

Construction d'un graphe de connaissances RDF à partir des données d'ORESM à l'aide de l'ontologie RiC-O : premiers résultats

par Jean-François Moufflet et Florence Clavaud (Archives nationales)

Journée d'étude ORESM, 26 novembre 2021





Introduction : la longue histoire des métadonnées de description des archives de l'université de Paris aux Archives nationales

par Jean-François Moufflet (Département du Moyen Age et de l'Ancien Régime, direction des Fonds, Archives nationales ; courriel : jean-françois.moufflet@culture.gouv.fr)



Les premiers inventaires de la série M/MM

M, MM		233	T9 E2		233. - † Inventaire sommaire, par M. Douët d'Arcq. - 1863.	Manuscrit. - 1 registre.
-------	--	-----	-------	--	--	--------------------------

M, MM		234	T9 E1		234. - * Répertoire, numérique, par M. Lecoy de la Marche. - 1895.	Manuscrit. - 1 registre.
-------	--	-----	-------	--	--	--------------------------

M		238		M 65-257	238. - Inventaire analytique des cartons M 65-190, par MM. Douët d'Arcq et Roulland.	2 boîtes de fiches.
---	--	-----	--	----------	--	---------------------

Refonte par Jean-Pierre Babelon (1979)

M 67^A Université

1
Lettres de Philippe le Hardi sur les réparations qu'il exige de l'abbaye de Saint-Germain-des-Prés, pour des excès commis contre des écoliers. Poissy, juillet 1278. Et lettres de Philippe le Bel qui établit une rente sur le Châtelet pour l'entretien de deux chapelles fondées par l'abbaye de Saint-Germain-des-Prés, et dont elle s'était délivrée en cédant au roi les Foires de Saint-Germain. Paris, juillet 1286. Dans un vidimus de l'official de Paris de 1318.

M 66^B (suite)

58 et 59

Privilège de Committimus maintenu à l'Université contre les cardinaux. 2 pièces. 1612.

60 et 61

Arrêt du parlement du 9 mars 1624, défendant au greffier en chef du Châtelet de donner des lettres de protection à d'autres qu'à ceux qui auront lettres de scolarité signées du recteur. 2 pièces.

M 67^A Université

1

Lettres de Philippe le Hardi sur les réparations qu'il exige de l'abbaye de Saint-Germain-des-Prés, pour des excès commis contre des écoliers. Poissy, juillet 1278. Et lettres de Philippe le Bel qui établit une rente sur le Châtelet pour l'entretien de deux chapelles fondées par l'abbaye de Saint-Germain-des-Prés, et dont elle s'était délivrée en cédant au roi les Foires de Saint-Germain. Paris, juillet 1286. Dans un vidimus de l'official de Paris de 1318.

1 bis

Lettres de l'official de Paris invitant le clergé de Paris à une manifestation contre le prévôt de Paris pour obtenir de lui le désaveu des attentats qu'il a commis contre l'église..7 septembre 1304. Orig. parch.

2

Acte par lequel Pierre Bertrand, clerc, reconnaît que son oncle Pierre Bertrand, archidiacre de Billom dans l'église de Clermont, l'a entretenu de tous pendant qu'il étudiait dans l'Université de Paris, et que c'est de lui qu'il tient tous ses livres. 1317. Orig. parch.

3

Lettres de rémission accordées par Charles V à deux sergents au Châtelet qui avaient tué un écolier en se défendant. Senlis, 2 août 1365. Dans un vidimus de la prévôté de Paris, du 23 août.

4

Arrêt de condamnation contre Charles de Savoisy. 22 août 1404. Orig. parch.

5

Lettre de Charles VI à l'Université sur la Ligue des Princes. Paris, 14 octobre 1411. Orig. parch.

Reprise bureautique par Isabelle Aristide (2004)

M 67A Titres. XIII^e-XVIII^e siècles.

Histoire de l'Université. 1286-1723.

1. Philippe le Hardi enjoint à l'abbaye Saint-Germain-des-Prés de fonder deux chapelles en réparation d'excès commis contre des écoliers au Pré-aux-Clercs et du meurtre de deux d'entre eux. Poissy, juillet 1278. Philippe le Bel établit une rente sur le Châtelet pour l'entretien des deux chapelles fondées par l'abbaye Saint-Germain-des-Prés, et dont elle s'était délivrée en cédant au roi les foires de Saint-Germain. Paris, juillet 1286. Dans un vidimus de l'official de Paris de 1318.

Orig. parch.

Du Boulay, tome III, p. 480. Jourdain, Index, n° CCLIII et CCLXXXV

1bis. L'official de Paris invite le clergé de Paris à manifester contre le prévôt de Paris pour obtenir de lui le désaveu des attentats qu'il a commis contre l'église. 7 septembre 1304.

Orig. parch.

Encodage en XML EAD (dtd ead SIA AN) – J.-F. Moufflet, 2020

The image shows a software interface for managing archival records. On the left, a 'Plan de classement' (classification plan) tree is visible, with the selected path being: Université de Paris | 1200-XVIIIe siècle (1200-1800) | Organisation, titres, privilèges | Date non renseignée | Titres | 19 avril 1232-19 juin 1723 | Histoire de l'Université | 7 septembre 1304-19... | Vidimus de l'official de Paris de deux acte... The right pane, titled 'Description', shows the following fields:

- Intitulé du parent:** Dossier - 7 septembre 1304-19 juin 1723 - Histoire de l'Université
- Audience:** SIV Internet
- Niveau de description:** Pièce
- Cote(s):** M//67/A
- Cote de consultation / pièce:**
- Précisions sur la cote:** n° 1
- Intitulé*:** Vidimus de l'official de Paris de deux actes de Philippe le Hardi et de Philippe le Bel relatifs à deux chapelles fondées par l'abbaye de Saint-Germain-des-Prés en réparation d'excès commis contre des écoliers. [Donné à Paris]
- Options:** Date en titre, Personne en titre, Lieu en titre, Collectivité en titre
- Dates extrêmes:** 1318
- Date normalisée:** 1318-01-01 / 1318-12-31
- Présentation du contenu:** Audience associée : SIV Internet
- Deux actes vidimés:**
 - 1°. Poissy, juillet 1278 : Philippe III le Hardi enjoint à l'abbaye Saint-Germain-des-Prés de fonder deux chapelles en réparation d'excès commis contre des écoliers au Pré-aux-Clercs et du meurtre de deux d'entre eux.
 - 2°. Paris, juillet 1286 : Philippe IV le Bel établit une rente sur le Châtelet pour l'entretien des deux chapelles fondées par l'abbaye de Saint-Germain-des-Prés, et dont elle s'était délivrée en cédant au roi les foires de Saint-Germain.

