



Archives & Linked Open Data

are our tools ready to "complete the picture"?

Anila Angjeli

{ BnF | Bibliothèque
nationale de France

*Some observations from
a metadata and standards specialist,
a librarian collaborating with archivists to
interconnect our data*

Whether texts, images,
websites, etc.

**No matter what
information you have**

Everything is interconnected
in a network of interrelated
entities

1. You are dealing with real things
2. You only have part of the picture
3. Your objects live in the graph



"Bringing History Alive : Telling stories with Linked Data and open source tools -
Australian War Memorial"

by **David Peterson**, at Drupal Down Under 13-15 January 2012, Melbourne

<http://2012.drupaldownunder.org/session/telling-stories-linked-data-australian-war-memorial>

Collective history building, in action

Pieces of "stories"

Pinned to places

Pieced together

Available to others
for reuse

to build new
perspectives:

- social,
- historical,
- ethnological,
- ...



“Records” from institutions meet those from individuals 3

Archival data

- Huge amount of historical records
 - providing documentary evidence of all domains of human activity
(people, their lives and activities, places, events, intellectual creations, artifacts, ...)
 - held in repositories, curated in organized collections
 - preserved in shelves, boxes, folders, digital files
- Represent interrelated pieces of historical evidence
 - ... as interrelated were the events and people that gave birth to them
- Challenge: reveal those interrelations!

Pro-active approach to access

- Reveal the existence of the records, describe
 - based on archival community standards
 - Content: ISAD(G), ISAAR(CPF), etc.
 - Communication structures: EAD, EAC-CPF, etc.
 - Other "knowledge organization systems": thesauri, etc.
- Make the Finding Aids accessible
 - available on the Internet
 - easily searchable
- Today the mission of access - extended
 - Reveal hidden treasures – inform about the context – highlight implicit connections – "actionable" links to other resources

Keep pace with the changing context

An ever-broadening vision of access to archival records and their interrelation with other resources

- Archival standards (both content and communication ones) are likely to change
- EAC-CPF – released, 2010
- EAD – under revision
- ICA content standards – process underway

a SAA communication standard for describing "people"

A step forward towards archival information as
Linked Open Data

- While ISAAR (CPF) focuses primarily on the notion of provenance
- EAC-CPF applies to **all "people" documented in the records** (contextual information essential to the understanding of records)
- Two focal points:
 - Identity resolution (accurately identify "people")
 - Make explicit relationships (key dimension of the context)
- Data centric
- Identifiers
- Other name spaces
- Linking mechanisms

fully based on the schema

http://archivi.ibc.regione.emilia-romagna.it/ontology/reference_document/referencedocument.html

Developed by:

Silvia Mazzini - [Regesta.exe](#)

Francesca Ricci – [IBC](#) (Istituto beni artistici culturali e naturali della Regione Emilia-Romagna)

"EAC-CPF standard is quite different from his relative EAD widely used in archival description, especially because EAC-CPF standard **has not the concept of "inheritance"** and is more **"data-centric"** than EAD."

For interoperability purposes,

- title, date, author, etc.
- are duplicated by RDF terms such as
- Dublin Core, FOAF, etc.

- use of **owl:sameAs**
for linking to external resources

External RDF vocabularies references

DC: <http://purl.org/dc/elements/1.1/>

DC Terms: <http://purl.org/dc/terms>

OWL: <http://www.w3.org/2002/07/owl#>

Bio: <http://purl.org/vocab/bio/0.1/>

FOAF: <http://xmlns.com/foaf/0.1/>

VIAF: <http://viaf.org/ontology/1.1/#>

Geonames:

<http://www.geonames.org/ontology#>⁸

instances expressed in properties and subproperties to encode records and show them as Linked Open Data.

