

# Implementing *ICA Records in Contexts-Ontology (RiC-O)* at the National Archives of France (ANF): first steps and prospects

*Florence Clavaud*

*(Archives nationales de France, Mission référentiels /  
ICA EGAD executive member, lead of RiC-O development team) ;*

email: [florence.clavaud@culture.gouv.fr](mailto:florence.clavaud@culture.gouv.fr)

Wed., February 3<sup>rd</sup>, 2021

**Study day on « The semantic web and cultural heritage »  
Université de Lille, laboratoire Gériico**

# Outline

- Records in Contexts (RiC-CM and RiC-O): the new international standard for describing archives
- Why use RiC and RiC-O at the ANF?
- A first step: the PIAAF prototype and its outcomes
- Next steps (ongoing): moving to a larger scale
- Conclusion: prospects



# 1. Records in Contexts (RiC-CM and RiC-O): the new standard for describing archives

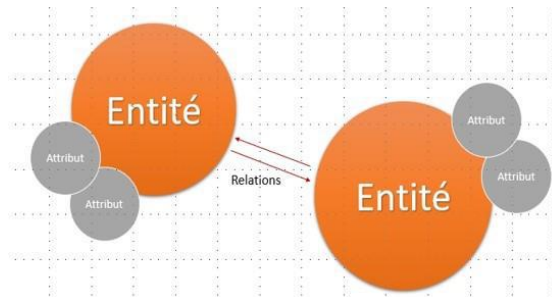
# What is Records in Contexts (RiC)?

A three-part standard:

- RiC-Conceptual Model or RiC-CM : **an international, generic and abstract, implementation independant, conceptual model for describing archives and their contextual entities**
- RiC-ontology or RiC-O: **an OWL 2 ontology, technical implementation of RiC-CM**, that defines the vocabulary and rules to be applied for creating RDF datasets (or generating them from existing archival metadata) that describe in a consistent way any kind of archival record resource. It can support publishing RDF datasets as Linked Data, querying them using SPARQL, and making inferences using the logic of the ontology.
- And (later), RiC-Application Guidelines or RiC-AG

# Why RiC?

- Need to replace the four existing international standards...



- ...and **move from description principles to a unique, entity-relationship based, model**
- To enable a better representation of records, of their history and the multiple layers of contexts in which they exist through time and space
- To enable a more precise, more nuanced, easier to process, multidimensional description of archives...
- ...therefore, to **facilitate access to archival metadata, thus to the archival resources they describe**

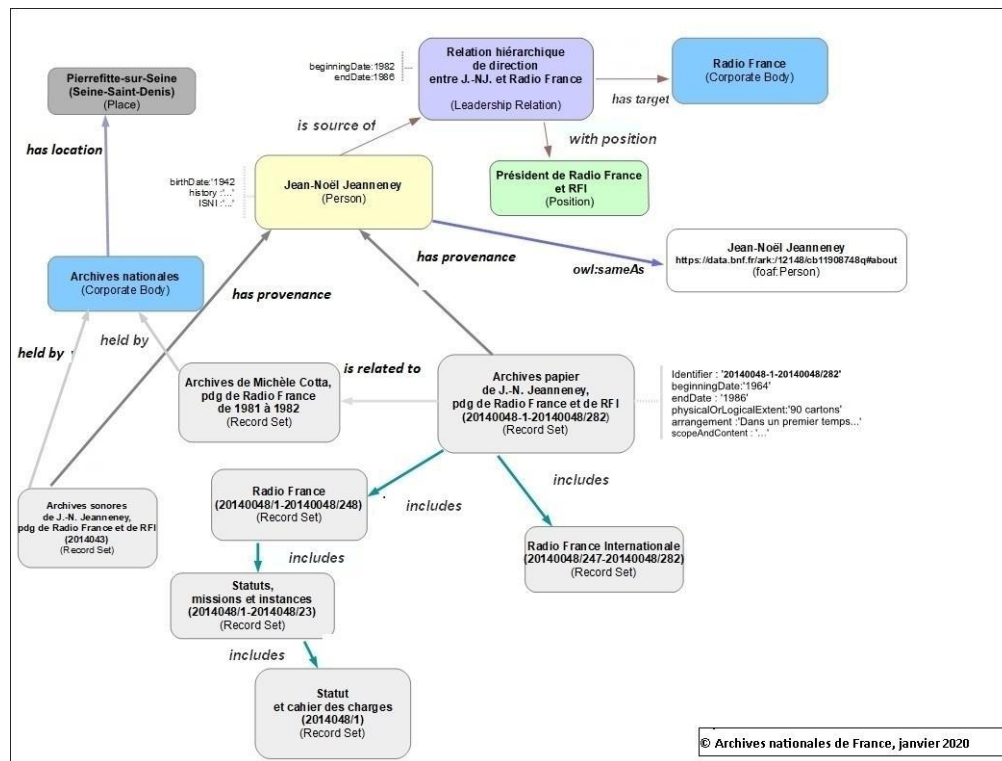
# The archival world seen as an oriented graph: a small example

A very partial view,  
conforming with RiC-O v0.1,  
of the graph formed by Jean-  
Noël Jeanneney's archives as  
the president of Radio  
France

This view *re-presents* the  
data contained in an existing  
XML/EAD 2002 finding aid.

See:

[https://www.siv.archives-nationales.culture.gouv.fr/siv/IR/Fran\\_IR\\_050629](https://www.siv.archives-nationales.culture.gouv.fr/siv/IR/Fran_IR_050629)



# RiC is being developed by ICA EGAD

- ICA: International Council on Archives
- ICA EGAD: ICA *Expert Group on Archival Description*
- Formed in 2013
- Mandate: 2013-2016, renewed for 2017-2021
- EGAD's work is regularly submitted to comments in order to involve the professional communities
- It must be continuously maintained after 2021
- See: <https://www.ica.org/en/egad-steering-committee-0>
- Chair: Daniel Pitti (University of Virginia, USA; lead of [Social Networks and Archival Context Cooperative](#) project)
- About 20 persons, from all over the world
- Different views on archives, different positions and skills

# RiC history and roadmap

- August 2016 : **RiC-CM v0.1 published with a call for comments**  
A lot of comments received from August 2016 to January 2017, taken into account since
- February-November 2019: beta versions of RiC-O sent to persons who applied as early reviewers
- December 12, 2019:
  - **first public release of RiC-O (v0.1)** with a call for comments
  - [public release of a preview of RiC-CM v0.2](#). This version replaces RiC-CM v0.1 which is obsolete
- March 2020: [RiC-O Git repository](#) opens on GitHub, along with [an information website](#)
- February 12, 2021: [public release of RiC-O v0.2](#)
- March or April 2021 : **public release of RiC-CM v0.2 full draft, with a call for comments**
- Autumn 2021: release of **RiC-CM and RiC-O v1.0**
- **The versions of RiC that precede v1.0 are drafts.**

However RiC-CM v0.2 and RiC-O v0.2 are close to v1.0 and can be used.



# RiC-CM 0.2: 22 entities

## RiC Entities Hierarchy

First Level	Second Level	Third Level	Fourth Level
RiC-E01 Thing	RiC-E02 Record Resource	RiC-E03 Record Set	
		RiC-E04 Record	
		RiC-E05 Record Part	
	RiC-E06 Instantiation		
	RiC-E07 Agent	RiC-E08 Person	
		RiC-E09 Group	RiC-E10 Family
			RiC-E11 Corporate Body
		RiC-E12 Position	
		RiC-E13 Mechanism	
	RiC-E14 Event	RiC-E15 Activity	
	RiC-E16 Rule	RiC-E17 Mandate	
	RiC-E18 Date	RiC-E19 Single Date	
		RiC-E20 Date Range	
		RiC-E21 Date Set	
RiC-E22 Place			

Four essential entities, that constitute the core of the model: **Record Resource, Instantiation, Agent, Activity.**

# RiC-CM entities have attributes

- 41 attributes
- Attributes common to all entities: identifier, name, descriptive note
- Attributes for each top level entity (inherited by the sub-entities)
- Attributes specific to some entities
- Some are repeatable or extensible

<b>ID</b>	Assigned identifier beginning with <b>RIC-A</b> and followed by a number.
<b>Name</b>	Natural language label.
<b>Definition</b>	Brief definition.
<b>Domain</b>	Entity or Entities which may have the Attribute
<b>Specifications</b>	The possible specifications, precisions, annotations or qualifiers of the meaning of the value an attribute in a description.
<b>Extensibility</b>	The possible differentiation of the attribute using extensions or specializations in sub-attributes.
<b>Repeatability</b>	Whether the attribute can be repeated or not in the description of a particular entity or relation.
<b>Value Schema</b>	The rules for selecting or formulating the value of the attribute.
<b>Scope</b>	It provides additional information to understand and use the attribute.
<b>Examples</b>	It presents examples of application of each attribute.