Encodage en XML EAD (dtd ead SIA AN) – J.-F. Moufflet, 2020

Contexte

Description physique

Original, parchemin.

Historique de la conservation

Audience associée :

SIV Internet

1624 : archives de l'Université de Paris, cote A. 6. d. (d'après l'inventaire BIS/MSAU 103).


Autres instruments de recherche

- Du Boulay, *Histoire de l'Université de Paris...*, t. III, p. 480.
- Jourdain, *Index chronologicus...*, n° CCLIII et CCLXXXV



Une preuve de concept pour démontrer la faisabilité et l'intérêt de la sémantisation des données d'ORESM

par Florence Clavaud (Lab des Archives nationales, direction du Numérique et de la Conservation, Archives nationales ; courriel : florence.clavaud@culture.gouv.fr)



Objectifs de la preuve de concept

Montrer :

- Que les données produites dans le cadre du projet peuvent être converties en graphe de connaissances
 - On parle ici de graphe d'entités liées en RDF (technologie du web de données)
- Que cette approche est intéressante pour le projet

Pour mémoire, parmi les objectifs scientifiques du projet :

- mieux connaître les personnes (organismes et surtout personnes physiques) liées à l'université de Paris et le réseau qu'elles forment
- mieux connaître la composition du fonds d'archives de l'université au Moyen Âge et sous l'Ancien Régime

Choix initiaux

- Temps disponible avant la conférence très limité (une vingtaine d'heures)
- Données pas toutes accessibles
> donc une preuve de concept au **périmètre et résultats concrets volontairement restreints**
- **Données prises en compte : descriptions très précises (par Louise Gousseau) des actes conservés aux Archives nationales dans la série M et provenant des archives du collège des Cholets.**
La version traitée du fichier de dépouillement décrit **188 pièces**.
- **Des données au volume réduit mais qui sont au cœur du projet (et seront au centre du graphe)**
- Inventaire de ces archives dans son état initial (fichier XML/EAD tel que présenté par Jean-François Moufflet) pas pris en compte (mais ce point ne posera pas de problème)

Aperçu des données dans leur format initial

Fichier Excel ; une ligne par pièce traitée ;
plusieurs dizaines de colonnes ; structure du
fichier et consignes de saisie définies par le
GT métadonnées

Ci-contre, en présentant les données
verticalement pour faciliter la lecture pour
les humains, l'exemple des données réunies
sur la première pièce analysée (cotée AN,
M//111 pièce 1)

```
identifiant : c-4hjwxyobu-lko65o85kwly↵
cote : FRAN M//111/1↵
analyse : Vidimus de l'official de Paris d'une bulle de Martin IV adressée à Jean [Cholet], ↵
cardinal prêtre du titre de Sainte-Cécile, par laquelle il lui permet de tester et de ↵
disposer de ses biens meubles ecclésiastiques↵
Type_de_document : Vidimus ↵
Date_forme_litteraire : 10 novembre 1289↵
Date_forme_normalisee : 1289-11-10↵
Lieu_de_passage_acte_forme_litteraire : Paris↵
Lieu_de_passage_acte_forme_normalisee : Paris↵
Lieu_de_passage_acte_code_INSEE_ou_code_postal : 75056↵
Lieux_cites_dans_analyse_forme_litteraire : Official de Paris↵
Lieux_cites_dans_analyse_forme_normalisee : Archevêché de Paris↵
Lieux_cites_dans_analyse_coordonnees_geographiques : 48.8497, 2.3501↵
Importance_materielle : 1 pièce↵
Support : Parchemin↵
Hauteur_x_largeur_mm : 182 x 195↵
Personnes_physiques : [auteur] Martin IV (pape ; 1220-1281) [0000 0000 5315 0353 ; 086043064 ; ↵
[ ] ; Q227719] | [destinataire] JEAN Cholet (cardinal ; 12.-1293) [0000000017204924 ; ↵
167364952 ; [ ] ; Q1772058]↵
Personnes_morales : [auteur] Officialité de Paris (tribunal ; 11.-17..) [ [ ] ; 185265235 ; [ ] ; ↵
Q99775030] ↵
Lieu_de_conservation_antérieur : Université de Paris ; [ ] ; [ ] ; [ ] ; [ ] ; ↵
Statut : Original↵
Date_temps_Acte_vidime : 1281-12-28↵
Date_lieu_Acte_vidime : Viterbe ; Viterbe (Italie) ; [ ] ; https://fr.wikipedia.org/wiki/Viterbe↵
Analyse_Acte_vidime : Bulle de Martin IV adressée à Jean [Cholet], cardinal prêtre du titre de ↵
Sainte-Cécile, par laquelle il lui permet de tester et de disposer de ses biens meubles ↵
ecclésiastiques↵
Cote_Acte_vidime : lost-0c91c21b-19ee-4a54-bb13-0f4fa35f1d52↵
Mentions_dorsales_anciennes_analyses : "Vidimus officialis parisiensis bulla Martini 4i qui ↵
permettoit au cardinal Cholet de disposer de ses biens meubles ecclésiastiques I.C." | "M ↵
111 n° 1" | "[?] domini Choleti [?]" | "Sancto dei vidimus Paris officialis presentiam [?]" ↵
| il y a un C barré | "troisième pièce de la 44e" | ↵
"D"↵
Autres_mentions : Non↵
Anciennes_cotations_non_identifiees : I.C. | D | «Troisième pièce de la 44e ↵
liasse»↵
Nombre_approximatif_de_noms_presentes : 3↵
Lieu_de_conservation_actuel : Archives nationales↵
Identifiant_du_lieu_de_conservation : FRAN↵
Langue_des_documents : Latin↵
Alphabet : Latin ↵
Liste_des_elements_bibliographiques : Élisabeth Rabut, «Les Cholets, Étude historique et ↵
topographique d'un collège parisien», Mémoires de la fédération des sociétés historiques et ↵
archéologiques de Paris et d'Île de France, 1970, t. 21, pp. 7-95, et 1971, t. 22, pp. ↵
119-231.↵
```