[illegible]



as Linked Open Data - live

 prototype at: <http://socialarchive.iath.virginia.edu/xtf/search>

```

    <span class="arcrole">creatorOf</span>
  </a>
</div>
</div>
<h3>
  <span>
    <a href="#">Linked Data (1)</a>
  </span>
</h3>
<div>
  <div class="related"
    about="http://socialarchive.iath.virginia.edu/xtf/view?docId=Bush+Vannevar+1890-1974-cr.xml#entity">
    <div class="arcrole">sameAs</div>
    <a xmlns:owl="http://www.w3.org/2002/07/owl#" rel="owl:sameAs"
      title="Virtual International Authority File"
      href="http://viaf.org/viaf/15572358">http://viaf.org/viaf/15572358</a>
    </div>
  </div>

```

RDFa owl:sameAs

HTML5 microdata in chronlist

```

<div class="bioHist">
  <div class="chronList">
    <div itemscope="itemscope">
      <time itemprop="date">1890, Mar. 11</time>
      <div itemprop="placeEntry">Everett, Mass.</div>
      <div itemprop="event">Born,
      Everett, Mass.
    </div>
  </div>

```

Also experimenting with italian RDF via xsst

```

<rdf:RDF xmlns="urn:isbn:1-931666-33-4" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:eac-cpf="http://archivi.ibc.regione.emilia-romagna.it/ontology/eac-cpf/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dcterms="http://purl.org/dc/terms/" xmlns:vial="http://viaf.org/ontology/1.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:gn="http://www.geonames.org/ontology#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2008/05/skos#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:bio="http://purl.org/vocab/bio/0.1/"
  xmlns:eac="urn:isbn:1-931666-33-4">
  <eac-cpf:person rdf:about="http://socialarchive.iath.virginia.edu/xtf/view?docId=franklin-benjamin-1706-1790-cr.xml#entity">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Franklin, Benjamin, 1706-1790
    .. </rdfs:label>
    <dc:date/>
    <eac-cpf:control rdf:parseType="Resource">
      <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Franklin, Benjamin, 1706-1790
      .. </rdfs:label>
    </eac-cpf:control>
    <eac-cpf:recordID>DLC.ms003039.r033</eac-cpf:recordID>
    <eac-cpf:otherRecordID>DLC.ms003048.c</eac-cpf:otherRecordID>
    <eac-cpf:otherRecordID>DLC.ms008139.r005</eac-cpf:otherRecordID>
    <eac-cpf:otherRecordID>DLC.ms006025.r008</eac-cpf:otherRecordID>
    <eac-cpf:otherRecordID>DLC.ms006027.r007</eac-cpf:otherRecordID>
    <eac-cpf:otherRecordID>DLC.ms007105.r016</eac-cpf:otherRecordID>
  </eac-cpf:person>

```

10

SNAC SPARQLer

- [Social Networks and Archival Context Project \(SNAC\)](#) project website
- [Raw Data](#) on google code
- [RDF Snippets](#) github:gist

RDF data made available by the [Social Networks and Archival Context Project \(SNAC\)](#) under [Open Data Commons Attribution License](#).

SPARQL Query forms

eac.rdf @ <http://socialarchive.iath.virginia.edu/sparql/eac>

General SPARQL query : input query, set any options and press "Get Results"

