

Utilisations de l'ontologie RiC-O aux Archives nationales de France (AnF)

Florence Clavaud, responsable du Lab, Archives nationales de France

courriel : florence.clavaud@culture.gouv.fr

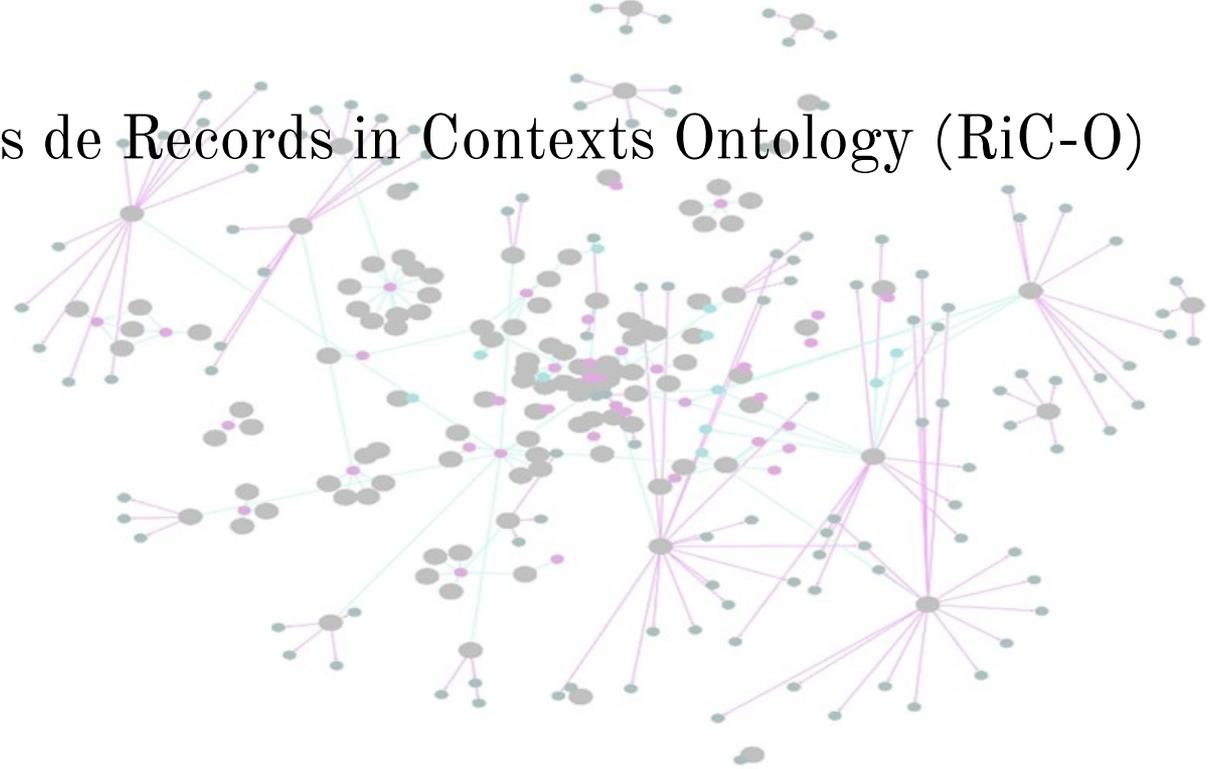
SemWeb.Pro 2024

Paris, 21 novembre 2024

Sommaire

- Actualités de Records in Contexts Ontology (RiC-O) :
 - une version 1.0
 - une communauté d'utilisateurs
 - la feuille de route
- Utilisations de RiC-O aux AnF :
 - le contexte
 - les projets, ce qu'ils font aux données, ce qu'ils en disent
 - perspectives

1. Actualités de Records in Contexts Ontology (RiC-O)



Une version 1.0 de RiC-O publiée fin décembre 2023

- RiC-O est la **troisième partie d'un standard métier pour la description des archives dans leurs contextes de production, de préservation et d'utilisation**, Records in Contexts
- RiC est élaboré par un groupe d'experts du Conseil International des Archives, ICA/EGAD.
- RiC-O 1.0 remplace la v0.2 publiée en février 2021.
- Première version complète et stable, une recommandation officielle de l'ICA, donc **une étape majeure**. Et un long chemin parcouru depuis 2017.

