

# FranceArchives : portail de référence pour les archives françaises dans le web de données liées

Mathilde Daugas, Manonmani Restif, Élodie Thiéblin, Katia Saurfelt

# FranceArchives

# FranceArchives : informations clés

RECHERCHE DANS LA BASE DE NOMS  
PAGES PROFESSIONNELLES  
RECHERCHE SPARQL  
MENU

Partez à la découverte de  
15 siècles d'archives  
Que trouverez-vous sur FranceArchives ?

Rechercher

Archives référencées Ressources du site Recherche avancée

RECHERCHER

DÉCOUVRIR

COMPRENDRE

FranceArchives mode d'emploi  
Qui peut consulter des archives ?

Pages d'histoire  
Expositions virtuelles

Organisation des Archives en France  
Données d'activité des services d'archives

- En ligne depuis 2017
- Piloté par le ministère de la Culture
- 270 000 visiteurs par mois : généalogistes, chercheurs, archivistes, curieux...
- Archives : état civil, registres, chartes, documents du 7<sup>e</sup> au 21<sup>e</sup> siècle, photographies, cartes et plans, cartes postales, films, témoignages oraux...

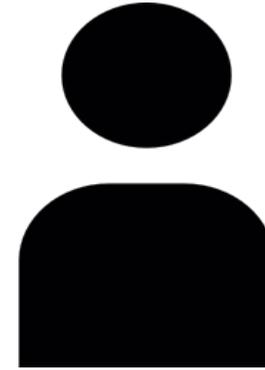
# FranceArchives : chiffres clés



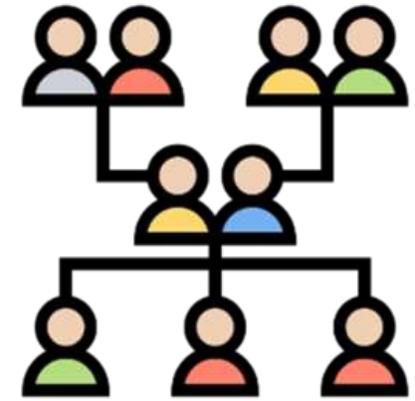
164 partenaires dont 138 contributeurs



83 301 inventaires soit  
21 630 023 unités de descriptions d'archives



16 000 fiches descriptives de producteurs

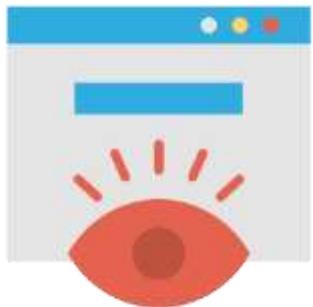


8 783 888 noms dans la base de noms



International Image Interoperability Framework

Environ 400 000 objets numériques en IIIF



XML-EAD, XML-EAC, Pdf, csv  
OAI-PMH / import manuel



Elastic Search



Environ 400 millions de triplets RDF

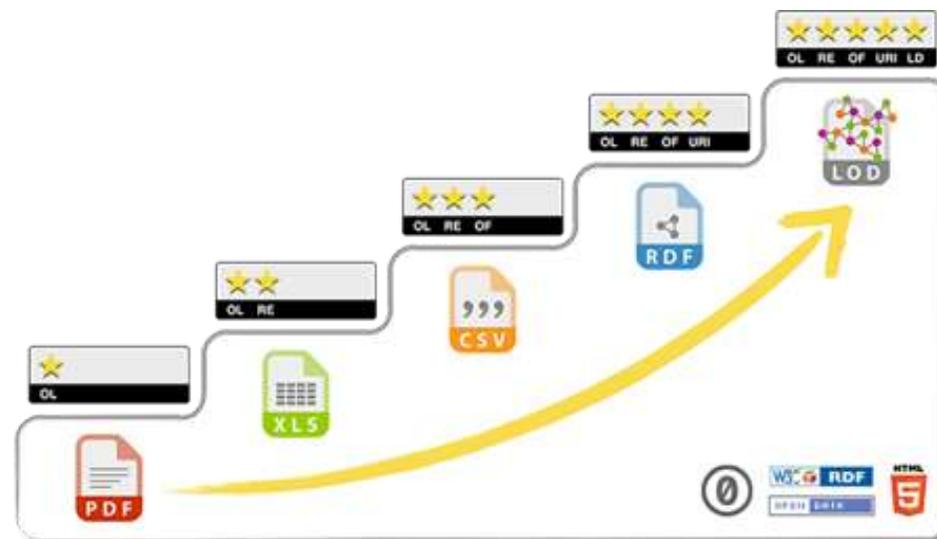
# Agréger pour publier des données ☆ ☆ ☆ ☆ ☆