En fait beaucoup plus d'objets que les 188 actes sont décrits ou mentionnés

Toujours pour cet exemple, les rectangles concernent :

- en rouge, les **deux actes décrits** (l'acte M//111/1 et l'acte vidimé)
- en bleu, des **cotes** actuelle ou anciennes, les **personnes physiques et morales** mentionnées et identifiées, les **lieux de passage des actes**
- en vert, des **catégories/concepts**

D'autres « objets » d'intérêt sont mentionnés (rectangles jaunes) ; pour cette rapide preuve de concept nous ne les avons pas pris en compte, mais c'est bien sûr possible

Identifiant : c-4hjwxkybu-lko65o85kwy
cote : **FRAN M//111/1**
analyse : Vidimus de l'official de Paris d'une bulle de Martin IV adressée à Jean [Cholet], cardinal prêtre du titre de Sainte-Cécile, par laquelle il lui permet de tester et de disposer de ses biens meubles ecclésiastiques
type de document : **vidimus**
Date_forme_litteraire : 10 novembre 1289
Date_forme_normalisee : 1289-11-10
Lieu_de_passage_acte_forme_litteraire : **Paris**
Lieu_de_passage_acte_forme_normalisee : Paris
Lieu_de_passage_acte_code_INSEE_ou_code_postal : 75056
Lieux_cites_dans_analyse_forme_litteraire : Official de Paris
Lieux_cites_dans_analyse_forme_normalisee : Archevêché de Paris
Lieux_cites_dans_analyse_coordonnees_geographiques : 48.8497, 2.3501
Importance_materielle : 1 pièce
Support : **Parchemin**
Hauteur_x_largeur_mm : 182 x 195
Personnes_physiques : **[auteur] Martin IV (pape ; 1220-1281) [0000 0000 5315 0353 ; 086043064 ; [] ; Q227719] | [destinataire] JEAN CHOLET (cardinal ; 12.-1293) [0000000017204924 ; 167364952 ; [] ; Q1772038]**
Personnes_morales : **[auteur] Officialité de Paris (tribunal ; 11.-17..) [[] ; 185265235 ; [] ; Q99775030]**
Lieu_de_conservation_antérieur : **Université de Paris** ; [] ; [] ; [] ; []
Statut : **Original**
Date_temps_Acte_vidime : 1281-12-28
Date_lieu_Acte_vidime : **Viterbe ; Viterbe (Italie) ; [] ; https://fr.wikipedia.org/wiki/Viterbe**
Analyse_Acte_vidime : **Bulle de Martin IV adressée à Jean [Cholet], cardinal prêtre du titre de Sainte-Cécile, par laquelle il lui permet de tester et de disposer de ses biens meubles ecclésiastiques**
Cote_Acte_vidime : lost-0c91c21b-19ee-4a54-bb13-0f4fa35f1d52
Mentions_dorsales_anciennes_analyses : **"Vidimus officialis parisiensis bulla Martini 4i qui permettoit au cardinal Cholet de disposer de ses biens meubles ecclésiastiques I.C." | "M 111 n° 1" | "[?] domini Choleti [?]" | "Sancto dei vidimus Paris officialis presentiam [?]" | il y a un C barré | "troisième pièce de la 44e" | "D"**
Autres_mentions : Non
Anciennes_cotations_non_identifiees : **I.C.** | **D** | **«Troisième pièce de la 44e**
liasse»
Nombre_approximatif_de_noms_presentes : 3
Lieu_de_conservation_actuel : **Archives nationales**
Identifiant_du_lieu_de_conservation : FRAN
Langue_des_documents : **Latin**
Alphabet : **Latin**
Liste_des_elements_bibliographiques : **Elisabeth Rabut, «Les Cholets, Étude historique et topographique d'un collège parisien», Mémoires de la fédération des sociétés historiques et archéologiques de Paris et d'Île de France, 1970, t. 21, pp. 7-95, et 1971, t. 22, pp. 119-231.**

Problématique

- Le format employé est pratique pour saisir les données, mais son contenu est difficile à exploiter en l'état
- Passer à un graphe d'entités liées (en regroupant les données par entité décrite), plus précisément à un graphe de connaissances / sémantique.

Qualités intéressantes de ce type de représentation technique pour le projet :

- cadre formel (RDF), formats de sérialisation, technologies d'exploitation (comme SPARQL) font l'objet de standards publics du W3C ;
- principes de base de RDF simples (des phrases en trois segments, sujet-prédicat-objet; les sujets et prédicats ont obligatoirement un URI) ;
- **on peut atteindre une très grande granularité**, ce qui rend les données finement exploitables
- **RDF est fait pour le liage de données et pour construire de l'interopérabilité**
- par définition **un graphe RDF peut être étendu ou affiné autant que nécessaire** (tout comme les données de ce projet de recherche, qui vont évoluer tout au long du projet), il n'a pas de borne
- le vocabulaire et les règles formelles utilisables dans le graphe peuvent être définis au moyen d'une ontologie, qui peut elle-même être écrite en utilisant les standards RDFS et OWL du W3C
- modélisation ontologique orientée objet, ce qui permet de produire un modèle extensible et d'articuler plusieurs ontologies entre elles
- **possibilités de raisonnement** (construction de nouveaux triplets RDF par inférence)