RiC-O : l'essentiel

- IRI : <https://www.ica.org/standards/RiC/ontology>
- Développement public sur GitHub depuis mars 2020
On trouve dans le dépôt beaucoup d'autres ressources (version modularisée, listes des composants en CSV, exemples, diagrammes) :
<https://github.com/ICA-EGAD/RiC-O>
- Dernière release en date : [v1.0.2](#) (correctifs apportés à la v1.0)
- Un [site web d'information](#)



INTERNATIONAL COUNCIL ON ARCHIVES
CONSEIL INTERNATIONAL DES ARCHIVES

EGAD
EXPERT GROUP ON ARCHIVAL DESCRIPTION
GROUPE D'EXPERTS SUR LA DESCRIPTION ARCHIVOLOGIQUE

International Council on Archives Records in Contexts Ontology (ICA RiC-O) version 1.0.2
Ontology IRI: <https://www.ica.org/standards/RiC/ontology>
This HTML view: https://www.ica.org/standards/RiC/RiC-O_1-0-2.html
OWL source file: https://www.ica.org/standards/RiC/RiC-O_1-0-2.rdf

Table of contents

- [Ontology metadata](#)
- [Namespaces used](#)
- [Introduction](#)
 - [RiC-O design principles](#)
 - [Understanding RiC-O: a quick overview of some features](#)
 - [From RiC-CM to RiC-O](#)
 - [RiC-O documentation and annotation properties](#)
 - [Next steps](#)
- [RiC-O classes](#)
- [RiC-O datatype properties](#)
- [RiC-O object properties](#)
- [RiC-O annotation properties](#)
- [RiC-O named individuals](#)
- [Appendix I: list of changes made since the release of RiC-O v0.1](#)

Ontology metadata

This version: Version 1.0.2 - 2024-09-04.

RiC-O 1.0.2 : l'essentiel

- Une ontologie OWL 2 de domaine, fondée sur un socle conceptuel solide, [RiC-Conceptual Model 1.0](#), qu'elle transpose, affine et étend
- Déjà riche (105 classes, 61 datatype properties, 400 object properties)
- **Flexible** et bien évidemment extensible
- Entièrement documentée en anglais (les labels des composants sont aussi proposés en français et en espagnol actuellement)
- Objectif principal : permettre aux professionnels des archives de pousser les métadonnées archivistiques vers le web de données, sans faire de compromis sur la perspective métier, tout en gardant une certaine souplesse (**ontologie de transition**, conçue pour être immédiatement utilisable)
- Une perspective métier renouvelée sous la pression des besoins des utilisateurs et d'autres facteurs
- Le même chemin que celui emprunté par les communautés voisines
- Cela donne un modèle très englobant, permettant de sortir du carcan finalement formé par les normes et règles de description que RiC remplace, sans parler de faciliter l'interopérabilité et la réutilisation de ces métadonnées décrivant des biens culturels

Une communauté d'utilisateurs en cours de constitution

- Un [groupe de discussion sur Google Drive](#) depuis avril 2023
 - 348 membres aujourd'hui
 - Outre les messages d'information, des échanges intéressants sur l'ontologie et ses usages
- Quelques contributions également significatives sur GitHub
- Des événements permettent aux utilisateurs de se réunir et d'échanger, à l'échelle internationale ou plus locale
- Un standard encore tout jeune, mais :
 - un certain nombre de projets pionniers assez avancés (ex. : AnF ; portail FranceArchives ; Archives de la ville d'Amsterdam ; plusieurs projets en Suisse dont le logiciel Docuteam context ; Archives nationales de Singapour)
 - d'autres porteurs de projets se sont manifestés plus récemment
 - aussi des projets de recherche en SHS