- **Agrégation ==> Harmonisation**

- Formats : Alignement avec le schéma de FranceArchives
- Autorités : **Gestion des autorités**

- **Publication de données 5 étoiles**

- Données alignées
- **Export RDF** avec ontologie standard



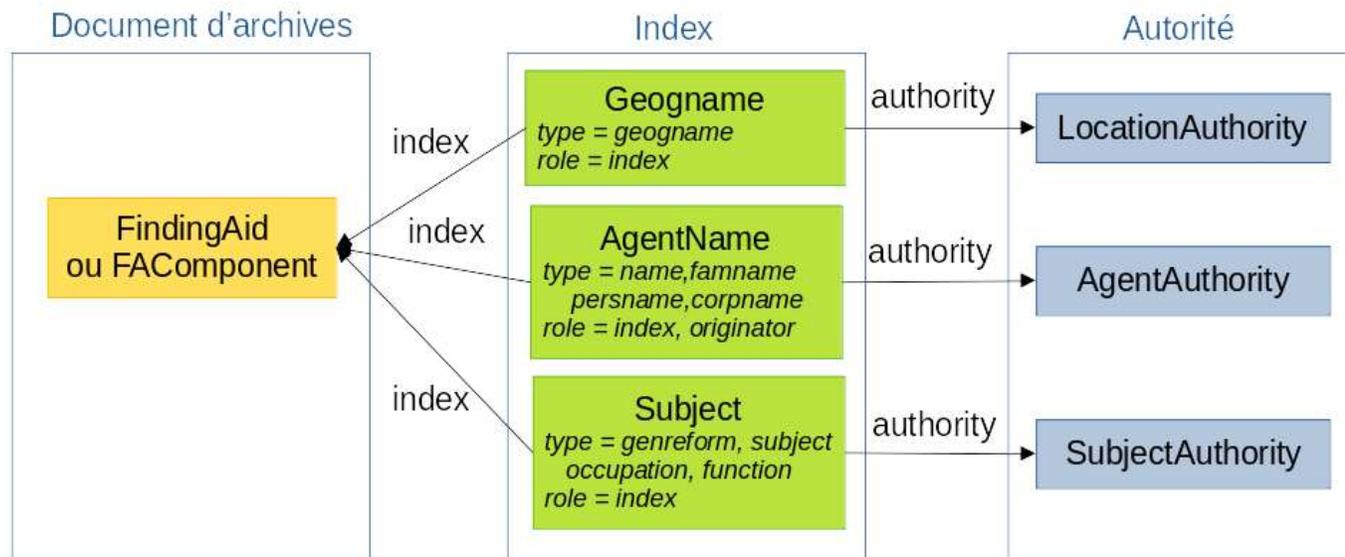
# Gestion des autorités

# Des index aux autorités

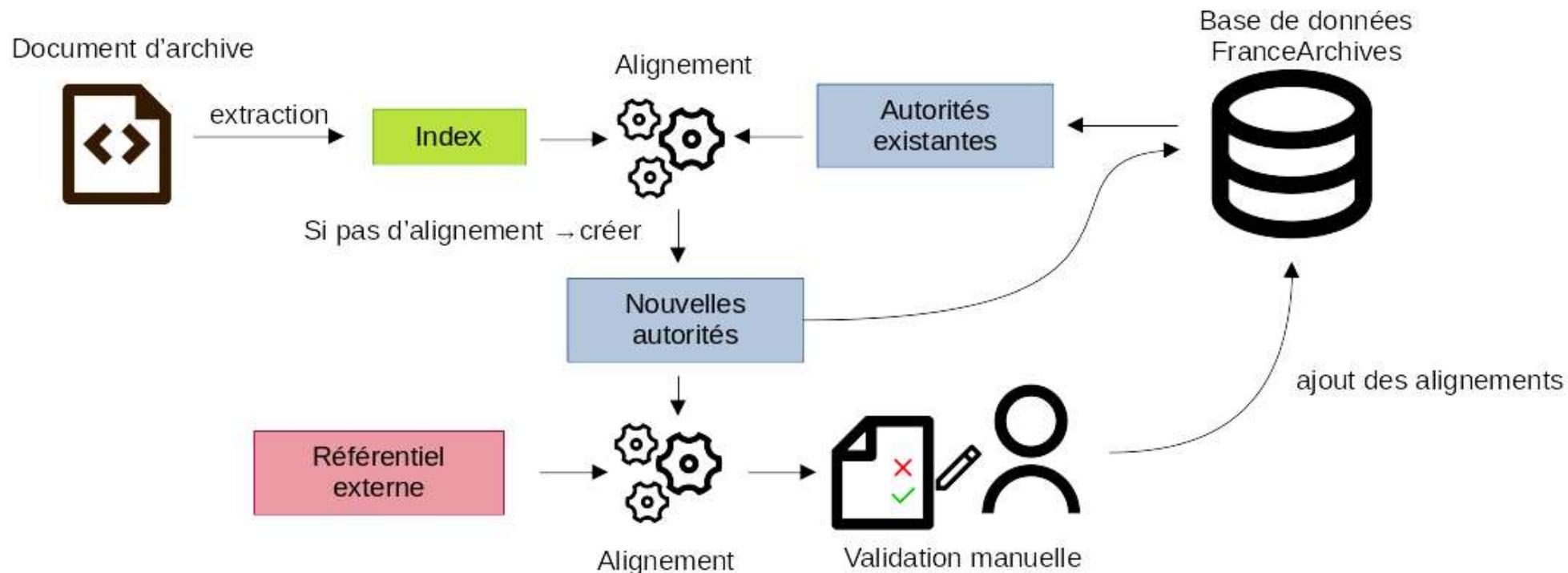
Lier les index des documents d'archives à 3 types d'autorités :

