

FranceArchives: A portal for French archives

Élodie Thieblin, **Fabien Amarger**, Katia Saurfelt, Mathilde Daugas

FranceArchives

FranceArchives : key informations

Discover 15 centuries of
archives

What can you find on FranceArchives ?

Search

Referenced archives

Site contents

Advanced search

- Online since March 2017
- Managed by the French Ministry of Culture
- For a wide audience (researchers, genealogists, archivists...)
- Various partners: national and local archives, university archives, public archival institutions and private archival institutions

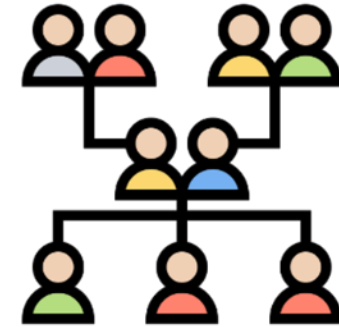
FranceArchives : key informations



137 content providers



82 590 findingaids and
21 357 059
components



8 552 308 names in the
base de noms



XML-EAD, XML-EAC, Pdf, csv
OAI-PMH / manual upload



About 270 000 visitors
per month

Aggregation requires Harmonisation

- **Format** harmonisation:
 - Input: EAD, Dublin Core, particular CSV formats, etc.
 - Mapping them to the Database Schema
- **Authority** harmonisation
 - Input: Documents' Indexes
 - **Authority Management**