Processus suivi

- Choix d'une ou plusieurs ontologies de référence
- Modélisation des données conformément à cette ontologie
- Conversion des données (itérations)
- Import des données obtenues dans une base RDF
- Premières requêtes SPARQL
- Mise en place d'une première interface de recherche graphique

Un modèle conceptuel et une ontologie tout indiqués pour le projet : ICA Records in Contexts (RiC)

- La dernière génération des standards internationaux pour la description des archives
- Élaboré par un groupe d'experts du Conseil international des Archives (ICA/EGAD)
- Version actuelle : RiC 0.2 (draft, à un niveau de maturité important et suffisant)
 - un modèle conceptuel entités-relations, RiC-CM (juillet 2021)
 - une ontologie OWL, transposition de RiC-CM, plus précise, RiC-O (février 2021)
 - un manuel de mise en œuvre et une introduction générale (à venir)
- Recommandation officielle 1.0 à paraître en 2022
- Initiative lancée en 2012, comparable, pour les archives, à celles des autres organismes internationaux représentant les GLAM ([IFLA LRM](#), [CIDOC CRM](#))

RiC-O

- Transposition en OWL 2 de RiC-CM
IRI : <https://www.ica.org/standards/RiC/ontology>
- Une ontologie de domaine pour la description des archives
- **Le modèle technique de référence pour publier des jeux de métadonnées archivistiques en RDF de qualité**
- Générique, extensible tout en étant précis et riche ; **permet de décrire aussi les entités contextuelles aux archives (agents producteurs, lieux...) et les relations entre toutes ces entités**
Nombreuses propriétés et relations applicables aux agents ; système de classes de relations n-aires
- Fichiers source, exemples, diagrammes, disponibles sous licence CC-BY 4.0 dans un dépôt public sur GitHub : <https://github.com/ICA-EGAD/RiC-O>
- Site web d'information qui donne notamment une liste des projets qui utilisent RiC-O
- **Bien connue du Lab des Archives nationales et déjà mise en œuvre dans le cadre de plusieurs projets**

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

OWL source file: 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

Creators:

Florence Clavaud (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 Pitt (University of Virginia, USA), chair of EGAD

Mia Herrala (National Archives of Finland), member of EGAD

Tobias Wildi (Docuteam 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 Records in Contexts Conceptual Model (RiC-CM).

Premier travail de modélisation

Objectif : définir les contours du graphe à produire

Points d'attention :

- prendre en compte toutes les données existantes dans le fichier CSV sinon s'assurer que les données non retenues pourront l'être
- identifier les lacunes dans l'ontologie de référence le cas échéant, définir les composants manquants
- évaluer le travail (ici, fait à l'occasion des opérations itératives de conversion)

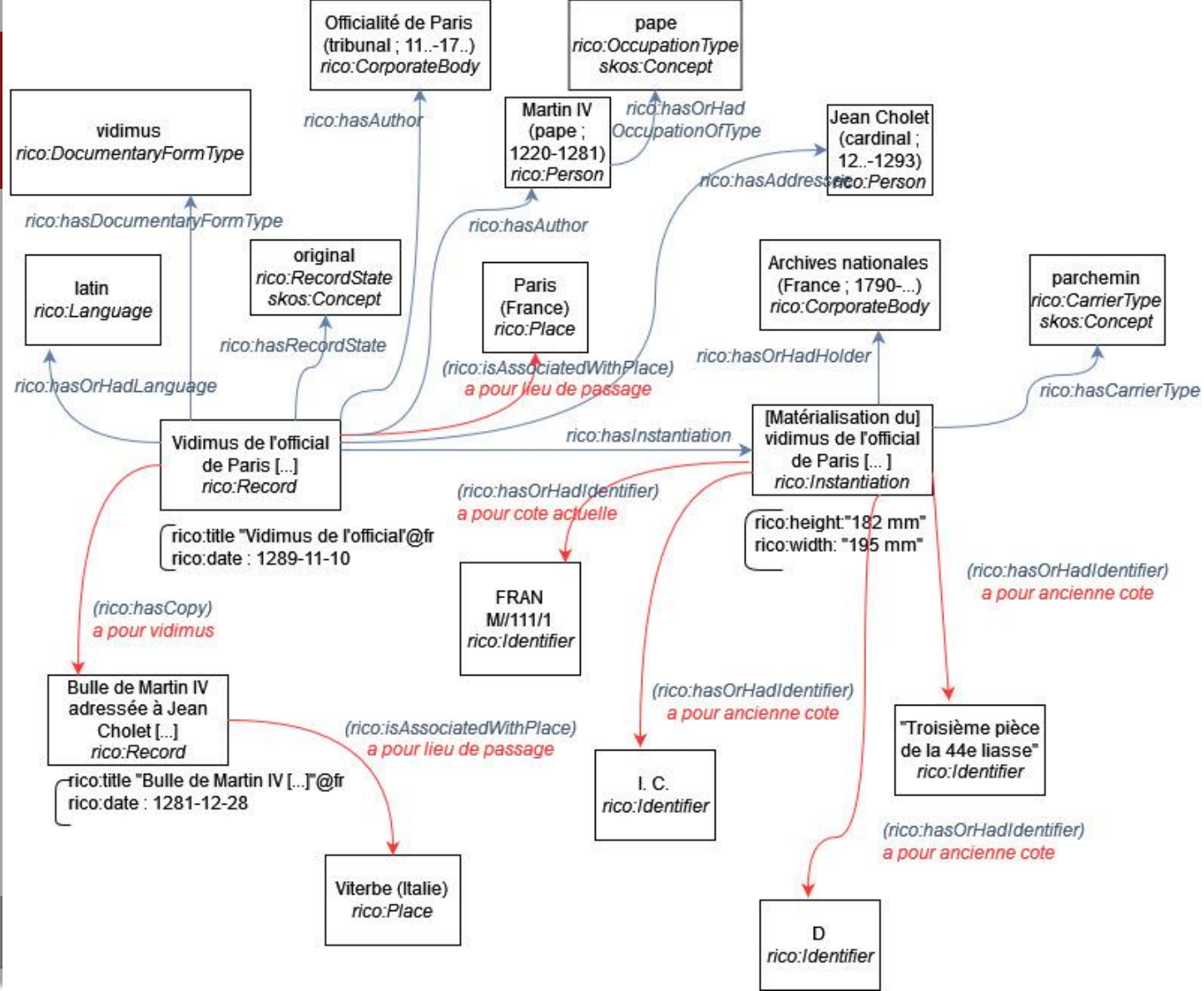
> Une première ontologie pour le projet ORESM :