- **Lieux**
- **Agents** (Person, Organisation, ...)
- **Sujets**

**Identifiants uniques**

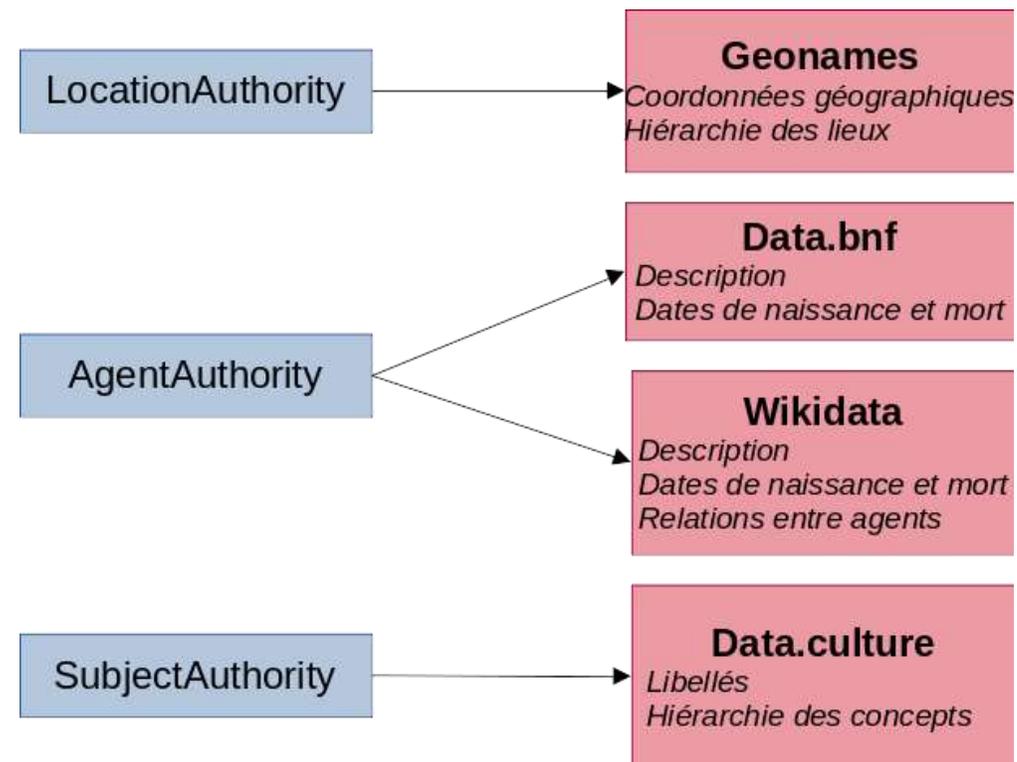


# Processus de gestion des autorités



# Alignement vers des référentiels externes

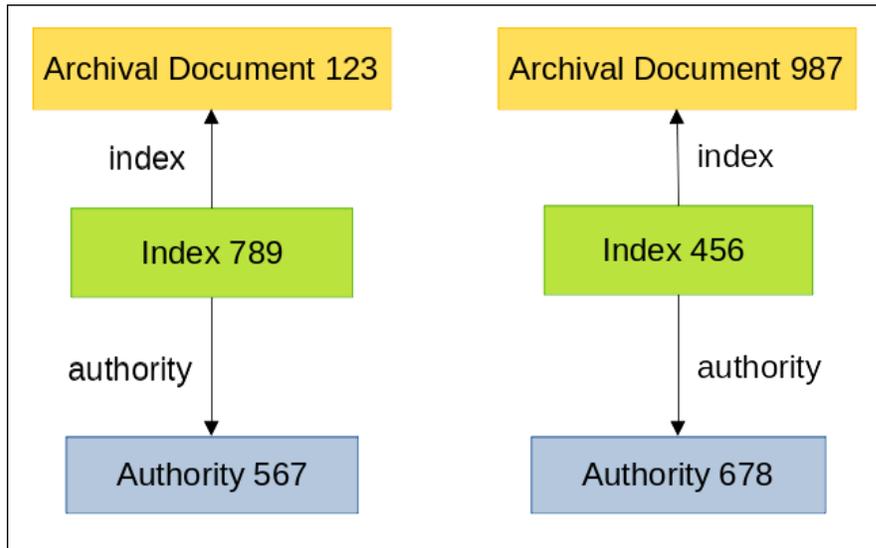
- Alignement (réconciliation) vers des bases externes
- **Récupération d'informations** depuis ces nouvelles sources
- Mise à jour régulière des informations récupérées