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

B. Tingle: http://socialarchive.iath.virginia.edu/NAAC_meeting1_slides.html



SNAC: The Social Networks and Archival Context Project

Find Corporate, Personal, and Family Archival Context Records

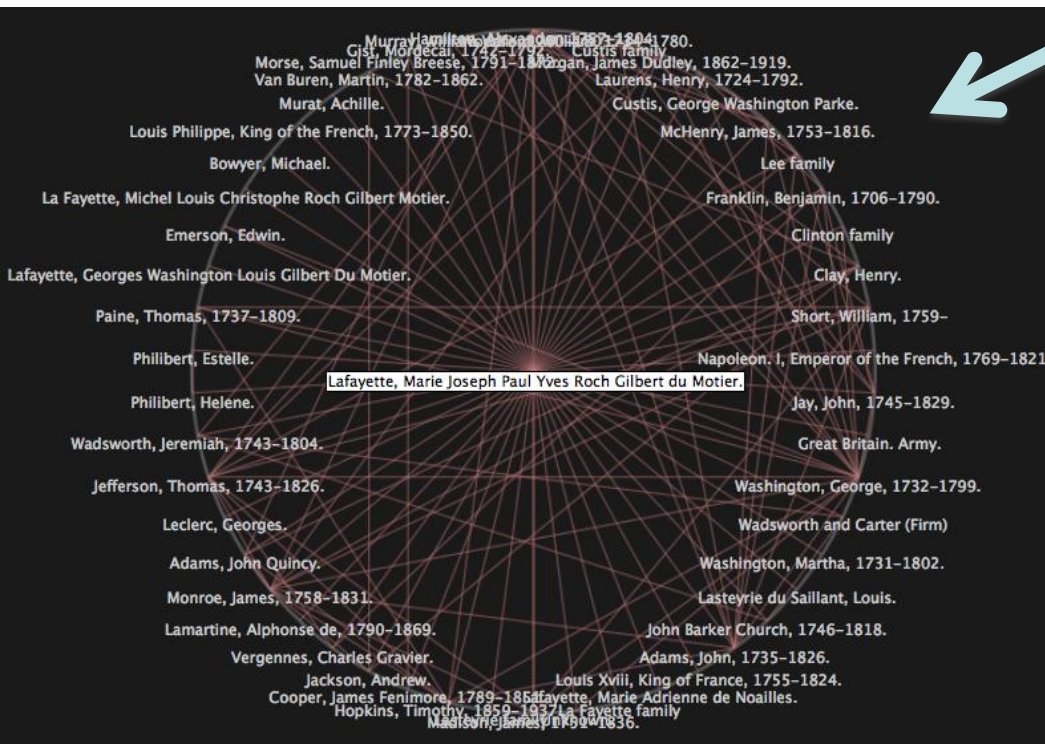
radial graph demo

note data issue ↗
view source EAC-CPF

Lafayette, Marie Joseph Paul Yves Roch Gilbert du Motier - DLC

(1757 – 1834) Male France French

→ Alternative forms of name



Related Entries

▶ Archival Collections (21)

▶ People (61)

▶ Corporate Bodies (5)

▶ Resources (102)

▼ Linked Data (1)

sameAs

<http://viaf.org/viaf/22144104>

feedback

Prosopography

<EAC-CPF>

and Linked Data

(from a post by Ethan Gruber (American Numismatic Society) recently on the EAD list)



Linked Ancient World Data Institute (LAWDI) May 31- June 2, 2012

http://wiki.digitalclassicist.org/Linked_Ancient_World_Data_Institute

hosted by New York University's Institute for the Study of the Ancient World

EG: "... growing interest in ancient prosopography among classicists. Several participants are now working together to build a network of Roman emperors and their relations. ..."

"Having straddled both the museum and library/archival worlds, it makes sense to me to use **EAC-CPF** for this endeavor."

XForms-based information management system for EAC-CPF records called xEAC (pronounced "zeke") – ongoing, beta version



<http://wiki.numismatics.org/xeac:xeac>

<https://github.com/ewg118/xEAC>

- Create/edit/validate EAC Records
- Integration with **Geonames**, **Nomisma**, and **Pleiades** for modern and ancient place names

<http://wiki.numismatics.org/xeac:placeentry>

- Import names and associated authorities from **VIAF**

<http://wiki.numismatics.org/xeac:viaf-import>)

- Easily maintain links between records and create record stubs

<http://wiki.numismatics.org/xeac:relations>

- Generic installation instructions

http://wiki.numismatics.org/xeac:generic_installation



chronologies

example

13

<http://admin.numismatics.org/xeac/record/augustus>

... all is about links

complete the "picture" – interrelate - rebuild the history

quote from Ethan Gruber's post:

"From a numismatic standpoint, I plan to:

- ↳ link emperors together with their families,
- ↳ link emperors to provincial governors,
- ↳ link these people to mints (corporate entities), and
- ↳ link mints to anonymous die carvers which are identifiable by their artistic style.

Each of these entities can be

- ↳ linked to affiliated coins to start, but we can eventually
- ↳ associate sculpture, epigraphy, and literary sources to them."

