HAL open science

Produire des données géohistoriques “ FAIR ”. L’application OntoME et le consortium Data for History.

Vincent Alamercery, Francesco Beretta

▶ **To cite this version:**

Vincent Alamercery, Francesco Beretta. Produire des données géohistoriques “ FAIR ”. L’application OntoME et le consortium Data for History.. La semaine des humanités numériques, Oct 2019, Lyon, France. . halshs-02307315

HAL Id: halshs-02307315

https://shs.hal.science/halshs-02307315

Submitted on 7 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Produire des données géohistoriques « FAIR »

L'application OntoME et le consortium Data for History

Vincent Alamercery – vincent.alamercery@ens-lyon.fr
Francesco Beretta – francesco.beretta@cnrs.fr

Le LARHRA est unité mixte de recherche du CNRS, des universités Lumière Lyon 2, Jean-Moulin Lyon 3, Grenoble-Alpes et de l'École normale supérieure de Lyon, spécialisé en histoire et en histoire de l'art, pour les périodes moderne et contemporaine, du XVIe siècle au temps présent.

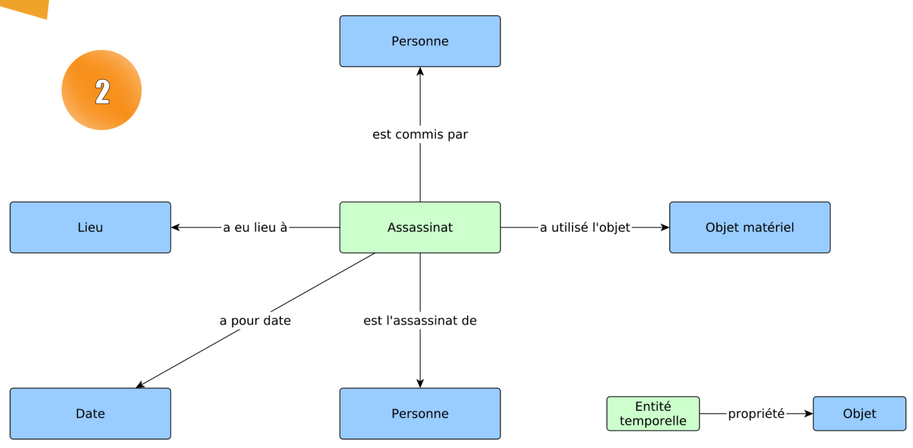
Il est organisé en six axes de recherche et trois transversalités dont l'équipe du *Pôle histoire numérique,* animé par Francesco Beretta, qui s'intéresse particulièrement aux questions méthodologiques de l'histoire numérique et mène une réflexion approfondie sur la modélisation des données géohistoriques. Le PHN travaille en collaboration étroite avec le *special interest group* du CIDOC CRM et est à l'initiative de la création du consortium international *Data for History* en 2017.

1

À partir du travail d'analyse critique de ses sources,
le plus souvent des documents d'archives,
l'historien dégage des faits
qui vont constituer son corpus, son matériau.

Pour pouvoir analyser les données ainsi produites,
il est nécessaire de penser
un modèle de données (ou « ontologie »).
qui doit expliciter la manière dont
les différentes entités (personnes, objets, lieux…)
interagissent entre elles.

2



Prenons un historien ou une historienne qui souhaite modéliser l'assassinat d'Henri IV par Ravaillac.

L'événement (ou entité temporelle), auquel on peut rattacher une date, est en vert.
Les objets, qui ont une existence physique (personnes, objets physiques…)
ou abstraite (idées, concepts…), sont en bleu.
Les propriétés que peuvent avoir ces objets sont représentées par des flèches,
et indiquent la manière dont les différentes entités interagissent.

3

data for history

Un modèle développé uniquement
pour un seul projet de recherche
empêche d'autres chercheurs de comprendre
et de réutiliser les données produites.

Le consortium européen Data for History
promeut une méthode commune de modélisation
et de gestion des données de la recherche historique.

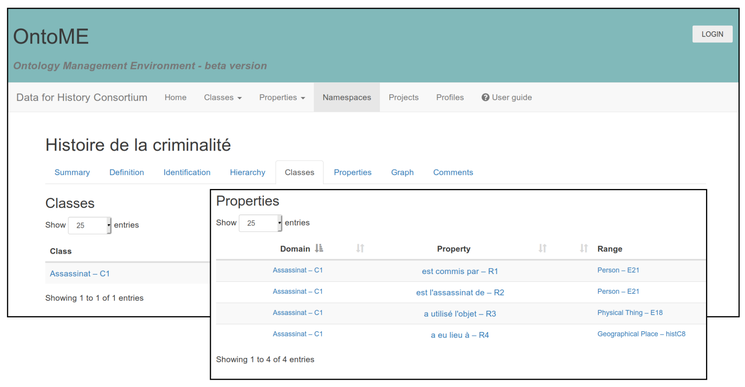
Il soutient le développement d'OntoME qui en est un élément clé.