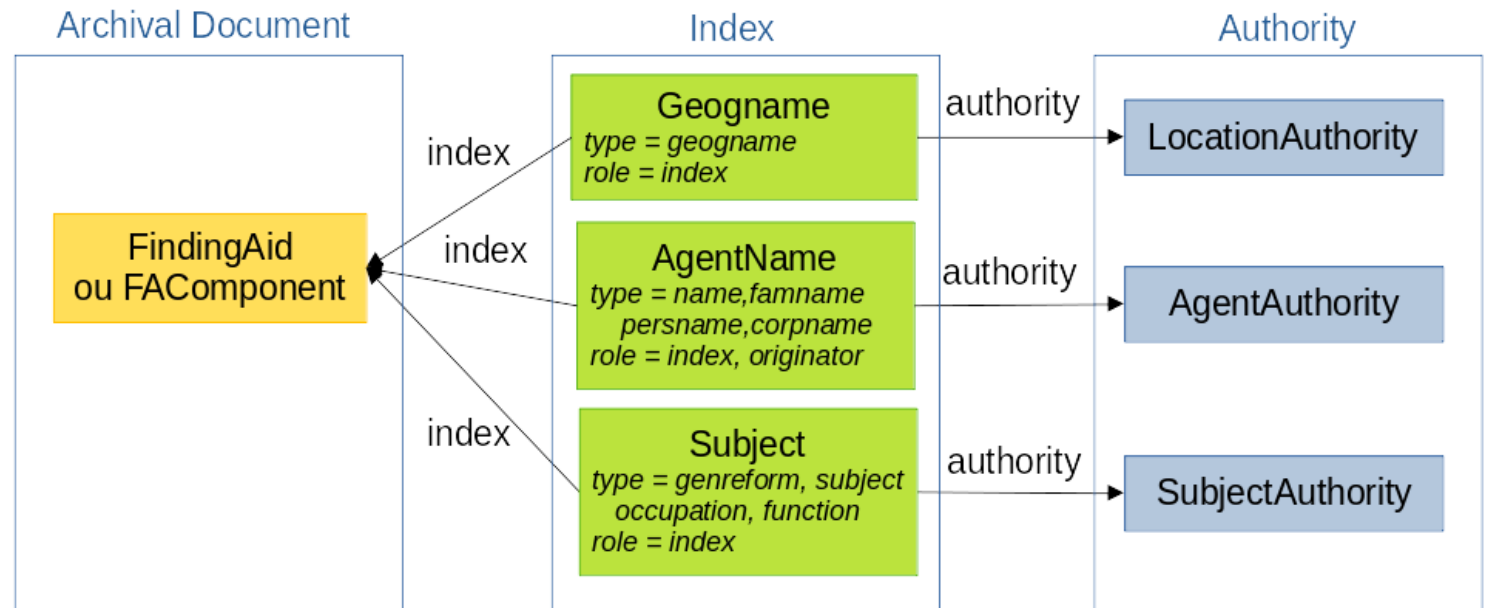
Authorities Management

From Indexes to Authorities

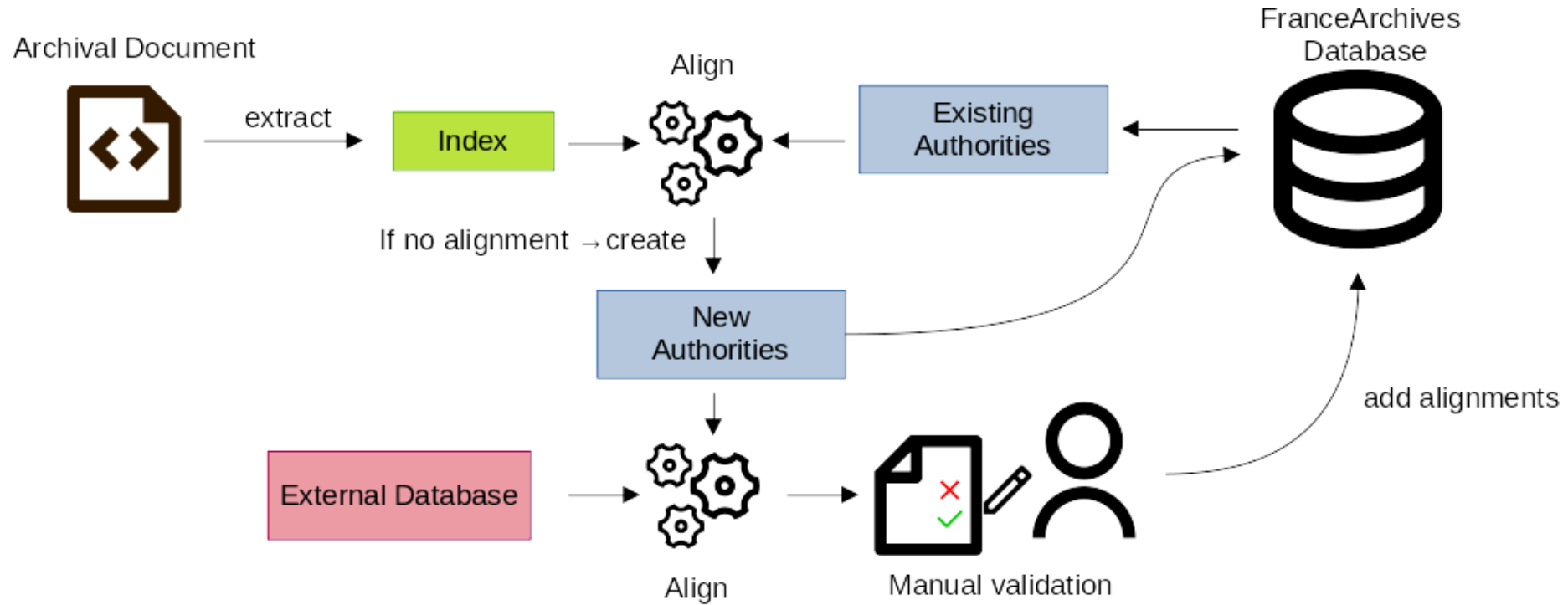
Linking FindingAid indexes to 3 types of authorities:

- **Locations**
- **Agents** (Person, Organisation, ...)
- **Subjects**

Unique identifiers

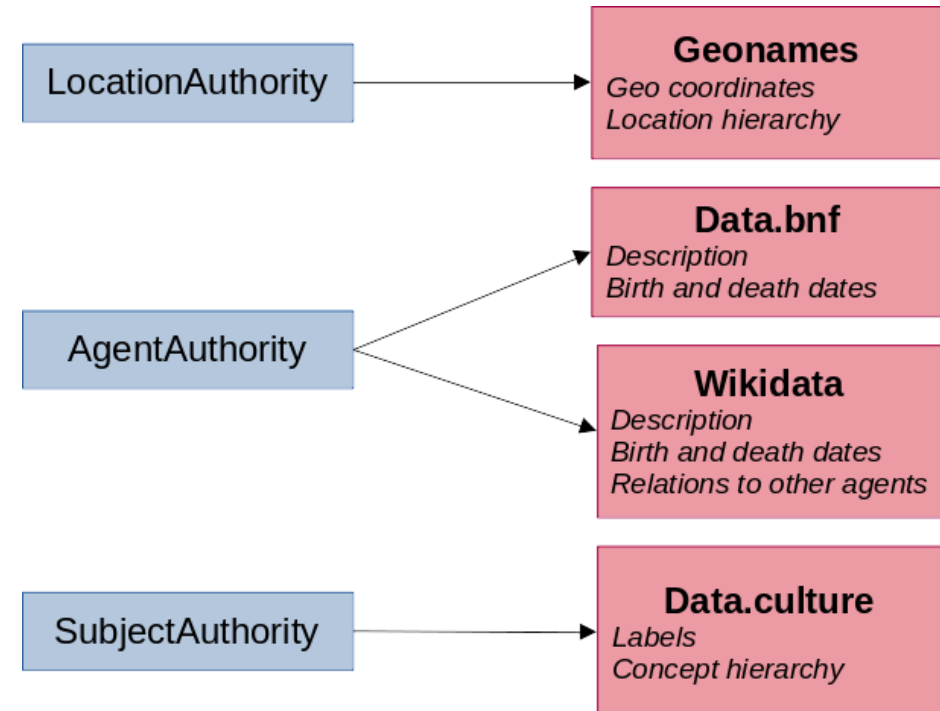


Authority workflow



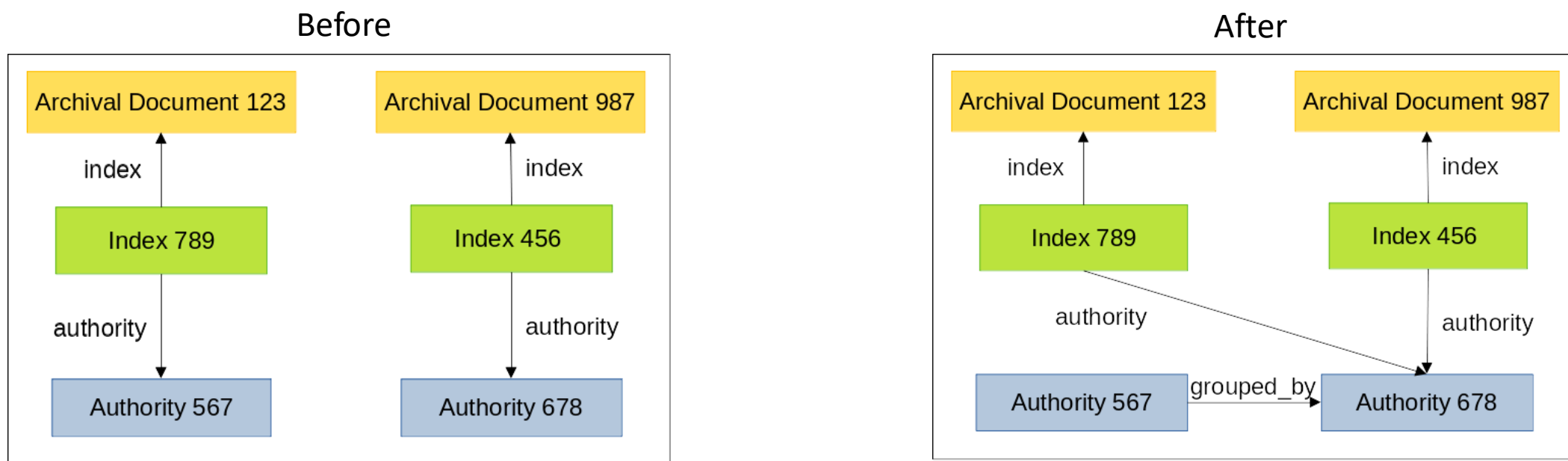
External Reference Mapping

- **Alignment** to external reference databases
- **Information retrieval** from the new sources
- **Periodical update** of the retrieved information



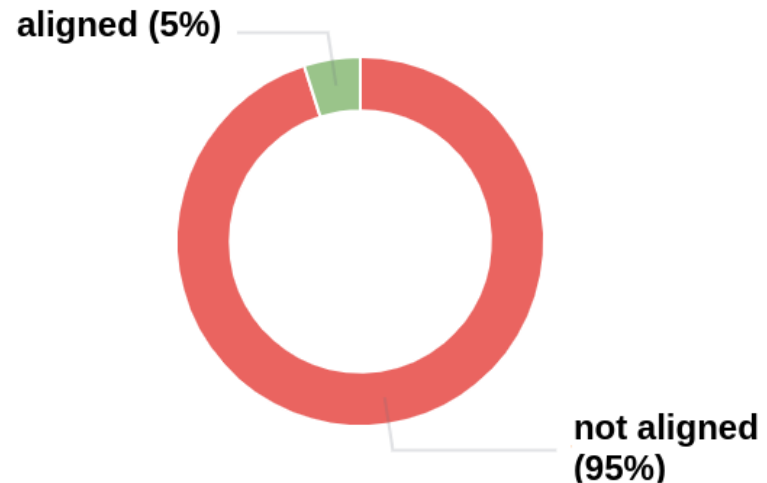