- extension de RiC-O (tout ce qui est défini dans RiC-O s'y retrouve, plus quelques propriétés spécifiques au monde médiéval étudié, elles-mêmes articulées aux propriétés RiC-O existantes)
- pourra être enrichie par la suite

Aperçu du travail de modélisation

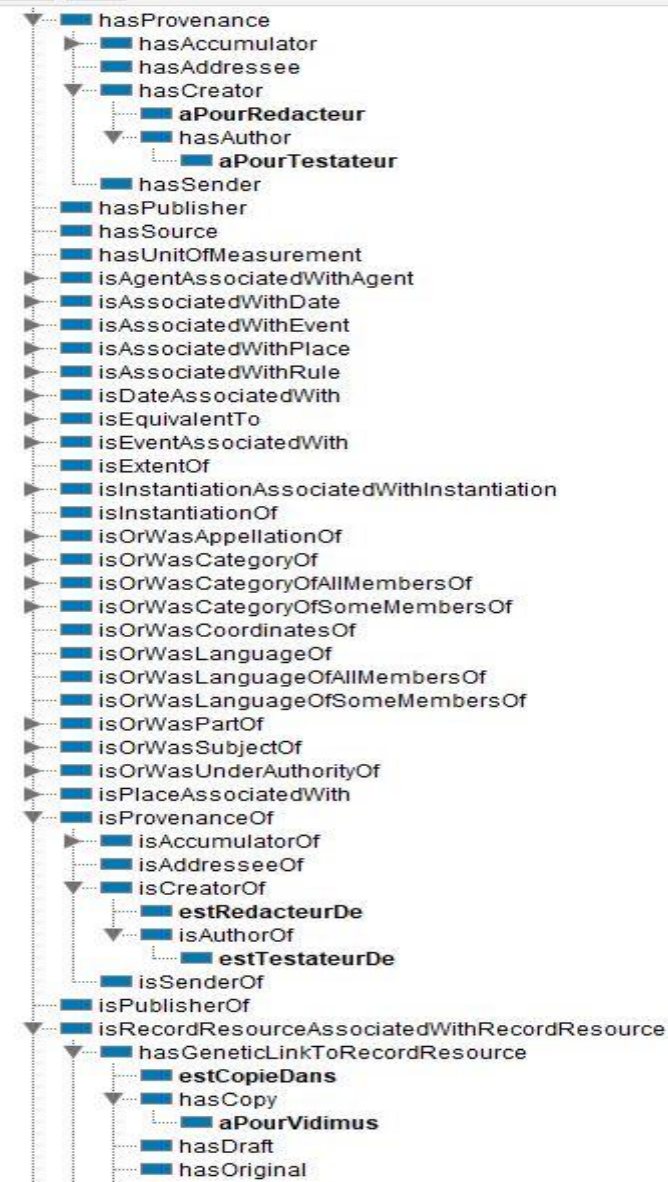
Diagramme montrant une première modélisation RDF/RiC-O des données de la notice décrivant la pièce AN M//111/1.

En rouge, les relations qui n'existent pas dans RiC-O 0.2



Création d'une ontologie pour le projet

- Extension de RiC-O (donc tous les composants de RiC-O y sont disponibles)
- Espace de noms (provisoire) : <http://data.oresm.fr/ontology#>
- 27 propriétés créées pour l'instant, toutes des sous-propriétés de propriétés RiC-O : 1 datatype property, 13 object properties et leurs propriétés inverses
- Concernent les relations entre documents, et entre documents et agents (personnes physiques ou organismes)
- Ci-contre une copie d'écran de l'éditeur open source d'ontologies Protégé, montrant la hiérarchie des propriétés, et en gras quelques-unes des propriétés ajoutées



Conversion des données en jeu de données RDF/RiC-O

Technologie utilisée : XSLT

Format de sortie : XML/RDF

URIs (provisoires) formés :

- à partir du numéro d'ordre de la ligne, pour les documents, leurs instantiations et les actes copiés/vidimés
- à partir des noms des entités, pour les agents, les lieux, les types de documents et autres concepts permettant de catégoriser les entités

Au total, pour l'instant, on trouve dans le graphe :

- 203 actes (dont 15 vidimus)
- 512 personnes physiques
- 22 organismes
- 29 lieux

Conversion des données en jeu de données RDF/RiC-O

Un seul agent (et un seul lieu) pour plusieurs blocs de description strictement identiques

```
Personnes physiques : [auteur] Martin IV (pape ; 1220-1281) [0000 0000 5315 0353 ; 086043064 ;  
[] ; Q227719] | [destinataire] JEAN Cholet (cardinal ; 12..-1293) [0000000017204924 ;  
167364952 ; [] ; Q1772058]
```

Donc à partir des données ci-dessus, qui respectent un formalisme précis ([role] nom (titre/activité ; dates d'existence) [ISNI ; Idref; id AN; id Wikidata]) et qu'on trouve plusieurs fois dans le fichier CSV, on obtiendra la description RDF de 2 personnes.

Le rôle joué par la personne est représenté par une propriété RDF.

