PIAAF

Pilot for Interoperable Archival Authorities in France

From Several Datasets to One Graph

Anila Angjeli - Bibliothèque nationale de France
Florence Clavaud - Archives nationales, France
Reminder from last year’s presentation
Who? What? Objectives

- **Participants**
  - Archives nationales, France
  - Bibliothèque nationale de France
  - Service interministériel des Archives de France (SIAF)

- **Objectives**
  
  demonstrate usefulness of interconnecting the archival resources from different data providers by using the semantic web technologies:
  
  - **represent in RDF** all datasets;
  - **interconnect** and **enrich** the resulting datasets, especially by expressing relationships that exist among the described objects;
  - **visualize** the datasets in a useful way, **graphically**, by enabling the user to navigate in the graph and access the source metadata.

- **Qualitative requirements; Rigorous methodological approach**
  
  Expressiveness | Possibility of reasoning (inference) | User-friendly interface
Datasets, Ontology, Prototype requirements

- **Dataset for the project**
  - a small selection of EAC-CPF records, plus some FA from each of the partners
  - relevant, controllable (amount, characteristics…), possible to interconnect with the data provided by the others
  - Selection from two functional domains
    - Management of historical monuments and “civil buildings”
    - Public reading

- **Ontology for the project**
  - Records in Contexts ontology (agreement with EGAD)

- **Requirements for the prototype**
  - Web application in compliance with standards and best practices
  - Input data analysis
  - Open source
  - Architecture based on a RDF triplestore
  - Functionalities: Alignments, SPARQL queries, Construction of new triples, Analytical cartographies (Graph navigation, Timeline…)
Datasets
rich <eac-cpf> descriptions and <ead> FA

- BnF
  - 55 <eac-cpf> records for corporate bodies
  - 12 <eac-cpf> records for persons
  - 6 <ead> FA

- SIAF
  - 14 <eac-cpf> records describing categories of corporate bodies
  - 92 <eac-cpf> records for corporate bodies
  - 2 <eac-cpf> records for persons
  - 2 <ead> FA

- Archives nationales
  - 90 <eac-cpf> records for corporate bodies
  - 10 <eac-cpf> records for persons
  - 29 <ead> FA

A total of:
275 <eac-cpf> rich records
37 <ead> FA
PIAAF on the move
PIAAF - Timeline

• Fall 2016 – Spring 2017
  ✓ Data preparation
  ✓ Selection of contractor -
  ✓ Data delivery

• Early summer 2017
  ✓ Start of developments for
    PIAAF Web application
    (user interface)

• Fall 2017
  o PIAAF prototype to go live

develops the Archives Portal France
Source data curation, harmonization and enrichment
Curation, harmonization, enrichment

- Each of the partners preserves its own perspective
- However, agreement on applying common encoding rules:
  - **Names** in `<nameEntry>`
    - same rules applied (authorized forms)
  - **Identity**
    - ISNI identifiers added in `<entityId>`
  - **Relations**
    - Same encoding rules and enrichment policy applied
    - Expressiveness of `@cpfRelationType` enhanced beyond “hierarchical” and “associative” with a variety of `@xlink:arcrole` values: `isControlledBy`, `isDirectedBy/isDirectorOf`, `hasPart/isPartOf`, `isFunctionallyLinkedTo`, etc.
    - enrichment with “identity” links to descriptions for the same entities in other datasets
  - **Functions** in `<function>`
    - Same encoding rules and enrichment policy applied
## Curation, harmonization, enrichment

- **“entity identifiers”**
  - ISNI, whenever available, added in `<entityId>`

From the **Archives Nationales authority record** for Jean Favier

```xml
</identity>
  <entityId>ISNI 0000 0001 2099 1284</entityId>
  <entityType>person</entityType>
  <nameEntry localType="preferredFormForProject" scriptCode="Latn" xml:lang="fre">
    <part>Favier, Jean (1932-2014)</part>
  </nameEntry>
</identity>
```

From the **BnF archival authority record** for Jean Favier

```xml
<identity>
  <entityId>ISNI 0000 0001 2099 1284</entityId>
  <entityType>person</entityType>
  <nameEntry localType="preferredFormForProject">
    <part>Favier, Jean (1932-2014)</part>
    <authorizedForm>AFNOR_Z44-061</authorizedForm>
  </nameEntry>
</identity>
```
• Refining relations

⇒ Expressiveness of @cpfRelationType enhanced beyond “hierarchical” and “associative” with a variety of @xlink:arcrole values:

isControlledBy/controls, isDirectedBy/isDirectorOf, hasPart/isPartOf, isFunctionallyLinkedTo, etc.