Authorities Grouping

When two authorities represent the same real-world entity: they are **grouped** (manual operation)

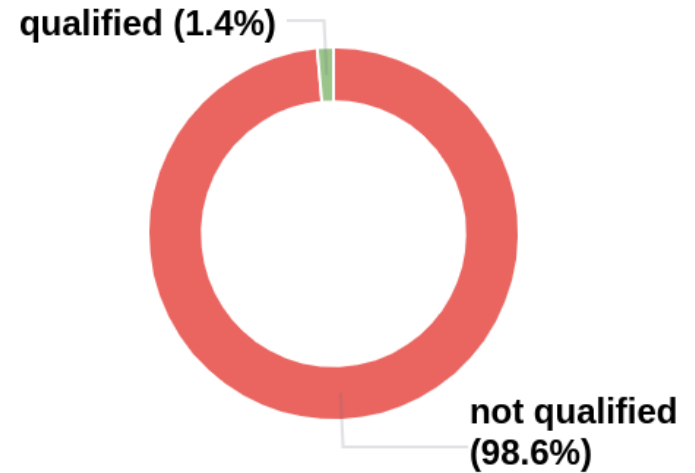


Authorities

- **Agent Authorities:** 3 729 074 (aligned: 29 556 - qualified: 24 541)
- **Location Authorities:** 641 916 (aligned: 189 879 - qualified: 34 426)
- **Subject Authorities:** 198 542 (aligned: 6 403 - qualified: 2 737)



