

# RiC aux Archives nationales de France : enjeux, réalisations, perspectives

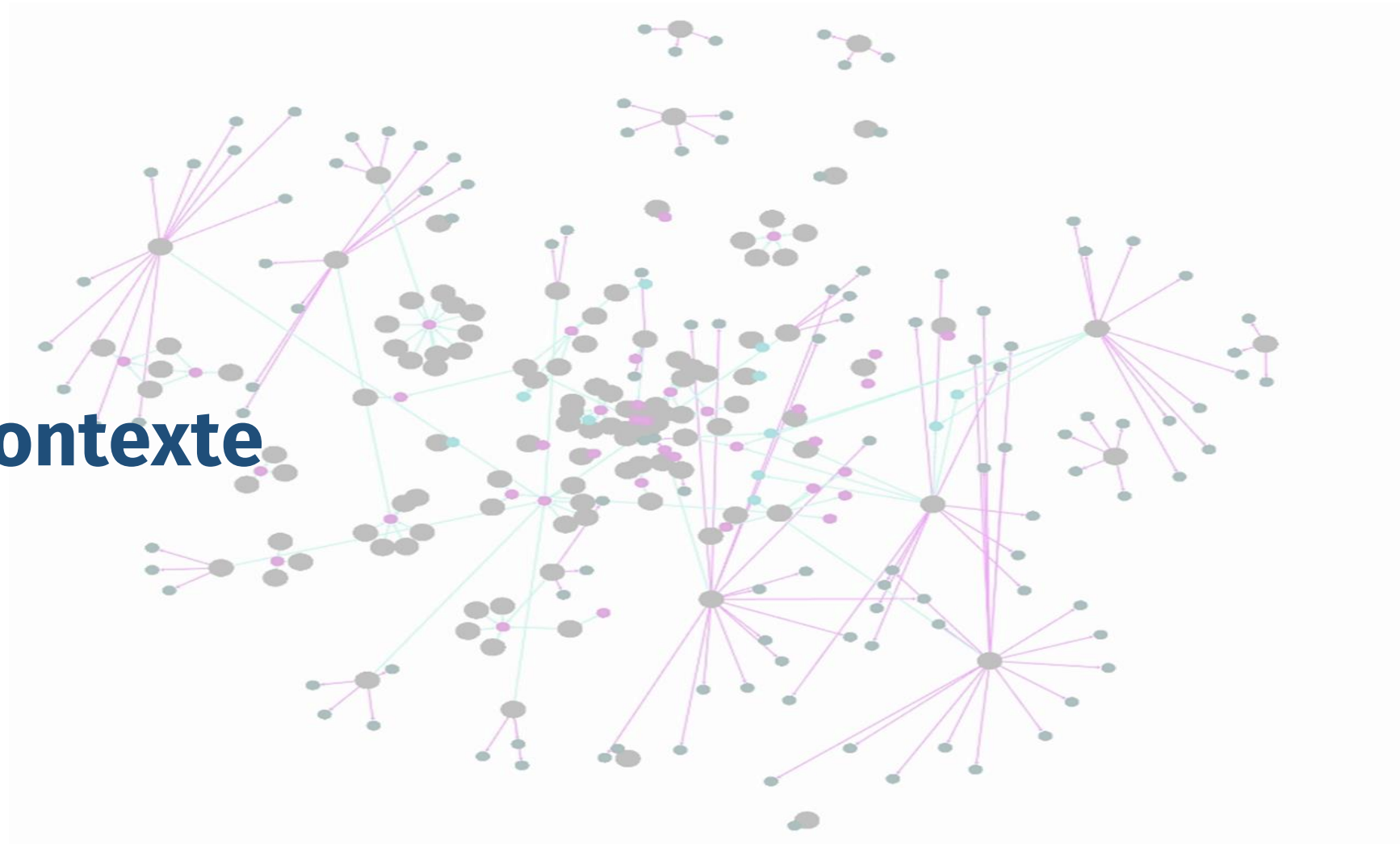
Florence Clavaud ([florence.clavaud@culture.gouv.fr](mailto:florence.clavaud@culture.gouv.fr))

Responsable du Lab, Archives nationales de France  
membre du groupe ICA/EGAD, responsable du développement de RiC-O

*RiC (Records in Contexts) : quels changements et quelles perspectives ?*  
Demi-journée de formation, Lausanne, 8 décembre 2022

- Éléments de contexte  
une situation à la fois classique et singulière
- Préparer avec RiC un futur proche : aperçu des premières réalisations
- Focus sur un projet en cours : les données liées au service des utilisateurs finaux
- En guise de conclusion : de nouvelles perspectives

# Éléments de contexte



# Les Archives nationales de France (ANF)

- Un service à compétence nationale créé en janvier 2007 par le ministère de la Culture (décret du 24 décembre 2006).
  - Mission : **collecter, traiter, décrire, préserver, communiquer, et valoriser les archives publiques des administrations centrales de l'Etat, des notaires de Paris, et des archives privées d'intérêt national.**
- En fait, une très vieille institution, née sous la Révolution française ; les archives qu'elles conservent datent du Moyen Âge à nos jours.
- Chiffres clés (2020) :
  - plus de 373 km linéaires d'archives papier, plus de 70 To d'archives nativement numériques
  - plus de 8,8 millions de documents numérisés accessibles en ligne
  - 8000 lecteurs différents, environ 115 000 articles communiqués en salle de lecture en 2019



- Un site web institutionnel (<https://www.archives-nationales.culture.gouv.fr/>), sur lequel on peut depuis 2013 rechercher et consulter les métadonnées archivistiques qui décrivent les collections conservées, et réserver les cartons utiles à sa recherche, en utilisant la **Salle des inventaires virtuelle (SIV)** (<https://www.siv.archives-nationales.culture.gouv.fr/>)

# Les métadonnées de description des archives aux ANF

- **Massives**  
Plus de 31000 instruments de recherche décrivant des millions d'items et d'agrégations, plus de 15400 notices, une vingtaine de vocabulaires, plusieurs bases de données textuelles de taille conséquente décrivant le plus souvent des séries ou groupes de dossiers
- **En constante évolution**
- **Hétérogènes**  
Différentes structures et formats  
Les 31000 instruments de recherche sont tous au même format (XML/EAD) mais cela cache une grande variabilité
- **De nombreux silos**  
D'où une **dispersion des données, des redondances, des difficultés à gérer et même à appréhender le tout, une sous-utilisation de certains réservoirs**
- **Plusieurs applications pour les usagers**
- **Pas d'accès distant pour les machines** (sauf partiellement via un entrepôt OAI-PMH), pas de service de connexion à des réservoirs distants  
Or **le besoin d'interopérabilité est fort, ne serait-ce que pour les ANF elles-mêmes et pour les entités formant son écosystème administratif**
- **Le cœur du SI est ancien (risques de l'obsolescence)**

# Faire évoluer ces métadonnées, une nécessité

- **Plusieurs défis à relever :**
  - la **rationalisation** (décloisonner, regrouper pour mieux gérer)
  - la **montée en qualité** (normaliser, harmoniser, enrichir)
  - l'**accessibilité (plus de points d'accès et des fonctionnalités de recherche et d'exploration plus intuitives pour les humains, des APIs pour les machines)**
- Atteindre le troisième objectif implique de travailler sur les deux premiers
- Prise de conscience à l'échelle de l'institution et par ses agents
- Démarche également entreprise par d'autres types d'institutions culturelles (cf. transition bibliographique en bibliothèque), et encouragée par le ministère de la Culture français : feuille de route données culturelles, stratégie sémantique, stratégie référentiels