Archives

not a self-contained world

- Proliferating projects with archival datasets
- Applying Semantic Web and Linked Open Data technologies

examples:

- Locah, Agrippa, Trove, Australian War memorial, HADOC

Show that:

- A lot can be done to inter-relate the existing archival data with other data from other cultural heritage domains
- By using the existing tools
- By remodeling the existing datasets

Interconnecting French cultural heritage treasures on the Web



Digital documents
gallica (DC)
BIBLIOTHÈQUE
NUMÉRIQUE

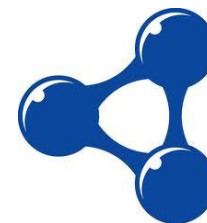
Other BnF
resources

External
resources



Web pages for
Internet users

Raw data for machines



data.bnf.fr

*Modeling
Matching
Clustering
Alignments
Semantic Web
technologies*



LICENCE OUVERTE
OPEN LICENCE

heterogeneous resources

**BnF Archives and
Manuscripts
catalogue
(EAD)**



BnF Main catalogue
{**BnF** (MARC)
*catalogue
général*

Modeling: an FRBR inspired ontology a choice : perspective appropriate to library materials



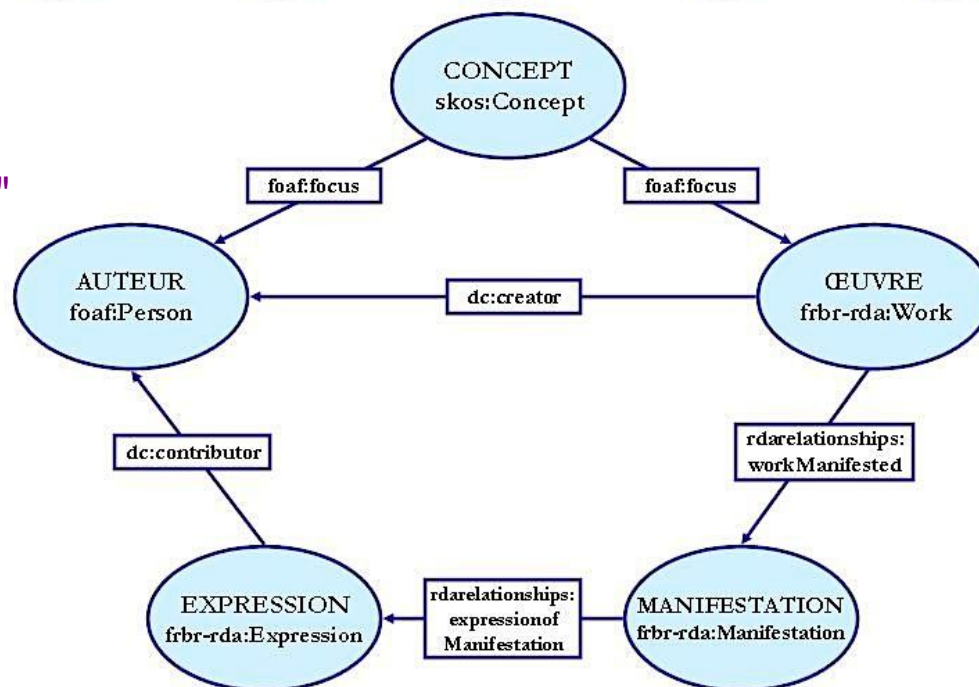
data.bnf.fr

Sources de
données BnF



Pages for "authors"

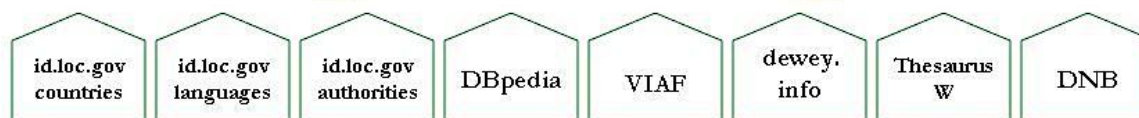
collocate/reorganize
heterogeneous
resources
by type of
contribution



Pages for "works"

collocate/reorganize
heterogeneous
resources
by "FRBR" level

Alignements
externes



for archival and
data.bnf.fr manuscript
descriptions in EAD



“patch together a solution”

Key issues

- Expose only the relevant elements to create meaningful clusters of data (FRBR: W-E-M-I) → “re-contextualize”
-
- EAD Finding Aid perspective – difficult to explore as such
 - Different FA (MS, performing arts, literary archives, etc.) → different EAD implementations
 - Difficult to devise common rules



Key issues “patch together a solution”

- Id perspective
 - anchors: call numbers of units
 - URIs of the authority records (ARK identifiers) occurring in the FA
 - URI of the FA
- For "authors":
 - trace authority URIs occurring in the FA
 - crawl upwards the hierarchy of the XML tree
 - inference rules to harvest the appropriate information at the higher levels
 - combine information and re-create inheritance

data.bnf.fr

rechercher dans data.bnf.fr

Émile Zola (1840-1902)

