

# ICA ROMA 2022



September 19<sup>th</sup>-23<sup>rd</sup>, Italy



From Classic Finding  
Aids to  
RiC-O Knowledge  
Graphs: Providing  
New Perspectives For  
End Users

## ARCHIVES: BRIDGING THE GAP

9<sup>th</sup> Annual Conference of the  
International Council on Archives

19-23 September 2022 | Rome, Italy

Florence Clavaud and  
Pauline Charbonnier,  
Lab of the Archives  
nationales de France (ANF)

# Outline

The Context: Research & Development projects on applying RiC in the ANF

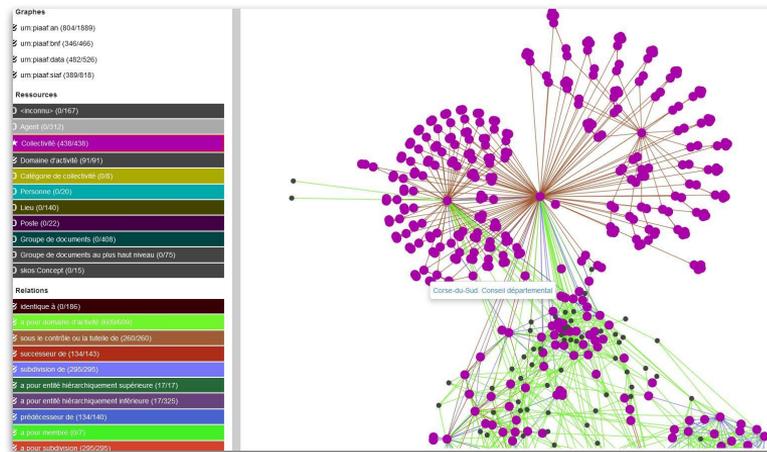
The 'ANF Sparnatural demonstrator' project:

- objectives
- the dataset produced: scope and content
- the search prototype developed

First assessment and prospects

# Research & Development projects on applying RiC in the ANF

- The ANF: a **huge collection of archival material, and of metadata**, in several silos
- Main formats used: XML/EAD 2002 (~ 31000 finding aids), EAC-CPF (~ 15400 authority records), and a local XML DTD for 20 authority lists and vocabularies
  - A lot of relations between all these files already, thus a dense oriented graph already; but not visible, not browsable, not searchable, in the current information system
- **An active participation to ICA EGAD work since 2013**
  - From 2017, an early version of RiC ontology available; now [RiC-O v0.2](#), a mature, rich and consistent model for producing high quality RDF archival metadata sets
- **First project: [PIAAF](#) (2015-2018): a qualitative RDF/RiC-O proof of concept, with a first data visualisation interface**
  - Very encouraging results (opened new paths), but for a very limited quantity of archival metadata



A screenshot of PIAAF data visualisation interface, from this page: <https://piaaf.demo.logilab.fr/ric/CorporateBody>

# Research & Development projects on applying RiC in the ANF: moving forward to a far larger scale

- Need of reliable, efficient, configurable tools for converting existing metadata
  - [RiC-O Converter open source software](#) (version 1.0 released in March 2020)
  - and a series of specific scripts to convert the vocabularies; the [resulting RDF files are now on GitHub](#)
- We thus have converted the whole of our XML files to RDF/RiC-O
- Next: need to **provide end users with tools for querying this RDF/RiC-O dataset**  
And to find answers to the following questions:
  - How to value a knowledge graph as a knowledge graph?  
(*and not behind a classic search interface, e.g. a faceted search engine or clickable navigation links*)
  - How to promote exploratory user interactions (test/error) to discover the data?
  - How to *reconcile a generic, complex, conceptual model and ontology like RiC with the specific points of view of the end users?*

# The 'ANF Sparnatural demonstrator' project

- Focuses on the development of an **efficient and user-friendly search interface**, to enable the general public to query this large scale RDF/RiC-O graph:
  - without knowing a word of the SPARQL query language
  - without knowing the underlying ontology and abstract model used
- Uses the [Sparnatural visual SPARQL query editor](#), an open source tool whose first version was developed in 2018 by Sparna, a private company
- Is a part of the project, funded by the French ministry of Culture, the French National Library and the ANF, that also aimed to improve the functionalities, user experience and configuration possibilities of Sparnatural
  - The project team also included Sparna. The demonstrator (and another one built by the BnF) allowed to test this new version of Sparnatural.