# Comment faire ? passer à des graphes d'entités liées

- Détecter les entités de l'univers concerné, leur assigner des identifiants uniques
- Regrouper les données décrivant chacune de ces entités
- Relier les entités entre elles  
De fait **les interfaces et l'architecture informatique cachent un graphe sous-jacent** (voir ci-contre), qu'il faut gérer comme tel, consolider, enrichir et révéler
- Utiliser pour cela **un ou plusieurs modèles publics**
- Exposer les graphes obtenus sous la forme de graphes de connaissances / sémantiques, ou données ouvertes liées  
rendre du coup les données conformes aux principes FAIR

Pour les 15484 notices de producteurs au format EAC-CPF validées au 1er mars 2022	
Relations entre notices	Nombre de relations
Relations hiérarchiques	9833
Relations temporelles	4434
Relations associatives	6455
Relations familiales	529
<b>Total</b>	<b>21252</b>
<b>Relations d'identité (avec des notices externes)</b>	<b>2345</b>
<b>Relations avec des instruments de recherche</b>	<b>34021</b>

Le décompte est fait en considérant la relation comme une entité à part entière (on compte une seule fois la relation existant entre deux entités et sa relation inverse)

# Records in Contexts, une norme pour accompagner cette deuxième (r)évolution

- Le modèle conceptuel **RiC-CM, un cadre global de référence**
  - un **arsenal complet et organisé de composants, du plus générique au plus précis**
  - **permettant une utilisation flexible et une transition par étapes**
  - extensible
- L'ontologie **RiC-O, une transposition technique de RiC-CM pour produire et exploiter des graphes sémantiques / des jeux de données liées en RDF**
  - une syntaxe normalisée
  - un vocabulaire et des règles unifiées, maintenus par et pour la communauté des archivistes
  - conforme à RiC-CM
  - conçue pour être immédiatement utilisable

Pour commencer à les mettre en œuvre aux ANF :

- Avant de modifier le SI lui-même, **travailler nécessairement hors de ce SI**
- **Produire du concret, donc utiliser RiC-O**
- **Commencer petit (preuve de concept), puis passer à l'échelle**
- Préparer des **briques pour l'avenir**
- Faire des choix, apporter des **éléments de réflexion pour l'élaboration d'une stratégie pour le plus long terme**
- Monter en compétences
- **Partager les résultats**
- Et bien sûr, faire bénéficier le groupe EGAD et la norme RiC elle-même de ces travaux

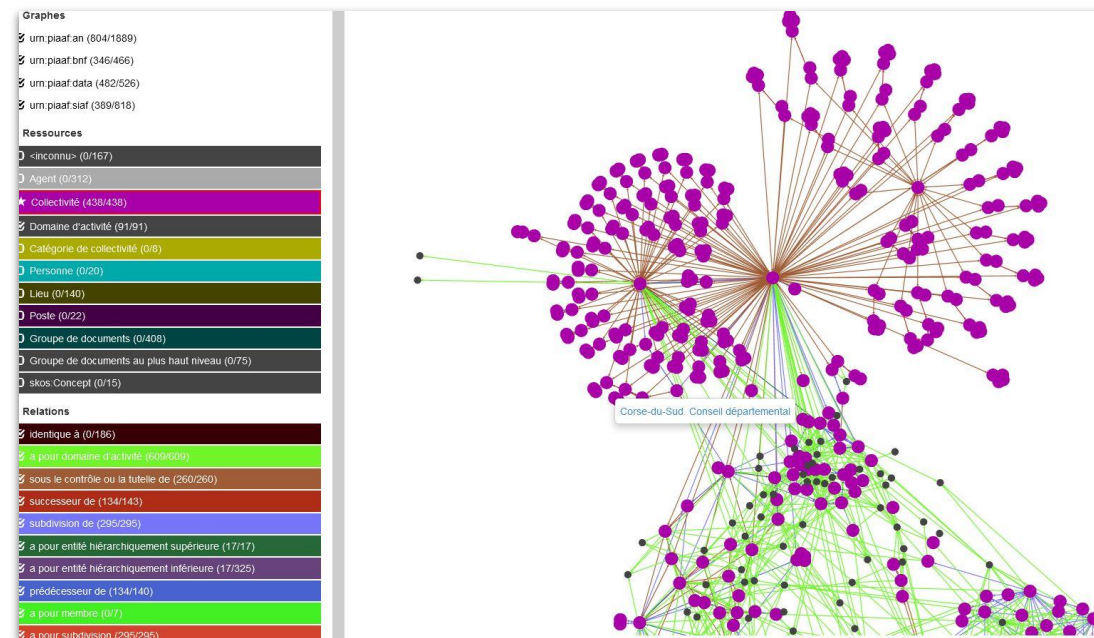


# Préparer avec RiC un futur proche : aperçu des premières réalisations



# La preuve de concept PIAAF (2014-2018)

- Un projet en partenariat avec la BnF et le ministère de la Culture
- Une **poc qualitative**, pour prouver la faisabilité et l'intérêt d'une sémantisation des métadonnées archivistiques existantes selon RiC-O, de l'interconnexion de jeux de données de provenances diverses, et travailler à la visualisation graphique des jeux de métadonnées obtenus
- Conversion en jeux de données RDF de l'ensemble des métadonnées fournies par les partenaires réalisée par les ANF, application web réalisée par la société Logilab



Copie d'écran de l'interface de data visualisation de PIAAF, graphe affiché à partir de la page : <https://piaaf.demo.logilab.fr/ric/CorporateBody>

## Résultats :

- **très encourageants (poc réussie !)**
- divers enseignements pour les partenaires et un certain retentissement notamment à l'étranger (une première mondiale de mise en œuvre de RiC)
- **PIAAF, laboratoire de test de RiC-O 0.1**

- Voir les pages de présentation des enjeux, de l'historique et de la réalisation du projet
- **Cependant, il ne s'agissait que d'une très petite quantité de métadonnées**
- Projet arrêté après la publication de l'application Web en février 2018

# Passer à l'échelle : réalisation d'un logiciel de conversion de fichiers EAD et EAC-CPF en données RDF/RiC-O (2019-2020)

- Besoin d'un logiciel **performant, fiable, facile à installer, configurable, documenté et open source**

- > **RiC-O Converter v. 1.0**

Développé par la société Sparna, avec une forte implication des ANF (notamment pour les **spécifications** et les **tests unitaires**)

Code source et documentation en anglais :

<https://github.com/ArchivesNationalesFR/rico-converter>

## Premiers résultats :

- Nous permet désormais en moins de 25 min de convertir en RDF/RiC-O 0.1 les 31000 fichiers EAD 2002 des ANF, et en moins de 15 min les 15400 notices EAC-CPF
- A déjà été adapté pour divers projets aux ANF
- Premiers **mappings, certes partiels, entre EAD 2002 et RiC-O 0.1, et EAC-CPF et RiC-O 0.1** (vont resservir)
- **Une meilleure connaissance du contenu des fichiers source**

## RiC-O Converter

Welcome to the **RiC-O converter documentation** ! RiC-O converter is a command-line tool to convert EAD and EAC files to RDF files expressed using [Records in Contexts ontology](#).

