u>MIX (NISO Metadata for Images in XML)</u> XML schema for encoding technical data elements required to manage digital image collections
- <u>PREMIS (Preservation Metadata)</u> Data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials.
- <u>TextMD (Technical Metadata for Text)</u> XML schema that details technical metadata for text-based digital objects

#### **Information Resource Retrieval Protocols**

- <u>CQL (Contextual Query Language)</u> Formal, user-friendly query language for use between information retrieval systems
- <u>SRU/SRW (Search and Retrieve URL/Web Service)</u> Web services for search and retrieval based on Z39.50 semantics
- <u>Z39.50</u> Supports information retrieval among different information systems

#### **Information Resource Retrieval Standards**

- <u>ISO/DIS 25577</u> Information and documentation (MarcXchange)
- ISO 20775 Schema for holdings information

## Simple HTML Example – Dublin Core http://www.dublincore.org/documents/2000/07/16/usageguide/simple-html/

```
<meta name = "DC.Creator"
     content = "Simpson, Homer">
  <meta name = "DC.Title"
     content = "The Communist Manifesto">
  <meta name = "DC.Creator"
     content = "Marx, K.">
  <meta name = "DC.Creator"
     content = "Engels, F.">
  <meta name = "DC.Title"
     content = "Capital">
```

# The Dublin Core Metadata Initiative DCMI http://www.dublincore.org/documents/2001/04/12/usageguide/generic/

Dublin Core	Examples	
Title	A Pilot's Guide to Aircraft Insurance	
Author/Creator	Federal Aviation Administration	
Subject and Keywords	Aircraft leasing and renting	
Description	Electronic illustrated guide to airport markings and lighting signals for airports with low visibility conditions	
Publisher	MIT	
Other Contributor	Aviation Safety Program	
Date	1998	
Resource Type	Type="text" Type="image"	
Format	Text/pdf image/gif 640 x 512 pixels	
Resource Identifier	"0385424728" [ISBN]	
Source	"RC607.A26W574 1996"	
Language	en	
Relation	Title="Electronic illustrated guide to airport markings and lighting signals for airports with low visibility conditions" Relation="IsFormatOf Illustrated guide to airport markings and lighting signals for airports with low visibility conditions, 2nd edition"	
Coverage	Boston, MA	
Rights Management	Access limited to members	