Pays : France
Langue : français
Sexe : masculin
Naissance : 02-04-1840, Paris
Mort : 29-09-1902, Paris
Biographie : Romancier, auteur dramatique, critique artistique et littéraire et journaliste
Domaines : Littératures, Histoire de la France

+ Plus de détails

Cœuvres les plus étudiées

Ses activités (780 documents)

Autres pages dans data.bnf.fr (11 pages)

Autres ressources

Cœuvres les plus étudiées

L'argent (1891)
Ce roman qui est le dix-huitième volume des "Rougon-Macquart", pénètre les mécanismes de la Bourse, s'inspirant de l'aventure du catholique Bontoux, de sa lutte avec les Rothschild et du krach de sa banque, l'Union générale

L'assommoir (1877)
Roman, septième de la série "Les Rougon-Macquart" d'abord paru en feuilleton dans le "Gil Blas" (avril)

Outils

> Voir les 421 documents au sujet de cet auteur

Imprimer la page
Exporter en PDF
Signaler un problème sur la page

Ressources BnF

Recherche dans :

Gallica
Catalogue général
BnF archives et manuscrits
CNLJ - La Joie par les livres

Sites extérieurs

Recherche dans :

data.bnf.fr

Author page

"build the story"

Voir tous les documents (780) Voir les documents numérisés (45)

Auteur du texte (609)
Auteur adapté (105)
Préfacier (25)
Auteur de lettres (6)
Contributeur (3)
Auteur de lettres

Auteur de la pièce jointe (2)
Dédicataire (2)
Illustrateur (2)
Adaptateur (1)
Ancien possesseur (1)

Éditeur scientifique
Lithographe (1)
Postfacier (1)

Manuscrits littéraires français du XXe siècle. De "Warnod, André (1885-1960)" à "Zola, Émile (1840-1902)". Correspondance. Correspondance adressée par à Émile Zola Émile Moreau
Cote : Mn 5.194 (1-2)
[catalogue BnF archives et manuscrits]

Manuscrits littéraires français du XXe siècle. De "Warnod, André (1885-1960)" à "Zola, Émile (1840-1902)". Zola, Émile (1840-1902). Correspondance. Correspondance d' à Émile Zola André Antoine
Cote : Ms -104
Destinataire de lettres : André Antoine
[catalogue BnF archives et manuscrits]

Compiled from different levels of EAD hierarchy "heritage context rebuilt"

BnF Archives et manuscrits EAD Finding Aid

Manuscrits littéraires français du XXe siècle
De "Warnod, André (1885-1960)" à "Zola, Émile (1840-1902)"
Zola, Émile (1840-1902)
Correspondance

Correspondance d'Émile Zola à André Antoine
Ms -104 (cote)

Unit call number

Links to BnF authority files

URL of FA

http://archivesetmanuscrits.bnf.fr/ead.html?id=FRBNFEAD000004233

data.bnf.fr in the RDF file

.....

```
- <rdf:Description rdf:about="http://archivesetmanuscripts.bnf.fr/ead.html?id=FRBNFEAD000004233">
```

```
  <rdf:type rdf:resource="http://rdvocab.info/uri/schema/FRBRentitiesRDA/Manifestation"/>
```

```
  <bnf-onto:cote>Ms -104</bnf-onto:cote>
```

```
  <bnf-onto:cote>Mn 37</bnf-onto:cote>
```

```
  <bnf-onto:cote>Mn 5/194 (1-2)</bnf-onto:cote>
```

```
  <dc:date>1878</dc:date>
```

```
  <dc:date>1890</dc:date>
```

```
  <rdfs:seeAlso rdf:resource="http://archivesetmanuscripts.bnf.fr/ead.html?id=FRBNFEAD000004233"/>
```

```
- <dc:title>
```

Manuscripts littéraires français du XX sièclee. De "Warnod, André (1885-1960)" à "Zola, Émile (1840-1902)". Zola, Émile (1840-1902). Correspondance. Correspondance d' à Émile Zola André **Antoine**

```
</dc:title>
```