### Table of Contents

The documentation is divided in these sections :

### About RiC-CM, RiC-O

- [RiC-CM and RiC-O](#): start here to learn more about Records in Contexts - Conceptual Model and Ontology ;
- [About the RiC-O converter project](#): get an understanding of the RiC-O converter project;
- [Mappings](#): learn how EAD and EAC files are mapped to Records in Contexts Ontology by this conversion tool.

### Technical documentation of RiC-O converter

- [Getting Started](#): start here to test drive the converter, understand the directory structure, learn how to print help message and adjust the command parameters;
- [Conversion](#): main documentation for the EAD and EAC conversion process;
- [Running unit tests](#): explains how to run the unit tests of the converter;
- [Customize the conversion](#): explains how to customize the conversion for your own files;
- [Common Problems](#): give answers to common problems;
- [Compile ricaoconverter](#): if you need to adjust something in the Java code and recompile, read this.

### Reference lists of error codes

- [Error Codes](#): reference list of the error codes of the application;
- [EAC XSLT Error Codes](#): reference list of the possible error codes during EAC conversion;
- [EAD XSLT Error Codes](#): reference list of the possible error codes during EAD conversion.

### Licence

This project is licensed under the terms of the CeCILL-B license (equivalent to MIT license and close to a public domain license) (see the [license file in english or in french](#)).

Copie d'écran de la page d'accueil de la documentation du logiciel

Voir au sujet de ce logiciel :

- le [webinaire organisé en juin 2020](#)
- l'[article](#) publié dans les actes du workshop *Linked Archives 2021*

# Passer à l'échelle : sémantisation, début d'enrichissement, et publication des référentiels des ANF (2020-....)

- Les référentiels des ANF, ce sont :
  - près de 20000 notices d'agents
  - près de 60000 notices de lieux
  - une vingtaine de vocabulaires d'indexation
  - dans le SI, **divers formats, divers états**
  - **des données sous-utilisées**
- Un rôle clé pour l'avenir : ils décrivent des **entités contextuelles des archives conservées aux ANF**, qui sont aussi autant de **clés de liage avec des réservoirs externes**
- Projet au long cours, beaucoup de travail ; **deux facettes** :
  - **normalisation** (conversion en données conformes à SKOS et RiC-O 0.2) et **divers travaux techniques d'alignement**
  - **des chantiers scientifiques d'enrichissement menés avec les agents de la direction des Fonds**

## Premiers résultats :

- Un **dépôt public sur GitHub**, plus de données dans ce dépôt (dans deux formats : RDF/XML et CSV) que dans la version interne des référentiels
- Utilisation des technologies RDF pour enrichir ces données
- Jeux de données déjà utilisés dans chacun des projets LOD des ANF
- Une **dynamique interne** et l'inscription dans le [plan stratégique 2021-2025](#) des ANF

## Référentiels des Archives nationales de France

Les référentiels des Archives nationales de France | the Archives nationales de France authority data and vocabularies

### Téléchargement

Télécharger la dernière release (la version 1.1, juin 2022): <https://github.com/ArchivesNationalesFR/Referentiels/releases/tag/1.1>

### Description

Formats :

- XML/RDF pour l'ensemble des référentiels. Principaux modèles utilisés : SKOS et Records in Contexts - Ontology (RiC-O) (dans sa dernière version officielle en date, la v0.2 de février 2021)
- XML/EAC-CPF pour les notices du référentiel des producteurs
- CSV (encodage UTF-8; séparateur de champs : virgule ; valeurs des colonnes encadrées par des guillemets ("")) pour les concepts et les lieux. Ces fichiers CSV ont été produits à partir des versions RDF. Pour les agents, les fichiers CSV fournis sont des listes, qui ne contiennent pas toute la substance des notices RDF des agents.

En cours de construction | a work in progress...

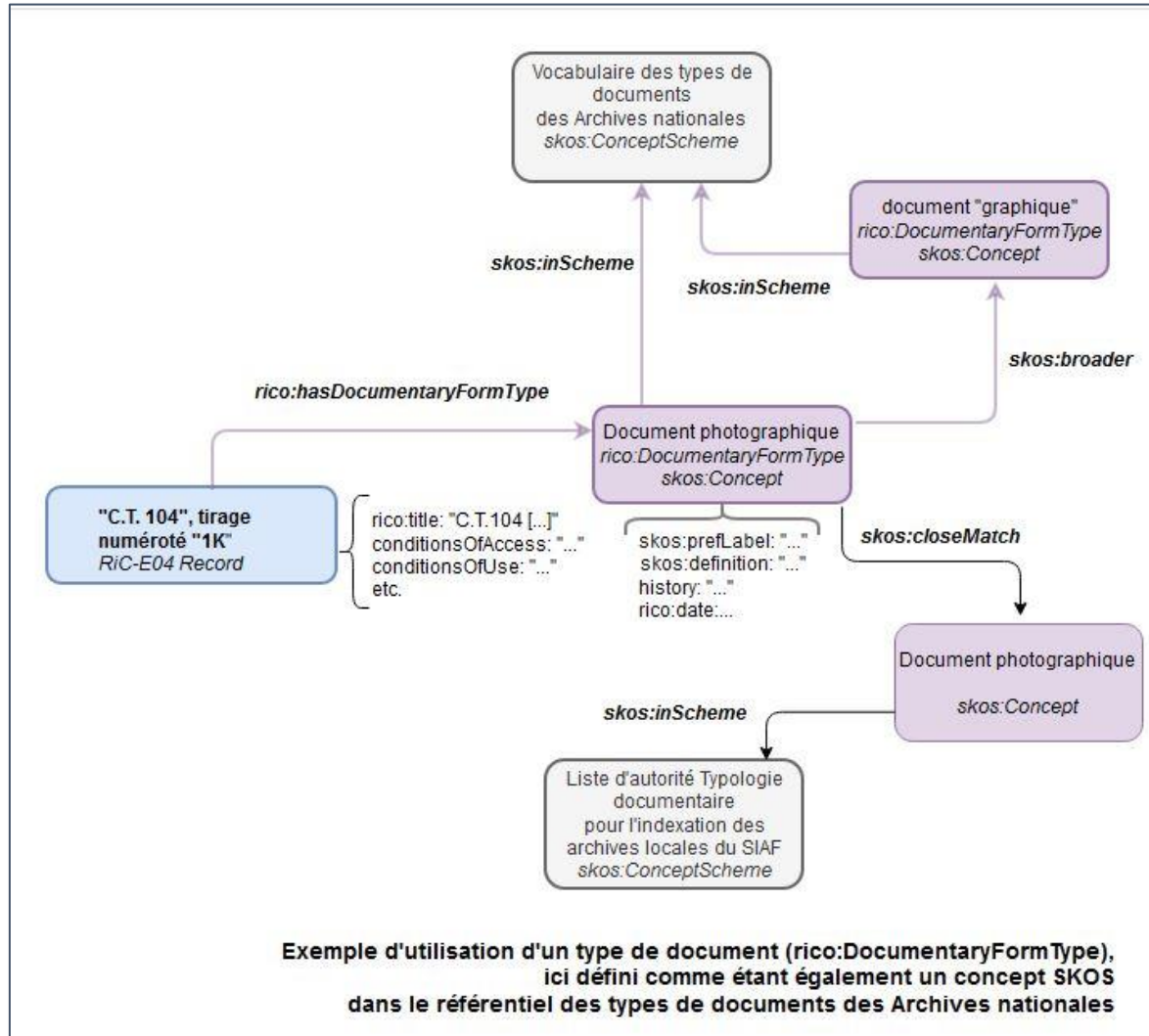
Licence | License : Ces métadonnées étant des informations publiques, l'utilisateur dispose d'un droit non exclusif et gratuit de libre « réutilisation » à des fins commerciales ou non, dans le monde entier et pour une durée illimitée. Il doit accompagner chaque rediffusion des informations de l'indication précise de l'origine des métadonnées : « Archives nationales (France) », date de ces métadonnées (mai 2022), nom du référentiel (fourni dans le fichier Excel qui en donne la liste). Voir à ce sujet la page : <https://www.archives-nationales.culture.gouv.fr/fr/web/guest/reutilisation-des-donnees-publiques>.