# Example: the attributes of Record and Instantiation

RiC-E04	Record
Attribute ID	Attribute Name
RiC-A16	Descriptive Note
RiC-A22	Identifier
RiC-A28	Name
RiC-A21	History
RiC-A35	Record Resource Extent
RiC-A38	Scope and content
RiC-A40	Structure
RiC-A03	Authenticity Note
RiC-A07	Classification
RiC-A08	Conditions of Access
RiC-A09	Conditions of Use
RiC-A10	Content Type
RiC-A17	Documentary Form Type
RiC-A24	Integrity
RiC-A25	Language
RiC-A26	Legal Status
RiC-A39	State

RiC-E06	Instantiation
Attribute ID	Attribute Name
RiC-A16	Descriptive Note
RiC-A22	Identifier
RiC-A28	Name
RiC-A03	Authenticity Note
RiC-A04	Carrier Extent
RiC-A05	Carrier Type
RiC-A08	Conditions of Access
RiC-A09	Conditions of Use
RiC-A21	History
RiC-A23	Instantiation Extent
RiC-A24	Integrity
RiC-A31	Physical Characteristics
RiC-A33	Production Technique
RiC-A34	Quality of Representation
RiC-A37	Representation Type
RiC-A40	Structure

# RiC-CM v0.2 entities are related to each other by relations

- Whole-part relations
- Sequential and temporal relations
- Subject relations
- Record Resource to Record Resource relations
- Record Resource to Instantiation relations
- Provenance relations
- Instantiation to Instantiation relations
- Management relations
- Agent to Agent relations
- Event relations
- Rule relations
- Date relations
- Spatial relations

<p>Type: <u>Record Resource to Record Resource relations</u></p> <p>RIC-R022: Record Resource <i>is record resource associated with record resource</i> Record Resource</p>	<p>RIC-R023: Record resource <i>has genetic link to Record Resource</i> Record Resource</p> <p>RIC-R013: Record Resource <i>has reply</i> Record Resource (see also above)</p> <p>RIC-R003: Record <i>has or had constituent</i> Record Part (see also above)</p> <p>RIC-R024: Record Set <i>includes or included</i> Record or Record Set</p>	<p>RIC-R010: Record <i>is original of</i> Record (see also above)</p> <p>RIC-R011: Record <i>is draft of</i> Record (see also above)</p> <p>RIC-R012: Record Resource <i>has copy</i> Record Resource (see also above)</p>	
<p>Type: <u>Record Resource to Instantiation relations</u></p> <p>RIC-R025: Record Resource <i>has instantiation</i> Instantiation</p>			

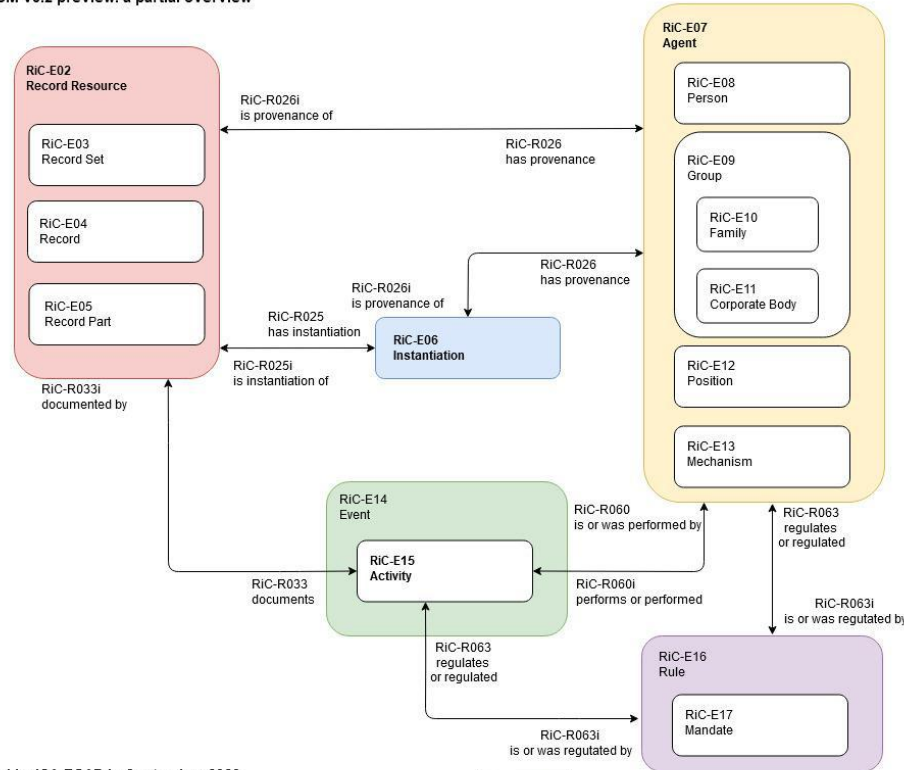
In RiC-CM v0.2, there are **78 relations** (and the inverse ones when they are not symmetric), grouped and organize in a poly-hierarchical system.

The top level relation is 'is related to'.

They have attributes (identifier, description, date, certainty, source).

# RiC-CM 0.2 main entities and some of the relations

RiC-CM v0.2 preview: a partial overview



Created by ICA-EGAD in September 2020.

This diagram is licensed under a Creative Commons Attribution 4.0 International License.



This diagram, and other ones, can be found in the RiC-O GitHub repository: [https://github.com/ICA-EGAD/RiC-O/tree/master/diagrams/diagrams\\_v0-2](https://github.com/ICA-EGAD/RiC-O/tree/master/diagrams/diagrams_v0-2)

# RiC-O: main information

- The latest public release is to be found at:  
<https://www.ica.org/standards/RiC/ontology>
  - It is fully documented in English
  - License:  
<https://creativecommons.org/licenses/by/4.0/>
  - The source code, examples, diagrams, are available at:  
<https://github.com/ICA-EGAD/RiC-O>
- > RiC-O is a public Git project. You can clone it, or create issues, comments, and even fork.
- See also the official website: <https://ica-egad.github.io/RiC-O/>



International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 0.2

Ontology IRI: <https://www.ica.org/standards/RiC/ontology>

This HTML view: [https://www.ica.org/standards/RiC/RiC-O\\_v0.2.html](https://www.ica.org/standards/RiC/RiC-O_v0.2.html)

OWL source file: [https://www.ica.org/standards/RiC/RiC-O\\_v0.2.rdf](https://www.ica.org/standards/RiC/RiC-O_v0.2.rdf)

Ontology metadata

This version: Version 0.2 - 2021-02-12.

See also the project on GitHub, where other resources can be found, and the accompanying website: <https://github.com/ICA-EGAD/RiC-O> (the website is at <https://ica-egad.github.io/RiC-O/>)

Prior version: [https://www.ica.org/standards/RiC/RiC-O\\_v0.1.rdf](https://www.ica.org/standards/RiC/RiC-O_v0.1.rdf)

Creators:

Florence Clavard (Archives nationales de France), member of EGAD, lead of EGAD RiC-O team  
International Council on Archives Expert Group on Archival Description (ICA EGAD)

Contributors:

Aaron Rubinstein (University of Massachusetts Amherst, USA), member of EGAD  
Daniel Pitts (University of Virginia, USA), chair of EGAD  
Mia Herralta (National Archives of Finland), member of EGAD  
Tobias Wildi (Docuseam GmbH, Switzerland), member of EGAD

Publisher: International Council on Archives

Rights: Copyright 2019-.... International Council on Archives (ICA)

License: <https://creativecommons.org/licenses/by/4.0/>

Abstract

RiC-O (Records in Contexts-Ontology) is an OWL ontology for describing archival record resources. As the second part of Records in Contexts standard, it is a formal representation of Recon. The current official version is v0.2. It is compliant with RiC-OM v0.2 full draft, that will be published in February or March 2021, and that is slightly different from [RiC-OM v0.2 preview](#), that was ; RiC-O provides a generic vocabulary and formal rules for creating RDF datasets (or generating them from existing archival metadata) that describe in a consistent way any kind of archival record, querying them using SPARQL, and making inferences using the logic of the ontology.