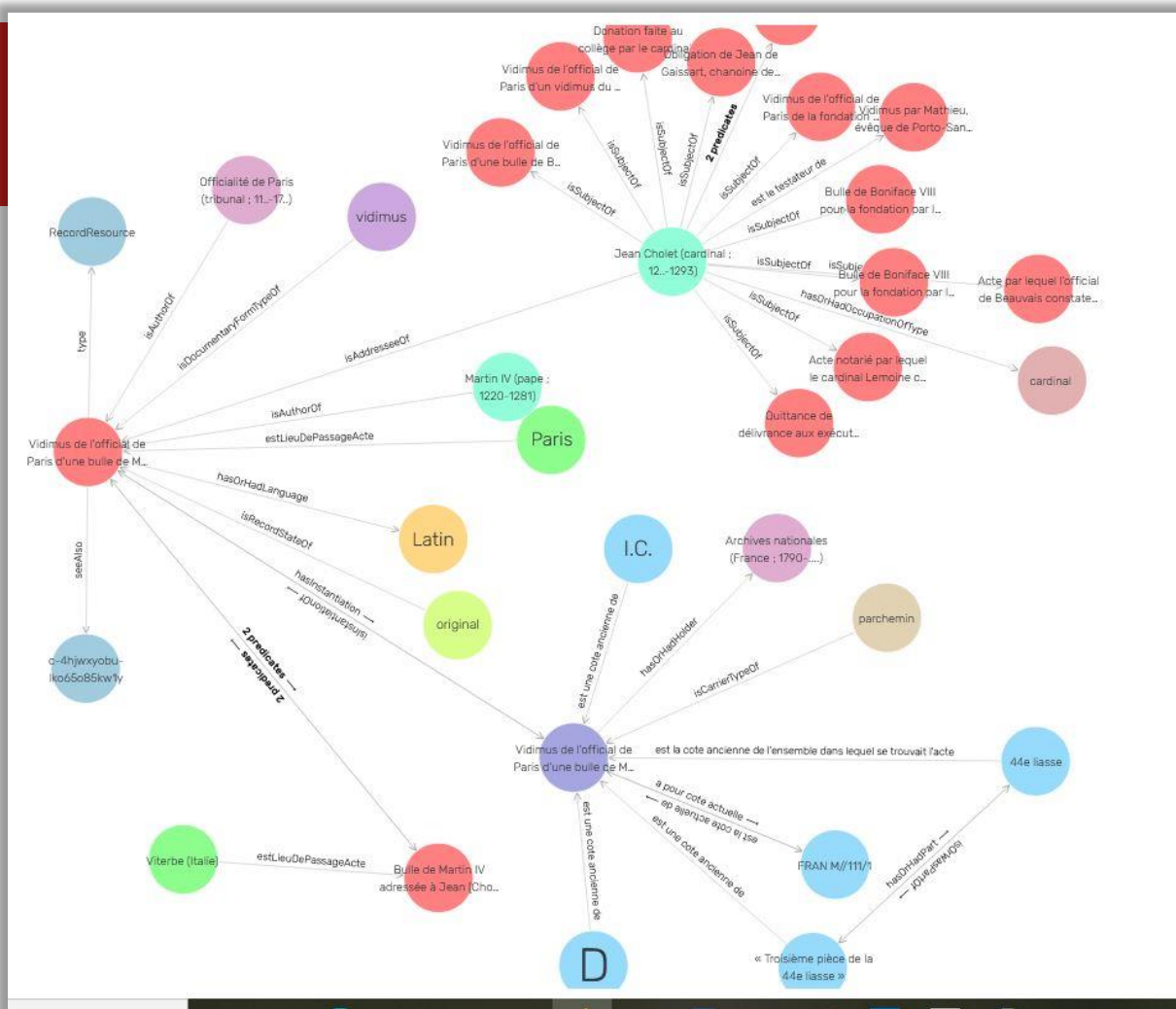
```
<rico:Person rdf:about="person/Jean%20Cholet%20%28cardinal%2%A0%3B%2012..-1293%29">  
<rdfs:label xml:lang="fr">Jean Cholet (cardinal ; 12..-1293)</rdfs:label>  
<rico:name xml:lang="fr">Jean Cholet (cardinal ; 12..-1293)</rico:name>  
<rico:hasOrHadOccupationOfType>  
  <rico:OccupationType rdf:about="occupationType/cardinal%2%A0">  
    <rdfs:label xml:lang="fr">cardinal</rdfs:label>  
    <rico:name xml:lang="fr">cardinal</rico:name>  
  </rico:OccupationType>  
</rico:hasOrHadOccupationOfType>  
<owl:sameAs rdf:resource="http://isni.org/isni/0000000017204924"/>  
<owl:sameAs rdf:resource="http://www.idref.fr/167364952/id"/>  
<owl:sameAs rdf:resource="http://www.wikidata.org/entity/Q1772058"/>  
<rico:isAddresseeOf rdf:resource="recordResource/or-antl-1"/>  
<rico:isAuthorOf rdf:resource="recordResource/or-antl-2"/>  
<oresm-onto:estTestateurDe rdf:resource="recordResource/or-antl-2"/>  
<oresm-onto:estTestateurDe rdf:resource="recordResource/or-antl-4"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-5"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-6"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-7"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-11"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-12"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-13"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-15"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-22"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-23"/>  
<rico:isOrWasSubjectOf rdf:resource="recordResource/or-antl-24"/>  
</rico:Person>
```

Visualisation du graphe obtenu

Après import dans une base (un triplestore) RDF - ici, en utilisant le logiciel libre [GraphDB Free](#), on peut *voir* le graphe et *le parcourir*.

L'image ci-contre n'est qu'une copie d'écran statique de la partie du graphe concernant la pièce AN M//111/1, à un moment de ce parcours.

En bas à gauche, les documents, en haut à droite Jean Cholet.



On peut requêter finement les données RDF obtenues, directement avec SPARQL (1)

Ci-contre, une requête pour obtenir les catégories et formes d'actes identifiées dans le corpus, et le nombre d'actes concernés.