La liste fournie (fichier [Liste-referentiels.xlsx](#)) donne diverses informations sur chacun de ces référentiels.

### Avertissements

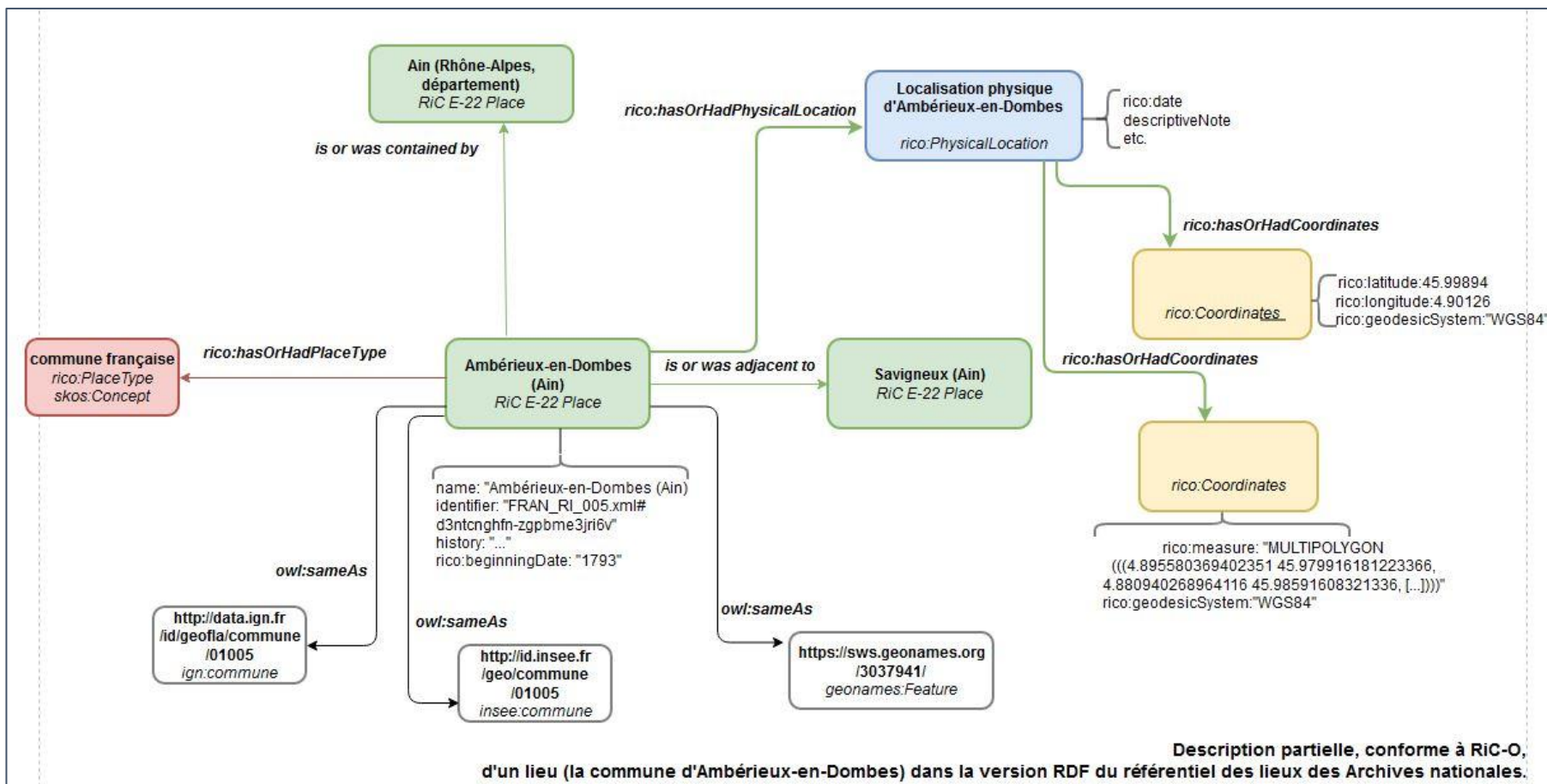
Les référentiels sont en construction, ce qui veut dire qu'ils sont en règle générale très incomplets : il y manque la description de nombreuses entités, et les descriptions fournies sont parfois pauvres. Ils sont cependant utilisés, dans un format propre au SI des Archives nationales, pour indexer la description des archives que l'institution conserve, ou pour

# Exemple d'entité dans les référentiels des ANF : une instance des classes RiC-O Documentary Form Type et SKOS Concept



- Le *Record* présent dans ce graphe (à gauche) est le même que celui décrit ici, dans le graphe de connaissances produit dans le cadre du projet de recherche ALEGORIA : <http://data.alegoria-project.fr/page/anf%2Fflapie%2Frecord%2F058220-c2vzyex71lzy-1rdwcqmm05btq>
- Et une des *Instantiations* numériques de ce *Record* est consultable [ici](#)
- En savoir plus les données RDF produites dans le cadre du projet :
  - <https://www.alegoria-project.fr/en/Metadata>
  - [Présentation](#) faite pendant le séminaire de clôture du projet il y a un an

# Exemple d'entité dans les référentiels des ANF : une instance de la classe RiC-O Place



Voir [les référentiels de lieux](#) dans le [dépôt GitHub](#) (fichier des communes du département de l'Ain)