4

OntoME

Ontology Management Environment

Développée par le Pôle histoire numérique du LARHRA,
OntoME, *Ontology management environment,*
est une application en ligne qui permet aux projets de recherche
de gérer leur ontologie d'une façon ouverte et collaborative.



Un projet peut créer des classes (objets et entités temporelles)
ainsi que des propriétés, dans un espace de noms dédié.

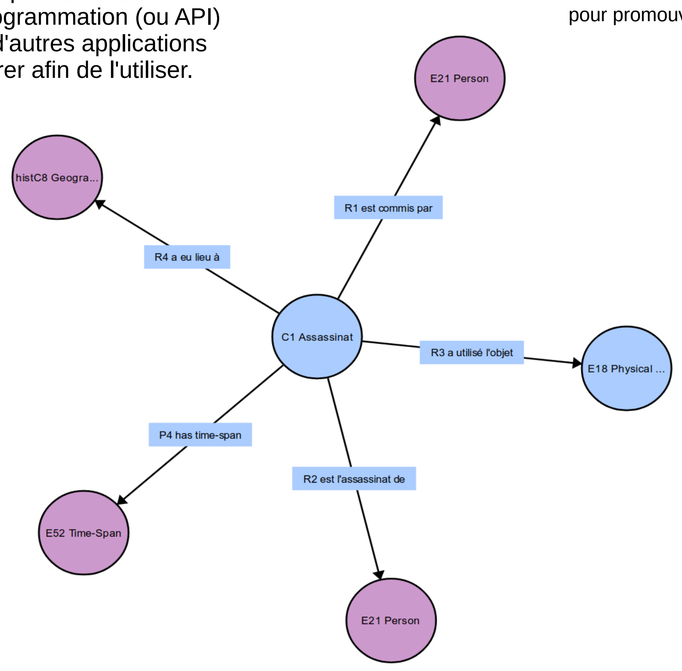
Un espace de noms désigne un ensemble conçu pour accueillir
des termes appartenant à un même domaine.
C'est un peu l'équivalent d'une langue :
il est important de connaître la langue d'origine d'un mot pour connaître son sens.

5

Les classes et propriétés peuvent ensuite être alignées,
dans une dimension hiérarchique ou d'équivalence,
avec des classes et propriétés du CIDOC CRM.

Le CIDOC CRM est une ontologie de haut niveau, norme ISO,
fournissant les définitions et la structure formelle
pour décrire les objets et leurs relations, dans le temps et l'espace.
C'est une « langue commune » qui permet aux données d'être interopérables,
c'est à dire d'être comprises quel que soit leur modèle d'origine.

6



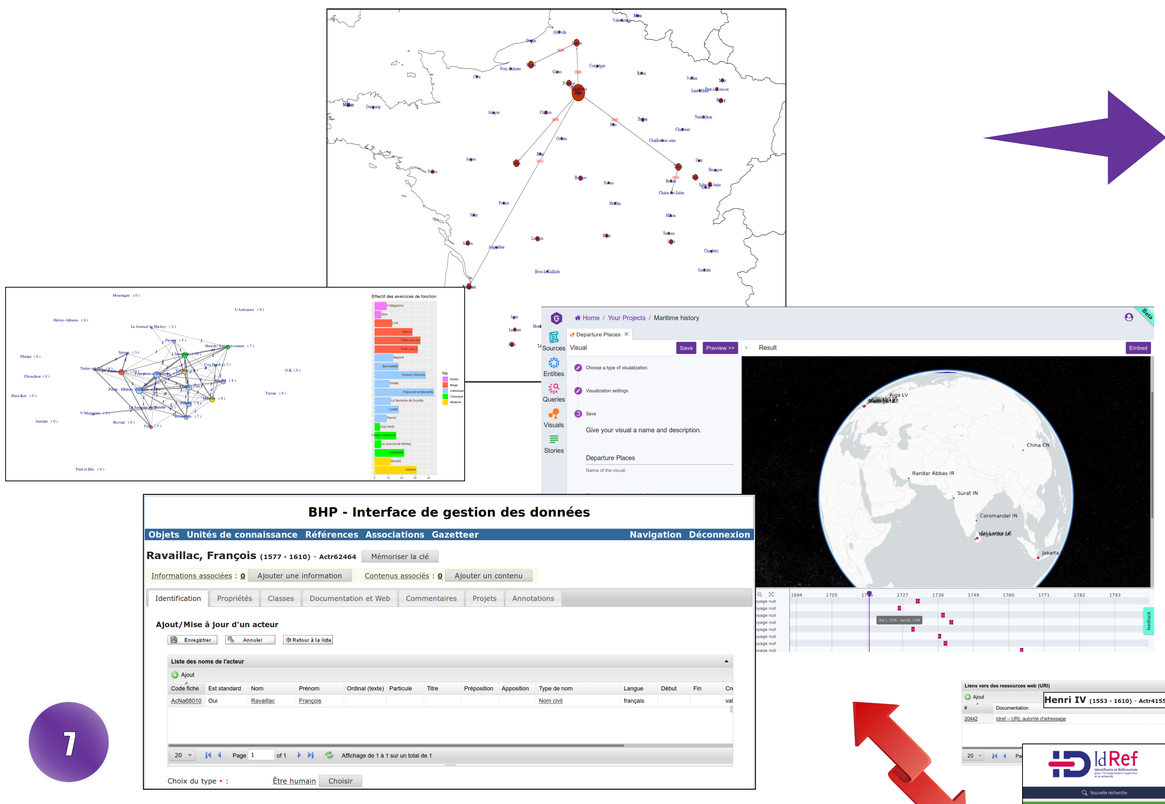
Dans OntoME, le projet de recherche peut créer un profil d'application
qui reproduit le modèle envisagé comme un ensemble cohérent de classes et de propriétés
de différents espaces de noms.