History note

A first beta version of RiC-O was developed in 2017 and used by the National Archives of France for building a proof of concept (<https://sialf.demio.loglab.fr>)

# RiC-O: design principles

- **A domain, reference ontology**  
mappings to the components of other ontologies will be done in 2021
- **Immediately usable (a 'transitional' ontology)**
- **Flexible**
- **Useful (functional)**
- **Extensible**

See also: [RiC-O design principles](#)

# RiC-O must be useful: querying a RDF/RiC-O graph

```
1 # Les entités qui ont succédé à une entité donnée (ici, n° 51120, France. Ministère de l'Intérieur. Direction des Beaux-Arts. Bureau des Monuments historiques (1839-1848))
2
3 PREFIX r1co: <http://www.ica.org/standards/RiC/ontology#>
4 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
6
7 * SELECT ?entite ?nom ?dateDebut ?dateFin WHERE {
8   <http://data.archives-nationales.culture.gouv.fr/agent/051120> (r1co:agentIsSourceOfAgentTemporalRelation/r1co:agentTemporalRelationHasTarget)+ ?entite .
9   ?entite rdfs:label ?nom; r1co:beginningDate ?dateDebut.
10  OPTIONAL {?entite r1co:endDate ?dateFin.}
11
12
13
14 .
```

Table Raw Response Pivot Table Google Chart Download as

Filter query results Showing results from 1 to 10 of 10. Query took 0.2s, today at 14:48.

	entite	nom	dateDebut	dateFin
1	http://data.archives-nationales.culture.gouv.fr/agent/051121	"France. Ministère de l'Intérieur. Direction des Beaux-Arts. Bureau des Beaux-Arts, des Musées et des Monuments historiques (1848-1852)" <sup>981</sup>	"1848-01-01" <sup>982</sup> xsd:date	"1852-12-31" <sup>983</sup> xsd:date
2	http://data.archives-nationales.culture.gouv.fr/agent/051122	"France. Ministère d'État. Division des Beaux-Arts. Service des Monuments historiques (1853-1858)" <sup>981</sup>	"1853-01-01" <sup>982</sup> xsd:date	"1858-12-31" <sup>983</sup> xsd:date
3	http://data.archives-nationales.culture.gouv.fr/agent/051123	"France. Ministère d'État. Bureau des Monuments historiques (1858-1863)" <sup>981</sup>	"1858-01-01" <sup>982</sup> xsd:date	"1863-12-31" <sup>983</sup> xsd:date
4	http://data.archives-nationales.culture.gouv.fr/agent/051124	"France. Ministère de la maison de l'Empereur. Surintendance des Beaux-Arts. Bureau des Monuments historiques (1863-1870)" <sup>981</sup>	"1863-01-01" <sup>982</sup> xsd:date	"1870-12-31" <sup>983</sup> xsd:date
5	http://data.archives-nationales.culture.gouv.fr/agent/051125	"France. Ministère de l'Instruction publique. Direction des Beaux-Arts. Bureau des Monuments historiques (1870-1907)" <sup>981</sup>	"1870-01-01" <sup>982</sup> xsd:date	"1907-12-31" <sup>983</sup> xsd:date
6	http://data.archives-nationales.culture.gouv.fr/agent/051126	"France. Ministère de l'Éducation nationale. Direction des Beaux-Arts. Bureau des Monuments historiques (1907-1939)" <sup>981</sup>	"1907-01-01" <sup>982</sup> xsd:date	"1939-12-31" <sup>983</sup> xsd:date
7	http://data.archives-nationales.culture.gouv.fr/agent/051149	"France. Ministère de l'Éducation nationale. Secrétariat général des Beaux-Arts. Direction des services d'architecture. Bureau des Monuments historiques et Sites (1940-1944)" <sup>981</sup>	"1940-01-01" <sup>982</sup> xsd:date	"1944-12-31" <sup>983</sup> xsd:date
8	http://data.archives-nationales.culture.gouv.fr/agent/051150	"France. Ministère de l'Éducation nationale. Direction de l'architecture. Sous-direction des Monuments Historiques et des Sites (18 août 1945-8 janvier 1959)" <sup>981</sup>	"1945-08-18" <sup>982</sup> xsd:date	"1959-01-08" <sup>983</sup> xsd:date
9	http://data.archives-nationales.culture.gouv.fr/agent/051151	"France. Ministère de la Culture et de la Communication. Direction de l'Architecture et du Patrimoine. Sous-direction des Monuments Historiques et des Espaces protégés (1959-2009)" <sup>981</sup>	"1959-01-08" <sup>982</sup> xsd:date	"2009-12-31" <sup>983</sup> xsd:date
10	http://data.archives-nationales.culture.gouv.fr/agent/051153	"France. Ministère de la Culture et de la Communication. Direction générale des patrimoines. Service du patrimoine. Sous-direction des Monuments historiques et des Espaces protégés (2009-...)" <sup>981</sup>	"2009-01-01" <sup>982</sup> xsd:date	

keyboard shortcuts

Querying the RDF graph of the archival creators (for now, 15000 EAC-CPF authority records) at the Archives nationales de France (screenshot from a GraphDB instance installed locally).

**New questions can be answered**, going through the arcs/relations, like shown here.



# Much more components in RiC-O than in RiC-CM

- Some attributes (e.g. those that have a controlled value in RiC-CM) are represented as classes
- Some RiC-CM entities represented by several classes (e.g. the Place entity)
- Other classes and properties introduced
- Sometimes several methods for representing an assertion (including a system of n-ary relations representing many of RiC-CM relations, apart from the corresponding object properties, that can be inferred from the longer path)
- More details can be found there:

[https://www.ica.org/standards/RiC/RiC-O\\_v0-2.html#fromRiCCM-to-RiCO](https://www.ica.org/standards/RiC/RiC-O_v0-2.html#fromRiCCM-to-RiCO)



## 2. Implementing RiC-O at the National Archives of France

# A few words about the Archives nationales de France (ANF)

- A national public service created on 1 January 2007 by a decree issued by the Ministry of Culture (dated 24 December 2006). Its mission is to collect, classify, inventory, conserve, restore, provide access to, and promote public **archives from the central administrations of State, the archives of Parisian notaries, and private records of national interest.**
- In fact, **an old institution, created as a result of the French Revolution; records kept dated from the Middle Ages to nowadays**
- **Key figures (2019)** (see also <https://www.archives-nationales.culture.gouv.fr/en/web/guest/chiffres-cles>):
  - more than 370 linear km of analogue records, 65,4 To of digital archives
  - about 7,8 million records digitized and accessible online
  - 8000 different readers, about 115 000 boxes accessed in 2019
- An institutional website (<https://www.archives-nationales.culture.gouv.fr/>), in which you can query and display the archival metadata that describe the holdings, using the Salle des inventaires virtuelle (SIV) (<https://www.siv.archives-nationales.culture.gouv.fr>)

# Archival metadata at the ANF today (1)

- A **huge amount** of heterogeneous metadata, created by generations of archivists and historians over centuries  
A significant effort for digitizing them from 1990, and for updating and completing this legacy
- In a **series of silos**; the main one contains more than **29000 archival finding aids (XML/EAD files, thus structured documents) and about 15000 authority records (XML/EAC-CPF files)** on the archival creators, as well as about 20 vocabularies in a specific XML format  
But it is not the only one... We have a lot of databases, and other repositories (among which a digital library, and the recently implemented digital archival system)
- **Already a lot of relations between the files in the first silo, but not really viewable and not searchable through the front-end web application** (the Salle des inventaires virtuelle) and, as concerns the relations between the finding aids and the authority records, of one category only (the provenance relation)
- The **vocabularies and other authority records of the main silo can be used for indexing the finding aids**, but are rather **poor, not standardized and quite rarely used till now**