Relation of Jean Favier to the Bibliothèque nationale de France in the AN record

<cpfRelation xlink:href="FRAN_NP_051355" xlink:type="simple" cpfRelationType="associative" xlink:arcrole="http://www.piaaf.net/ontology/piaaf#isDirectorOf">  
<relationEntry>Bibliothèque nationale de France (1994-....)</relationEntry>  
<dateRange>  
  <fromDate standardDate="1994-01-01">1994</fromDate>  
  <toDate standardDate="1997-12-31">1997</toDate>  
</dateRange>  
</cpfRelation>

Relation of Jean Favier to the Bibliothèque nationale de France in the BnF record

<cpfRelation xlink:href="FRBNF_NP_protoLD_DGEARC_0006" xlink:type="simple" cpfRelationType="associative" xlink:arcrole="http://www.piaaf.net/ontology/piaaf#isDirectorOf">  
<relationEntry>Bibliothèque nationale de France</relationEntry>  
<dateRange>  
  <fromDate standardDate="1994">1994</fromDate>  
  <toDate standardDate="1997">1997</toDate>  
</dateRange>  
<descriptiveNote><p>Jean Favier est le premier Président de la Bibliothèque nationale de France de 1994 à 1997.</p></descriptiveNote>  
</cpfRelation>
• Relations: “identity” links

<cpfRelation cpfRelationType="identity"
xlink:href="http://data.bnf.fr/ark:/12148/cb11996420w"
xlink:type="simple">
  <relationEntry>Jean Favier (1932-2014)</relationEntry>
</cpfRelation>

BnF archival authority record for Jean Favier
Link to the Jean Favier description in the BnF LOD dataset from the BnF General catalog
Curation, harmonization, enrichment

• Relations: “identity” links

From the EAC-CPF Archives Nationales authority record for Jean Favier

```xml
  <relationEntry>Jean Favier</relationEntry>
  <dateRange>
    <fromDate standardDate="1932-04-02">2 avril 1932</fromDate>
    <toDate standardDate="2014-08-12">12 août 2014</toDate>
  </dateRange>
  <descriptiveNote><p>Notice Wikipédia</p></descriptiveNote>
</cpfRelation>

<cpfRelation cpfRelationType="identity" xlink:href="http://catalogue.bnf.fr/ark:/12148/cb11996420w" xlink:type="simple">
  <relationEntry>Favier, Jean (1932-2014)</relationEntry>
  <dateRange>
    <fromDate standardDate="1932-04-02">2 avril 1932</fromDate>
    <toDate standardDate="2014-08-12">12 août 2014</toDate>
  </dateRange>
  <descriptiveNote><p>Notice d'autorité BnF</p></descriptiveNote>
</cpfRelation>
```
Curation of Functions

• Creation (from scratch) of two separate controlled vocabularies for Functions terms, covering:
  • **Functional domains** (areas of action) of archives’ creators
  • **Activities** performed by archives’ creators

• Using an open source tool for thesauri management
  • Developed by the Ministry of Culture and made available to all cultural institutions
  • ISO 25964 compliant
  • Exportable in SKOS
  • Enables alignments to other vocabularies
Curation of Functions

• Each of the partners indexed its EAC-CPF source instances combining terms from these vocabularies
  - Pattern applied: “functional domain” – “activity”
  - EAC-CPF <function> element was slightly modified for the project:
    - <term> element was made repeatable
    - attributes were added
  - Use of URIs of concepts from the Function vocabularies as generated by GINCO

Example for indexing corporate bodies with the function :
“preservation of cultural objects”
(in French « objets culturels -- conservation")

```xml
<function localType="piaafFunction">
  <term vocabularySource="http://data.culture.fr/thesaurus/resource/ark:/67717/51232822-adac-4a33-aa14-29e2c701a5ee" piaaf:rank="2" piaaf:conceptId="http://data.culture.fr/thesaurus/resource/ark:/67717/b06a81ff-bbe0-4bf3-b33e-e51d26cb0431">conservation</term>
</function>
```
Functions in more detail