.....

```
- <rdf:Description rdf:about="http://archivesetmanuscripts.bnf.fr/ead.html?id=FRBNFEAD000004233#frbr:Expression">
```

```
  <bnfroles:r110 rdf:resource="http://data.bnf.fr/ark:/12148/cb119297200#foaf:Person"/>
```

```
  <rdf:type rdf:resource="http://rdvocab.info/uri/schema/FRBRentitiesRDA/Expression"/>
```

```
  <rdarerelationships:expressionofManifestation rdf:resource="http://archivesetmanuscripts.bnf.fr/ead.html?id=FRBNFEAD000004233"/>
```

```
  <dc:type rdf:resource="http://purl.org/dc/dcmitype/Text"/>
```

```
  <marcrel:aut rdf:resource="http://data.bnf.fr/ark:/12148/cb119297200#foaf:Person"/>
```

```
  <dc:contributor rdf:resource="http://data.bnf.fr/ark:/12148/cb12276039q#foaf:Person"/>
```

```
  <dc:contributor rdf:resource="http://data.bnf.fr/ark:/12148/cb119297200#foaf:Person"/>
```

URI of FA

Unit call number

Identification of
level in data.bnf.fr

Relator code: author of letters

URI of Authority record for Zola



Subject headings as Linked Open data

Renaissance



Titre original : Renaissance

Vedettes : Noms communs (et noms de peuples, de batailles, etc.)
Noms géographiques

Origine : RAMEAU

Notes : Voir aussi la subdivision chronologique affranchie Renaissance

Voir aussi la subdivision Contribution au concept de Renaissance aux personnes

Voir aussi la subdivision Influence de la Renaissance aux sujets noms communs appropriés (artistiques, culturels, littéraires et de civilisation en général)

Domaines : **Histoire**

Autres formes : Civilisation de la Renaissance
Civilisation (Renaissance)
Europe -- Civilisation (Renaissance)
Renaissance -- Europe
Renaissance -- Histoire

Sources : Encycl. universalis, 1989, 19, 779

Pages équivalentes

Bibliothèque du Congrès
<http://id.loc.gov/authorities/sh85112806#concept>

Deutsche Nationalbibliothek
<http://d-nb.info/gnd/4049450-0>

Thésaurus W
<http://www.archivesdefrance.culture.gouv.fr/thesaurus/resource/T4-57>

Url catalogue : <http://catalogue.bnf.fr/2418/sh13318531w>



data.bnf.fr

Outils

Imprimer la page
 Exporter en PDF
 Signaler un problème sur la page

Ressources BnF

Recherche dans :

Gallica
Catalogue général
BnF archives et manuscrits
CHILJ - La Joie par les livres

Sites extérieurs

Recherche dans :

Catalogue collectif de France
Europeana
OCLC WorldCat
Sudoc

Ce thème dans :
Wikipedia

Thésaurus for indexing local archives

Alignments with other vocabularies in Linked Open Data

Standards, vocabularies, ontologies, ... a conglomeration of tools

- Technological standards
 - Web standards, e.g. RDF, OWL, SKOS, SPARQL
- Content conveying, broadly used vocabularies
 - Developed by
 - cross domain professional initiatives, driven by metadata interoperability aspiration
 - e.g. [Dublin Core \(DC Terms\)](#)
 - special interest professional initiatives
 - e.g. [Geonames](#), [Bio](#), [RDA Vocabularies](#)
 - movements for creation of social networks on the web (“non-professionals”)
 - e.g. [FOAF](#)

What about the specific archival concepts?

