

# Sparnatural, un éditeur graphique souple et intuitif pour explorer des graphes de connaissances

Florence Clavaud

Thomas Francart

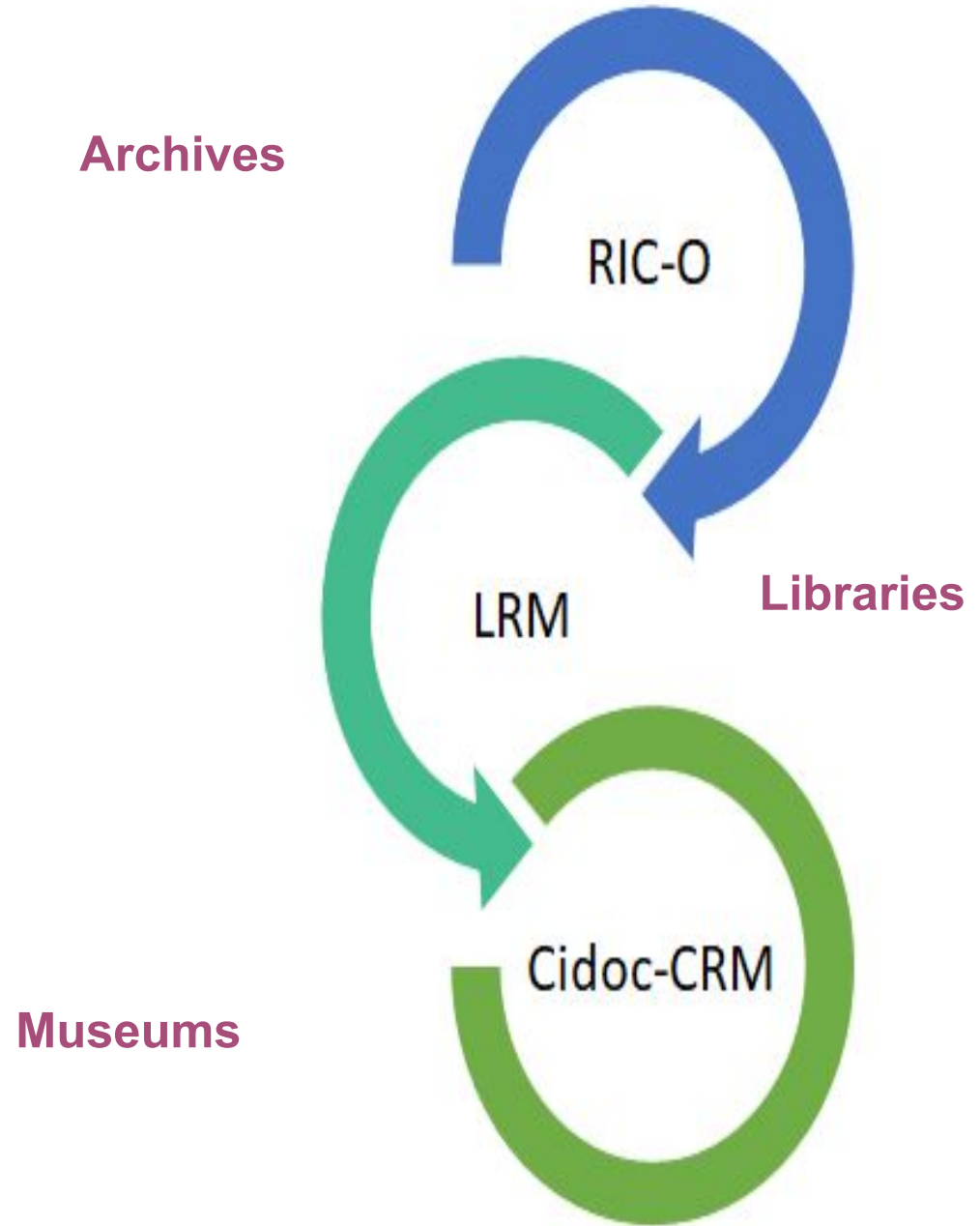


Colloque Humanistica 2022  
samedi 21 mai 2022

From 01-01-1500 to 31-12...