End-User interface for your SPARQL endpoint

- Adapt to your ontology**  
Sparnatural can be configured to adapt to your graph structure. You can hide some of the complexity of your graph, and simplify the navigation for your users. Configuration is made through an **OWL ontology**, fully documented.
- Client-only (no additional server required)**  
Sparnatural is pure client javascript and can be **plugged to any existing SPARQL endpoint**, without additional server required.
- Free and open-source**  
Sparnatural license is **LGPL-3.0**, which means you are free to integrate it in your website. Any modification you make *inside the source code* of Sparnatural needs to be distributed with the same license, or contributed to the original project.
- Open-Source on Github**  
Sparnatural development is open and happens on the **GitHub repository**.
- Fully customizable**  
Sparnatural offers different **widgets for value selection** : dropdown lists, ordered by occurrence count or alphabetically, autocomplete search fields, date pickers, tree widgets... this is all part of Sparnatural configuration.
- Fully documented**  
Sparnatural comes with complete **documentation**.

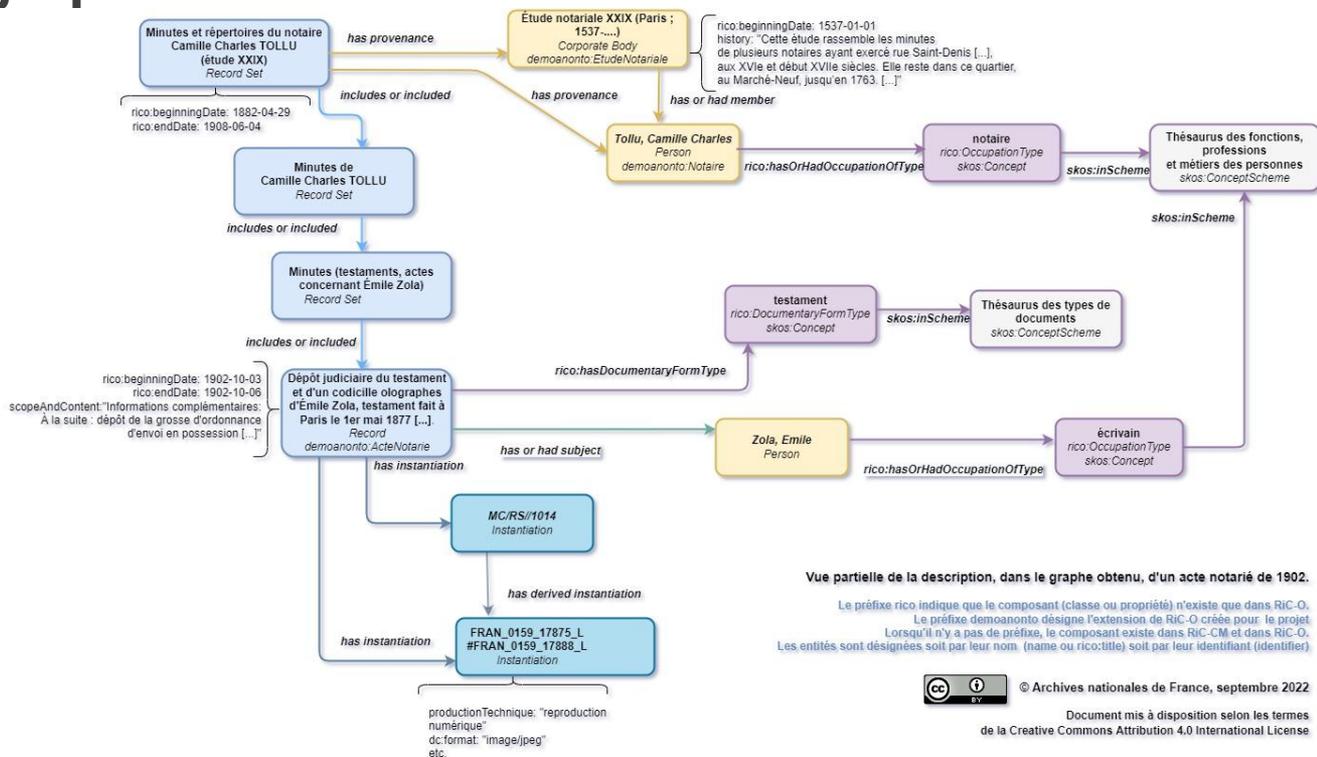
Sparnatural features  
(<https://sparnatural.eu/#features>)