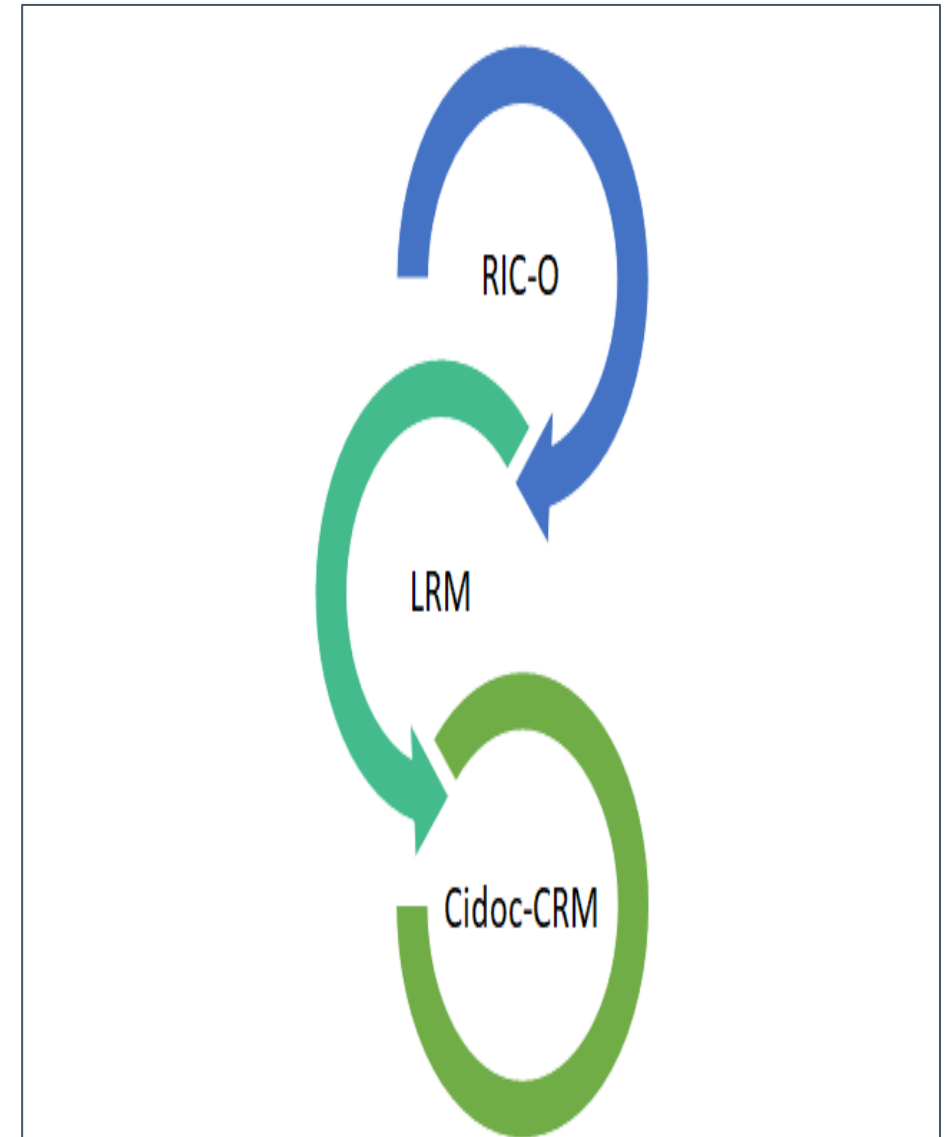
Dans ce diagramme, lorsqu'il n'y a pas de préfixe, cela signifie que le composant de RiC qui est utilisé (par exemple identifier) existe dans RiC-CM et dans RiC-O. Le préfixe `rico` indique un composant présent uniquement dans RiC-O (par exemple `rico:hasOrHadPlaceType`).

**Focus sur un projet en cours : les données liées au service des utilisateurs finaux**



# Construire une interface d'exploration et de recherche dans un graphe de connaissances RDF/RiC-O de taille significative

- Objectifs :
  - Valoriser le graphe en tant que graphe
  - En permettre la découverte, favoriser les interactions exploratoires
  - Montrer que la sémantisation ouvre de nouvelles possibilités de recherche aux utilisateurs
- Deux verrous à lever :
  - la plupart des utilisateurs finaux d'archives ne connaissent pas SPARQL
  - la plupart ne connaissent pas nos modèles métier (encore moins RiC-O !)
- Des enjeux communs à de nombreuses institutions, notamment culturelles
  - les graphes de connaissances qu'elles produisent sont vastes, décrivent une grande variété d'entités, et leurs contenus sont hétérogènes (*les modèles sont flexibles et peuvent être appliqués de différentes manières, les métadonnées culturelles peuvent être incomplètes, etc.*)





# L'idée : se servir de Sparnatural, après l'avoir amélioré

- <https://sparnatural.eu/>
- Un éditeur graphique de requêtes SPARQL / un outil d'exploration de graphes de connaissances
- Un composant Javascript **open source** (côté client uniquement)
- **Agnostique** (peut être configuré pour requêter n'importe quel jeu de données RDF)
- Première version développée en 2018 pour le projet OpenArchaeo porté par le consortium MASA

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**.

Projet porté par les ANF en partenariat avec le ministère de la Culture, la BnF et Sparna

Deux objectifs concrets :

- **développer une nouvelle version de Sparnatural**
- **utiliser cette version pour construire deux démonstrateurs, un pour la BnF et un pour les ANF**

# Le démonstrateur Sparnatural des ANF : d'abord, sélectionner et préparer un jeu de métadonnées

- Les métadonnées retenues, en concertation avec le Département du Minutier central des notaires de Paris (DMC), service des AN qui en est responsable : **celles qui décrivent les archives des 40 premières études sur les 122 études parisiennes**, soit :
  - 40 notices de collectivités décrivant les études (et une autre, celle de l'étude LIII) ;
  - 1079 notices décrivant les notaires qui ont exercé au sein de ces études ;
  - 1577 instruments de recherche archivistiques (32 % des 4902 instruments de recherche existants au Département du Minutier central des notaires de Paris - DMC - et validés en mars 2022)
  - et les référentiels des ANF utilisés pour décrire ces producteurs ou indexer les instruments de recherche (notices des personnes et organismes sujets des documents ; vocabulaires des types de documents, mots-matières, activités, lieux de Paris).

Ce corpus correspond au **tiers des fichiers de métadonnées décrivant les archives notariales aux Archives nationales** (qui forment un ensemble de plus de **26 km linéaires, 194 500 cartons**, dans lesquels on estime que sont conservés **20 millions d'actes** ; voir la [fiche d'aide à la recherche sur les chantiers d'inventaires au DMC](#)).

- Pourquoi ce choix ?
  - Des métadonnées **cohérentes**, des descriptions souvent riches
  - **Des documents très consultés par tous les publics des ANF**
  - **Une valeur d'exemple pour le réseau des services publics français d'archives**

- **Un processus commencé via RiC-O Converter...**
- **....puis complété par quelques posts-traitements :**
  - mise en conformité avec la version 0.2 de RiC-O
  - quelques enrichissements (pour indexer la catégorie de document – Record – ou de partie de document – Record Part notamment)
- **...et poursuivi à l'aide de SPARQL Update, après import du graphe obtenu dans la base de graphes choisie, pour créer des classes spécifiques, sous-classes des classes Record, Record Part, Person, Corporate Body, Place, et pour les « peupler »**  
Ci-contre un exemple – création et peuplement de la classe Repertoire (un clic sur l'image vous permettra d'accéder à une version plus lisible).

```
PREFIX demoanonto: <https://sparnatural-demo-anf.huma-num.fr/ontology#>␣
PREFIX sparnademoan: <https://sparnatural-demo-anf.huma-num.fr/insertedData/>␣
PREFIX owl: <http://www.w3.org/2002/07/owl#>␣
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>␣
PREFIX rico: <https://www.ica.org/standards/RiC/ontology#>␣
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>␣
PREFIX dc: <http://purl.org/dc/elements/1.1/>␣
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>␣
␣
␣
# Déclaration de nouvelles classes␣
INSERT DATA { GRAPH sparnademoan: {␣
␣
demoanonto:Repertoire a owl:Class ;␣
rdfs:subClassOf rico:Record ;␣
rdfs:label "Répertoire"@fr ;␣
rdfs:comment "Liste, en général chronologique, se présentant sous forme de registre␣
ou d'ensemble de feuillets, que le notaire a obligation de tenir, aujourd'hui jour par jour,␣
en y faisant figurer date, nature de l'acte [...], espèce de l'acte (minute ou brevet),␣
noms des parties et toutes autres mentions prescrites par la loi ou les règlements.␣
Une fois versé avec les minutes dans un service public d'archives, le répertoire sert principalement␣
à repérer un acte donné [...], mais [...],␣
comme la minute, il comporte intrinsèquement une valeur probatoire."@fr .␣
}␣
}␣
;␣
␣
# calcul des instances des nouvelles classes ␣
␣
INSERT {␣
GRAPH sparnademoan: {␣
?rep a demoanonto:Repertoire␣
}␣
}␣
WHERE {␣
?rep a rico:Record ;␣
rico:hasDocumentaryFormType <http://data.archives-nationales.culture.gouv.fr/documentaryFormType/FRAN_RI_001-d-a04b1hkno-92bbw9jou9ed> .␣
}␣
} ◆
```