On peut exporter les résultats dans différents formats (cf. bouton « download as » à droite).

```
1 PREFIX rico: <https://www.ica.org/standards/RIC/ontology#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX oresm-onto: <http://data.oresm.fr/ontology#>
5
6 # Les catégories juridiques et formes d'actes représentées dans le corpus
7 SELECT ?typedoc (COUNT (?a) AS ?nbactes)
8 WHERE {
9   ?tdoc a rico:DocumentaryFormType; rdfs:label ?typedoc .
10  ?tdoc rico:isDocumentaryFormTypeOf ?a .
11 }
12
13 GROUP BY ?typedoc
```

Table Raw Response Pivot Table Google Chart

Filter query results Showing results from 1 to 37 of 37. Query tool

	typedoc	nbactes
1	"acte de juridiction contentieuse"@fr	"13"^^xsd:integer
2	"acte de juridiction gracieuse"@fr	"45"^^xsd:integer
3	"acte de tabellionnage"@fr	"1"^^xsd:integer
4	"acte de vente"@fr	"3"^^xsd:integer
5	"acte notarié"@fr	"47"^^xsd:integer
6	"assignation"@fr	"1"^^xsd:integer
7	"bulle pontificale"@fr	"4"^^xsd:integer
8	"commission"@fr	"1"^^xsd:integer
9	"cédule"@fr	"1"^^xsd:integer
10	"dénombrement"@fr	"1"^^xsd:integer

Download as
JSON
XML
CSV
TSV
Binary RDF Results

Nota : ici on utilise directement le service SPARQL (*SPARQL end point*) disponible sur l'instance GraphDB locale.

On peut requêter finement les données RDF obtenues, directement avec SPARQL (2)

Ci-contre, une requête pour obtenir la liste des actes dont le bénéficiaire est le collège des Cholets.

Nota : SPARQL est un langage de recherche, et c'est aussi un *protocole de recherche* pour le web. Les requêtes peuvent donc être lancées par une application distante ; des APIs peuvent aussi être construites sur la base de ce protocole.

```
1 PREFIX r1co: <https://www.ica.org/standards/R1C/ontology#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4 PREFIX oresm-onto: <http://data.oresm.fr/ontology#>
5 # Actes dont le bénéficiaire est le collège des Cholets
6 SELECT ?a ?analyse ?date ?n WHERE
7 {
8   ?a a r1co:RecordResource ; r1co:title ?analyse ; r1co:date ?date .
9   ?a {oresm-onto:estBeneficiaireActeConsigneDans} ?b .
10  ?b r1co:name ?n .
11  FILTER (REGEX(?n, "collège des Cholets", "i"))
12 }
13 }
```

Table Raw Response Pivot Table Google Chart

Filter query results Showing results from 1 to 11 of 11. Query took 0.1s, minutes ago.

	a	analyse	date	n
1	http://data.oresm.fr/recordResource/or-ant1-46	*Testament de Mathieu de Beronna, chanoine de Chartres, par lequel il fonde un obit et 60 livres de rente et lègue ses livres au collège des Cholets ^{ORFR}	"1402-01-03" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
2	http://data.oresm.fr/recordResource/or-ant1-51	*Testament de Nicolas Cavache, exécuteur testamentaire de Martin Loutrel, qui fonde sa sépulture en la chapelle du collège du Cardinal-Lemoine et fonde pour Martin Loutrel deux messes et donne 20 livres de rente au collège des Cholets ^{ORFR}	"1418-01-21" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
3	http://data.oresm.fr/recordResource/or-ant1-69	*Constitution par Jean de Charcy, marchand et hostelier, au collège des Cholets, de 32 sous paris de rente sur une maison assise à Verberie à l'enseigne du Heaume et du Faucon ^{ORFR}	"1458-06-19" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
4	http://data.oresm.fr/recordResource/or-ant1-149	*Sentence du prévôt de Crépy-en-Valois pour le paiement d'une rente de 10 sous paris ^{ORFR}	"1530-06-08" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
5	http://data.oresm.fr/recordResource/or-ant1-143	*Testament de Maître Anthoine Meyeure, ou La Mieuure, prêtre ^{ORFR}	"1543-03-07" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
6	http://data.oresm.fr/recordResource/or-ant1-150	*Sentence du Châtelet pour le collège contre Jacques Varlet, propriétaire à Verberie ^{ORFR}	"1550-06-20" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
7	http://data.oresm.fr/recordResource/or-ant1-150	*Sentence du Châtelet pour le collège contre Jacques Varlet, propriétaire à Verberie ^{ORFR}	"1550-06-20" ^{xsd:date}	"Nicole Fauchet (procureur du collège des Cholets)" ^{ORFR}
8	http://data.oresm.fr/recordResource/or-ant1-161	*Condamnation de Nicolas Panchehardy à payer 16 sous paris de rente pour une maison à Verberie ^{ORFR}	"1560-04-18" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
9	http://data.oresm.fr/recordResource/or-ant1-67	*Reconnaissance de la rente annuelle envers le collège des Cholets par le tuteur de Jean de Sercy ^{ORFR}	"1564-09-06" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
10	http://data.oresm.fr/recordResource/or-ant1-172	*Condamnation de Guillaume Le Nattier et Jean Sauvaige, habitants de Rhuis, à payer des rentes dues au collège ^{ORFR}	"1580-04-18" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}
11	http://data.oresm.fr/recordResource/or-ant1-68	*Condamnation de Pierre de Villiers et de Emery de Sercy à payer la rente annuelle au collège des Cholets ^{ORFR}	"1580-04-18" ^{xsd:date}	"Collège des Cholets (collectivité ; 1292-1763)" ^{ORFR}

Des requêtes fédérées interrogeant plusieurs services SPARQL publics en même temps (par ex. celui d'Idref ou celui de Bibliissima lorsqu'il sera ouvert) peuvent également être construites et exécutées.

On peut concevoir, réaliser et déployer des interfaces de recherche et de consultation plus intuitives qu'un éditeur SPARQL brut

Ci-contre : copie d'écran statique d'une première interface de recherche intuitive et dynamique, configurée sur la base de l'ontologie ORESM, à l'aide de l'outil Sparnatural (en cours d'évolution dans le cadre d'un projet Archives nationales – BnF – ministère de la Culture).

