

reLOAD

repository for
linked open
archival data

Archivio Centrale dello Stato, 26 giugno 2012

Un progetto di

Archivio Centrale dello Stato, Istituto beni culturali Regione Emilia-Romagna e regesta.exe

Sommario

- ❖ Gli obiettivi del progetto, *Giovanni Bruno*
- ❖ Strumenti del Web semantico per l'utilizzo dei LOD in ambito archivistico, *Silvia Mazzini*
- ❖ L'ontologia della descrizione archivistica: OAD, *Chiara Veninata*
- ❖ L'ontologia EAC-CPF, *Francesca Ricci*

reLOAD

Gli obiettivi del progetto

Giovanni Bruno

Lavorare con i Linked Data

- ❖ Raw Data Now
- ❖ Il Web of Data
- ❖ LOD : LODLAM

Gli obiettivi del progetto_Giovanni Bruno

Lavorare con i Linked Data

Raw Data Now

Gli obiettivi del progetto_Giovanni Bruno

- ☛ Tim Berners-Lee, TED, 2009
(http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html)
- ☛ Open Data: trasparenza | condivisione
- ☛ Open Linked Data: riuso | flessibilità



Lavorare con i Linked Data

Il Web of Data

Gli obiettivi del progetto_Giovanni Bruno

-  Strumenti e tecnologie per consentire un'interazione affidabile tra i dati sulla rete
(<http://www.w3.org/standards/semanticweb/>)
-  Liberare i dati rinchiusi in database
-  RDF: definisce le regole di sintassi per pubblicare e linkare i propri dati
-  Vocabolari per organizzare i dati:
 - ontologie basate su OWL
 - thesauri formalizzati in SKOS
-  Endpoint SPARQL
-  URI stabili

Lavorare con i Linked Data

 LOD : LODLAM

Gli obiettivi del progetto_Giovanni Bruno



Linked Open Data in Libraries,
Archives & Museums

- <http://lod-lam.net>
- GLAMS su CKAN
(<http://datahub.io/group/open-glams>): 57 istituti e progetti



Library Linked Data Incubator Group
del W3C

- Final Report
(<http://www.w3.org/2005/Incubator/lld/XGR-lld-2011025/>)
- Report on Use Cases
(<http://www.w3.org/2005/Incubator/lld/XGR-lld-usecase-2011025/>)
- Report on Datasets, Value
Vocabularies, and Metadata
Element Sets
(<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-2011025/>)

LOD per il patrimonio culturale

-  Alcuni progetti
-  Archivi e LOD
-  Principi

Gli obiettivi del progetto_Giovanni Bruno

LOD per il patrimonio culturale

Alcuni progetti

-  Library of Congress Subject Headings (<http://id.loc.gov/authories/subjects.html>)
-  British National Bibliography (<http://talis-systems.com/case-study/british-library-explores-linked-data/>): 2.6 milioni di descrizioni (84,961,180 triples) di libri pubblicati in UK negli ultimi 60 anni
-  EDM: Europeana Data Model (<http://pro.europeana.eu/web/guest/linked-open-data>): 2,4 milioni di metadata forniti da 8 providers diretti e provenienti da 200 istituti di 15 paesi
-  WorldCat Linked Data (<http://www.oclc.org/worldcat/default.htm>): il progetto di OCLC dà accesso a 271 milioni di record bibliografici, codificati con Schema.org, per oltre 1 miliardo di singoli elementi detenuti da istituzioni partecipanti

Gli obiettivi del progetto_Giovanni Bruno

LOD per il patrimonio culturale

Archivi e LOD

-  LOCAH Project | Linking Lives : i due progetti utilizzano un subset dei dati di oltre 200 archivi inglesi, scozzesi e gallesi di Archives Hub (<http://archiveshub.ac.uk/linkinglives>)
-  Thésaurus pour l'indexation des archives locales publié par les Archives de France: contesto storico, tipologie documentarie, materie (<http://www.archivesdefrance.culture.gouv.fr/thesaurus/>)
-  SNAC: The Social Networks and Archival Context Project (<http://socialarchive.iath.virginia.edu/>)
-  Conversion of EAD into EDM Linked Data (<http://ceur-ws.org/Vol-801/paper7.pdf>)

Gli obiettivi del progetto_Giovanni Bruno

LOD per il patrimonio culturale

Principi



favorire l'interoperabilità

- riuso sistematico di vocabolari e ontologie esistenti
- facile integrazione di ontologie pubbliche e vocabolari specifici



arricchire le descrizioni

- collegamenti molteplici a risorse esterne
- granularità e flessibilità nella descrizione delle risorse



riutilizzo selettivo dei dati

- per la creazione di nuovi servizi e prodotti da parte di soggetti diversi
- riuso creativo dei materiali digitali | nuovi mercati



rischi

- Loss of Attribution
- Loss of potential Income

Gli obiettivi del progetto_Giovanni Bruno

Questo progetto

- 👉 Obiettivi
- 👉 Risorse
- 👉 Un nuovo modello organizzativo?