# Contenu actuel du graphe

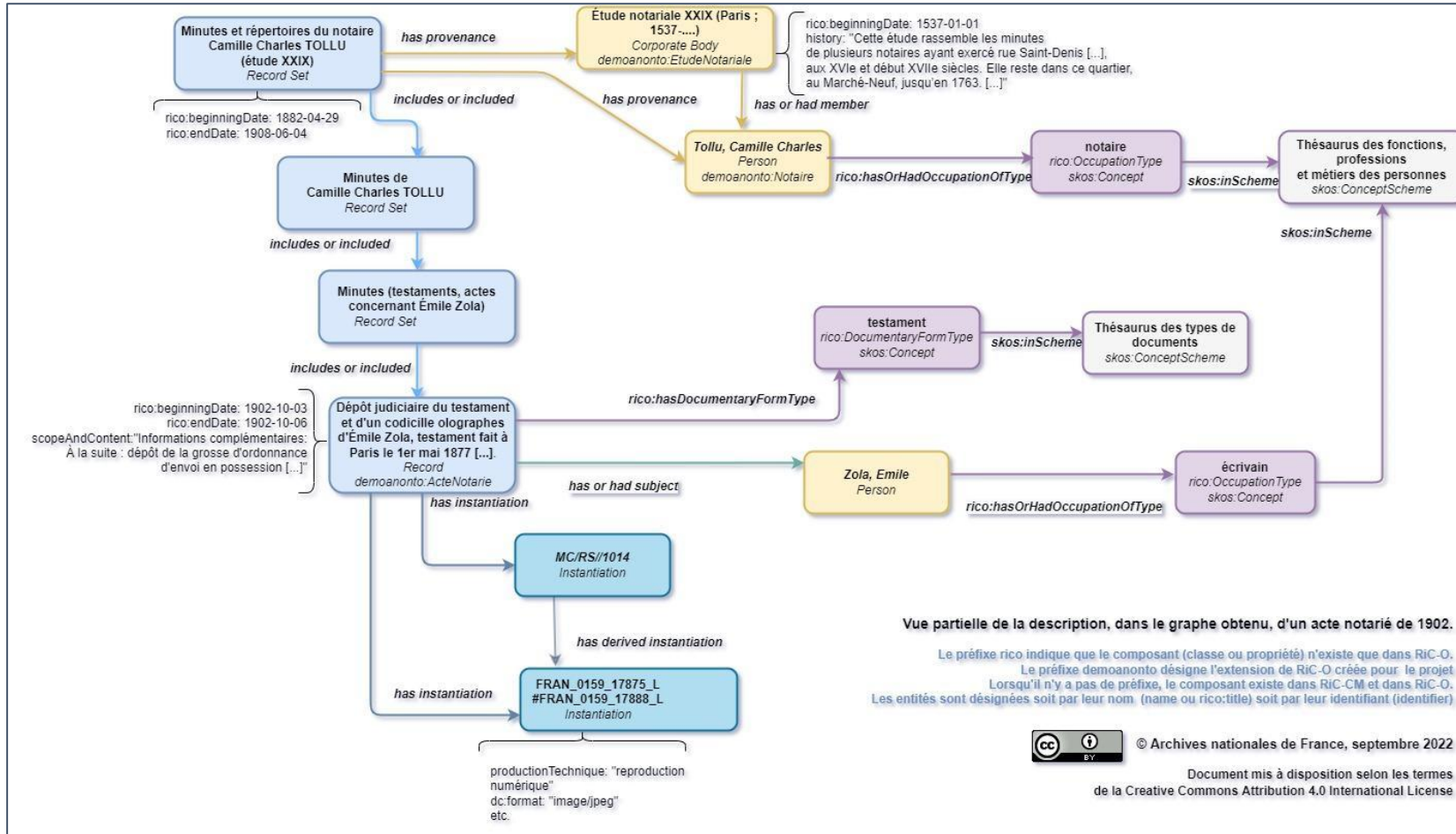
- Après trois itérations on obtient les statistiques suivantes :
  - **57,9 millions de triplets RDF**, dont environ 37 millions de triplets inférés
  - Voici les composants de RiC-O 0.2 que ce graphe instancie :
    - 34 classes (catégories d'entités) sur 106 (48 en comptant les triplets inférés) ;
    - 25 *datatype properties* (relations dont la cible est une chaîne de caractères) sur 62 (28 en comptant les triplets inférés) ;
    - 79 *object properties* (relations dont la cible est une entité) sur 423 (162 en comptant les triplets inférés).

Le jeu de données obtenu est déposé sur GitHub dans un dépôt rendu public tout récemment :  
[https://github.com/ArchivesNationalesFR/Sparnatural\\_prototype\\_data](https://github.com/ArchivesNationalesFR/Sparnatural_prototype_data)

**Les statistiques ci-contre, à l'heure actuelle, ne peuvent pas être obtenues à partir des métadonnées source via la salle des inventaires virtuelle des ANF, et il est difficile de les calculer pour un professionnel compétent dans le domaine des technologies XML. Une simple requête SPARQL permet aujourd'hui de les obtenir facilement.**

- Plus précisément on trouve dans le graphe :
  - 72 665 groupes de documents (fonds, séries, liasses, dossiers de clients...);
  - **400 570 actes notariés** (décrits à ce jour) au sein des liasses ;
  - **104 383 mentions d'actes** relevées au sein des répertoires ;
  - 3135 répertoires ;
  - **13 127 listes d'actes au sein des répertoires**, dont 11 670 numérisées ;
  - 629 246 personnes (dont 3 213 personnes dotées d'un IRI) ;
  - 530 organismes (dont 528 dotés d'un identifiant) ;
  - 15 559 lieux (dont 15463 dotés d'un IRI).

# Un diagramme montrant la description d'un acte notarié dans le graphe obtenu



- Un clic sur l'image vous permettra d'accéder à une version plus lisible.
- Dans le démonstrateur, voir la page suivante dans laquelle les mêmes métadonnées sont affichées sous la forme de texte : <https://sparnatural-demo-anf.huma-num.fr/lodview/recordResource/041835-c1p6y1e0w26r--183v6ez58hjun.html>

# Le démonstrateur Sparnatural des ANF : concevoir et réaliser le démonstrateur

- Cela consiste à **produire un modèle ontologique en OWL pour la recherche, et à définir ses correspondances, simples ou plus élaborées, avec les classes et propriétés de l'ontologie métier**

Voir la documentation à : <https://docs.sparnatural.eu/OWL-based-configuration>

À partir de ce fichier OWL, Sparnatural génère automatiquement l'interface de recherche

- Si on dispose d'un SPARQL endpoint, que l'on connaît HTML, le modèle métier employé, le ou les jeux de données à explorer et le langage OWL, **on peut donc construire et faire évoluer son (ses) interface(s) de façon autonome.**
- Comme il s'agit d'un fichier OWL, **une ontologie Sparnatural de recherche élaborée dans le cadre d'un projet donné peut être partagée pour être réutilisée dans d'autres projets utilisant le même outil.**