Avancer vers une communauté de pratiques

- Cela passera par :
 - des rencontres
 - des ateliers, des formations
 - la co-construction d'un document de type *Primer* ou de pages de type « how to »Ex. : Comment décrire une activité métier ? Comment décrire un manifeste IIF ?
- Le manuel de mise en œuvre de RiC (RiC-AG) devrait aussi intégrer, au moins dans une version 2, des sections sur la mise en œuvre de RiC-O
- L'analyse des usages pourra se traduire à la fois par des enrichissements de l'ontologie, et à moyen terme par une moins grande souplesse (moins d'options alternatives)

La feuille de route, de RiC-O 1.0.2 à RiC-O 2.0

- **RiC-O 1.1** (en cours ; publication vers la fin 2024) : des ajouts, quelques ajustements
Ex. :
 - une propriété pour relier un `rico:Record` à l'entité qu'il représente (pour les documents graphiques)
 - de nouveaux 'raccourcis' (object properties définis au moyen de property chain axioms)
- **RiC-O 2.0** (2025) ; une nouvelle version majeure, avec (entre autres) :
 - la version modularisée comme version de référence
 - des vocabulaires pour les types d'activité (`rico:ActivityType`) et de règles (`rico:RuleType`)
 - un rapprochement au moins partiel des classes `rico:Relation` avec `rico:Event` et `rico:Activity`
 - une articulation avec PREMIS, CIDOC-CRM et Schema.org (sinon des recommandations pour une utilisation conjointe de ces ontologies)
- Cette feuille de route sera publiée

En savoir plus

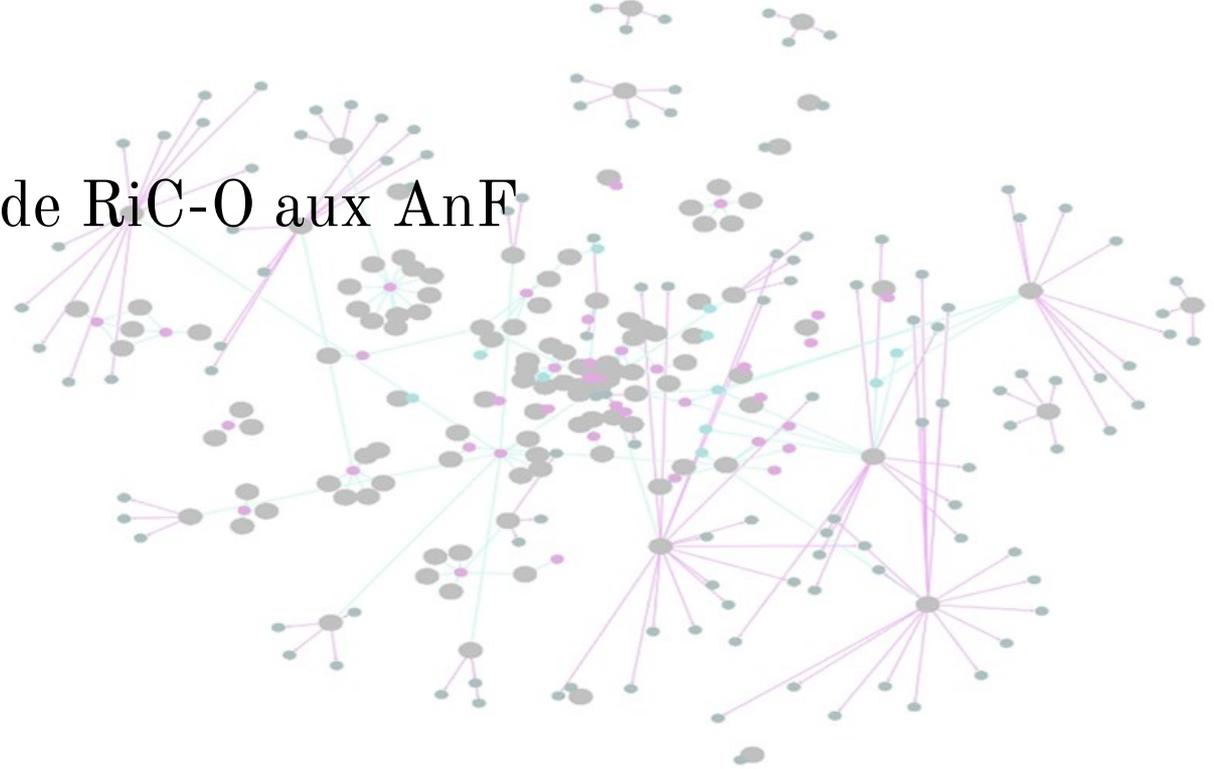
- Voir notamment :