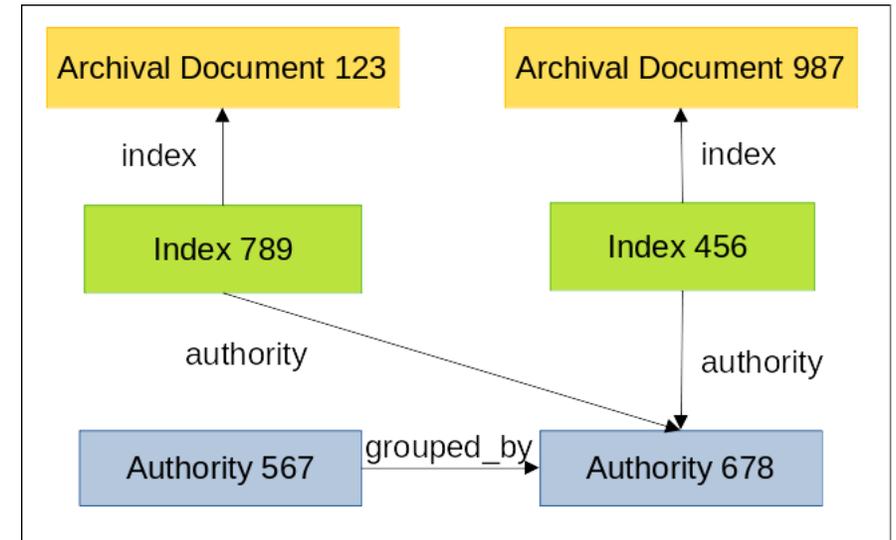
# Groupement d'autorités

Quand deux autorités représentent la même entité réelle : elles sont **groupées** (opération manuelle)

Avant

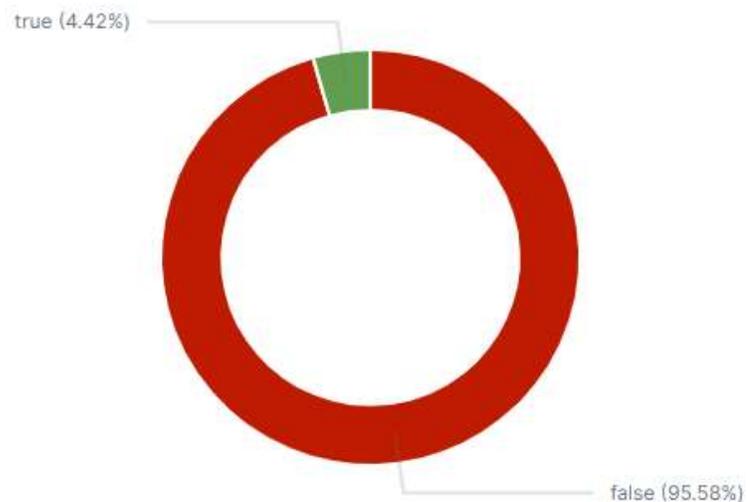


Après

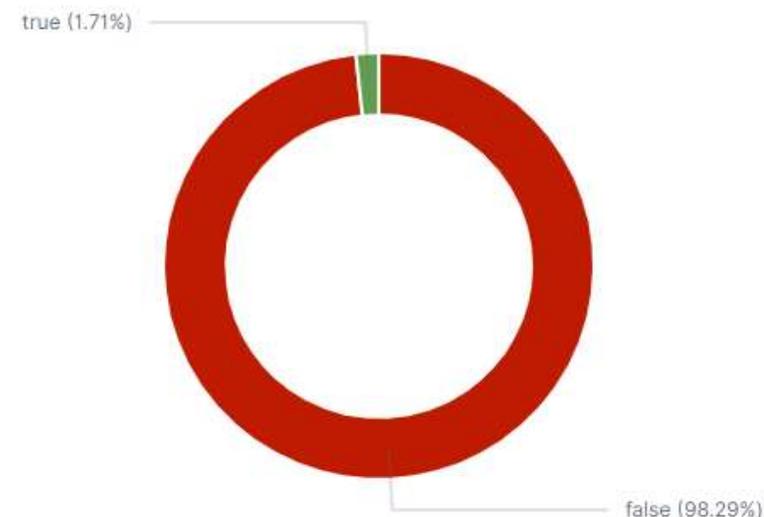


# Autorités

- **Autorités Agent** : 3 012 313 (alignées : 29 463 - qualifiées : 25 885)
- **Autorités Lieu** : 448 186 (alignées : 125 197 - qualifiées : 33 049)
- **Autorités Thème** : 114 529 (alignées : 3199 - qualifiées : 2 738)