- L'enjeu a donc été (et est, de façon générale) d'atteindre par ce biais un **compromis raisonnable entre la complexité du modèle métier et la nécessité de produire une interface compréhensible et performante pour les utilisateurs.**

- Nous avons conçu l'interface par itérations courtes, et **avons impliqué les utilisateurs finaux via 2 ateliers de test organisés fin 2021.**

# Conception d'une interface de recherche Sparnatural pour un graphe de données RiC-O : exemple

- Définition, à l'aide de l'éditeur open source d'ontologies Protégé, de la classe **Archives** dans l'ontologie de la [configuration A](#).
  - Cette classe sera un des points d'entrée dans le graphe. L'annotation *order* spécifie l'ordre d'affichage des points d'entrée (ici, le premier).
  - L'annotation *sparqlString* spécifie la correspondance (ici, très simple) entre la classe Sparnatural et la classe utilisée dans les données RDF (ici une classe spécifique au projet, Archive, qui réunit toutes les instances de `rico:RecordResource` sauf les instruments de recherche et les notices).

The screenshot shows the Protégé ontology editor interface. The left pane displays the class hierarchy with 'Archives' selected under 'SparnaturalClass'. The right pane shows the 'Annotations: Archives' section, which includes:

- rdfs:label** [language: fr]: Archives
- rdfs:label** [language: en]: Records
- defaultLabelProperty**: aPourIntitule
- faIcon**: fad fa-archive
- order** [type: xsd:integer]: 1
- sparqlString**: <<https://sparnatural-demo-anf.huma-num.fr/ontology#Archive>>
- tooltip** [language: fr]: Tout type de ressource archivistique, qu'il s'agisse d'un item (comme une minute ou un répertoire) ou d'un groupe de documents (série des minutes ou des répertoires d'un notaire, liasse de minutes, dossier de client...).
- tooltip** [language: en]: Any type of archival resource, whether it is an item (such as a minute or a registry) or a group of documents (series of minutes or registry of a notary, bundle of minutes, client file...).

The bottom pane shows the 'Description: Archives' section, which includes:

- Equivalent To**: SparnaturalClass
- SubClass Of**: SparnaturalClass
- General class axioms**

At the bottom right, there is a status bar with the text 'To use the reasoner click Reasoner > Start reasoner' and a checked 'Show Inferences' checkbox.

# Concevoir une interface de recherche basée sur Sparnatural

- **Nous avons pu tester et implémenter les fonctionnalités de Sparnatural développées par Sparna dans le cadre du projet :**
  - support de la clause OPTIONAL (on peut ainsi mieux prendre en compte les variations dans la complétude et la précision des données du graphe) ;
  - support de FILTER NOT EXISTS ;
  - sélection de valeurs dans une liste arborescente (de type thésaurus) ;
  - ajout d'aides contextuelles (*tooltips*) ;
  - choix des colonnes de résultats à afficher ;
  - ajout de requêtes pré-enregistrées ;
  - déclinaison de l'interface en plusieurs langues (en l'occurrence, le français et l'anglais).
- Nous avons aussi beaucoup travaillé à **l'ergonomie de l'outil**
- Démonstrateur mis en ligne en juin 2022 :  
<https://sparna-git.github.io/sparnatural-demonstrateur-an/>  
Son code source est ici : <https://github.com/sparna-git/sparnatural-demonstrateur-an>
- La BnF a également mis en ligne le sien (<https://data.bnf.fr/sparnatural/>).



# La première configuration, destinée à la découverte

The screenshot displays a query builder interface for Archives Nationales. The query is structured as follows:

- Condition 1: Archives ont pour type de document Type de document inventaire après décès (...)
- Condition 2: Archives concernent Personne
- Condition 3: Personne a pour activité Activité orfèvre
- Condition 4: Archives ont pour date de création Date de 1600 à 1700

Below the query builder, there is a table view with a search bar, sorting options (Tri: A↓ Z↓), and a toggle for 'Montrer le nom des colonnes'.

**Simple et généraliste ; elle peut quasiment s'appliquer à tout jeu de données archivistiques conformes à RiC-O**, dans la mesure où s'y trouvent indexées des entités de contexte comme les producteurs, des agents sujets des documents, des lieux.

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

# La deuxième configuration, pour explorer le graphe plus en profondeur

Le formulaire de recherche disponible sur la présente page propose une interface d'exploration dans ces métadonnées qui s'appuie sur les spécificités des archives notariales.

Groupes de documents produits par des notaires ayant exercé dans le quartier S ▾

Service SPARQL: <https://sparnatural-demo-anf.huma-num.fr/graphdb/repositories/sparnatural-demo-anf>

où

et

et

▶

Elle exploite les spécificités des archives notariales et de leur description.

Plutôt adaptée aux besoins d'une personne familière de cet univers, archiviste ou usager habitué.

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

## Démonstrateur Sparnatural des Archives nationales

### Français

Ce démonstrateur, construit avec l'outil [Sparnatural](#) vous permet d'interroger une partie des métadonnées décrivant les archives des notaires de Paris conservées aux Archives nationales. Ces métadonnées ont été sémantisées conformément à l'ontologie [Records In Contexts \(RIC-O\)](#). Le démonstrateur est composé de deux interfaces de recherche. Le [démonstrateur A](#) propose une interface simple et généraliste d'exploration dans ces métadonnées, qui pourrait être utilisée pour d'autres métadonnées archivistiques. Le [démonstrateur B](#) propose une interface d'exploration dans ces métadonnées qui s'appuie sur les spécificités des archives notariales. Par ailleurs les IRIs des ressources RDF ne sont actuellement pas déréférencables.

### English

Through this prototype, built with [Sparnatural](#), you can query a part of the metadata describing the archives of Paris notaries kept at the French national Archives. These metadata have been semantized according to the ontology [Records In Contexts \(RIC-O\)](#). The prototype is composed of two research interfaces. [Interface A](#) offers a simple and general exploration interface in this metadata, which could be used for other archival metadata. [Interface B](#) offers an exploration interface in these metadata based on the specificities of notarial archives. Furthermore, the IRIs of RDF resources are currently not dereferenceable.

Le site web inclut une documentation détaillée [en français](#) et [en anglais](#).

Nous n'avons pas eu le temps de faire une démo pendant la présentation. Vous pouvez regarder le [screencast relatif à ce site web](#), préparé pour présenter Sparnatural lors de la conférence [SemWeb.Pro](#) le 8 novembre 2022. Puis lancez-vous ! Et **n'hésitez pas à nous contacter pour toute question ou remarque.**