- les [supports de présentation](#) et [enregistrements vidéo](#) de la journée d'études internationale sur les premières implémentations de RiC (AnF, 15 novembre 2023) :

- les supports de présentation et enregistrements vidéo des trois webinaires en français sur RiC, organisés par le Lab des AnF, en juin et septembre 2024 (en cours de publication sur Google Drive :

<https://drive.google.com/drive/folders/1da3re6fqZ4psEaywWI3DX6jbasDxFree?usp=sharing>)

2. Utilisations de RiC-O aux AnF



Le contexte

- **Un opérateur national parmi les plus importants services d'archives du monde**, très actif, avec des moyens humains limités
- Un SI ancien, complexe, siloté, etc...
- Des métadonnées en très grande quantité, dans différents formats (notamment 32000 fichiers XML/EAD contenant plus de 11 000 000 d'unités de description, 16000 fichiers XML/EAC-CPF), dispersées ; mais en fait un immense graphe sous-jacent
- Plusieurs interfaces web, pas toujours d'accès aux données brutes pour les humains, pas d'API
- **Les projets sémantiques des AnF ne peuvent actuellement être menés qu'hors de ce SI**
- **De ce fait, ils sont portés par le Lab**, la petite équipe qui coordonne et mène les projets de R&D et d'innovation numérique
- **Donc, plusieurs contraintes :**
 - **peu de moyens**
 - pas d'échanges directs de données avec le SI (et les données sémantisées restent *pour l'instant* hors du SI)
 - **les projets sont 'isolés' du SI ; ils doivent être valorisés en tant que tels**
- Mais aussi **beaucoup de liberté** et finalement du temps, pour réfléchir et explorer diverses pistes
 - Quelques principes : ce qui est produit (données ou outils) doit servir l'institution, ses agents et ses publics, est évalué et documenté, puis publié sous licence libre

Une refonte du SI s'impose. Elle prendra des années. **En attendant, impossible d'ajouter une nouvelle brique ou de toucher le cœur du système.**

Un logiciel pour la sémantisation des métadonnées archivistiques ‘classiques’

Accueil

RiC-O Converter

Bienvenue dans la **documentation de RiC-O converter** ! RiC-O converter est un outil en ligne de commande permettant de convertir des fichiers EAD et EAC en fichiers RDF exprimés à l'aide de l'ontologie Records in Contexts.

Cet outil est téléchargeable depuis la section *release* du dépôt Github du convertisseur.

Table des matières

La documentation se compose des rubriques suivantes :

Au sujet de RiC-CM et RiC-O

- **RiC-CM et RiC-O** : consultez cette page pour en savoir plus sur le modèle conceptuel et l'ontologie Records in Contexts ;
- **À propos du projet RiC-O converter** : comprendre le projet RiC-O Converter ;
- **Mappings** : découvrez comment les fichiers EAD et EAC sont mappés sur l'ontologie RiC par le convertisseur.