Aligned authorities



Qualified authorities

Exporting the data

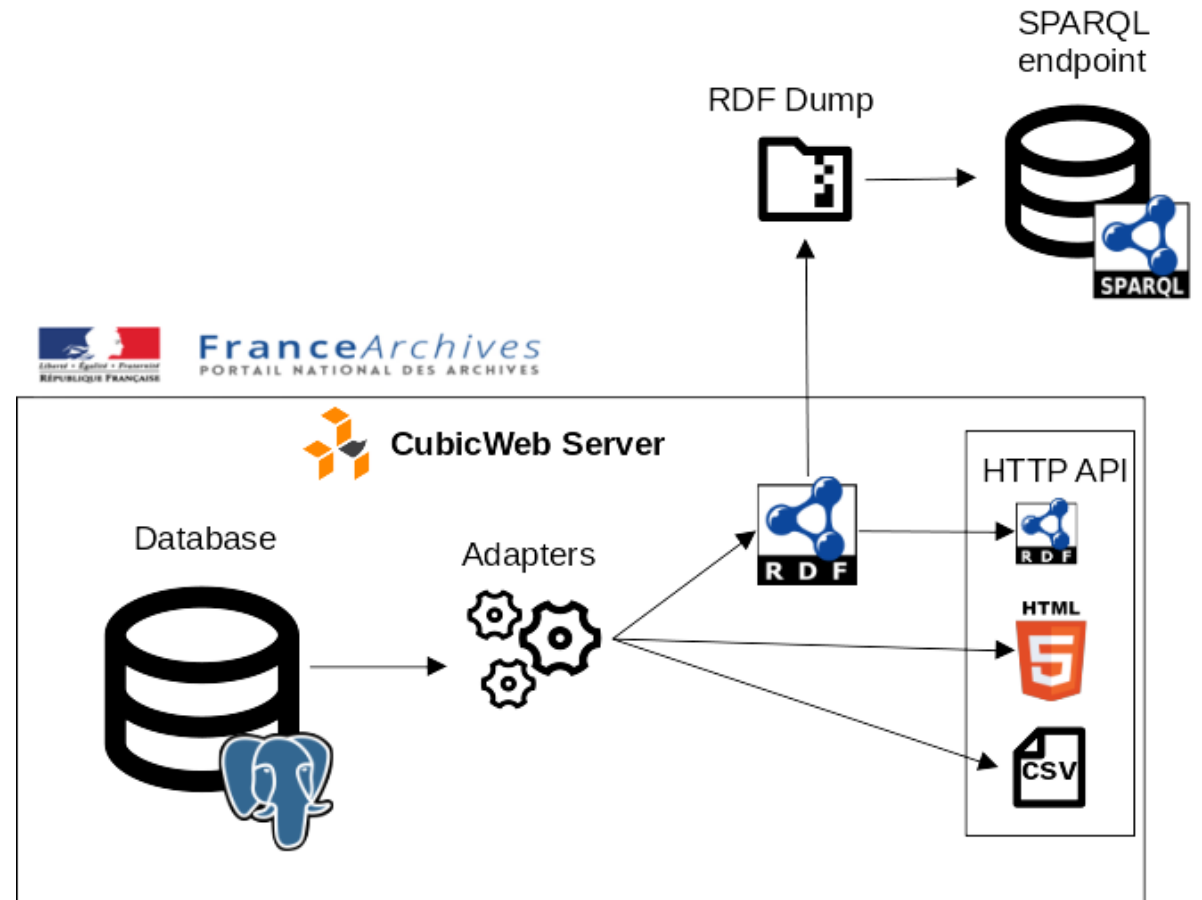
Records In Context - Ontology

- Based on Records In Context – Conceptual Model
- Created and promoted by the [International Council on Archives](#)
- Good coverage of EAD (archives) and EAC-CPF (authorities) standards
- Similar translation performed by the National Archives of France
- Other [projects](#) and tools implement this ontology
- Good support for ontology usage ([documentation](#), direct contact)



RDF Export Workflow

- Adapters to transform database entities
- Content negotiation on-the-fly
- Dump generated every 6 months
- SPARQL endpoint from the dump