# The dataset prepared

- A **subset of the whole of the ANF metadata had to be selected**
- Decision to work on the **metadata describing the archives of Parisian notaries produced from the 15<sup>th</sup> to the 20<sup>th</sup> century, collected and kept by the ANF**
  - i.e. the records created by 40 out of the 122 Parisian notarial offices (i.e. about **9 linear km out of 26 km of notarial records, out of 375 linear km of analogue records held by the ANF**)
  - a consistent and rich whole
  - documents widely consulted by various audiences
- - **1577 XML/EAD 2002 finding aids**
- - **1120 XML/EAC-CPF authority records** about the notaries and notarial offices which created the documents
- - **Vocabularies and small records on persons and corporate bodies, used to index these metadata**
- A few enrichments made in the EAC-CPF and vocabularies
- Source files converted to RDF/RiC-O using:
  - RiC-O Converter
  - + specific scripts, in particular to add some information to the RDF result (e.g. categorization of many record resources as being registers, notarial acts, lists or mentions of acts)
- The whole was imported in a RDF base (triplestore), then enriched again using SPARQL INSERT operations
  - An extension of RiC-O was thus created, that defines a few classes, subclasses of RiC-O Record Resource, Record, Record Part, Person, Corporate Body and Place

**> The resulting graph includes about 58 million triples for now**

# A diagram showing the description of a notarial act in the graph obtained



In the demonstrator, see the following page where the same metadata are displayed in text form:  
<https://sparnatural-demo-anf.huma-num.fr/lodview/recordResource/041835-c1p6y1e0w26r--183v6ez58hjun.html>

# Developing the demonstrator

- Two search interfaces were then developed by the ANF, with the help of Sparna, using Sparnatural
  - each one configured by a OWL file, therefore an **ontology that defines the entities and relations that the end user can query, and maps them to RiC-O ontology**
    - > compromise between the complexity of the model and the need to produce an understandable interface
  - These interfaces integrate the new functionalities added to Sparnatural, and the team project could use them to test these new functionalities
  - **The interfaces were tested during two workshops, by 20 persons of various backgrounds** (archivists, researchers, genealogists, DH engineers...) and **their feedback taken into account**



The screenshot shows the website for the 'Démonstrateur Sparnatural des Archives nationales'. The header is blue with navigation links: Accueil, Démonstrateur A, Démonstrateur B, Documentation, Mentions légales, and Sparnatural. The main title is 'Démonstrateur Sparnatural des Archives nationales' with a language selector set to 'Français'. The text explains that the demonstrator is built with Sparnatural and allows querying a knowledge graph of Paris notary archives. It mentions two interfaces: 'démonstrateur A' (simple and general) and 'démonstrateur B' (specific to notarial archives). The date of the last update is 25 août 2022. An English version is also available, with a similar explanation. The footer includes logos for the 'MINISTÈRE DE LA CULTURE' and 'SPARNA'.

The demonstrator was released in June 2022:

<https://sparna-git.github.io/sparnatural-demonstrateur-an/>

# How it works: the first search interface (interface A)

The screenshot displays a query builder interface with the following components:

- Top Row:** A sequence of criteria: "Records" (with a folder icon), "have documentary form type", "Documentary form type", and "Inventaire après décès (...)" (with a document icon). A refresh button is on the right.
- Logic:** An "And" connector links the top row to the second row.
- Second Row:** "Records" (with a folder icon), "have subject", and "Person" (with a person icon).
- Logic:** A "Where" connector links the second row to the third row.
- Third Row:** "Person" (with a person icon), "has occupation", "Occupation" (with a briefcase icon), and "orfèvre" (with a plus sign).
- Logic:** An "And" connector links the third row to the fourth row.
- Fourth Row:** "Records" (with a folder icon), "creation date", "Date" (with a calendar icon), and "de 1600 à 1700".
- Bottom Panel:** A search bar with a "this" dropdown, a "Sort" section with "A|Z" and "Z|A" options, and a "Show column names" toggle switch.

**Simple and generalistic.** It can be applied to **almost any RiC-O compliant archival dataset**, insofar as contextual entities such as archival creators, document subject agents, locations are indexed.

It is a priori rather intended for **the discovery of metadata.**

<https://sparna-git.github.io/sparnatural-demonstrateur-an/index-A.html?lang=en>

# How it works: the second search interface (interface B)

The screenshot displays a query builder interface with three main clauses:

- Clause 1:** A document of type "Documentary form type" is of type "Notarial act".
- Clause 2:** A "Notarial act" is created on a "Date" between "01/01/1848" and "31/12/1...".
- Clause 3:** A "Notarial act" has a "Reference code" of "Any".