La requête porte sur les cotes anciennes (sans avoir cherché à réconcilier les données sur ces cotes) connues pour des actes datant des années 1250-1350.

Rechercher [Afficher/Masquer éditeur SPARQL](#) | [Partager](#) | [Exporter](#)

Table Response 4 results in 0.024 seconds Simple view Ellipse Filter query results Page size: 100

	nom	page
1	33e liasse	<http://localhost:7200/resource?uri=http%3A%2F%2Fdata.oresm.fr%2Fidentifieur%2F33e%2520liasse>
2	44e liasse	<http://localhost:7200/resource?uri=http%3A%2F%2Fdata.oresm.fr%2Fidentifieur%2F44e%2520liasse>
3	quarante quatrième liasse	<http://localhost:7200/resource?uri=http%3A%2F%2Fdata.oresm.fr%2Fidentifieur%2Fquarante%2520quatrieme%2520liasse>
4	trente troisième liasse	<http://localhost:7200/resource?uri=http%3A%2F%2Fdata.oresm.fr%2Fidentifieur%2Ftrente%2520troisieme%2520liasse>

On peut aussi utiliser SPARQL pour inférer de nouvelles assertions (raisonner) à partir de triplets existants

Pour cela on utilise une requête de type CONSTRUCT. Exemples très simples de requêtes SPARQL auxquelles on peut penser :

- Si telle personne est la même qu'une personne décrite dans un référentiel externe, alors elle a les propriétés (dates d'existence, noms, activités, relations etc.) spécifiées dans ce référentiel externe

- Lorsque deux personnes a et b sont témoins du même acte, alors elles se connaissent (ci-contre)

- Lorsqu'un acte porte une mention de cote ancienne donnant une indication de liasse, alors il a été inclus dans cette liasse

```
1 PREFIX rico: <https://www.ica.org/standards/RiC/ontology#>
2 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
3 PREFIX oresm-onto: <http://data.oresm.fr/ontology#>
4 |
5 # Si deux personnes distinctes sont témoins du même acte, alors elles se connaissent.
6 |
7 CONSTRUCT { ?a rico:knows ?b }
8 WHERE {
9   ?a a rico:Person ; oresm-onto:estTemoInDe ?doc .
10  ?b a rico:Person ; oresm-onto:estTemoInDe ?doc .
11
12   FILTER (str(?a) > str(?b))
13 }
14
```

Table Raw Response Pivot Table Google Chart

Filter query results

Showing results from 1 to 5 of 5. Query took 0.1s, moments ago

	subject	predicate	object
1	http://data.oresm.fr/person/Jehan%20Guilidon%20%28%5B%5D%29	rico:knows	http://data.oresm.fr/person/Francoys%20Gombert%20%28%5B%5D%29
2	http://data.oresm.fr/person/Jehan%20Louaryn%20%28%5B%5D%29	rico:knows	http://data.oresm.fr/person/Francoys%20Gombert%20%28%5B%5D%29
3	http://data.oresm.fr/person/Jehan%20Louaryn%20%28%5B%5D%29	rico:knows	http://data.oresm.fr/person/Jehan%20Guilidon%20%28%5B%5D%29
4	http://data.oresm.fr/person/Radulphus%20de%20Rogy%20%28%5B%5D%29	rico:knows	http://data.oresm.fr/person/Anthonius%20Caignet%20%28%5B%5D%29
5	http://data.oresm.fr/person/Stephanus%20Paguet%20%28%5B%5D%29	rico:knows	http://data.oresm.fr/person/Guilelmus%20Frenyn%20%28%5B%5D%29

Nota : une personne ayant les droits ad hoc pourrait stocker les assertions obtenues dans la base de connaissances RDF, augmentant ainsi son contenu (utilisation de SPARQL comme langage de mise à jour des données, et d'une requête de type INSERT).

En guise de conclusion

La preuve de concept (la première mise en œuvre de RiC-O, à notre connaissance, pour la description fine d'archives médiévales et de leurs contextes) montre la faisabilité de la sémantisation des données produites, et l'intérêt de cette approche, notamment :

- l'expressivité et les potentiels d'exploitation et de liage du graphe
- la plasticité et l'évolutivité des données RDF

En guise de conclusion : aller plus loin

Beaucoup de choses restent à faire :

- **Consolidation du processus de conversion** (certaines données ont été mises de côté dans le fichier CSV source ; de plus un premier travail de réconciliation est à faire entre les données relatives à certaines personnes physiques et morales)
- **Extension du processus aux autres données descriptives des archives étudiées** (inventaires en XML/EAD, autres fichiers CSV)
- **Articulation des données obtenues avec les vocabulaires et référentiels existants, notamment ceux des AN** en cours d'enrichissement (vocabulaires des types et statuts des documents, des activités des personnes, notices d'agents et de lieux...) ; donc, premiers enrichissements des données ORESM sur les personnes et les lieux
- **Réconciliation des données obtenues sur les personnes physiques, organismes et lieux avec les données *Studium Parisiense*** (donc nouveaux enrichissements des données ORESM, dédoublonnages, etc.)
- Modélisation des incertitudes et hypothèses pour l'identification, la datation, la catégorisation des agents, l'organisation ancienne des documents, et toute autre propriété (RiC-O possède les briques de base permettant cette extension)

- Mise en place d'une architecture appropriée pour rendre accessible le jeu de données RDF sur un serveur web public
- Conception, réalisation et mise en production d'une interface complète de recherche et de consultation
- Mise en place des dispositifs (règles d'inférence, etc.) permettant de générer simplement les assertions incertaines ou hypothétiques, et de documenter ces assertions (date, auteur, degré d'incertitude, etc.)
-

Merci de votre attention !

À votre disposition pour en reparler.

Il est aussi possible de créer un dépôt public de sources sur GitHub pour y placer les données produites (ontologie, graphe, etc.).