

Les Carnets du LARHRA

ISSN : 2648-1782

Publisher : Laboratoire de recherche historique Rhône-Alpes

2023 | 1 | 2023

Numéro spécial "Les 20 ans du LARHRA"

INFRASTRUCTURE NUMÉRIQUE DE RECHERCHE EN SCIENCES HUMAINES ET SOCIALES

Francesco Beretta, Julien Caranton and Vincent Alamercery

 <https://publications-prairial.fr/larhra/index.php?id=799>

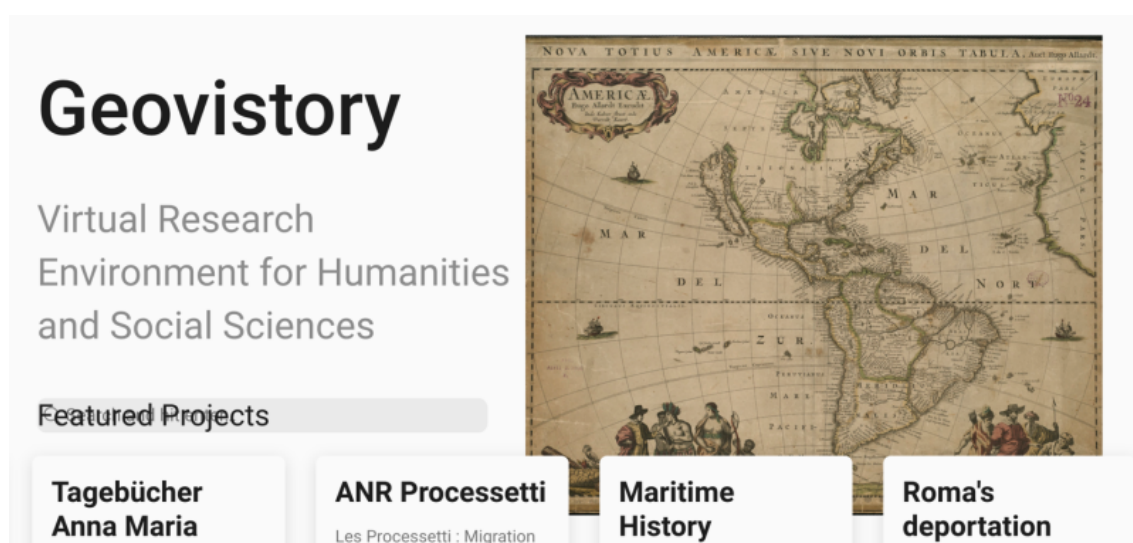
Electronic reference

Francesco Beretta, Julien Caranton and Vincent Alamercery, « INFRASTRUCTURE NUMÉRIQUE DE RECHERCHE EN SCIENCES HUMAINES ET SOCIALES », *Les Carnets du LARHRA* [Online], 2023 | 1 | 2023, Online since 18 décembre 2023, connection on 22 décembre 2023. URL : <https://publications-prairial.fr/larhra/index.php?id=799>

INFRASTRUCTURE NUMÉRIQUE DE RECHERCHE EN SCIENCES HUMAINES ET SOCIALES

Francesco Beretta, Julien Caranton and Vincent Alamercery

TEXT



Capture d'écran, <https://www.geovistory.org/>, 15/12/2023

- 1 Le LARHRA a une longue tradition dans l'utilisation d'outils numériques pour la recherche. Le laboratoire se dote d'un Pôle méthodes en 2007, sous la responsabilité de Jean-Pierre Dedieu, puis de Francesco Beretta en 2009. L'idée de départ est d'offrir à la communauté une méthodologie permettant la production collaborative et cumulative de données de la recherche, afin d'en faciliter l'analyse, avec des outils adaptés, et la réutilisation. En 2013, le Pôle méthodes devient le Pôle histoire numérique, se plaçant ainsi dans la grande famille des humanités numériques. Afin d'insister sur une démarche disciplinaire et épistémologique propre, il devient en 2021 un axe spécifique de Recherche en histoire numérique, sous la responsabilité de Julien Caranton et Vincent Alamercery.
- 2 Dans ce contexte est né en 2008 le projet symogih.org (Système modulaire de gestion de l'information historique)¹ visant la production collaborative de données. Ainsi, le projet ANR SIPPAF a utilisé cette infrastructure pour mettre en place un système d'information prosopographique consacré au patronat français (XIX^e-XX^e siècles²).