Gli obiettivi del progetto_Giovanni Bruno

Questo progetto

Obiettivi

-  Rete di dati archivistici collegati
-  Condividere le informazioni descrittive di dettaglio
-  Accrescere visibilità delle risorse archivistiche sul Web
-  Favorire il riuso dei dati archivistici anche in altri contesti, in collegamento con altre fonti informative

Gli obiettivi del progetto_Giovanni Bruno

Questo progetto

Risorse

Gli obiettivi del progetto_Giovanni Bruno



Ontologia della descrizione archivistica

- Crescita di risorse digitali
- Diffusione di EAD
- Condivisione dei principi delle ISAD(G)



Strumenti e metodologie per l'annotazione semantica

- Thesauri SKOS
- Ontologie pubbliche: Dbpedia, Geonames, ecc.

Questo progetto

 Un nuovo modello
organizzativo?

 Decentramento

 Partecipazione

 Competenze

 Efficienza

Gli obiettivi del progetto_Giovanni Bruno

reLOAD

Strumenti del Web semantico per l'utilizzo dei LOD in
ambito archivistico

Silvia Mazzini

Risorse e standard del semantic web

-  RDF
-  SKOS
-  SPARQL

Risorse e standard del semantic web

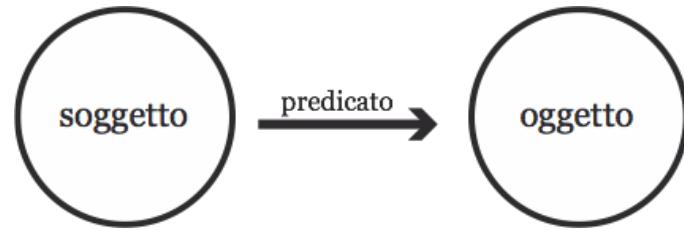


RDF

- ◆ RDF is a standard model for data interchange on the Web
- ◆ RDF è lo strumento base proposto dal W3C per la codifica, lo scambio e il riutilizzo di metadati strutturati
- ◆ RDF consente l'interoperabilità tra applicazioni che si scambiano informazioni sul Web



RDF



Tripla

La tripla è l'unità informativa minima che si articola in soggetto (subject), relazione (predicate) e oggetto (object).

Risorse e standard del semantic web



RDF

-  OAD Ontology
-  EAC-CPF Ontology
-  Ontologie pubbliche

Risorse e standard del semantic web



SKOS

- ❖ è un modello di organizzazione dei dati finalizzato alla condivisione e al collegamento di "sistemi di organizzazione della conoscenza"
- ❖ ISAD skos
- ❖ titolario astengo skos