La documentation technique de RiC-O converter

- **Pour commencer** : commencez par consulter cette page pour tester le convertisseur, comprendre la structure du répertoire, apprendre à afficher un message d'aide et ajuster les paramètres de l'outil ;

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

- Avec la société Sparna, développement depuis 2019, puis évolutions du **logiciel open source RiC-O Converter**, pour convertir en données RDF/RiC-O les fichiers XML des AnF
 - Un logiciel performant, fiable, entièrement documenté (y compris via des mappings et tests unitaires), facile à installer et à exécuter, adaptable
- **Version 3.0 en cours de réalisation ; elle permettra d'obtenir des jeux de données conformes à RiC-O 1.0.2.**
- Elle inclura aussi des jeux de règles SHACL pour produire des rapports de validation et permettre ainsi un contrôle de qualité des données résultat.
 - Règles de structure et règles métier, générales ou propres aux AnF.
 - A terme : idée de pouvoir intégrer à ces rapports de validation l'indication de la source EAD ou EAC-CPF dans le cas d'erreurs
 - Il sera en effet très utile de pouvoir identifier et compter les descriptions incomplètes, ou non conformes aux bonnes pratiques métier et aux règles institutionnelles

« Libérer » les métadonnées des AnF, en faire autre chose et/ou faire plus avec elles

- **Publication, accompagnée d'une interface d'exploration et de consultation intuitive pour les usagers, d'un jeu de données RDF obtenues à partir du tiers des métadonnées décrivant les archives notariales parisiennes.**
Soit en tout actuellement 60 millions de triplets RDF (voir https://github.com/ArchivesNationalesFR/Sparnatural_prototype_data)

The screenshot displays the Sparnatural demonstrator interface. At the top, it shows a SPARQL query: "Types de documents des actes notariés datant de 1848, avec les cotes de ces acte". The query is visualized as a graph with nodes and edges. The nodes include "Type de document", "Acte notarié", "Date", and "Cote". The edges represent relationships like "est le type de document de l'acte ou de la mention d'acte", "a pour date de création", and "a ou ont pour cote". Below the query, there are options to "Afficher/Masquer éditeur SPARQL", "Partager", and "Exporter". The results are displayed in a table format, showing 679 results in 32.288 seconds. The table has columns for "this", "Date_2_begin", "Date_2_end", "ActeNotarie_1", and "Cote_3".

this	Date_2_begin	Date_2_end	ActeNotarie_1	Cote_3
1 renonc...	"1848-01-01"	"1848-01-31"	Acte de renonciation par Marie Françoise Oberson	MC/ET/1/...
2 renonc...	"1848-01-01"	"1848-01-31"	Acte de renonciation par Marie Françoise Oberson	MC/ET/1/...
3 renonc...	"1848-04-26"	"1848-04-26"	Renonciation par Louis Gervais Serize, notaire, à la succession de son épouse, Marie Madeleine C...	MC/ET/1/...
4 accord	"1848-12-18"	"1848-12-18"	Procès-verbal de remise des opérations de liquidation de la succession d'Isaac Rodrigues-Henriq...	MC/ET/1/...
5 liquida...	"1848-12-18"	"1848-12-18"	Procès-verbal de remise des opérations de liquidation de la succession d'Isaac Rodrigues-Henriq...	MC/ET/1/...
6 partage	"1848-12-18"	"1848-12-18"	Procès-verbal de remise des opérations de liquidation de la succession d'Isaac Rodrigues-Henriq...	MC/ET/1/...