# Contexte du projet

- En France, une **stratégie nationale** pour la mise en œuvre des principes des données liées ouvertes (Linked Open Data) pour les métadonnées culturelles
- Les métadonnées culturelles évoluent vers des **graphes de connaissances** ; l'un des objectifs est de mieux répondre aux besoins des utilisateurs et de leur procurer des métadonnées structurées selon les normes et principes FAIR
- **Besoin de nouvelles interfaces**, en particulier **pour faire des recherches** dans ces jeux de données liées



# Sparnatural - <http://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

Un outil pour **améliorer la “découvrabilité” des contenus culturels** et pour **lever plusieurs verrous** :

- peu d'utilisateurs connaissent SPARQL ;
- peu connaissent les modèles et ontologies sous-jacents pour le patrimoine culturel, qui sont souvent complexes ;
- ces graphes de connaissances 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.*).

# Développement d'une nouvelle version de Sparnatural

- Travail réalisé dans le cadre d'un projet associant le ministère de la Culture français, la Bibliothèque nationale de France (BnF), les Archives nationales de France (ANF) et la société Sparna
  - élaboration des spécifications en 2020 ; développements commencés en août 2021, quasiment achevés aujourd'hui
  - deux démonstrateurs mis en place (travail en cours)
  - retours des utilisateurs pris en compte (plusieurs ateliers fin 2021)
- Parmi les principales améliorations résultant de ce projet :
  - une plus grande facilité de **configuration** (au moyen d'une ontologie OWL) ;
  - **nouvelles fonctionnalités de recherche** (prise en compte de l'opérateur SPARQL OPTIONAL, sélection des colonnes faisant partie des données résultat, sélection de concepts ou entités dans une arborescence... ) ;
  - une grande attention apportée à l'**UX design et à l'ergonomie** ('tooltips', possibilité de décliner l'interface en plusieurs langues, ateliers utilisateurs... ) ;
  - une **documentation complète**, en anglais et en français (en cours).

Soutenu  
par



MINISTÈRE  
DE LA CULTURE

Liberté  
Égalité  
Fraternité

ARCHIVES  
NATIONALES

{ BnF

Bibliothèque  
nationale de France

SPARNA

# Les deux démonstrateurs

- *Travail quasiment achevé* (sera terminé dans un mois)
  - **Le démonstrateur de la BnF**
    - Permet d'explorer l'ensemble des données RDF disponibles dans <https://data.bnf.fr> et recherchables via le SPARQL endpoint <https://data.bnf.fr/sparql/>. Les jeux de données sont conformes à une ontologie basée sur IFLA [LRM](#)
  - **Le démonstrateur des ANF**
    - Permet d'explorer un jeu de données conformes à la récente ontologie ICA [Records in Contexts ontology](#) (v0.2, février 2021)
    - Ce jeu de données est la **première publication à grande échelle, par les ANF (et en France), de métadonnées archivistiques en RDF conformes à RiC-O\***
      - Les métadonnées source sont des instruments de recherche archivistiques (env. 1600) en XML/EAD 2002, des notices décrivant des agents, des concepts et des lieux ; le tout concerne les archives des 40 premières études notariales parisiennes conservées aux ANF (soit le tiers de ces archives, environ 9 km linéaires de documents, décrits en environ 50 millions de triplets RDF).
- \*Noter aussi que des jeux de données RDF/RiC-O ont été publiés en décembre 2021 par l'équipe ANR du projet ALEGORIA ; voir <https://www.alegoria-project.fr/en/Metadata>

Comment valoriser un graphe de connaissances  
*en tant que graphe de connaissances ?*

(et non pas derrière un moteur de recherche à facettes ou derrière des liens de navigation cliquables)

Comment favoriser les interactions exploratoires (essai/erreur) des utilisateurs pour *partir à la découverte des données* ?