Risorse e standard del semantic web



SPARQL

- 💡 È un linguaggio standard per interrogare le risorse RDF
- 💡 ENDPOINT SPARQL

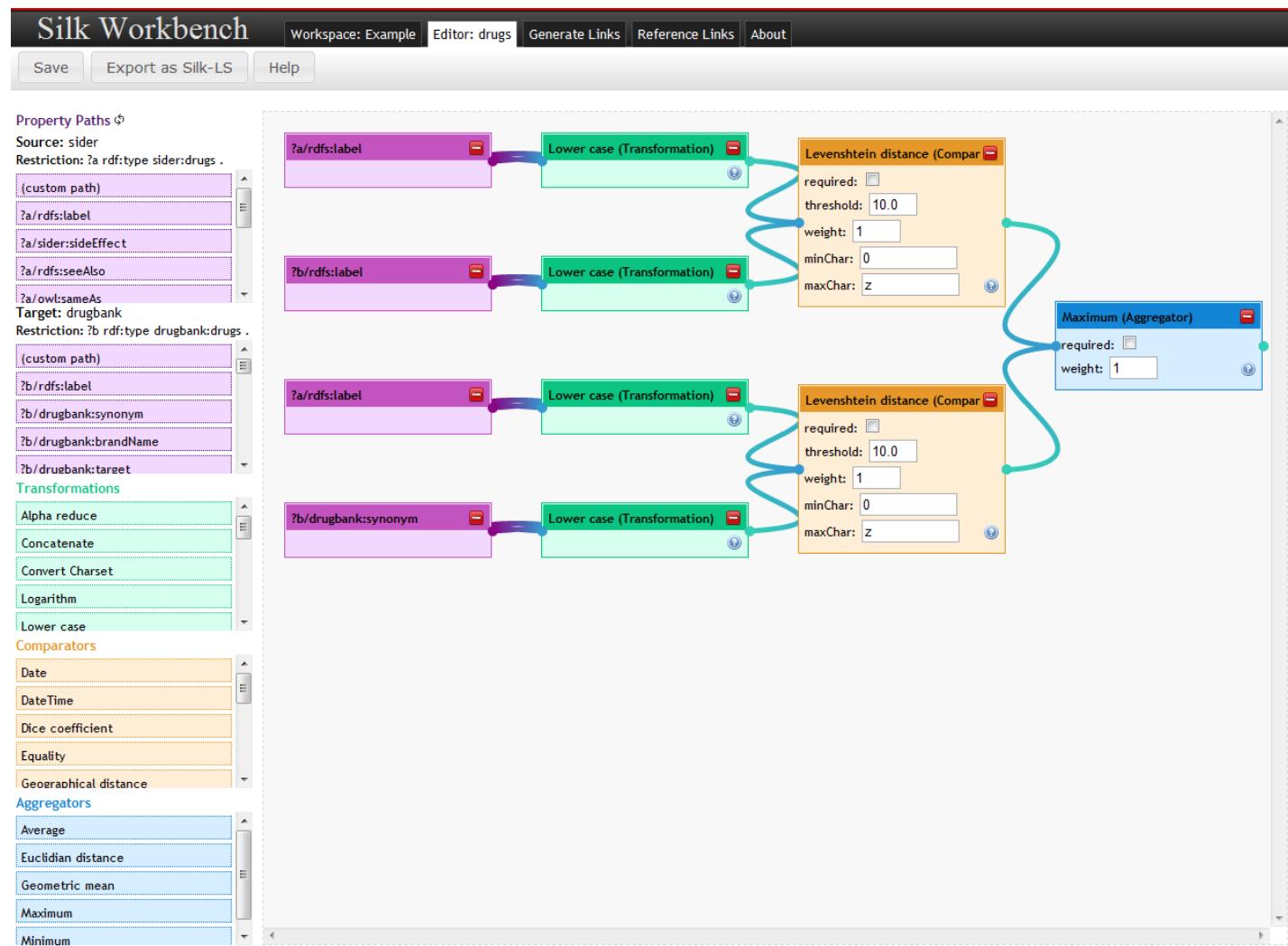
Strumenti

-  mush-up
-  semantic annotation
-  grafi

Linked data per gli archivi_Silvia Mazzini

Strumenti

 mush-up



Linked data per gli archivi_Silvia Mazzini

Strumenti

 semantic annotation

```

<rdf:RDF xml:base="http://lod.xdams.org/reload/oad">
  -<skos:ConceptScheme rdf:about="titolarioMAIC/titolario">
    <dc:creator>regesta</dc:creator>
    <dc:language>it</dc:language>
    <dc:rights>All rights reserved..</dc:rights>
    <skos:prefLabel>Thesaurus del titolario MAIC</skos:prefLabel>
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Thesaurus del titolario MAIC</rdfs:label>
  </skos:ConceptScheme>
  -<skos:Concept rdf:about="titolarioMaic/T">
    <skos:inScheme rdf:resource="titolarioMAIC/titolario"/>
<c id="IT-ER-IBC-AS00064-0000156">
  <unittitle>Categoria XI: Agricoltura - Industria e Commercio</unittitle>
  -<scopecontent>
    Contiene: Classe 1: Fasc. 1, Agricoltura - Provvedimenti ed iniziative agricole - Battaglia del grano Fasc. 2, Caccia -
  -<subject>
    <content>Pesca</content>
    <content>Battaglia del grano</content>
    <content>Caccia</content>
<oad:uod rdf:about="uod/IT-ER-IBC-AS00235-0000170">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Miniera di rame dell'Idice. Carteggio</rdfs:label>
  <dc:title>Miniera di rame dell'Idice. Carteggio</dc:title>
  <oad:has_level rdf:resource="levelOfDescription/series"/>
  <dcterms:isPartOf rdf:resource="uod/IT-ER-IBC-AS00235-0000000"/>
  <dcterms:provenance rdf:resource="organization/ibc"/>
  <dcterms:subject rdf:resource="titolarioMaic/XXX"/>
  <dc:date>04 gennaio 1900 - 18 luglio 1906</dc:date>
  <oad:startDate>19000104</oad:startDate>
  <oad:endDate>19060718</oad:endDate>
  -<oad:scopeAndContent>
    La serie è costituita dal carteggio riguardante le ricerche minerarie e le miniere di Bisano e Gurlano, nei comuni di Monterenzio e Monghidoro.
  </oad:scopeAndContent>
  <oad:languageScriptsOfMaterials rdf:resource="http://id.loc.gov/vocabulary/iso639-1/it"/>
  <dcterms:isReferencedBy rdf:resource="http://archivi.ibc.regione.emilia-romagna.it/ead-str/IT-ER-IBC-AS00235-0000170"/>
</oad:uod>

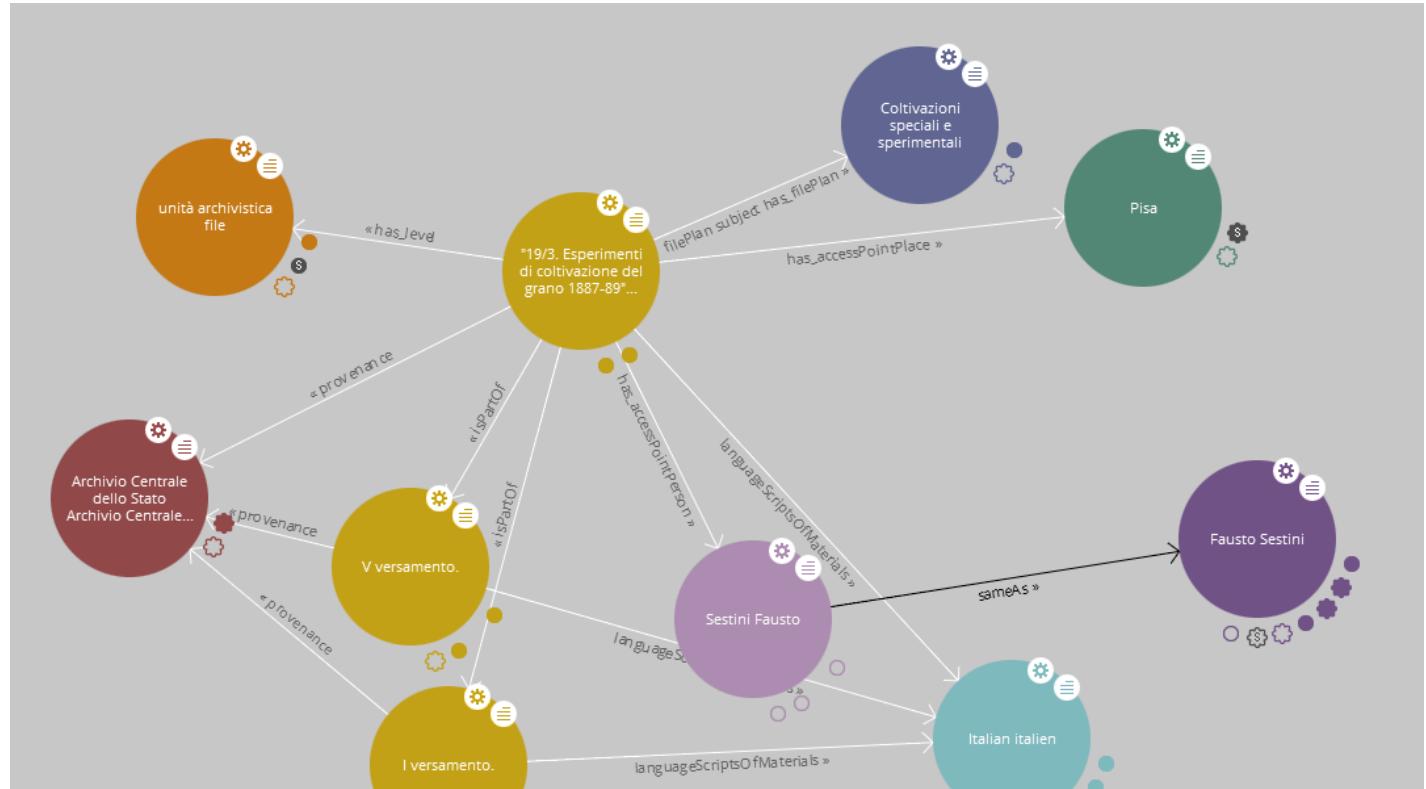
```