Première version publiée à l'été 2022 : <https://sparnatural.github.io/sparnatural-demonstrateur-an/>

Le prototype d'exploration et de recherche sur les archives notariales : quelques enseignements

- RiC-O à la fois utilisée partiellement et étendue
- **L'interface permet de poser des questions qu'il n'est pas possible de poser au SI actuel (passage à un graphe unique d'entités liées)**
- L'utilisateur s'engage plus.
- Le caractère disruptif de l'interface peut aussi gêner certains utilisateurs.
- En savoir plus : <https://sparnatural.github.io/sparnatural-demonstrateur-an/presentation-fr.html>

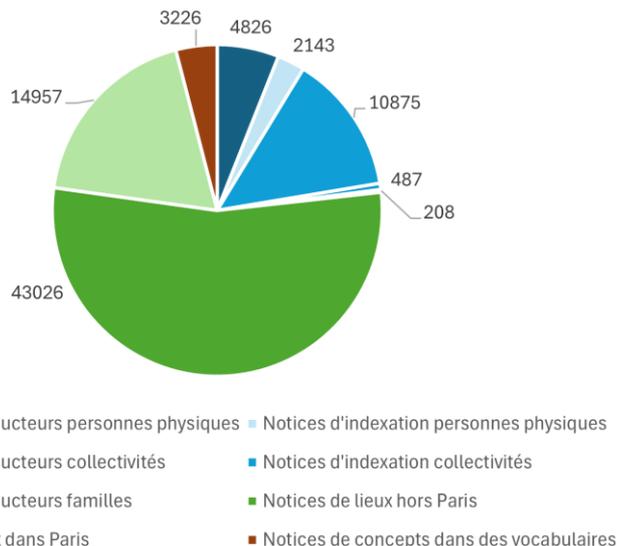
The screenshot displays the Sparnatural search interface. At the top, there is a navigation bar with links for 'Accueil', 'Démonstrateur A', 'Démonstrateur B', 'Documentation', 'Lang.', 'Mentions légales', and 'Sparnatural'. The main area features a query graph with nodes for 'Type de document' (set to 'Acte notarié'), 'Date' (set to 'de 01/01/1848 à 31/12/...'), and 'Acte notarié'. Below the graph, there are options to 'Afficher/Masquer éditeur SPARQL', 'Partager', and 'Exporter'. The search results section shows 66 items in the result set, with 680 results in 28.574 seconds. Three result cards are visible, each with a title and a list of items:

- dépôt de pièces**
 - est le type de document de l'acte ou de la mention d'acte : **Dépôt par Théophile Hippolyte Fontaine, clerc de notaire demeurant au 17 rue Bleue, d'une pièce concernant la succession de Charles François Guillaume marquis de Chanaleilles.** (Acte notarié)
 - est le type de document de l'acte ou de la
- partage**
 - est le type de document de l'acte ou de la mention d'acte : **Dépôt par Théophile Hippolyte Fontaine, clerc de notaire demeurant au 17 rue Bleue, d'une pièce concernant la succession de Charles François Guillaume marquis de Chanaleilles.** (Acte notarié)
 - est le type de document de l'acte ou de la
- décharge**
 - est le type de document de l'acte ou de la mention d'acte : **Décharge par Julien Alexandre Monique Lelasseux, ancien professeur demeurant à La Flèche, à Théophile Hippolyte Fontaine, clerc de notaire demeurant au 8 place Sainte-Opportune.** (Acte notarié)

Version 2 en cours de réalisation. Les données RDF seront mises à jour et rendues conformes à RiC-O 1.0.2. Cette version utilisera la dernière version en date de l'éditeur visuel de requêtes SPARQL Sparnatural.