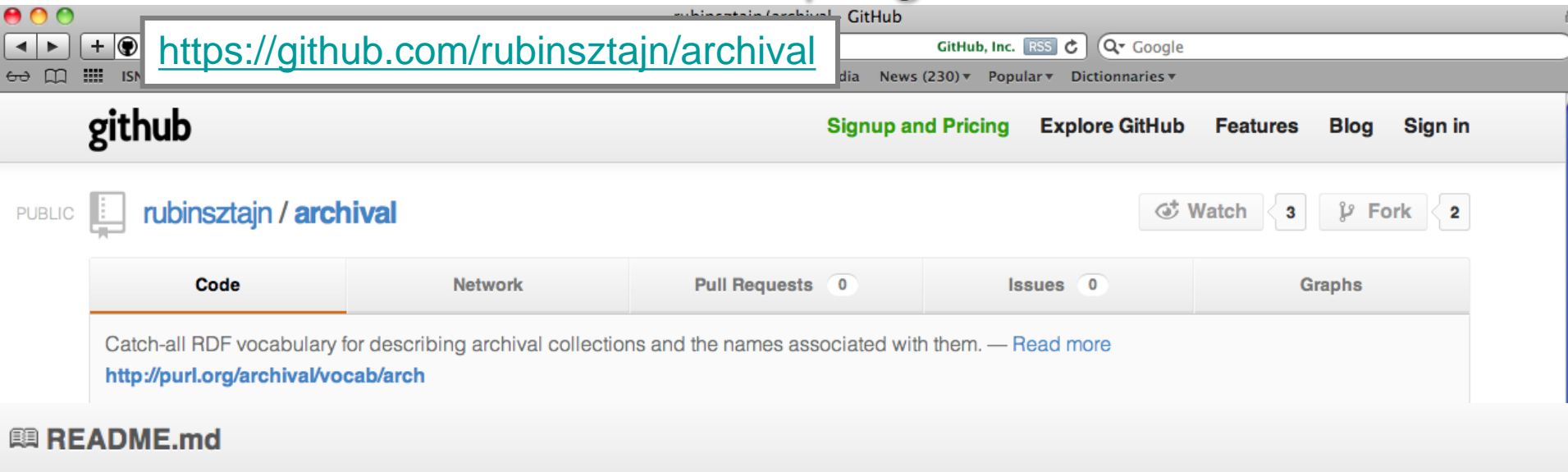
- Are the existing tools/vocabularies/ontologies enough
 - to correctly bring the archival datasets in the LOD cloud?
 - to interconnect resources appropriately, and not distort “the story”?

In the experiences with LOD

- Each of the projects has created its own data model (ontology)
 - Local ([LOCAH](#), [data.bnf.fr](#))
 - International cross-domain ([EDM- Europeana Data Model](#))
- Other “bottom-up” experiments
 - [Archival ontology](#)

Archival Ontology

a work in progress



Archival vocabulary

An OWL/RDF vocabulary for describing archival collections and the names associated with them.

This vocabulary is a work in progress and is currently being used as a catch-all for creating classes and properties for concepts and relationships that don't exist in other vocabularies.

The vocabulary URI is <http://purl.org/archival/vocab/arch#>

Contents

- arch
- Classes (8)
- Object Properties (15)
- Annotation Properties (4)
- Datatypes (2)

OWL HTML inside

Entities (31)

- Thing
- Agent
- appearsWith
- Archive
- bulkEnd
- bulkStart
- Collection
- comment

- correspondedWith
- created
- creator
- description
- Family
- Group
- hasProvenance
- heldBy
- holds
- inclusiveEnd
- inclusiveStart

- knows
- Manuscript
- Nothing
- Organization
- Person
- PlainLiteral
- primaryProvenanceOf
- referencedIn
- relatedSubject
- string
- subject
- title

arch

Annotations (3)

- creator "Aaron Rubinstein" ()
- description "An RDF vocabulary for describing archival collections and the names associated with them" ()
- title "Archival" ()

References

- Classes (8)
- Object Properties (15)
- Annotation Properties (4)
- Datatypes (2)

OWL HTML inside

arch: classes (10)

- Thing
- Agent
- Archive
- Collection
- Family
- Group
- Manuscript
- Nothing
- Organization
- Person

An archival high level ontology, a conceptual model?

Why?

- to address correctly the key concepts of the domain, the interrelationships among objects
- to help in building good systems
- to give to others the key of understanding archival data
- to enable them to use the data in appropriate ways, without distortion
- The revision of EAD is raising such issues

By whom?

Common agreement among the archivists that the best professional forum is



ICA/CBPS is now taking up the challenge of building an archival conceptual model

Two preliminary reports are published

- ***Progress report for revising and harmonising ICA descriptive standards***
- ***Relationship in archival descriptive systems***

<http://www.ica.org/13155/standards/cbps-progress-report-for-revising-and-harmonising-ica-descriptive-standards.html>

You know your world better than anyone?

Model it

and open your data to the world