

# Le projet Sparnatural en quelques mots : enjeux, déroulement, principaux résultats

Florence Clavaud (Lab des Archives nationales)

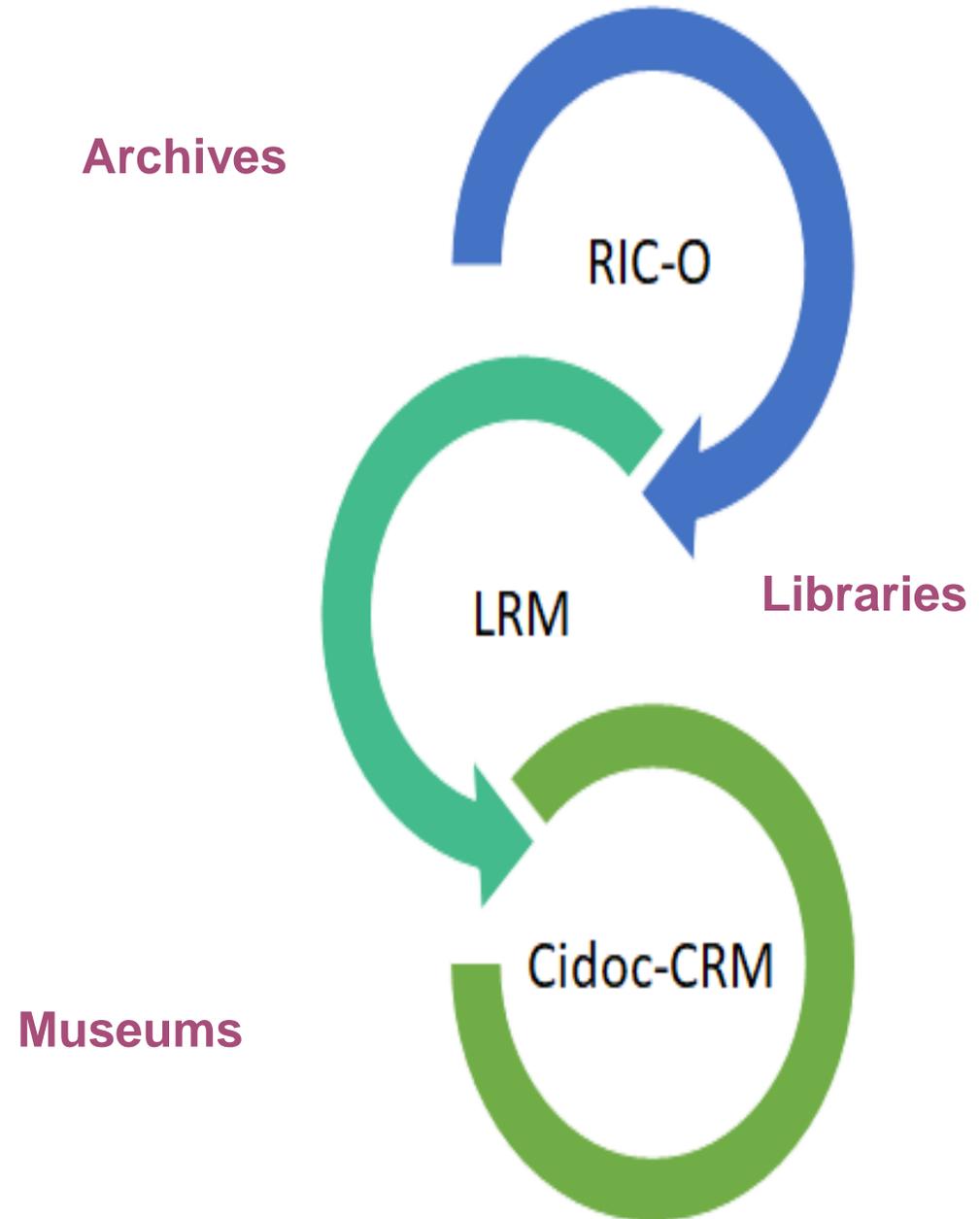
*Demi-journée de restitution du projet Sparnatural*

17 juin 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** conformes à des modèles métier ; l'un des objectifs est de **mieux répondre aux besoins des utilisateurs** et de leur procurer des métadonnées structurées selon les 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 la plateforme sémantique [OpenArchaeo](#), un projet du 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

- Projet associant le ministère de la Culture français, la Bibliothèque nationale de France (BnF), les Archives nationales de France (ANF) et la société Sparna
- 2020 : co-construction des spécifications
  - trois entités publiques, des différences dans les contextes et enjeux
  - mais un consensus atteint sur les principaux besoins et objectifs
  - deux idées retenues par ailleurs : réaliser deux démonstrateurs ; approche agile, associant les utilisateurs finaux
- Premier semestre 2021 :
  - Constitution d'un groupement de commandes (convention tripartite signée en juin 2021) pour pouvoir administrativement monter le projet, désigner un coordinateur qui passe, exécute et paye le marché (les Archives nationales)
  - Marché public (MAPA) adjugé à la société Sparna en juillet 2021
- Début des travaux de développement en août 2021 ; travaux achevés récemment
- Retours des utilisateurs pris en compte (plusieurs ateliers fin 2021)

Soutenu  
par



MINISTÈRE  
DE LA CULTURE

*Liberté  
Égalité  
Fraternité*

ARCHIVES  
NATIONALES

{ BnF

Bibliothèque  
nationale de France

SPARNA

# Résultats du projet : une nouvelle version de Sparnatural

- Code source disponible ici : <https://github.com/sparna-git/sparnatural.eu>
  - Inclut des exemples
- 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 (encore en cours de rédaction).

# Résultats du projet : deux démonstrateurs

- **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/>.
  - Soit actuellement environ 600 millions de triplets conformes à une ontologie basée sur IFLA [LRM](#)
- **Le démonstrateur des Archives nationales**
  - Permet d'explorer un jeu de données conformes à la récente ontologie ICA [Records in Contexts ontology](#) (v0.2, février 2021).
  - Soit actuellement environ 50 millions de triplets, produits dans le cadre du projet et chargés dans une base RDF dédiée, hébergée sur la grille de services de la TGIR Huma-Num du CNRS
  - Il s'agit de la **première opération de sémantisation (conversion en RDF) conformément à RiC-O et de publication à grande échelle, par les ANF (et en France), de métadonnées archivistiques « classiques »**

# Ce que nous allons vous présenter aujourd'hui

- Nous avons déjà présenté très rapidement le projet, dans son état du moment :
  - le 1er mars dernier, lors de la conférence organisée en ligne par Europeana (voir [la vidéo produite et publiée à cette occasion](#))
  - le 22 mai dernier, lors du [colloque Humanistica 2022](#)
- Nous allons présenter aujourd'hui **plus en détail** :
  - l'**outil Sparnatural** dans sa version actuelle
  - et **pour la première fois aujourd'hui les démonstrateurs dans une première version finalisée, consultable sur internet** :  
<https://data.bnf.fr/sparnatural/>  
<https://sparna-git.github.io/sparnatural-demonstrateur-an/>

# Cette demi-journée ne clôt pas le sujet

- Il y aura d'autres présentations, notamment des travaux réalisés pour les Archives nationales
- Et surtout, l'outil et les démonstrateurs vont continuer à évoluer, grâce aux retours des utilisateurs notamment

Après la journée, si l'outil vous intéresse et que vous avez un peu de temps et quelques moyens humains, n'hésitez pas à le tester sur vos données RDF !

Merci de votre attention !