Un autre chantier en cours : la FAIRisation des référentiels des AnF

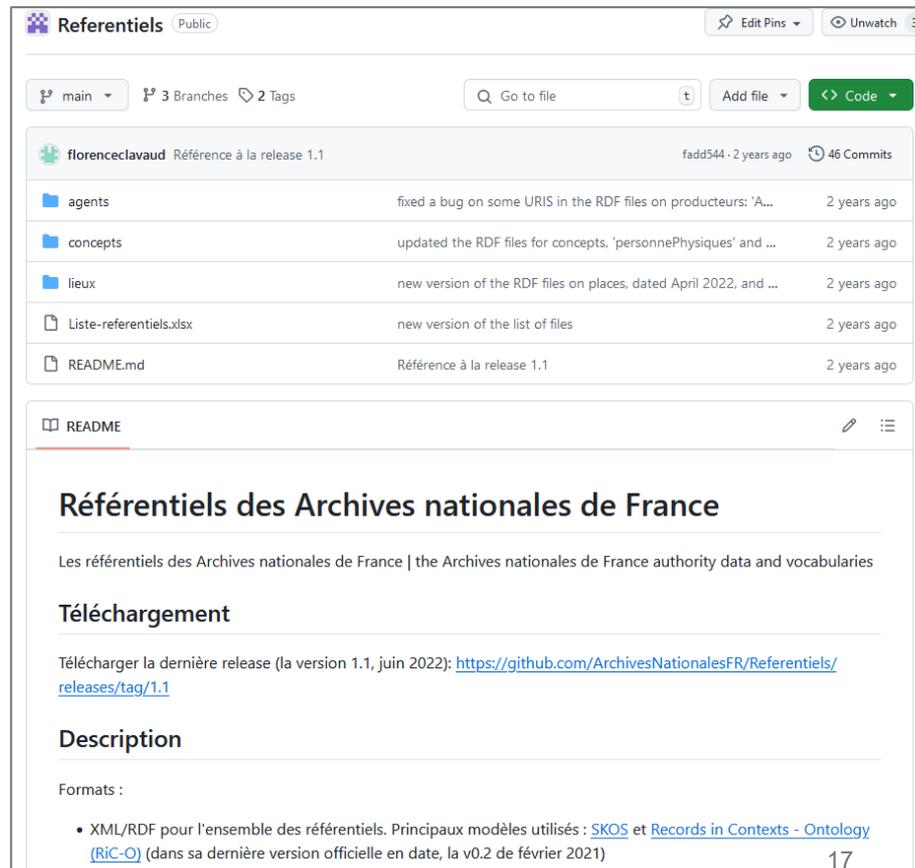
Les référentiels des AnF en avril 2024 :
informations quantitatives globales



- À ce jour, ces données de référence décrivent environ 18 500 agents, 58 000 lieux et 3 250 concepts, entités de contexte des archives conservées par l'institution.
- Un réservoir de connaissances incomplet et imparfait mais sans équivalent, compte tenu du domaine d'intervention des AnF et de l'expertise nécessaire pour décrire ces entités.
- Des données inaccessibles aux usagers (sauf pour les notices de producteurs). Impossible par ailleurs d'en exploiter la nature même (celle d'un graphe orienté : les notices sont liées entre elles).

Un autre chantier en cours : la FAIRisation des référentiels des AnF (2)

- Depuis 2021, les référentiels sont disponibles sur GitHub, au format RDF (SKOS et RiC-O) et CSV :
<https://github.com/ArchivesNationalesFR/Referentiels>
- **Des enrichissements des données source ont déjà été réalisés** (par alignement avec des réservoirs externes notamment, tels que ceux de l'INSEE ou de l'IGN, également exprimés en RDF).
D'autres enrichissements conséquents sont en cours, en particulier sur le vocabulaire dit des 'types de documents' et sur la description des 13189 voies de Paris.
- **Nouvelle release prévue vers janvier 2025** (conforme à RiC-O 1.1)



The screenshot shows the GitHub repository page for 'Referentiels' by 'florenceclavaud'. The repository is public and has 3 branches and 2 tags. The file list includes folders for 'agents', 'concepts', and 'lieux', and files for 'Liste-referentiels.xlsx', 'README.md', and 'agents'. The README file is selected and displays the following content:

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)