Autorités alignées



Autorités qualifiées

# Exporter les données en RDF

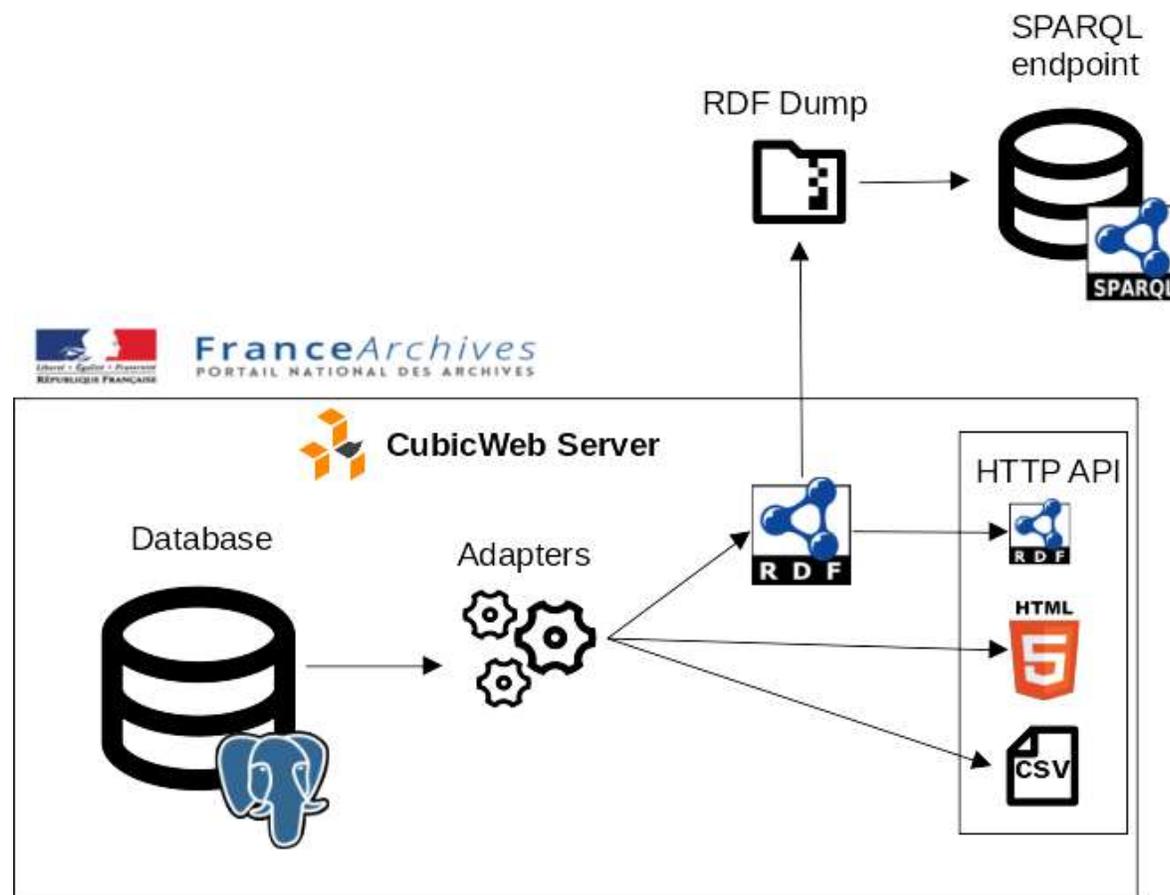
# Records In Contexts - Ontologie

- Basée sur le modèle contextuel de Records In Contexts – RiC-CM
- Créée et portée par le [Conseil international des Archives](#)
- Bonne adaptation des modèles EAD (description des documents) et EAC-CPF (notices descriptives de producteurs d'archives)
- Conversion antérieure par les Archives nationales de France
- Ontologie implémentée par d'autres [projets](#) et outils
- Bon support pour l'utilisation de l'ontologie ([documentation](#), contact direct)



# Processus d'export RDF

- Adapteurs pour transformer les données en RDF
- Négociation de contenu (HTTP) sur le moment
- [Archive RDF](#) tous les 6 mois
- [SPARQL Endpoint](#) peuplé à partir de l'archive



# Sparnatural sur les données de "qualité"

- Outil d'interrogation de SPARQL
- Seulement sur les données "de qualité"
- Modèle d'interrogation conçu par le SIAF