From the EAC-CPF instance

```xml
<function localType="piaafFunction">
  <term vocabularySource="http://data.culture.fr/thesaurus/resource/ark:/67717/f14e8183-5885-46d6-8fc9-17ebd83c27e" piaaf:rank="1"
    piaaf:conceptId="http://data.culture.fr/thesaurus/resource/ark:/67717/4002e166-eb11-446c-a777-cb73b6def641">objets cultures</term>

  <term vocabularySource="http://data.culture.fr/thesaurus/resource/ark:/67717/51232822-adac-4a33-aa14-29e2c701a5ee" piaaf:rank="2"
    piaaf:conceptId="http://data.culture.fr/thesaurus/resource/ark:/67717/b06a81ff-bbe0-4bf3-b33c-e51d26eb0431">conservation</term>
</function>
```

concept in SKOS

```xml
<skos:Concept rdf:about="http://data.culture.fr/thesaurus/resource/ark:/67717/b06a81ff-bbe0-4bf3-b33c-e51d26eb0431">
  <skos:prefLabel>
    <iso-thes:PreferredTerm rdf:about="http://data.culture.fr/thesaurus/resource/ark:/67717/9ec6b27a-da4e-4b9e-95a6-f5b3e48db0b2"/>
  </skos:prefLabel>
  <skos:prefLabel xml:lang="fr">conservation</skos:prefLabel>
</skos:Concept>
```
Data processing

**RiC-Ontology** (in its current state of development)
Transformation into RDF
Delivery of files to the contractor
Preamble

One of the objectives of PIAAF

• Test the feasibility of expressing archival data by applying the concepts of Records in Contexts as formalized by RiC Ontology

• RiC-O is under development by EGAD
  • Team leader Florence Clavaud from the Archives nationales France

• RiC-O development – currently sufficiently advanced, especially for classes and properties PIAAF focuses on

• In agreement with EGAD, PIAAF would apply RiC-O and serve as testbed for RiC-O solutions

Note however that RiC-O will continue to evolve. Even the names of classes and properties that PIAAF uses are likely to change.
Definition of classes for PIAAF
(for the described “objects”)

• Main RiC-O classes used
  (note that these represent only part of RiC-O classes)

  RiC:CorporateBody
  RiC:Person
  RiC:FunctionAbstract
  **RiC:Position**  to express the position occupied by a Person within a Corporate Body
  RiC:Mandate
  RiC:Event
  **RiC:AgentName**  note the distinction between an Agent and its Name
  RiC:Record
  RiC:RecordSet
  **RiC:FindingAid**  for links to the extant Finding Aids in their original environment
  (a FA is viewed as a document in itself (PDF file, )
  RiC:Description
  RiC:Place
  **RiC:Relation** (and its subclasses)  note that Relation is a class in itself.
  This enables qualification of a relationship with properties such as nature of the relationship, date of the relationship, certainty, etc.

  **RiC:GroupType**  to describe in detail the category or the type of groups or corporate bodies. Enables full historical and contextual definitions of such types.
  E.g. define and describe what are “Departemental Archives” in France, the establishment by law/decree of such type of institutions, the mandates they are given, their jurisdictional scope of activity, etc.
RDF transformation of source files

Processing

• A suite of XSLT scripts was written to generate XML/RDF files from the source data
  • They are run in a given order
    • as indicated by the number in the name of each script

1-processFunctions.xsl
2-processLegalStatuses.xsl
3-processMandates.xsl
4-processEvents.xsl
5-processAgentNames.xsl
6-processPlaces.xsl
7-processPositions.xsl
8a-processRelations-between-group-types.xsl
8b-processRelations-between-agents.xsl
8c-processProvenanceRelations.xsl
8d-processNotDescribedRecordsAndRecordSets-fromEACfiles.xsl
9-processSIAFGroupTypes.xsl
10-processAgents.xsl
11-processNotDescribedAgents.xsl
12-processRecordSetsAndRecords.xsl
13-processProxies.xsl
RDF transformation of source files

Output

• XML/RDF files are generated for each and every individual “object”