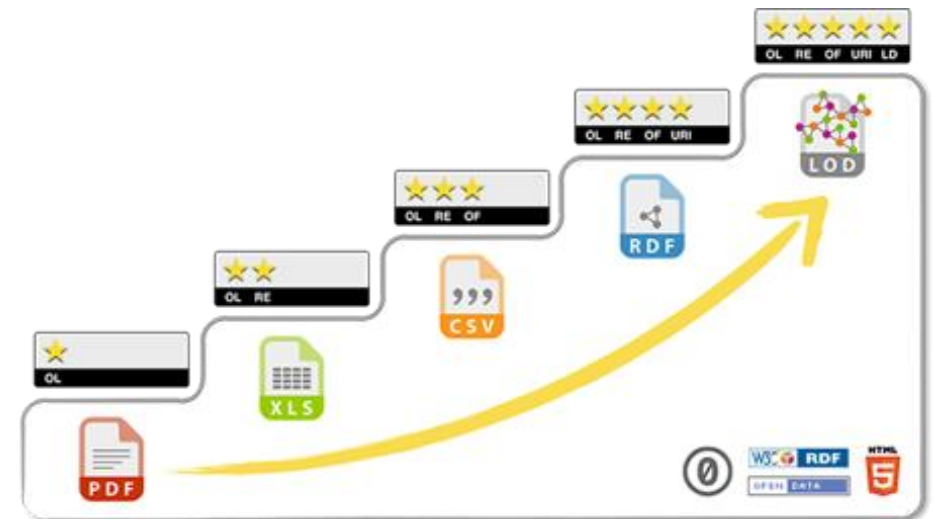
Access to RDF data

- Content negotiation : HTTP Response
- Button from the entity's page to download in the formats: xml, ttl, nt, n3 et jsonld, calculated on the spot.
- [RDF Dump](#) generated every 6 months.
- [SPARQL Endpoint](#) updated every 6 months.
- Classic SPARQL interrogation tool : Yasgui.
- User-friendly SPARQL interrogation tool : [Sparnatural](#).

Conclusion and Perspectives

Conclusion

- Archival documents searchable by all
- Quality indexation
- LOD compliant data
- Large-scale RDF data management



Perspectives

- Exporting authorities links to the external databases:
 - Data.bnf
 - Wikidata
 - Persée
- Extend the retrieved data from external databases to allow more queries
- Extend the RDF translation to other documents (personal records, etc.)
- Optimize the workflow for larger data
- Follow RIC-O's evolutions

Thank you !