# Comment réconcilier des modèles conceptuels génériques avec les points de vue spécifiques des utilisateurs ?

Modèle conceptuel d'interopérabilité :  
FRBR, CIDOC-CRM, RiC-O, etc.

- Niveau d'abstraction élevé
- Vocabulaire conceptuel
- Logique métier



Utilisateurs

- Niveau d'abstraction "usuel dans la communauté"
- Vocabulaire précis
- Logique de recherche



# Démo

Voir la vidéo diffusée sur YouTube, à la suite de la présentation faite lors de la conférence Europeana du 1<sup>er</sup> mars 2022 (durée 6 min 50 s) :

<https://youtu.be/93XV-SZD6tk> (démos à 3:15)

# Concevoir une interface de recherche basée sur Sparnatural

Se fait essentiellement en produisant une **ontologie**, qui s'appuie sur deux ontologies génériques incluses dans le code source du logiciel.

On définit donc *un modèle ontologique pour la recherche et 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>

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 son (ses) interface(s) de façon autonome.

L'enjeu est alors d'atteindre par ce biais un **compromis raisonnable** entre la **complexité du modèle métier** (qu'on peut aussi étendre) et la **nécessité de produire une interface compréhensible et performante** pour les utilisateurs.

# Concevoir une interface de recherche basée sur Sparnatural

Ci-contre, un exemple de configuration très simple pour une classe.

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.

Cela peut être intéressant dans le cas où l'ontologie utilisée pour produire les jeux de données RDF est un modèle générique.

C'est ce qui est envisagé pour des jeux de métadonnées archivistiques conformes à RiC-O.

The screenshot displays the Sparnatural web interface for configuring an ontology class. The browser address bar shows the URL: `https://sparna-git.github.io/sparnatural-demonstrateur-an/ontology#Archives`. The interface is divided into several sections:

- Class hierarchy:** A tree view on the left shows the class hierarchy starting from `owl:Thing`. The `Archives` class is highlighted under the `SparnaturalClass` category.
- Annotations:** A table lists various annotations for the `Archives` class, including:
  - `rdfs:label` (language: fr) with the value `Archives`
  - `rdfs:label` (language: en) with the value `Records`
  - `defaultLabelProperty` with the value `aPourIntitule`
  - `falcon` with the value `fad fa-archive`
  - `order` (type: xsd:integer) with the value `1`
  - `sparqlString` with the value `<https://sparnatural-demo-anf.huma-num.fr/ontology#Archive>`
  - `tooltip` (language: fr) with the value `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) with the value `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...)`
- Description:** A section below the annotations showing:
  - Equivalent To:** A plus sign (+) indicating no equivalents are currently listed.
  - SubClass Of:** A plus sign (+) and a list containing `SparnaturalClass`.

# Démo de la configuration

## Articulation entre configuration et interface

- libellés, tooltips, propriétés, etc. : Modifier un libellé

## Mapping entre éléments de la configuration et éléments du modèle

- exemple : liens inverses “a pour provenance” vs. “a produit”

## Ajouter / modifier un lien existant (organisation/activité)

## Les Datasources qui alimentent les listes ou les champs d'autocomplétion

# Facilité de déploiement de la solution :

100% côté client, une simple page web suffit

Pas d'API supplémentaire à déployer

Seul prérequis : avoir des données interrogeables en SPARQL

# Pour en savoir plus

Participez à la demi-journée de restitution détaillée du projet, le 17 juin après-midi, sur place aux Archives nationales (site de Pierrefitte-sur-Seine, auditorium) ou à distance en visioconférence.

Pour s'inscrire :

<https://framaforms.org/inscription-a-la-demi-journee-de-restitution-du-projet-spar-natural-17-juin-2022-1652342598>

Vous pouvez aussi nous contacter :

Florence Clavaud (Lab des Archives nationales) : [florence.clavaud@culture.gouv.fr](mailto:florence.clavaud@culture.gouv.fr)

Thomas Francart (société Sparna) : [thomas.francart@sparna.fr](mailto:thomas.francart@sparna.fr)



**Merci de votre attention**

Des questions ?