Linked data per gli archivi_Silvia Mazzini

Strumenti



grafi



Linked data per gli archivi_Silvia Mazzini

reload

L'ontologia della descrizione
archivistica: OAD

Chiara Veninata

Navigazione

- 👉 RDF browser
- 👉 a “faccette”
- 👉 ReLoad

Linked data per gli archivi_Silvia Mazzini

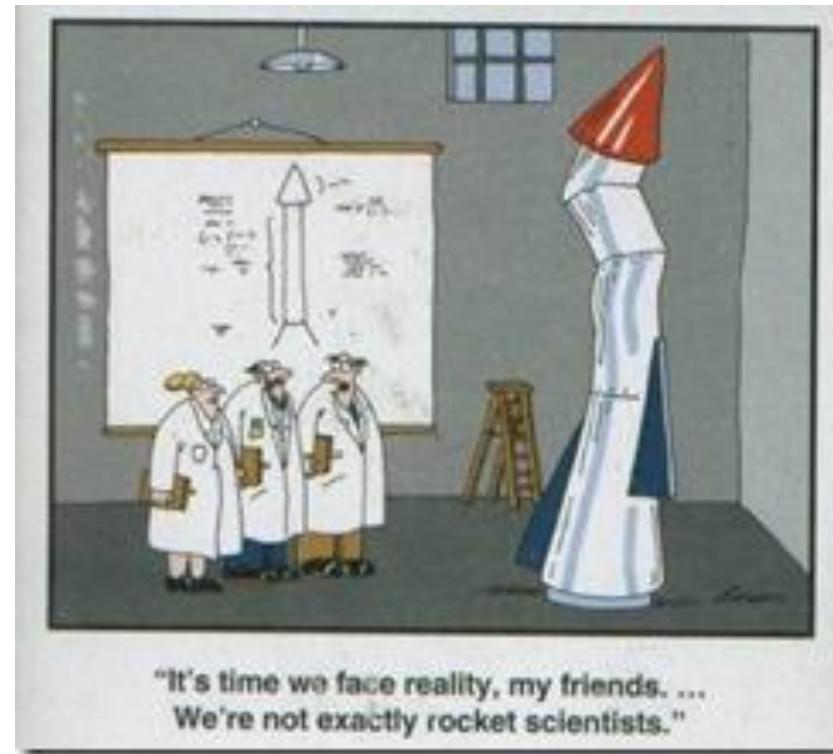
L'Ontologia OAD

- 👉 La sperimentazione
- 👉 L'ontologia OAD
- 👉 I risultati (meglio: le premesse!)

La sperimentazione

- 👉 Berners-Lee 4 principles
- 👉 Berners-Lee 5 stars deployment schema
- 👉 I risultati...

- 👉 Verificare il valore aggiunto dei LOD in ambito archivistico
- 👉 Sperimentare nuovi percorsi di ricerca
- 👉 Aggiungere informazioni ai dati
- 👉 Presentare ai ricercatori i dati archivistici come una parte (significativa) di un insieme di fonti di vario tipo
- 👉 Immettere direttamente nella rete dati ad alto valore informativo (record archivistici) e ... vedere cosa succede ...(?!?)