The clauses are connected by "Where" and "And" operators. A play button is visible at the bottom right of the query builder.

Below the query builder, there are links for "Toggle SPARQL editor", "Share", and "Export".

The search results are displayed in a table view, showing 679 results in 24.343 seconds. The table has columns for "this", "Date\_2\_begin", "Date\_2\_end", "Date\_2\_exact", "ActeNotarie\_1", and "Cote\_3".

this	Date_2_begin	Date_2_end	Date_2_exact	ActeNotarie_1	Cote_3
1 accord	"1843-07-19"	"1851-07-25"		Obligation par les époux Louis Prosper Lalande, compositeur d'imprimerie demeurant au 5 rue d...	MC/ET/VIII...
2 accord	"1848-05-03"	"1861-03-08"		Procès verbal d'ouverture de la liquidation de Charles François Guillaume marquis de Chanaleille...	MC/ET/VIII...
3 accord	"1848-09-21"	"1848-10-19"		Procès-verbal d'ouverture de la liquidation des reprises d'Agathe Elisa Gougibus, en son vivant po...	MC/ET/VIII...
4 accord	"1847-09-20"	"1848-04-24"		Contrat de mariage entre Pierre Tharaud, artiste musicien demeurant au 21 rue de Montholon, a...	MC/ET/VIII...

Exploits the specificities of notarial archives.

It is a priori rather adapted to the needs of a person familiar with this universe, archivist or accustomed user.

<https://sparna-git.github.io/sparnatural-demonstrateur-an/index-B.html?lang=en>

# First assessment

- We could prove that a large-scale conversion of XML/EAD and XML/EAC-CPF source data to RDF compliant with RiC-O is possible
  - Such an operation **will only use a subset of the RiC-O classes and properties**
  - Carrying out such a project may encourage, in particular in the case of a homogeneous archival corpus such as ours, to extend RiC-O very generic ontology
- **Very positive results as concerns the new possibilities for exploring the graph and the new research opportunities.**
  - Among the main points identified by users during the workshops, one can mention **the intuitive, flexible and interactive nature of this type of interface**
  - Such a query interface should be integrated into the future the ANF web application that enables to search, browse and display the whole RDF graph of our metadata.
- **The Lab of the ANF is now able on its own to design a Sparnatural search interface for other data sets and projects.**

- **We now also better know the content of the metadata** (statistical elements obtained) and **also can have a new perspective on them and the quality issues we have to solve**

- 72,665 groups of records (fonds, series, bundles, clients files, etc.);
- **400,570 notarial acts** (described to date) within the bundles;
- **104,383 mentions of acts** recorded in the registers;
- 3135 registers;
- **13,127 lists of acts within the registers**, including 11,670 digitized lists;
- 629,246 persons (including 3,213 persons with an IRI);
- 530 organizations (including 528 with an IRI);
- 15,559 places (including 15,463 with an IRI).

# As a conclusion: prospects

- We will continue to work on the data and on the interfaces
  - The two interfaces may be merged into one **by hierarchizing classes and properties, to give to the user the choice, within a single search interface, of a quick or more precise exploration** (e.g. select Person > Notary in a single interface).
  - This will involve changing the configuration possibilities offered by Sparnatural.
- The Lab of the ANF plans to produce and publish a generic Sparnatural search ontology for RiC-O compliant RDF data.
- Finally, we now want to work on the visualisation of such a huge knowledge graph.

# Thank you!

- Useful links:
  - the demonstrator: <https://sparna-git.github.io/sparnatural-demonstrateur-an/>
  - its documentation in English: <https://sparna-git.github.io/sparnatural-demonstrateur-an/presentation-en.html>
  - its public repository on GitHub (that includes the two configuration files of the interfaces): <https://github.com/sparna-git/sparnatural-demonstrateur-an>
- See also: other presentations and video recordings about Sparnatural and the whole project (mainly in French): <https://sparnatural.eu/#bibliography>

**For any question or remark, or if you would like to get a dump of the RDF dataset itself, do not hesitate to contact us.**

You can send an email to [florence.clavaud@culture.gouv.fr](mailto:florence.clavaud@culture.gouv.fr) or [le-lab.archives-nationales@culture.gouv.fr](mailto:le-lab.archives-nationales@culture.gouv.fr)