Les données continuent à être enrichies et utilisées plus de dix ans après la fin du financement (en 2010), notamment dans le cadre du projet Siprojuris consacré aux professeurs de droit en France de 1804 à 1950³. Ces deux projets disposent chacun d'un site Web dédié, bien que la production des données se fasse dans l'environnement virtuel de recherche unique symogih.org, ce qui favorise la réutilisation, l'amélioration de la qualité des données et le prolongement de leur durée de vie. Trois thèses de doctorat⁴ et une trentaine d'autres projets ont bénéficié de cette infrastructure de recherche.

- 3 Dès 2013, une épreuve de concept a été menée avec les données du projet Siprojuris afin de les transformer en ontologie (ensemble de concepts permettant de donner un sens aux informations) et les publier sous forme de données liées ouvertes, en les mettant à disposition sur un point d'accès SPARQL qu'on peut librement interroger⁵. Pour l'alignement des entités, nous avons choisi IdRef, le référentiel de l'Agence bibliographique de l'enseignement supérieur : non seulement la liste des publications de chaque professeur est établie en récupérant en temps réel les données des notices du catalogue Sudoc⁶, mais encore les données du projet peuvent être intégrées avec celles d'autres ressources du web sémantique et réutilisées pour de nouvelles recherches, en réalisant ainsi la vision exprimée par les principes FAIR⁷.
- 4 En 2016, il a été décidé d'inscrire l'expérience de modélisation acquise au fil des années dans l'univers du CIDOC CRM, une ontologie largement utilisée dans le domaine des musées et des humanités numériques. Un écosystème d'extensions pour la recherche a été créé dans le projet Semantic Data for Humanities and Social Sciences (SDHSS)⁸, afin de produire une sémantique extensible, qui permet d'assurer la cohérence entre les modèles de recherche de différents projets et domaines des SHS⁹. Pour soutenir et faciliter ce processus de modélisation ouverte et collaborative, un nouveau service en ligne a été créé, OntoME (Ontology Management Environment¹⁰), adopté par différents projets notamment ERC¹¹.
- 5 Le consortium Data for History a été créé à Lyon en novembre 2017 à l'initiative du LARHRA afin de promouvoir cette vision¹². Il a été suivi par un deuxième atelier lyonnais en 2018, puis par deux colloques internationaux, respectivement à Leipzig en 2019 et à Berlin en 2021

(en ligne)¹³, et un workshop d'experts à l'Université de Bologne en 2022¹⁴. Dans ce contexte, s'inscrit aussi un nouveau projet de sémantisation et de mise à disposition de données issues de projets terminés, Data for Humanities¹⁵, qui s'appuie, entre autres, sur une convention de collaboration de recherche entre l'ABES et le LARHRA, notamment dans le projet ANR HisArc-RDF (Partage et réutilisation de données archéologiques et historiques : une description en RDF appuyée sur les référentiels et les normes du Web sémantique, 2019-2022).