Oggetto della sperimentazione

L'ontologia OAD_Chiara Veninata

ACS



Inventario della Direzione
Generale dell'Agricoltura

IBC

- Inventario dell'archivio storico della Provincia di Piacenza
- Inventario dell'archivio storico comunale di Alfonsine
- Carte Andrea Costa
- Archivio Giovanni Codronchi jr

Criticità

“expose yourself,
be vulnerable”



L'ontologia OAD_Chiara Veninata

- ➔ Assenza di una ontologia “ufficiale” di riferimento
- ➔ Modello dati complesso
- ➔ Metodologia ancora poco diffuse in ambito culturale (LOD LAM)
- ➔ Occorre sperimentare: necessità di poter disporre di *use cases* concreti
- ➔ Motivare altri *content providers* a contribuire coi loro dati

Il primo passo



L'ontologia OAD_Chiara Veninata



Ontologia per la descrizione archivistica (OAD)



Quali oggetti trattare come “risorse” cui assegnare degli URI, cosa considerare “classe” dell’ontologia, cosa “proprietà”



Archival description, EAD & Linked Data

OAD Ontology



Classes

uod
 levelOfDescription
 administrativeBiographicalHistory
 findingAid
 publicationNote
 eadElement
 place



Datatype Properties

referenceCode
 title
 date
 extentAndMedium
 archivalHistory
 immediateSourceOfAcquisitionOrTransfer
 scopeAndContent
 appraisalDestructionAndSchedulingInformation
 accruals
 systemOfArrangement
 conditionsGoverningAccess
 conditionsGoverningReproduction
 physicalCharacteristicsAndTechnicalRequirements
 existenceAndLocationOfOriginals
 existenceAndLocationOfCopies
 otherlevel
 note
 archivistsNote



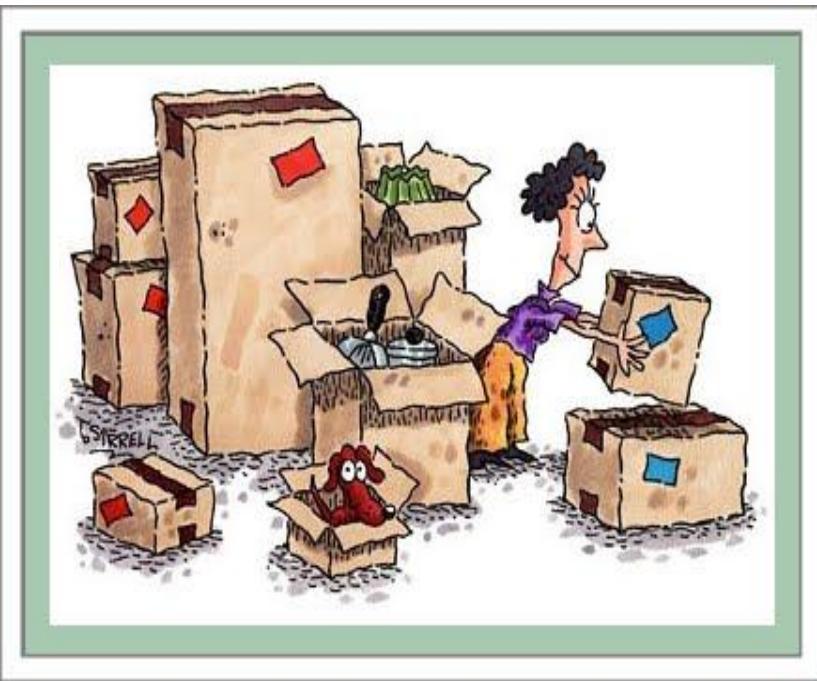
Object Properties

has_level
 has_nameOfCreator
 has_administrativeBiographicalHistory
 has_filePlan
 has_languageScriptsOfMaterial
 has_findingAid
 has_relatedUnitsOfDescription
 has_publicationNote
 has_repository
 use_eadElement
 has_accessPoint
 has_accessPointPerson
 has_accessPointPlace
 has_accessPointCorporateBody
 has_accessPointFamily

- 💡 il data model RDF si fonda sui dettami dello standard ISAD (G)
- 💡 non si propone come standard alternativo ma semplicemente come una diversa formalizzazione utile al *semantic web* e favorevole all'interoperabilità
- 💡 ISAD (G) è stato integrato per la parte concernente le c.d. “voci d’indice” e i collegamenti con il soggetto conservatore e il soggetto produttore

- 💡 rapporto con EAD
- 💡 integrazione con altre ontologie
- 💡 nella scheda RDF alcune informazioni, soprattutto i metadati descrittivi classici come titolo, data, autore, sono state duplicate
- 💡 per agevolare il collegamento con risorse esterne ed alimentare i LOD → owl:sameAs

Data model



When moving to a new place we bring loads of baggage and stuff from our old house that we feel will be necessary in our new abode. (...) Just like moving house, we initially grasp for the familiar and that might not always be helpful.