Utilisation des référentiels sémantisés pour la recherche en SHS/DH

- Ainsi sémantisés, les référentiels ont déjà été réutilisés dans le cadre de projets de recherche dans lesquels les AnF sont partenaires :
 - ALEGORIA (2018-2021)
 - ORESM (2021-....)
- Voir aussi les résultats d'un travail réalisé à l'automne 2023 (ci-dessous) dans le cadre d'un atelier sur les graphes de connaissances géo-historiques de l'ANR Soduco (https://soduco.geohistoricaldata.org/atelier_graphes_geohistoriques_annuaires/)

Visualisation des graphes géohistoriques construits à partir des entrées d'annuaires du commerce de Paris (XIXème siècle)

[SPARQL Endpoint](#) | [Dépôt Git-Hub](#) | [Aide](#)

Dataset
Notaires

Filtres

Propriétés

Raison sociale
morel

Description
Ex : photo

Adresse
Ex : rivoli

Les champs du formulaire acceptent les expressions régulières. Deux regex utiles :

- La propriété contient mot1 ET mot2 (?= 'mot1(?= 'mot2)
- La propriété contient mot1 OU mot2. motifmot2

Tester une regex complexe.

Période

1850 1883

1790 1821 1853 1884 1915

Utilisation des graphes de connaissances RiC-O des AnF dans un projet de NER

- NER4Archives (Named Entity Recognition For Archives) (2020-2024)
- Objectif : élaborer une méthode et mettre en place, sous la forme d'un prototype fonctionnel, un écosystème/workflow pour la reconnaissance des entités nommées dans les fichiers XML/EAD des AnF (et potentiellement d'autres services d'archives)
- Enjeux : améliorer semi-automatiquement la qualité des métadonnées des AnF, sous-indexées ; enrichir en même temps les référentiels d'indexation, utilisés par les agents pour indexer ; partager les résultats
- Partenaires : AnF (Lab) ; Inria (équipe [ALMAnaCH](#))
- Financement : ministère de la Culture ; Inria; AnF (convention-cadre MC-Inria)
- Deux phases. Dans la 2^e phase (solutions pour l'identification des entités nommées préalablement reconnues et classifiées), les entités candidates dans les référentiels des AnF sont classées en exploitant les données de contexte des fichiers EAD (alignement des nœuds des deux graphes de connaissances, celui des référentiels et celui des fichiers EAD, après leur encodage et leur transformation en embeddings). En cours

Voir à ce sujet la présentation faite en mai 2024 (<https://ner-for-historical-docs.github.io/ressources/20240530-Clavaud-Graiff-NER4Archives.pdf>)

Reste à finaliser puis à évaluer selon la perspective des AnF, à rendre public et à valoriser

Quelques perspectives pour 2025 et au-delà

- Une **version 4 de RiC-0 Converter**, pour prendre en compte les descriptions des archives nativement numériques
- **Poursuite des travaux sur les référentiels sémantisés**
 - ajout de descriptions relatives à des agents
 - **conception et réalisation d'une application web de recherche et de consultation** des référentiels, qui valorise le tout en tant que graphe et utilise la dernière version en date de Sparnatural
- **Poursuite des travaux sur les descriptions des archives**
 - analyse des rapports SHACL de validation des données RDF issues des fichiers EAD
 - preuve de concept exploratoire : réalisation d'un pipeline pour réduire certaines redondances désormais bien visibles au cœur du graphe sur les archives notariales (un exemple des conséquences de la sédimentation des descriptions, très présente aux AnF)
 - un autre travail exploratoire sur la production de relations entre (descriptions de) documents (un autre point faible des métadonnées 'classiques')
 - autant de travaux à faire, autant que possible, sur les métadonnées source dans le SI
- Exploration de pistes pour l'utilisation conjointe d'IA et de graphes de connaissances (LMM; RAG)
- Le tout en collaboration avec les utilisateurs (en premier lieu les agents des AnF) et nos ₂₀ partenaires extérieurs

Merci de votre attention !