# Archival metadata at the ANF today (2)

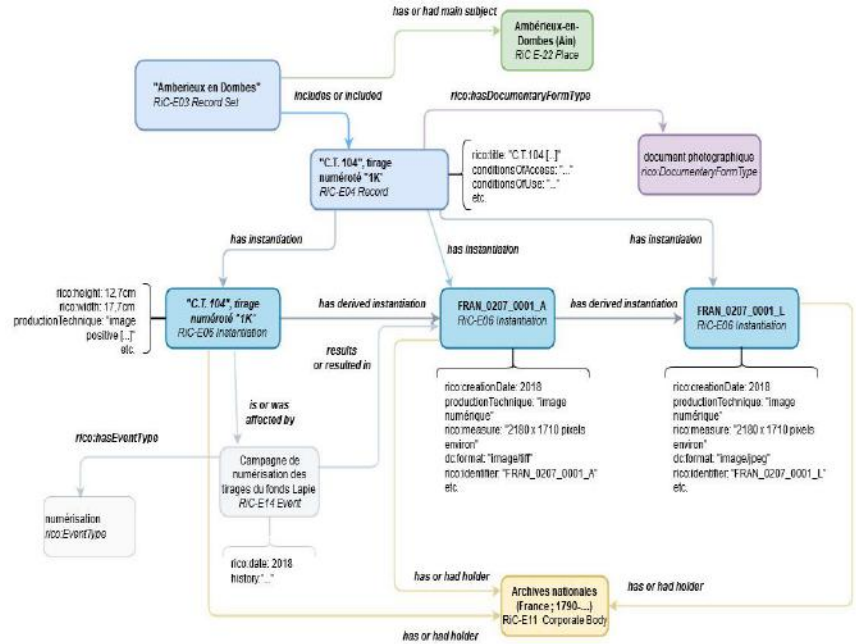
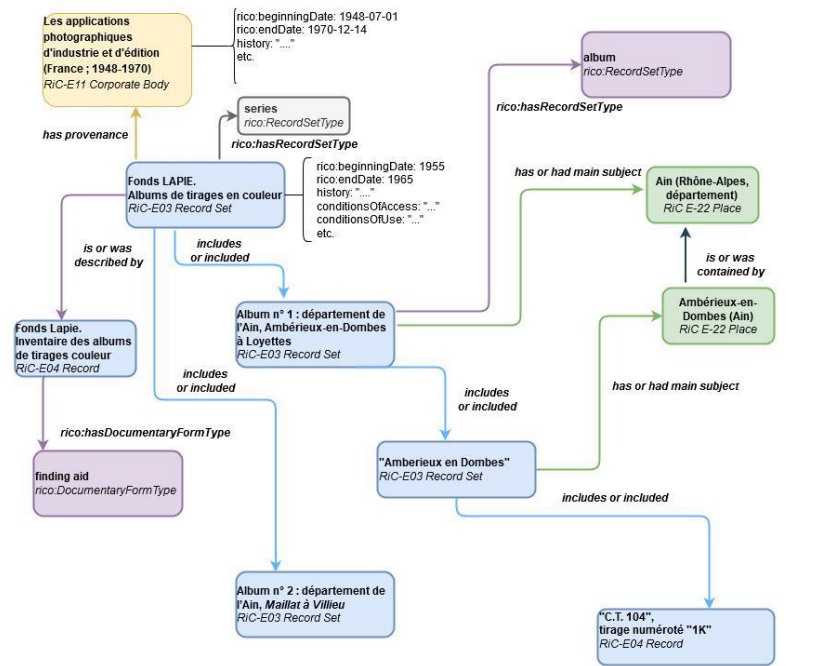
- Several end-user interfaces
  - **Very few intuitive access points (who, when, where, what...)**
  - **A lot of redundancies, from one silo to another and within the same silo** (particularly the main one, where the same group of records or same record may have been described several times in several finding aids)
  - **The end user finds it difficult to understand the result lists and how the results are displayed (in the context of a finding aid, as subcomponents of it)**
  - Very few bridges to other information systems
- > In fact, a classic situation, at a huge scale

# Needs, and a hypothesis

- Need to:
  - **Improve the consistency and granularity of the metadata, and facilitate their management**
  - **Break the internal silos**, link the data sets to each other through **shared authority data**
  - **Improve the access to metadata**, develop a new unique web interface
  - Move towards interoperability with external datasets
- The hypothesis: RiC can help to do this:
  - RiC-CM, as the global, conceptual, reference framework for designing the future graph of linked entities
  - RiC-O, as the technical model for building RDF knowledge graphs from the existing metadata and exposing them as Linked Open Data

This assumption has become a certainty over the first achievements.

# An example of how RiC (and RiC-O) helps to design the global target graph



Description partielle, conforme à RiC, de la série d'albums de tirages en couleur du fonds de la société LAPIE (A.N. 1PH/C/1 à 1PH/C/158) (instrument de recherche en cours de réalisation).

Description partielle, conforme à RiC, de l'un des tirages photographiques contenus dans l'album 1PH/C/1 (A.N. fonds LAPIE) (instrument de recherche en cours de réalisation).

Les relations sont présentées uniquement dans un sens (les relations inverses existent mais ne figurent pas dans le diagramme).  
Le préfixe rico: indique que le composant (classe ou propriété) n'existe que dans RiC-O.  
Lorsqu'il n'y a pas de préfixe, le composant existe dans RiC-CM et dans RiC-O.

Les relations sont présentées uniquement dans un sens (les relations inverses existent mais ne figurent pas dans le diagramme).  
Le préfixe rico: indique que le composant (classe ou propriété) n'existe que dans RiC-O.  
Lorsqu'il n'y a pas de préfixe, le composant existe dans RiC-CM et dans RiC-O.

# What RiC can help to do (a small, not exhaustive list)

- Focus on the objects that are to be described
- Make the relations between archival aggregations, or between records, explicit
- Consider the finding aids as... records
- Consider the digital images of analogue records as physical instantiations (among others) of a unique intellectual content
- Consider when needed some curation events, and more generally speaking, the history of the records
- Consider other relations between records and agents than the provenance relation
- Consider places as geo-historical entities
- Etc.



# A first proof of concept: PIAAF

- PIAAF : « Pilote d'Interopérabilité pour les Autorités Archivistiques Françaises »
- A prototype to prove that:
  - real-world existing (XML/EAD and XML/EAC-CPF) archival metadata sets can be converted to good quality RDF/RiC-O datasets, without losing any information;
  - RDF/RiC-O datasets coming from several institutions can be interconnected and linked to other datasets;
  - new web interfaces can be built for these sets, that enable to query, visualize and browse the graph thus generated
- A collaborative project lead by the ANF, with the French National Library and the Ministry of Culture, and a French private company, Logilab
- Qualitative project, on a small quantity of metadata, using an early version of RiC-O
- Project started in 2015; datasets selected, then accurately checked; converted to RDF by the ANF; prototype published online in February 2018 : <https://piaaf.demo.logilab.fr>

# PIAAF: the conversion results

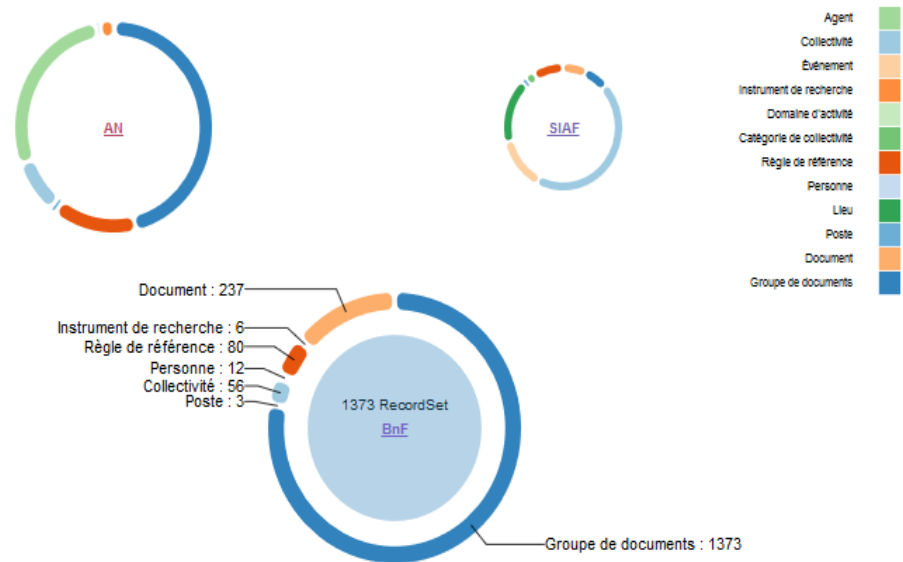
The entities/individuals described in the RDF sets that have been produced:

- 940 agents (from 262 EAC-CPF authority records)
- 20 corporate body types (from 14 EAC-CPF authority records)
- 547 agent names
- 335 mandates
- 114 events
- 140 places
- 28 positions
- 16 legal statuses
- 91 activity types
- 2062 'record sets', 291 'records'
- About 1600 relations (as full entities)

See also:

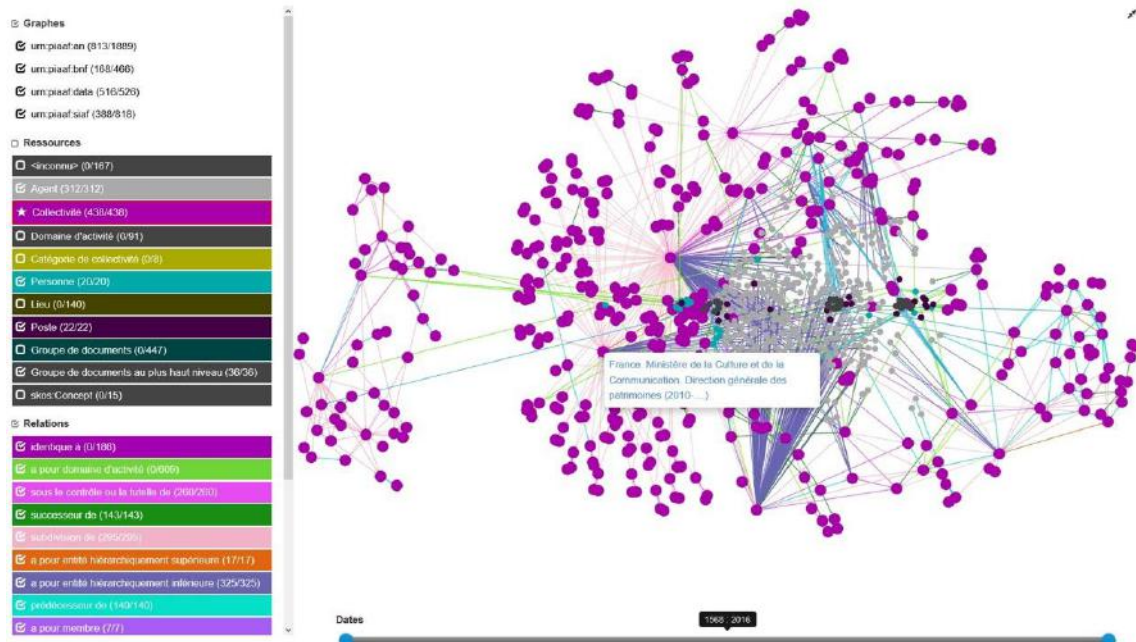
<https://piaaf.demo.logilab.fr/stats/>

Répartition des ressources par catégorie

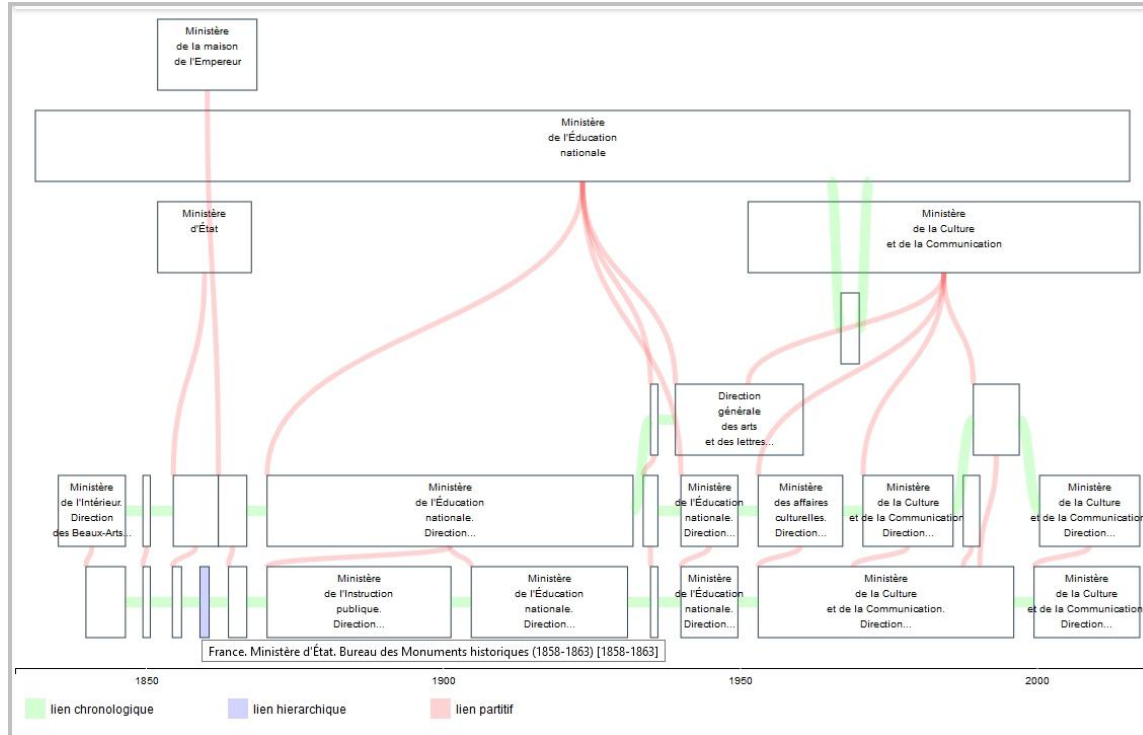


# PIAAF interface: one can see the underlying graph

A screenshot from the Corporate bodies (Collectivités) tab (<https://piaaf.demo.logilab.fr/ric/CorporateBody>), once other entities selected as well as the relations between these entities



# PIAAF: an example of an automatically generated chart (« diagramme chronologico-hiérarchique »)



A partial copy of the web page on the corporate body whose name is « France. Ministère d'État. Bureau des Monuments historiques (1858-1863) » ([https://piaaf.demo.logilab.fr/resource/Fran\\_corporate-body\\_051123](https://piaaf.demo.logilab.fr/resource/Fran_corporate-body_051123))

PIAAF SPARQL endpoint :  
here, a query that goes  
through the graph (query on  
the persons and the corporate  
bodies that they have lead).

Screenshot from the page  
<https://piaaf.demo.logilab.fr/sparql>.

The screenshot shows the PIAAF SPARQL endpoint interface. At the top, there are several input fields containing the following text: "1800 et 1970", "Alignement d'entités fondé sur une relation d'équivalence des 2 entités avec une entité tierce", "e des documents et agrégats documentaires", "s organisations et types d'organisations, avec pour chaque fonction le nombre d'organisations qui l'exercent", "Institutions succédant à", "es d'organisations ayant comme domaine fonctionnel "objets culturels" ou un sous-domaine de ce domaine", "Producteurs des documen", "ayant un certain label dans tous les jeux de données importés dans le triplestore (AN, BnF, SIAF, et le référentiel de l'organisation admin", "herche sur les entités ayant un statut donné", "Relations entre des personnes et les institutions qu'elles ont dirigées", and "igrégats documentaires". Below these fields is a dropdown menu with ".logilab.fr/virtuoso/sparql" selected. The main area displays a SPARQL query: 

```
select ?personLabel ?rel ?desc ?bDate ?eDate ?cbodylabel ?cbody ?positionLabel
where {
  ?rel ric:Person.
  ?rel ric:label ?personLabel.
  ?rel ric:agentHasLeadershipRelation ?rel.
  ?rel ric:leadershipOn ?cbody.
  ?rel ric:fs:label ?cbodylabel.
  {?rel ric:description ?desc .}
  {?rel ric:beginningDate ?bDate .}
  {?rel ric:endDate ?eDate .}
```

# PIAAF: the project results

- Project successful! A lot of not trivial work, but its outcomes were high quality ones, and beyond them **the methodology proved to be of great help for the next steps.**
- A very encouraging experience of an implementation of RiC-O, which confirmed that:
  - RiC-O can be used for expressing accurately the complexity of the world of archives
  - **interesting outcomes can be generated from real-world existing metadata.** However, **several aspects must be carefully taken into account:**
    - the existing information systems and the data models and processes they use have to be assessed, and if necessary changed, in order **to manage unique identifiers for the objects that may emerge;**
    - and more generally speaking, **the quality of the source metadata could be enhanced and more accurately controlled**
    - as concerns **the visualisation of the graph and the queries, very interesting new pathes had been opened, and could to be further investigated**
- See also, in the project documentation, **the following page:**  
<https://piaaf.demo.logilab.fr/editorial/contexte-technique>

## Next (ongoing) steps: moving forward to a far larger scale

- Generating RDF sets from larger metadata sets
  - **from the whole of the ANF finding aids and authority records (2019-2020): developing a software, RiC-O Converter**
  - **from the whole of the other authority data (vocabularies) (ongoing):** defining a new, richer, SKOS and RiC-O based target data model, converting the source files to data conforming to this data model, and **enriching the resulting datasets**
  - from the other metadata silos (from 2021): database per database, and also from the digital archival system

- A software for **converting into RDF/RiC-O datasets all the EAD and EAC-CPF files held by the ANF**
- **Takes into account various scenarios, since the granularity of the files, and sometimes the content model of the elements, change from one file to another, or within the same file**
- **A strategy had to be defined and applied in order to deduplicate the relations between agents, and to generate unique persistent identifiers when the source files did not provide them**
- **Developed from March 2019 to April 2020, by Sparna and the ANF; funded by the Ministry of Culture**
- **Source code and documentation (in English), as well as unit tests, available since April 2020 at <https://github.com/ArchivesNationalesFR/rico-converter>**
- **Today the ANF can convert into RiC-O sets, using this software, the most important set within their metadata system, in less than 30 min on a personal computer.**  
The process generates about 155 million RDF triples for now.



- The software can easily be configured, and is written mainly in XSLT 2; the ANF can update it according to their needs.
- RiC-O Converter is **released under a free open source license**. It can be reused and modified by any person or institution.
- It comes along with **[EAD 2002 to RiC-O](#)** and **[EAC-CPF to RiC-O mappings](#)**, that can be used as a basis for developing any other tool of the kind.
- For more information on the tool and the outcomes of the project:
  - the presentation dated December 2020 for the SemWeb Pro 2020 event:  
<https://peertube.semweb.pro/videos/watch/c195e540-4691-4b38-9619-0a59300d1adc> ;
  - the video recording of the webinar dated 2020, June 20th (duration: about 2h):  
<https://www.dailymotion.com/playlist/x6x1d0>

# Converting to a new, richer, data model, and enriching, the vocabularies and authority data

For now, we have converted the ANF 20 vocabularies, whose source files had a very simple, most often flat, XML structure, to RDF/SKOS and RiC-O files.