• Each of the folders contains all the RDF files for the individuals of one class

• Licence:
  • For now all files are in a private repository
  • By the end of the project the source code will be made available under GPL licence
Properties

• Reminder of the objectives defined
  – Qualitative requirements | Expressiveness (high granularity) | Accurate relations between the datasets | Usefulness of interconnection | Graph navigation | Possibility of inference…

• Hence, precise properties
  – To describe characteristics of each “object”
  – To accurately express relationships between them
  – To accurately interconnect “objects” to one another

• RiC-O properties are used (at the current state of development)
Expressiveness of relations with RiC-O

Example of a dated and described whole/part relation (n-ary relation)

Le Département des manuscrits est un département de la Bibliothèque nationale.

RDF code for this relation:

```xml
<rdf:Description rdf:about="http://www.piaaf.net/relations/FRBNF_agent_relation_0094-d20e557">
  <rdfs:label xml:lang="fr">Relation du tout à une partie entre les agents "Bibliothèque nationale de France. Département des manuscrits" et "Bibliothèque nationale (France)"</rdfs:label>
  <Ric:description xml:lang="fr">Le Département des manuscrits est un département de la Bibliothèque nationale. </Ric:description>
  <Ric:beginningDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1792</Ric:beginningDate>
  <Ric:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1994</Ric:endDate>
  <Ric:wholeHasAgentPart rdf:resource="http://www.piaaf.net/agents/FRBNF_corporate-body_0004"/>
</rdf:Description>
```
Expressiveness of relations with RiC-O

Example of a dated whole/part relation (n-ary relation)

1994
RiC:endDate

RiC:AgentWholePartRelation

1792
RiC:beginingDate

RiC:Agent

RiC:groupHasPart

Bibliothèque nationale (France)

RiC:agentPartOf

RiC:wholeHasAgentPart

RiC:AgentWholePartRelation

RiC:agentHasWholePartRelation

Département des manuscrits

RDF code

```xml
<rdf:Description rdf:about="http://www.piaaf.net/relations/FRBNF_agent_relation_0094-d20e557">
  <rdf:type rdf:resource="http://www.ica.org/standards/RiC/ontology#AgentWholePartRelation"/>
  <rdfs:label xml:lang="fr">Relation du tout à une partie entre les agents "Bibliothèque nationale de France. Département des manuscrits" et "Bibliothèque nationale (France)"</rdfs:label>
  <RiC:description xml:lang="fr">Le Département des manuscrits est un département de la Bibliothèque nationale.</RiC:description>
  <RiC:beginningDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1792</RiC:beginningDate>
  <RiC:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#gYear">1994</RiC:endDate>
  <RiC:wholeHasAgentPart rdf:resource="http://www.piaaf.net/agents/FRBNF_corporate-body_0003"/>
  <RiC:agentHasWholePartRelation rdf:resource="http://www.piaaf.net/agents/FRBNF_corporate-body_0094"/>
</rdf:Description>
```
Source files transformed into RDF

example of Agent

Jean Favier

From the BnF record
The experience with data processing by implementing RiC-O in its current state of development is extremely encouraging.

It shows that RiC-O works to express the complexity of archival description!

To learn more (on modeling issues, data analysis and processing, implementation of RiC-O in PIAAF...)

attend the conference: Florence Clavaud gives a detailed presentation on August 9
Instructions for the development of the web application prototype

• Definition of navigation “nodes”
  • Persons
    Directors of institution, Architects
  • Organizations
    BnF and its subdivisions, Archival services under the control of SIAF, National Archives and all entities-creators of archives held by the NA
  • Types of organizations
    The abstract categorization of organizations such as “Departmental archives” (instantiated then in the individual archival services that belong to this category)
  • Positions
    President, Director,
  • Legal statuses
  • Functions
  • Records and Records Sets
    Related to the Agents

Relations between these resources: enable different perspectives on datasets

• Timeline (using temporal information in the source data)
• Alignments between partners and other external datasets
PIAAF Web application

• Currently under development (not yet ready to be shown)

• As mentioned in **Timeline**
  will be open to public early fall 2017

• We will communicate and invite professionals to consult and comment
Thank you!

Anila Angjeli - Bibliothèque nationale de France
Florence Clavaud - Archives nationales, France