- 6 Un laboratoire de recherche comme le LARHRA ne pouvant héberger, dans le long terme, une infrastructure de recherche de l'envergure de celle du projet symogih.org, une migration des données est en cours vers Geovistory¹⁶, un nouvel environnement virtuel de recherche et de plateforme de publication des données, actuellement porté par le LARHRA, la société KleioLab¹⁷ et la chaire d'humanités numériques de l'Université de Berne¹⁸, dans le cadre d'un partenariat public-privé destiné à pérenniser l'infrastructure, en constituant un consortium d'institutions promouvant son développement dans le long terme. Geovistory reprend et développe de manière significative l'approche du projet symogih.org, tout en apportant des améliorations et des modernisations concernant les technologies utilisées. Le modèle de données est géré dans la plateforme OntoME et profite de l'écosystème d'ontologies du projet SDHSS. Dès leur production, le sens des données est ainsi explicité en utilisant une sémantique qui garantit leur interopérabilité et leur réutilisation pour de nouvelles recherches.
- 7 Une interface de requêtes graphique facilite l'exploration et l'exportation des données dans différents formats. Les projets peuvent disposer d'un site Web dédié, ainsi que d'un point d'accès SPARQL, sur lequel sont publiées les données du projet selon sa perspective propre. De plus, un entrepôt générique publie les données de l'ensemble de la communauté, avec toute sa richesse et ses points de vue différents. Les entités de Geovistory sont reliées, tout comme celles du projet symogih.org, avec les référentiels du Web sémantique, notamment les notices d'autorité IdRef de l'ABES, et s'inscrivent ainsi dans la logique des données de la recherche ouvertes.

- 8 Dès le départ, et avec une plus grande maturité aujourd'hui, l'infrastructure co-portée par le LARHRA a été mise en place, afin de permettre aux projets de recherche en sciences historiques et, plus largement, en sciences humaines et sociales, de gérer une problématique de recherche propre. Cette infrastructure participe aussi à la production d'un graphe géant d'information au service de la recherche et du public, fondé sur les méthodes et les technologies du Web sémantique, afin de contribuer au renouvellement de la recherche grâce aux outils numériques¹⁹.
- 9 La construction et le développement de cette infrastructure n'aurait pas été possible sans la contribution active et engagée de nombreux collègues du LARHRA, enseignants-chercheurs et ingénieurs, dont Bernard Hours, Pierre Vernus, Djamel Ferhod, Sylvain Boschetto, Charlotte Butez, Christine Chadier, Séverine Gedzelman, Alexandre Perraud, Morgane Pica, que nous tenons à remercier pour leur implication.

NOTES

- 1 <http://symogih.org> — Francesco Beretta et Pierre Vernus, « Le projet SyMoGIH et la modélisation de l'information : une opération scientifique au service de l'histoire », *Les Carnets du LARHRA*, 1, 2012, 81-107.
- 2 <http://www.patronsdefrance.fr>
- 3 <http://siprojuris.symogih.org>
- 4 Marie Delcourte, Université de Valenciennes, 2016 ; Jessica Kohn, Université Paris 3, 2018 ; Aurélie Philipe, Université Paris 13, 2021.
- 5 <http://symogih.org/?q=rdf-publication>
- 6 <http://siprojuris.symogih.org/siprojuris/enseignant/44315/>
- 7 <https://sdhss.org>
- 8 <https://sdhss.org>
- 9 Francesco Beretta, « A Challenge for Historical Research : Making Data FAIR Using a Collaborative Ontology Management Environment (OntoME) », *Semantic Web*, 12, 2, 2021, 279-294.
- 10 <https://ontome.net>

11 En particulier, deux projets financés sur fonds européens ont utilisé OntoME pour la préparation du modèle de données : *Silknow* et *Read-it*.

12 <http://dataforhistory.org>

13 <https://d4h2020.sciencesconf.org>

<https://dhistory.hypotheses.org/category/forschungskolloquium/data-for-history-lectures>

14 <https://data4history-unibo.github.io/meeting2022/>

15 <http://dataforhumanities.org>

16 <https://www.geovistory.org>

17 <https://kleiolab.ch>

18 https://www.dh.unibe.ch/index_eng.html

19 Francesco Beretta, « Données ouvertes liées et recherche historique : un changement de paradigme », *Humanités numériques*, 7, 2023

INDEX

Mots-clés

outils numériques, données numériques, sémantisation

Keywords

digital tools, digital data, sementization

AUTHORS

Francesco Beretta

Julien Caranton

Vincent Alamercery