Ici, on retrouve, les classes (Cx) et propriétés (Rx) de l'espace de noms « Histoire de la criminalité »
et des classes (Ex) et propriétés (Px) complémentaires du CIDOC CRM et d'autres espaces de noms.

Le même processus peut être réalisé par n'importe quel autre projet souhaitant
rendre ses données réutilisables par d'autres chercheurs.

Le profil est publié dans une
interface de programmation (ou API)
permettant à d'autres applications
de le récupérer afin de l'utiliser.

7



Connecté à l'API d'OntoME, un environnement virtuel de recherche (ou VRE)
applique aux données le modèle construit par les chercheurs :
L'assassinat (C1) du 14 mai 1610 (P4 -> E52) rue de la Ferronerie (R4 -> histC8)
est l'assassinat d'Henri IV (R2 -> E21) commis par François Ravaillac (R1 -> E21)
où a été utilisé un couteau (R3 -> E18).

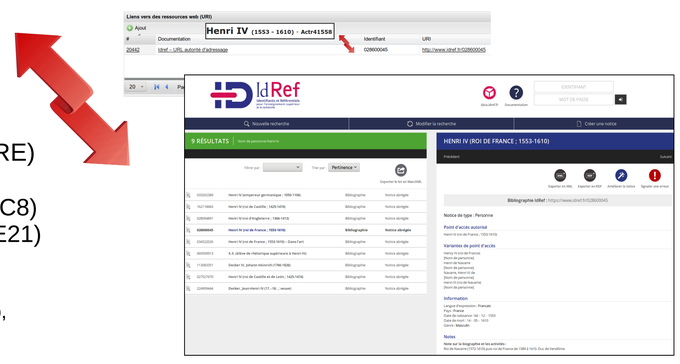
Un environnement virtuel de recherche (ou VRE), comme Geovistory de la startup KleioLab,
est une application en ligne permettant d'importer, de produire, de gérer et d'analyser
les données de la recherche.

8

SPARQL endpoint
FAIR data

À la fin du processus,
les données sont mises à disposition sur une page web
ou un point d'accès SPARQL
qui permettent à quiconque de les interroger
et de faire sa propre analyse.

L'ensemble de ces étapes permet de disposer de données
conformes aux principes **FAIR** :
findable, accessible, interoperable, reusable,
qui sont un des piliers de la science ouverte.



Les environnements virtuels de recherche sont connectés
à des serveurs de notices d'autorité qui servent à identifier sans ambiguïté
des personnes, des institutions, des lieux ou d'autres concepts,
en leur attribuant un identifiant unique.
Ainsi, Henri IV, roi de France, ne sera pas confondu
avec Henri IV, roi de Castille.

Les autorités sont maintenues par les bibliothèques nationales de chaque pays
ou de grandes institutions comme l'Agence bibliographique de l'enseignement supérieur (ABES)
qui gère les identifiants et référentiels (IdRef) du catalogue unifié des bibliothèques universitaires.
Le LARHRA travaille en coopération avec l'ABES
pour promouvoir l'interconnexion des identifiants d'autorité et leur enrichissement.

@phn_larhra larhra.fr – dataforhistory.org – ontome.dataforhistory.org

fête de la Science
Semaine des humanités numériques 7 – 11 octobre 2019

Crédits
- Vignette : Archives départementales des Vosges, 8 T 990 3189 - Histoire de France, pl. 9 (1864).
- Logo Data for History : Annett Riechert – annett-riechert-design.de

Attribution - Partage dans les Mêmes Conditions 4.0 International (CC BY-SA 4.0)