The source files, fortunately, assign a unique identifier to each concept or entity described.

And we are enriching them, aligning them to other RDF authority data, or using other resources.

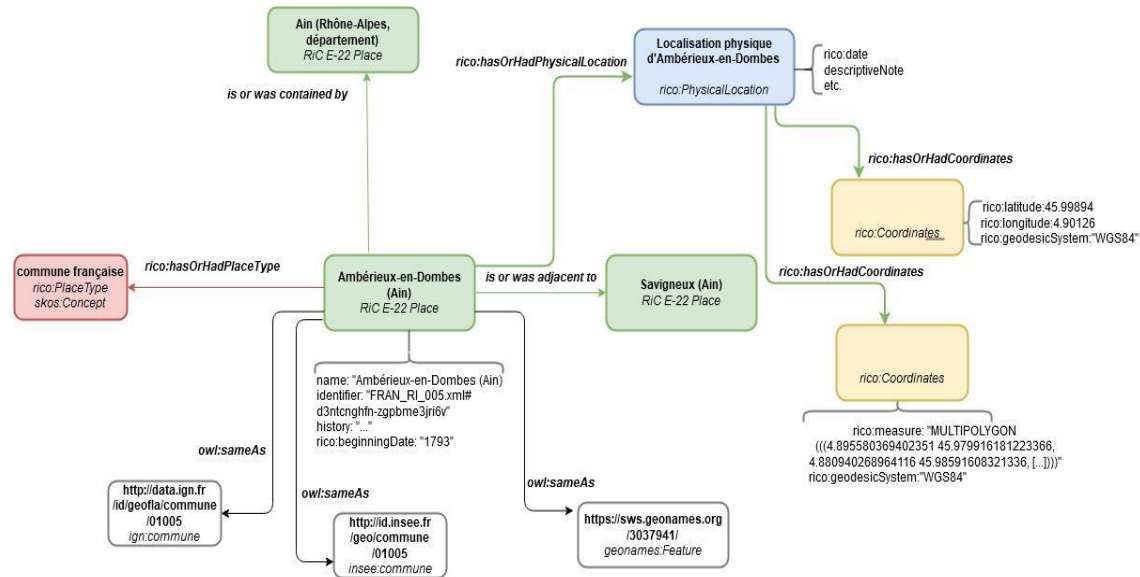
We have also begun to build new authority data, the main one being a vocabulary on the corporate bodies' activity domains and processes. This project is a collaborative one.

## A specific example: working on the description of places

```
<d id="d3ntcnghfn-zgpbme3jri6v">
<terme>Ambérieux-en-Dombes (Ain)</terme>
  <noticel>
    <forms>
      <f>Ambérieux-en-Dombes</f>
    </forms>
    <geo>C01-101#001-0005#N01-1</geo>
    <reg>Rhône-Alpes</reg>
    <dpt>Ain</dpt>
    <arr>12</arr>
    <canton>Saint-Trivier-sur-Moignans</canton>
    <insee>01005</insee>
  </noticel>
</d>
```

A French local administrative area (the ‘commune d’Ambérieux-en-Dombes’) in the source file that describes the places in the ANF information system for now.

# A specific example: places in RiC-O are considered as geo-historical entities



- It is also possible to add links to the name(s) of the place, to places that preceded it, or that are located within it. And, of course to agents whose location or jurisdiction is the place, or records that have subject the place, etc.

Description partielle, conforme à RiC-O,  
d'un lieu (la commune d'Ambérieu-en-Dombes) dans la version RDF du référentiel des lieux des Archives nationales.

Les relations sont présentées uniquement dans un sens (les relations inverses existent mais ne figurent pas dans le diagramme).  
Le préfixe riko: indique que le composant (classe ou propriété) n'existe que dans RiC-O.  
Lorsqu'il n'y a pas de préfixe, le composant existe dans RiC-CM et dans RiC-O.

© Archives nationales de France, janvier 2021

Document mis à disposition selon les termes  
de la Creative Commons Attribution 4.0 International License