“Non sottostimare la pulizia dei dati”

Linking Lives



L'ontologia OAD_Chiara Veninata

*The first steps given for
publishing your content as
LOD are:*

- 1. Analyse your data*
- 2. Clean your data*
- 3. Model your data*

*Linked Open Data: The Essentials
(Bauer, Kaltenbock)*



http://issuu.com/andreas_blumauer/docs/linked-open-data-essentials

*...moving to a
new place...*

Il processo di trasformazione dei dati

- XML EAD -> (XSLT) -> RDF
- Triple store OAD
- Semantic annotation (Stanbol)
- Link a risorse “esterne”

```
<c id="IT-ACS-AS0002-0026914" level="file">
  <did>
    <unitid countrycode="IT" encodinganalog="ISAD 1 - 1 reference code" repositorycode="ACS" type="segnatura definitiva">489</unitid>
    <unitid type="segnatura originale">289</unitid>
    <unitid type="numero di classificazione">15.5<title linktype="simple">Comizi agrari</title>
  </unitid>
    <note label="altre segnature"><p>Vecchi numeri di busta e di fascicolo utilizzati negli elenchi di versamento: b. 25 fasc. 219</p></note>
    <container encodinganalog="ISAD 1 - 5 extent and medium of the unit of description" type="busta">0047</container>
    <unitdate calendar="gregorian" encodinganalog="ISAD 1 - 3 date(s)" era="ce" normal="18920101-18961231">1892-1896</unitdate>
    <unittitle encodinganalog="ISAD 1 - 2 title">"15/5. Trapani".</unittitle>
  </did>
</c>
```

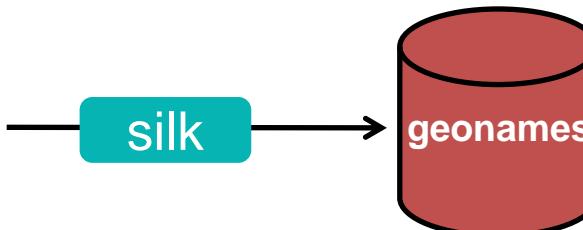
XSLT

1

```
<oad:uod rdf:about="uod/IT-ACS-AS0002-0026914">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">"15/5. Trapani".</rdfs:label>
  <dc:title>"15/5. Trapani".</dc:title>
  <oad:has_level rdf:resource="levelOfDescription/file"/>
  <dcterms:isPartOf rdf:resource="uod/IT-ACS-AS0002-0000002"/>
  <dcterms:provenance rdf:resource="organization/acs"/>
  <oad:referenceCode>489</oad:referenceCode>
  <dc:date>1892-1896</dc:date>
  <oad:startDate>1892</oad:startDate>
  <oad:endDate>1896</oad:endDate>
  <oad:extentAndMedium>busta 47</oad:extentAndMedium>
  <oad:has_filePlan rdf:resource="titolarioMaic/15.5"/>
  <dcterms:subject rdf:resource="titolarioMaic/15.5"/>
  <oad:has_accessPointCorporateBody rdf:resource="entiACS/Comizi_agrari"/>
  <oad:has_accessPointPlace rdf:resource="luoghiACS/Trapani"/>
  <oad:languageScriptsOfMaterials rdf:resource="http://id.loc.gov/vocabulary/iso639-1/it"/>
  <dcterms:isReferencedBy rdf:resource="http://search.acs.beniculturali.it/OpacACS/inventario/IT-ACS-AS0002-0026914"/>
</oad:uod>
```

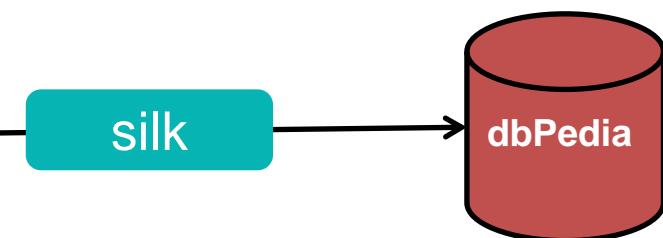
2

```
<oad:place rdf:about="luoghiACS/Trapani"/>
<skos:Concept rdf:about="luoghiACS/Trapani">
|  <skos:prefLabel>Trapani</skos:prefLabel>
|  <rdfs:label xml:lang="it">Trapani</rdfs:label>
|  <dc:title xml:lang="it">Trapani</dc:title>
</skos:Concept>
```



3

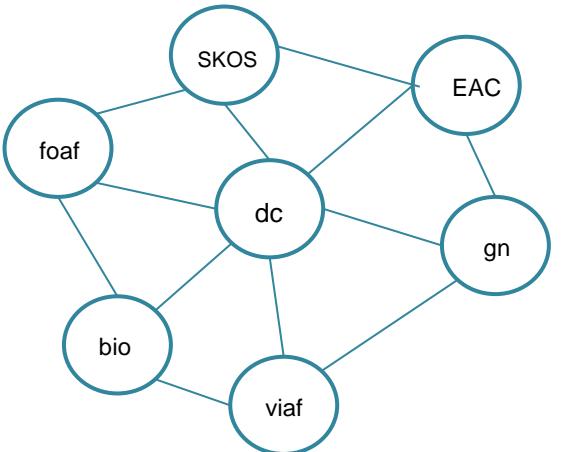
```
<foaf:Organization rdf:about="entiACS/Comizi_agrari">
|  <foaf:name>Comizi agrari</foaf:name>
|  <rdfs:label xml:lang="it">Comizi agrari</rdfs:label>
|  <dc:title xml:lang="it">Comizi agrari</dc:title>
|  <owl:sameAs rdf:resource="http://it.dbpedia.org/page/Comizi_agrari"/>
</foaf:Organization>
<skos:Concept rdf:about="entiACS/Comizi_agrari">
|  <skos:prefLabel>Comizi agrari</skos:prefLabel>
|  <rdfs:label xml:lang="it">Comizi agrari</rdfs:label>
</skos:Concept>
```