Archives antérieures à 1715 concernant Pamiers

Document d'archives a pour date de fin Date Jusqu'à 1715

et

Document d'archives a pour sujet Lieu

où

Lieu a pour nom Nom Pamiers

Afficher/Masquer éditeur SPARQL

Table Response 6 results in 0.09 seconds Page size: 50

**Archives\_1**

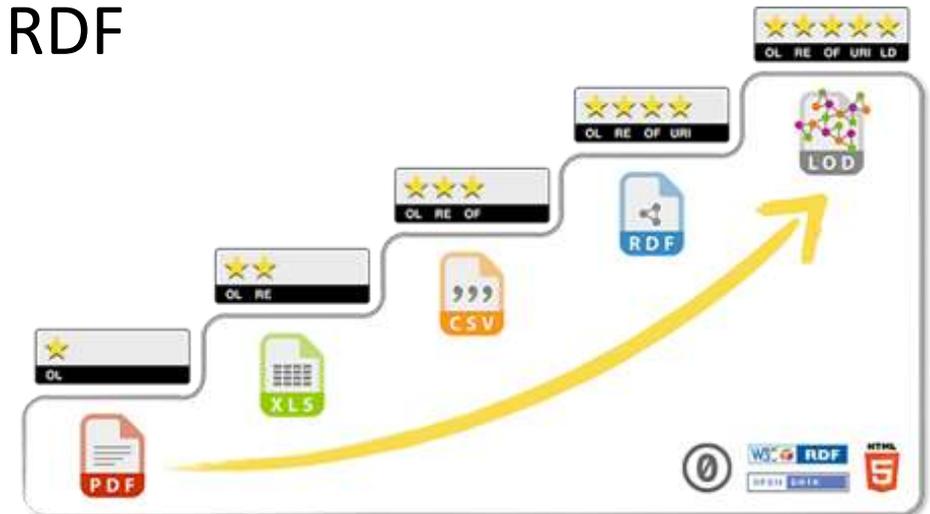
- 1 [Ratification par Paul Fellisson-Fontarrier d'une transaction passée à Pamiers, le 14 mai 1689, entre son procureur et les représentants du Conseil politique de la ville au suje](#)
- 2 [Cour des consuls de Pamiers \(XIVe-XVe siècles\) : cause criminelle contre Bernard Ademar, accusé de meurtre \(1325\)](#)
- 3 [Cour des consuls de Pamiers \(XIVe-XVe siècles\) : Registre des causes civiles et criminelles \(1475-1476\)](#)
- 4 [Reliquaire de sainte Eugénie, vierge, et de quatre saints martyrs \(Castule, Cyrille, Charles et Cyran ?\) - Cliché Jean Teulière, couleur](#)
- 5 [Donation à Jean Calais, religieux de l'abbaye Vermand, étudiant au collège de Prémontré \(Paris\), de droits sur coupes de bois à l'abbaye, par l'évêque de Pamiers, \(abbé de v](#)
- 6 [Jean Barbançon, évêque de Pamiers. Bail du revenu de son évêché moyennant 8000 lt. par an. Et autres actes.](#)

Showing 1 to 6 of 6 entries

# Conclusion et Perspectives

# Conclusion

- Des documents d'archives recherchables par tou·te·s
- Indexation de qualité
- Données qui suivent les recommandations du Linked Open Data \*\*\*\*\*
- Gestion de grands volumes de données RDF



# Perspectives

- Export des alignements d'autorités dans les référentiels externes :
  - Data.bnf
  - Wikidata
  - Persée
- Étendre les données récupérées des référentiels externes
- Étendre l'export RDF aux autres documents (état-civil, etc.)
- Optimiser le processus pour la gestion de données encore plus volumineuses
- Suivre les évolutions de RiC-O

# Merci !

[elodie.thieblin@logilab.fr](mailto:elodie.thieblin@logilab.fr)

[mathilde.daugas@culture.gouv.fr](mailto:mathilde.daugas@culture.gouv.fr)