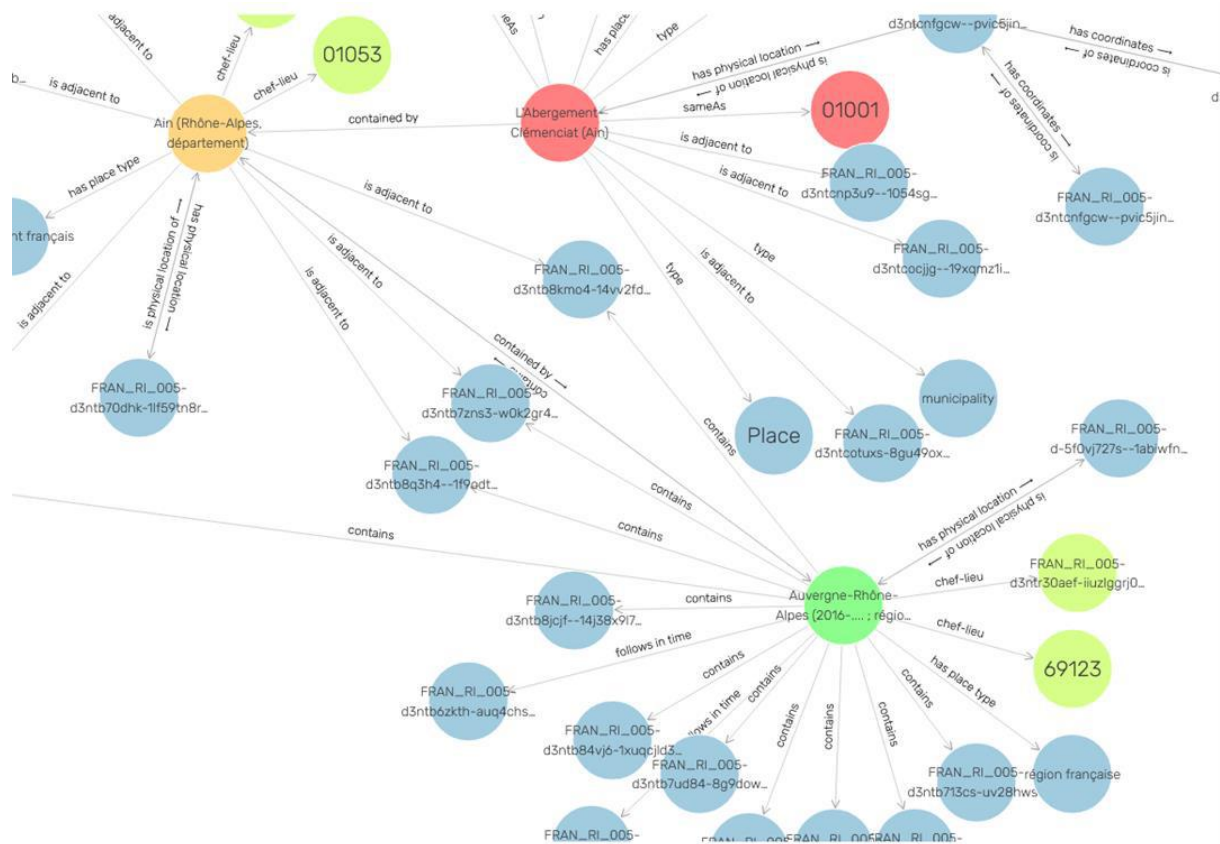


# The description of the commune of Ambérieux-en-Dombes, before and after the conversion and enrichment process

```
<d id="d3ntcnghfn-zgpbme3jri6v">
  <terme>Ambérieux-en-Dombes (Ain)</terme>
  <noticiel>
    <forms>
      <f>Ambérieux-en-Dombes</f>
    </forms>
    <geo>C01-101#001-0005#N01-1</geo>
    <reg>Rhône-Alpes</reg>
    <dpt>Ain</dpt>
    <arr>12</arr>
    <canton>Saint-Trivier-sur-Moignans</canton>
    <insee>01005</insee>
  </noticiel>
</d>
```

```
<rico:Place rdf:about="place/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v">
  <rdfs:label xml:lang="fr">Ambérieux-en-Dombes (Ain)</rdfs:label>
  <skos:prefLabel xml:lang="fr">Ambérieux-en-Dombes (Ain)</skos:prefLabel>
  <skos:altLabel xml:lang="fr">Collex-Bossy</skos:altLabel>
  <skos:altLabel xml:lang="fr">Ambérieux-en-Dombes</skos:altLabel>
  <rdf:type rdf:resource="http://data.ign.fr/de/geoFla#Commune"/>
  <rico:hasPlaceType rdf:resource="placeType/commune"/>
  <rico:identifier>Ancien identifiant SIA : FRAN_RI_005.xml#d3ntcnghfn-zgpbme3jri6v</rico:identifier>
  <owl:sameAs rdf:resource="http://id.insee.fr/geo/commune/01005"/>
  <owl:sameAs rdf:resource="http://data.ign.fr/id/geoFla/commune/01005"/>
  <insee-geo:codeCommune>01005</insee-geo:codeCommune>
  <rdf:type rdf:resource="http://rd.f.insee.fr/de/geo#Commune"/>
  <owl:sameAs rdf:resource="http://data.alegoria-project.fr/id/ign/unitesadministratives/commune/01005"/>
  <geoFla:dpt rdf:resource="http://data.alegoria-project.fr/id/ign/unitesadministratives/departement/01"/>
  <geoFla:dpt rdf:resource="http://data.ign.fr/id/geoFla/departement/01"/>
  <geoFla:dpt rdf:resource="http://id.insee.fr/geo/departement/01"/>
  <rico:containedBy rdf:resource="place/FRAN_RI_005-d3ntb70dnh-11f59tm8zrvbw"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcpcsqf-134n6r9bab3u"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcpcv92-1jvkvye54q3m"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcpev3-q3jo8lqulcig"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcc77jdn--n3n8zrznu57"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcqvg1g-Laons5oaam3Eh"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcc24b5-rm3up223m"/>
  <rico:isAdjacentTo rdf:resource="place/FRAN_RI_005-d3ntcqdodu-16m3yd49lc9a5"/>
  <rico:hasPhysicalLocation>
    <rico:PhysicalLocation rdf:about="physicalLocation/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v-2020">
      <rico:isPhysicalLocationOf rdf:resource="place/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v"/>
      <rico:hasCoordinates>
        <rico:Coordinates rdf:about="coordinates/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v-202003-RGF93">
          <rico:isCoordinatesOf rdf:resource="physicalLocation/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v-2020"/>
          <rico:measure rdf:datatype="http://www.opengis.net/ont/geosparql#wktLiteral">MULTIPOLYGON(((850413.2 6544061,850356.6 6544025.5,850208.4 654401
6544320,849593.9 6544265.5,849595.3 6544232,849593.8 6544211.2,849548.9 6544180.9,849294.3 6544107.1,849405.4 6544053.8,849247.4 6544051.4,849245.4 6544044.4,84921
65444509.6,846440.9 6544500.6,846264 6544430.7,846297.5 6544427.9,845993.6 6544413.8,845817.6 6544351.1,845583 6544262.6,845617.6 6544713.9,845628.4 6544796.3,84556
6545164,845389.5 6545495.6,845439.2 6545452,845411.2 6545711.5,845419.7 6545727.4,845516.5 6545780.9,845552.3 6545944.9,845615.8 6545975.1,845612.4,84558
6546486.7,846031.9 6546622.3,846007.9 6546621.5,846016.8 6546709.8,846060.9 6546703,846059.5 6546730,846059.5 6546829.1,846184 6546964.8,846188.6 6546927,846059 6546948.4,846069.1
6547709.7,845954.4 6547926.2,845947 6547998,846057.5 6548050.6 6486074.5 6548051.6,846126.4 6548022.2,8462329.5 6547980.1,846332.5 6547904.7,846441.7 6547912,846512
6548134.2,847036.2 6548138.5,847094.1 6548121.7,847139.5 6548092.8,847221.1 6548203.4,847215 6548342.2,847328 6548320.7,847365.7 65487347.4,847368.8 6548346.8,84748
6547962.2,847815.2 6547936.7,847785.1 6547880.4,847716.2 6547859.1,847716.2 6547841.4,847923.9 6547794.4,847927.2 6547894.6,848064.4 6547897.4,848155.5 6547854.3,8
6547177.1,848271.2 6547166.2,848281.3 6547160.1,848310.2 6547152.2,848267.2 6547154.2,848426.4 6547115.7,848472.8 6547064.9,848534.9 6547054.8,848574 6547059.9,848
6547081,848263.6 6547096.8,848201.3 6547106.6,848268 6547098.6,8484916.8 6547107,848492.8 6547035.9,848432 6546976.5,848442.8 6546930,848522.9 6546761.3,8495970.9 65
6546372.1,850697.4 6546090.6,850664.8 6545524.1,850726.4 6545839.4,850775.3 6545823.8,850769.3 6545725.2,850638.3 6545399.1,850607.1 6545367.9,850721.5 6545177.3,8
6544250.5,850416.3 6544339.1,850421.9 6544260.2,850381.9 6544154.8,850387.4 6544120.7,850413.2 6544061.1))</rico:measure>
        <rico:geodesicSystem>RGF93 (projection Lambert 93) [http://data.ign.fr/id/ign/fr/crs/RGF93LAMB93]</rico:geodesicSystem>
        </rico:Coordinates>
      </rico:isPhysicalLocationOf>
      <rico:hasCoordinates>
        <rico:PhysicalLocation>
          <rico:PhysicalLocation>
            <rico:hasPhysicalLocationOf>
              <rico:Coordinates>
                <rico:isCoordinatesOf rdf:resource="physicalLocation/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v-2018pts-RGF93">
                  <rico:measure>POINT(847200 6546000)</rico:measure>
                  <rico:geodesicSystem>RGF93 (projection Lambert 93) [http://data.ign.fr/id/ign/fr/crs/RGF93LAMB93]</rico:geodesicSystem>
                </rico:Coordinates>
              </rico:isCoordinatesOf>
            </rico:PhysicalLocation>
          </rico:PhysicalLocation>
        </rico:hasPhysicalLocationOf>
      </rico:hasPhysicalLocation>
    </rico:PhysicalLocation>
  </rico:isPhysicalLocationOf>
  <rico:hasCoordinates>
    <rico:Coordinates rdf:about="coordinates/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v-2018pts-RGF93">
      <rico:isCoordinatesOf rdf:resource="physicalLocation/FRAN_RI_005-d3ntcnghfn-zgpbme3jri6v-2018pts"/>
      <rico:measure>POINT(847200 6546000)</rico:measure>
      <rico:geodesicSystem>RGF93 (projection Lambert 93) [http://data.ign.fr/id/ign/fr/crs/RGF93LAMB93]</rico:geodesicSystem>
    </rico:Coordinates>
  </rico:isPhysicalLocationOf>
  <rico:hasPhysicalLocation>
  </rico:PhysicalLocation>
</rico:Place>
```

# Overview of the graph generated



## Auvergne-Rhône-Alpes (2016-....; région administrative)

Auvergne-Rhône-Alpes (2016-....; région administrative)<sup>fr</sup> • Auvergne-Rhône-Alpes (2016-....; région administrative)<sup>fr</sup>

Types:

- geofla:Region
- igeo:PaysOuTerritoire
- igeo:Region
- irdfs:Resource
- rico:Place

RDF rank:

0

Search instance properties

geofla:codeReg  
**84**

igeo:codeRegion  
**84**

rico:beginningDate  
**2016**

rico:history  
En 2016, la région Auvergne-Rhône-Alpes est créée suite au regroupement des régions Auvergne et Rhône-Alpes. La région Auvergne-Rhône-Alpes conserve sa dénomination. Le chef-lieu de la région Auvergne-Rhône-Alpes est fixé à Lyon.

rico:identifier  
Ancien identifiant SIA : FRAN\_RI\_005.xml#d-5f0vj727s--1abiwn0419rn

# Working on the authority data: first results and outcomes

- **Some of the RDF datasets built are already used by other projects**  
Example: the RDF sets describing places, and RDF sets created from a few finding aids, are being used by the [ALEGORIA](#) research project (web platform to be released in 2021).
- **A lot of work still to be done, and need to collaborate with other projects and institutions**
- **Though all this is done out of the current information system, we now have a standardized version of these authority data, and this work enabled us to begin to define a new model for their future version within the information system**
- **Need of new vocabularies > these too will have to be integrated later on (need of more flexibility)**
- **We plan to publish the datasets, in RDF and, when appropriate, other formats (TSV) as open data, at least in a Git repo (waiting for a triplestore and web interface)**

Two main directions for now:

- as already said, working on the future data models (as already said) and functionalities (with a special care, not only to the authority data model, but also to the evolution of the EAD and EAC-CPF formats)

Building a model for the ANF, selecting what we need in RiC and RiC-O, and extending it if necessary. Need to update the model through time.

- working on research and visualisation tools for the RDF sets



# Querying RDF knowledge graphs more easily

Two barriers have to be removed:

- an end user will most often know nothing of SPARQL language
- an end user will most often know nothing of the domain model (here, RiC-O)

Project aiming to improve Sparnatural, a visual SPARQL query builder  
(<https://github.com/sparna-git/Sparnatural>)

Four partners: the ANF, the BnF, the ministry of Culture and Sparna private company  
Should begin soon and last about one year.

A demonstrator (SPARQL endpoint + web interface integrating Sparnatural) that should enable to query a significant amount of the ANF RDF datasets, should be put online at the end of the project.