L'Ontologia

- RDF Vocabularies

- SKOS: <http://www.w3.org/2004/02/skos/core#>
- FOAF: <http://xmlns.com/foaf/0.1/>
- DC: <http://purl.org/dc/elements/1.1/>
- BIO: <http://purl.org/vocab/bio/0.1/>
- VIAF: <http://viaf.org/ontology/1.1/#>
- GN: <http://www.geonames.org/ontology#>
- EAC: <http://archivi.ibc.regione.emilia-romagna.it/ontology/eac-cpf/#>



Navigare i dati



reLOAD

L'ontologia EAC-CPF

Francesca Ricci

L'Ontologia

-  EAC-CPF Schema
-  EAC-CPF Descriptions
-  RDF Vocabularies

L'Ontologia

EAC-CPF Schema



Classes

element
attribute
controlled_value



Properties

mayContainElement
containRequiredElement
hasAttribute
hasRequiredAttribute
mayContainValue
reference
isElementOf
isRequiredElementOf
isAttributeOf
isRequiredAttributeOf
isControlledValueOf
mayContainDatatype
diagram_ref
occurrence

L'Ontologia

EAC-CPF Descriptions

L'ontologia EAC-CPF_Francesca Ricci



Classes

entity, person, corporateBody,
family , controlArea,
descriptionArea, nameArea,
language, place, relation



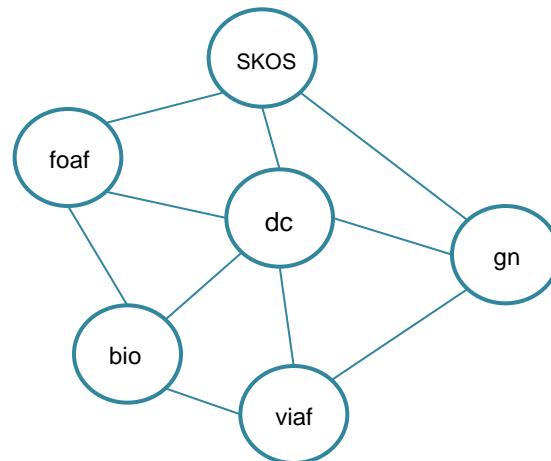
Properties

authorizedForm, biogHist,
control, conventionDeclaration,
cpfRelation, cpfRelationType,
description, existDates, function,
generalContext,
languageDeclaration,
languageUsed, legalStatus,
localTypeDeclaration,
maintenanceAgency,
maintenanceHistory,
maintenanceStatus, mandate,
nameEntry, occupation,
publicationStatus, recordID,
resourceRelation,
resourceRelationType, source,
structureOrGenealogy

L'Ontologia

☞ RDF Vocabularies

- ☞ SKOS: <http://www.w3.org/2004/02/skos/core#>
- ☞ FOAF: <http://xmlns.com/foaf/0.1/>
- ☞ DC: <http://purl.org/dc/elements/1.1/>
- ☞ BIO: <http://purl.org/vocab/bio/0.1/>
- ☞ VIAF: <http://viaf.org/ontology/1.1/#>
- ☞ GN: <http://www.geonames.org/ontology#>



Lo stato dell'arte

- 👉 Il progetto
- 👉 Concetti e principi
- 👉 La sperimentazione

Lo stato dell'arte

Il progetto

- ◆ l'ontologia EAC-CPF è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse
- ◆ il 25 ottobre 2011 il Library Linked Data Incubator Group del W3C ha pubblicato un report sui Datasets, Value Vocabularies e Metadata Element Sets ed ha inserito ufficialmente l'ontologia EAC-CPF tra i vocabolari disponibili

Lo stato dell'arte

Concetti e principi

-  il data model RDF si fonda sullo schema ufficiale dello standard EAC-CPF, non si propone come standard alternativo ma semplicemente come una diversa formalizzazione utile al semantic web e favorevole all'interoperabilità
-  non sono stati aggiunti concetti nuovi, non definiti nello schema XML
-  non sono stati introdotti nel data model gli elementi usati nello schema XML per parcellizzare le informazioni descrittive, mirando ad un accorpamento delle informazioni che favorisse la semplificazione e l'astrazione della struttura

Lo stato dell'arte

La sperimentazione

-  nella scheda RDF alcune informazioni, soprattutto i metadati descrittivi classici come titolo, data, autore, sono state duplicate utilizzando altri vocabolari RDF universalmente noti e diffusi come Dublin Core e FOAF per permettere una naturale interoperabilità con altre risorse simili
-  per agevolare il collegamento con risorse esterne ed alimentare i Linked Archival Data, per tutte quelle risorse di cui era possibile reperire URI alternative o informazioni alternative su altri siti web o su altri authority sono stati inseriti i riferimenti
-  http:// qui va messo l'indirizzo web
ibc/semantic

Navigare i dati

 Elda

 Da EAC a dbpedia

Elda è un linked data api open source

Navigare i dati



Elda

ISTITUTO PER I BENI ARTISTICI, CULTURALI E NATURALI
Soprintendenza per i beni librari e documentari