elodie.thieblin@logilab.fr

mathilde.daugas@culture.gouv.fr

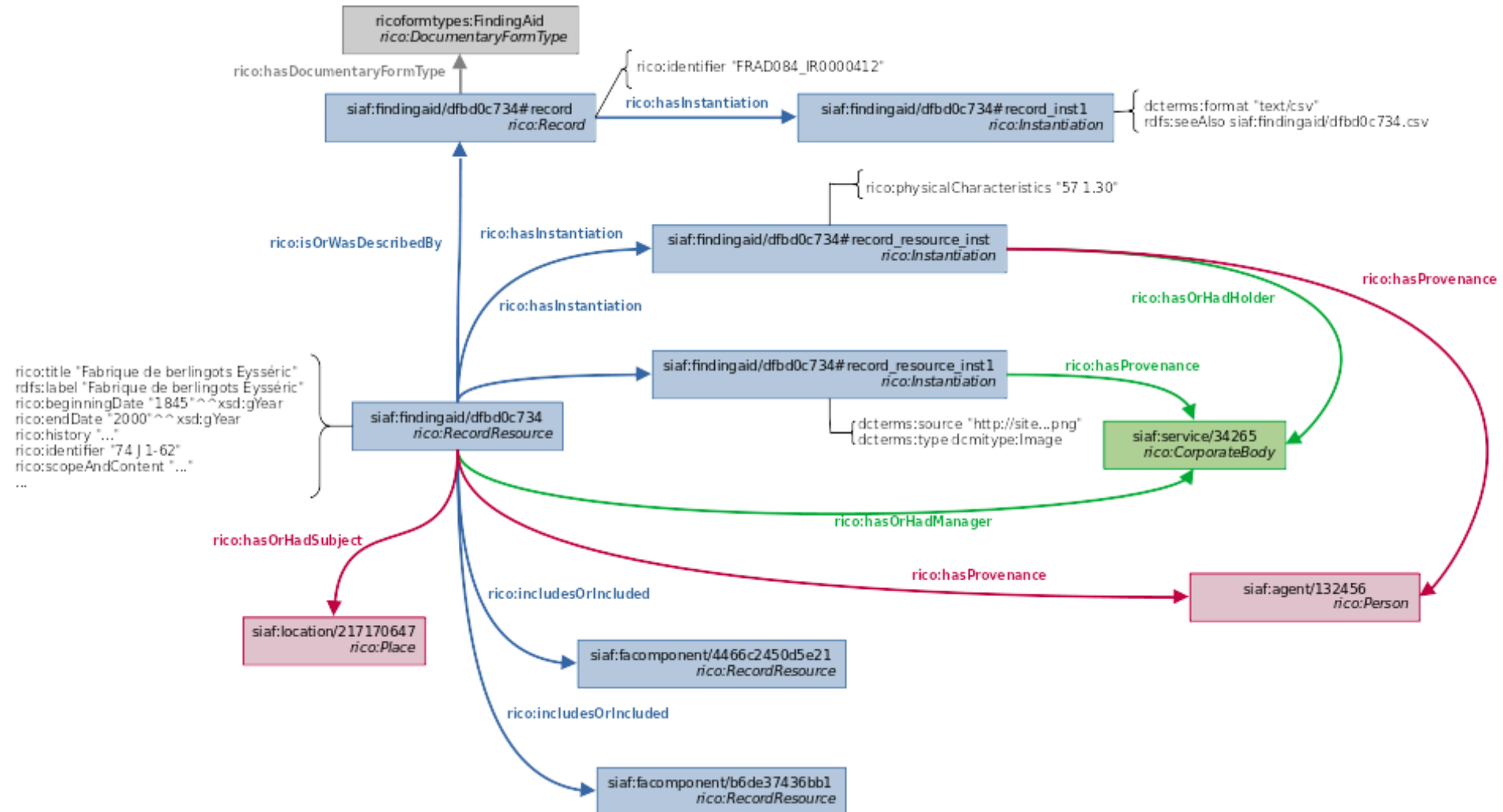
Annex

RIC-O

Other candidate ontologies

- OAD (Ontology for Archival Description)
- Locah
 - Not maintained
- The National Archives vocabulary (TNA) and the Digital Records Infrastructure vocabulary (DRI)
 - Not on the LOV
 - Not extensive enough

RIC-O



A look back at the implementation steps

- 2018- : processing of agent, location and subject authorities ;
- June-november 2020: study to adopt the best process ;
- November 2021: order & realization in 18 months
 - January 2022: qualification of authorities.
 - January-June 2022: conversion in RDF.
 - July-August 2022: installation of Virtuoso and 1st conversion (279 millions of triplets).
 - January 2023: 2nd conversion in RDF (316 millions de triplets).
 - February-May 2023: implementation of Sparnatural.
 - July 2023: persistent storage, process of conversion (about 2.5 days) & ingress on Virtuoso (about 6h) and generation of the RDF archive.
 - May 2023: 3^e conversion (385 millions of triplets):
 - ✓ 274 millions linked to qualified authorities
 - ✓ 109 millions with no link
- Cost (only RDF linked tasks) : about 88 000€ TTC or 10% of the total.

Sparnatural

Locations subject of archival documents linked to the finding aid « Fabrique de berlingot Eysséric » ▾