The project will include representatives of end users (through workshops) (tests and feedback).

# Sparnatural: to build a SPARQL query on RiC-O datasets



```
1 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
2 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
3 PREFIX rico: <https://www.ica.org/standards/RiC/ontology#>
4 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
5 SELECT DISTINCT (STR(?label) AS ?nom) ?link ?graphdb WHERE {
6   ?this <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
7     <https://www.ica.org/standards/RiC/ontology#RecordResource>;
8     <https://www.ica.org/standards/RiC/ontology#hasProvenance>
9     ?CorporateBody1.
10    ?CorporateBody1 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
11    <https://www.ica.org/standards/RiC/ontology#CorporateBody>.
12    ?CorporateBody1 <https://www.ica.org/standards/RiC/ontology#beginningDate>
13    ?Time2.
14    FILTER((?Time2 >= "1815-01-01"^^xsd:date) && (?Time2 <=
15    "2020-12-31"^^xsd:date))
16    ?CorporateBody1
17    (^rico:thingIsSourceOfRelationToType|rico:relationToTypeHasTarget) |rico:hasI
18    egalStatus <http://data.culture.fr/thesaurus/resource/ark://67717/63ff6870-
19    c6dd-44e9-8774-c9c4faa9a35c>.
```

Response Table Pivot Table Geo Google Chart Timeline </>

Showing 1 to 13 of 13 entries

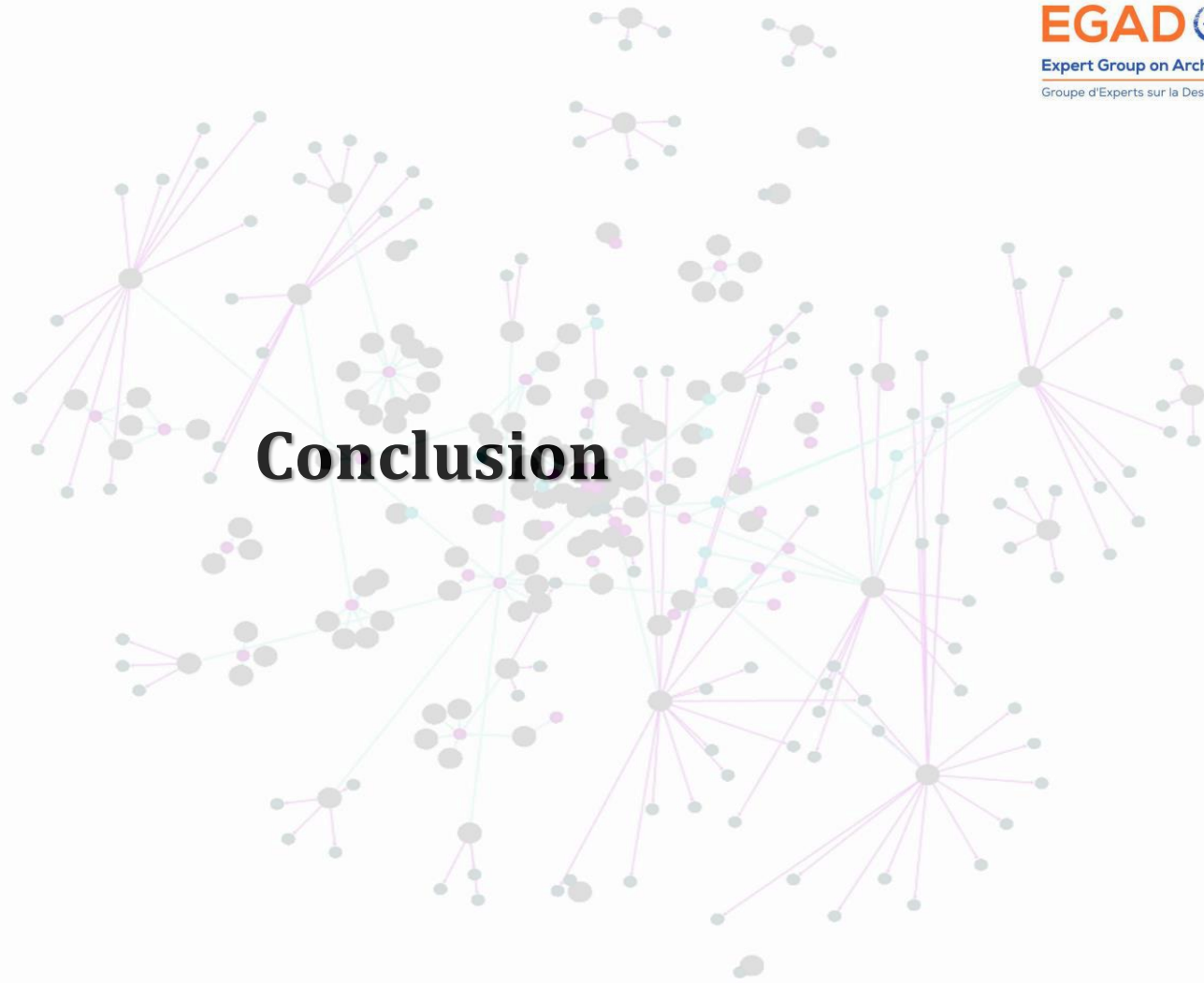
Search:  Show 50 entries

nom	link	graphdb
1 Administration générale de la BPI		<a href="http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F054094-top">http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F054094-top</a>
2 Archives de l'École nationale supérieure des bibliothèques (1931-1970)		<a href="http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F023643short-top">http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F023643short-top</a>
3 Archives de la bibliothèque publique d'information (1970-2012)		<a href="http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F054335-top">http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F054335-top</a>
4 Archives du département des bibliothèques et de la lecture de la direction du livre et de la lecture		<a href="http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F015882short-top">http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F015882short-top</a>
5 Bibliothèque publique d'information, expositions : diapositives et négatifs		<a href="http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F054639-top">http://localhost:7200/resource?uri=http%3A%2F%2Fdata.archives-nationales.culture.gouv.fr%2FRecordResource%2F054639-top</a>

Defining and applying a **global, institutional, strategy for improving the quality of the metadata** :

- principles and objectives (both realistic and ambitious, aiming to address the needs and expectations of the end users)
- going further and assessing the existing files more accurately
- defining and settling the methods and resources needed to reach the objectives, including human resources and tools

In 2021 : project to develop, with Inria, a Named entities recognition tool (using machine learning methods and external datasets e.g. Wikidata), for processing sets of EAD finding aids, indexing them and enriching the existing authority data (**NER4Archives project**)



# Conclusion

- A lot of work done, from first qualitative experiments to a far larger scale, whose results are encouraging and stimulating.  
These results also include some tools and methods.
- **The ANF projects also have been, and are, an opportunity and a way to test and enhance the reference standard**, which is still a draft
- **A lot of work still to be done, a lot of questions to be answered**
- A lot of tasks can only be done through **collaborative programs**, without losing our domain perspective  
Also need to ensure a technological watch
- Moving to a graph of linked entities is far from being only a technical issue; it fits within a **global issue: enhancing the quality and usability of archival metadata**  
**End users must be involved.**

- **Time to work on the information system itself**

- Need to include this work in a **global trajectory**

Strategic plan for 2021-2015 being designed just now.

A scenario that would result in building a semantic layer/module (with a web interface and APIs for machines), among other services for end users, is to be investigated and assessed.

A lot of questions, among which: what about the current SIV? How would this layer be articulated with it? And with the projects of French libraries (e.g. the FNE)?

## Let us mention that...

- Their position of a national operator gives the ANF an important responsibility within the current ecosystem of French archival institutions.
- Other projects have started; let us quote the decision of the French Archival portal (<https://francearchives.fr/>) to move to RiC-O, and projects in Switzerland and in the USA.

So, in a few words:

**a step by step, both ambitious, cautious and pragmatic approach**

Clavaud (Florence), Transformer les métadonnées des Archives nationales en graphe de données : enjeux et premières réalisations, in *Les Archives nationales, une refondation pour le XXIe siècle, La Gazette des Archives*, n°254 (2019-2), Association des Archivistes Français, Paris, 2019, p. 59-88.

See also (in French): the slides used during the study day that took place a year ago, at the ANF, on « **Les métadonnées archivistiques en transition : le nouveau cadre normatif, les enjeux et les premières réalisations**» (<https://labarchiv.hypotheses.org/1495>);

On the RiC-O website (in English):

- Why use RiC-O: <https://ica-egad.github.io/RiC-O/why-use-RiC-O.html>
- Projects and tools: <https://ica-egad.github.io/RiC-O/projects-and-tools.html>

Workshops and conferences to be held on RiC and its uses in 2021 (hopefully)





**Merci de votre attention !**  
**Thank you for your attention!**

Questions?

Email: [florence.clavaud@culture.gouv.fr](mailto:florence.clavaud@culture.gouv.fr)