Semantic Environment for IBC Archivi Authority Records

Linked Data API - IBC

html | json | plainhtml | rdf | text | ttl | xml

Search Results

Show Search Form

Anselmo Marabini, 1865-1948

| | | | |
|-------------|------------|---|-----------------|
| description | bio birth | date | 16 ottobre 1865 |
| | luogo | http://geonames.org/3175537/ | |
| | bio death | date | 09 ottobre 1948 |
| | luogo | http://geonames.org/3175537/ | |
| | name entry | cognome | Marabini |
| | | nome | Anselmo |

same as

<http://viaf.org/viaf/90281289/>

<http://www.treccani.it/enciclopedia/anselmo-marabini/>

Romeo Galli, 1872-1945

| | | | |
|-------------|-----------|---|------------------|
| description | bio birth | date | 10 dicembre 1872 |
| | luogo | http://geonames.org/3175537/ | |

mappa

Mappa ▾

An interactive map of Europe and the Mediterranean region. Two specific locations are marked with red pins: one near the city of Parma in Italy and another near the city of Istanbul in Turkey. The map includes labels for various countries like France, Germany, Poland, and Russia, along with the Mediterranean Sea and surrounding bodies of water.

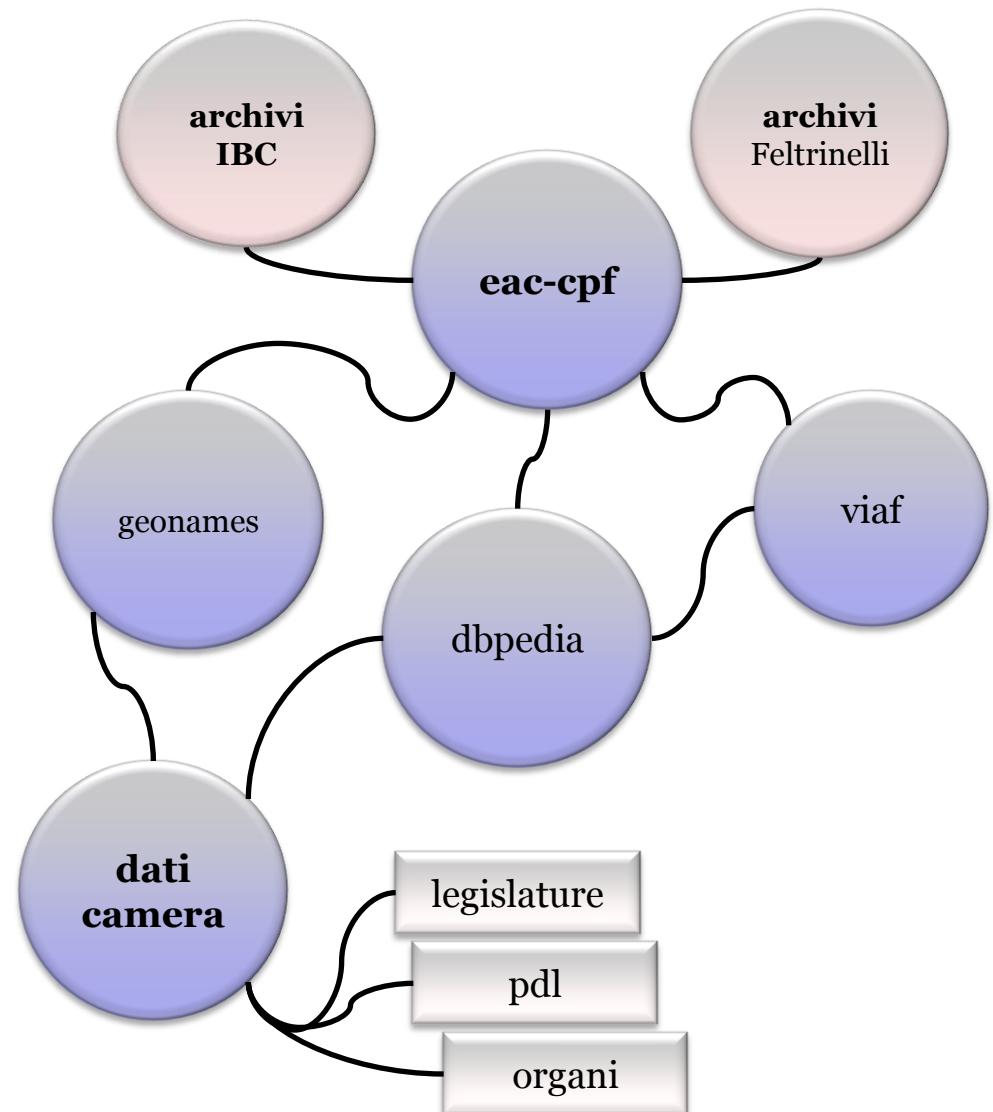
On This Page

Anselmo Marabini, 1865-1948
Romeo Galli, 1872-1945
Luigi Sassi, 1853-1902
Alfredo Xella, 1874-1950
Anna Kuliscioff, 1854-1925
Serafino Mazzotti, 1843-1925
Angelo Negri, 1855-1932
next >

L'ontologia EAC-CPF_Francesca Ricci

Navigare i dati

👉 Da EAC a dbpedia



grazie.