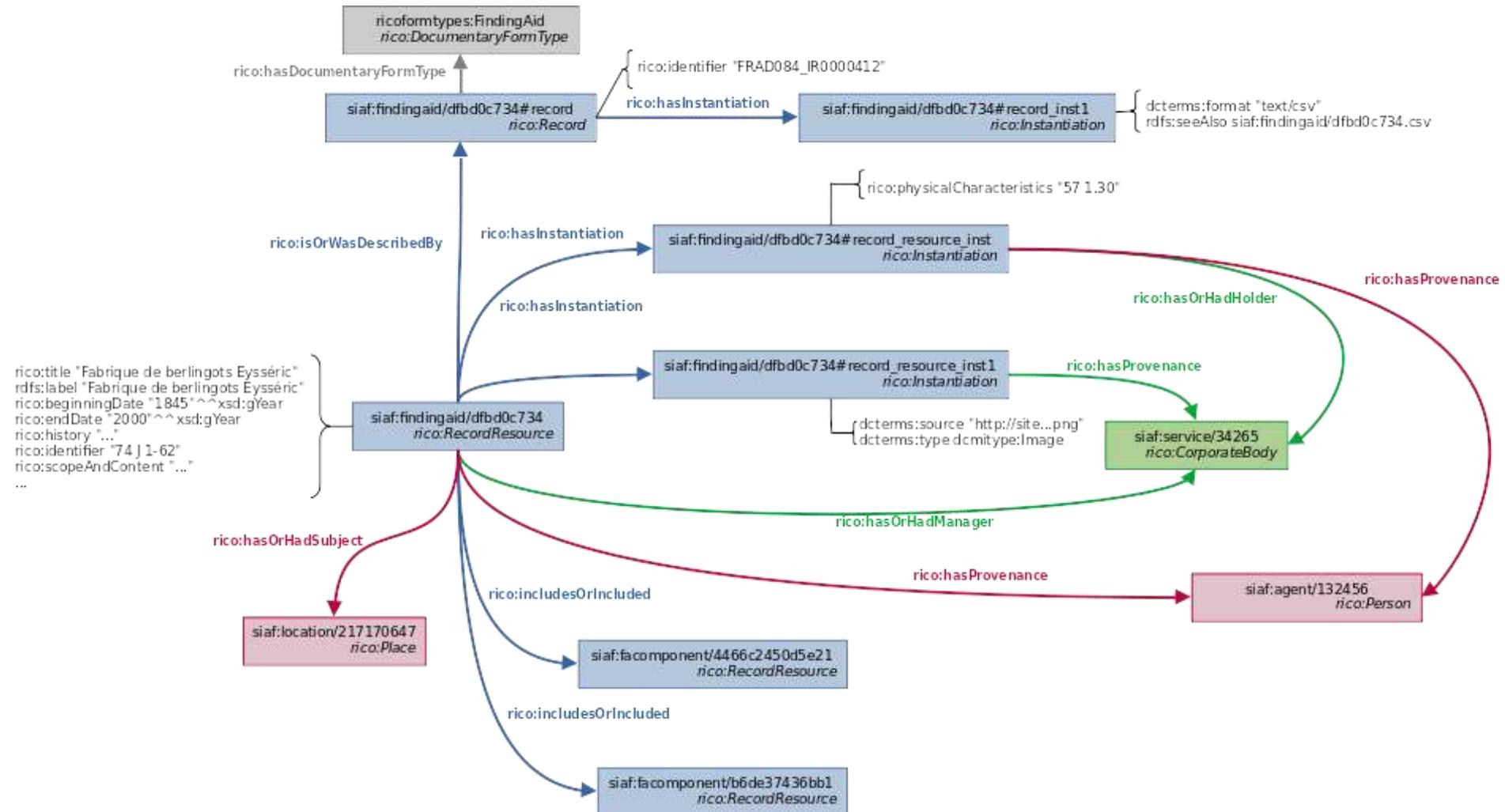
# Annexes

# 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: 2<sup>nd</sup> 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<sup>e</sup> 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 » ▾