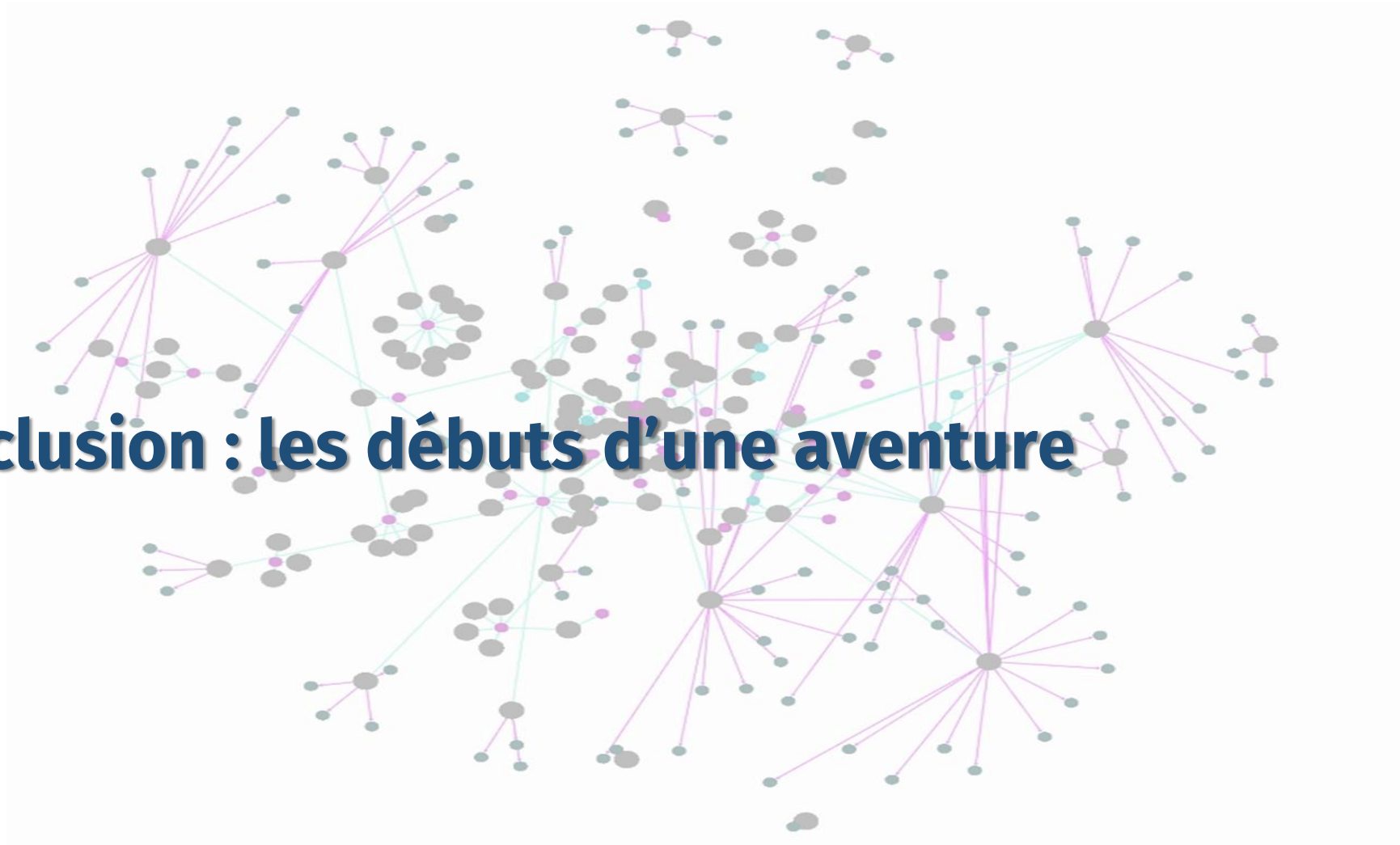
- **Bilan très positif :**
  - en ce qui concerne la **preuve de la faisabilité de la conversion de grandes quantités de métadonnées archivistiques de nature “classique” en données RDF/RiC-O** ;
  - en ce qui concerne les **nouvelles possibilités d’exploration et de recherche** offertes par l’outil.  
**Bien sûr nous espérons recueillir d’autres réactions, remarques, suggestions et retours d’expérience que ceux déjà reçus**
- **Autres résultats concrets du projet :**
  - Travail préalable à la mise à jour de RiC-O Converter réalisé ;
  - Travail de traduction en français des libellés anglais des classes et propriétés RiC-O, va être intégré dans les sources de RiC-O.

- **Le projet va se poursuivre :**
  - on l’espère, pour développer une nouvelle version de Sparnatural afin notamment **de fusionner les deux interfaces/configurations**
  - pour **améliorer les métadonnées source (donc les données RDF)**, afin notamment **de réduire les redondances liées à la longue histoire de la description des archives aux Archives nationales**, qui affectent le graphe et rendent aussi en SIV certains résultats de recherche difficiles à comprendre et à exploiter.
- Le Lab des ANF va aussi produire et **publier une ontologie Sparnatural générique** de recherche pour les données RDF conformes à RiC-O en prenant pour point de départ la configuration A.

# En savoir plus sur le projet

- La documentation (en français et en anglais) : <https://sparna-git.github.io/sparnatural-demonstrateur-an/presentation-fr.html>
- Les présentations et l'enregistrement vidéo de la demi-journée d'étude organisée le 17 juin 2022 : <https://sparnatural.eu/#bibliography>

**En guise de conclusion : les débuts d'une aventure**



# Beaucoup d'enseignements et de bénéfices tirés de ce qui a déjà été fait

## D'un point de vue général

- **La preuve que RiC-O, ça marche et que cela peut être utile pour tous**, même à partir de métadonnées « classiques »
- Des idées pour RiC-O 1.0 (et même les premiers mappings et la traduction en français des noms des composants)
- **Des outils et des briques réutilisables par tous**

## Pour les ANF

- **Une meilleure connaissance des métadonnées des ANF ; de quoi produire certaines préconisations, et aussi ouvrir de nouveaux chantiers**
- **De nouvelles données de qualité** (des référentiels enrichis)
- Un approfondissement des compétences, techniques et aussi méthodologiques
- Une expérience très bénéfique de la collaboration avec divers partenaires

# La suite : des programmes pour 2023

- **Version 2 de RiC-O Converter** (essentiellement mise en conformité avec RiC-O 0.2 et 1.0), **puis peut-être une version 3** (prise en compte des fichiers SEDA de description des archives nativement numériques)
- Probablement une **nouvelle version de Sparnatural** (discussions en cours)
- Travail sur les métadonnées et l'interface du démonstrateur Sparnatural
- Travail sur d'autres projets, avec des partenaires (comme, entre autres, ORESM) ou en interne (plusieurs services impliqués)



# Des programmes au long cours

- **Poursuite des travaux sur les référentiels :**
  - enrichissements : en 2023, des projets sur les agents, les lieux et les types de documents ; la conception d'un référentiel des relations entre entités, basé sur RiC-O
  - phase 2 d'un projet d'intelligence artificielle (reconnaissance, réconciliation et liage d'entités nommées dans les fichiers EAD ; voir à ce sujet la [présentation faite en mars 2022](#))
  - refonte du SI devrait commencer par les référentiels -> modélisation en cours, basée sur RiC-O et SKOS
- **Vers un module sémantique complet pour toutes les métadonnées descriptives des ANF**
- **Si possible, lancement d'un projet relatif à la data visualisation des graphes de métadonnées RiC-O**

# RiC ailleurs en France actuellement

- Ce qui est sûr, c'est que **l'équipe du portail FranceArchives :**
  - **a choisi RiC-O, et avance bien sur la mise en œuvre de RiC-O et la mise en place d'un SPARQL end point et de services de négociation de contenu ;**
  - souhaite adosser une interface de recherche Sparnatural à ce SPARQL end point
- La BnF a aussi inscrit RiC-O dans [la feuille de route de data.bnf.fr](https://data.bnf.fr)
- D'après les informations recueillies, les autres services d'archives attendent la publication de RiC-CM et RiC-O 1.0 (prévue en mars 2023)

**Merci de votre attention !**

N'hésitez pas à me contacter

