

## The symogih.org Project: Towards an International Consortium

Francesco Beretta, Vincent Alamercery, Djamel Ferhod

(CNRS - Université de Lyon, France)

Accepted poster : Digital Humanities Conference 2016 – Krakòw, 11-16 July 2016

### *symogih.org*, a project in constant evolution

The collaborative platform for historical research developed as part of the *symogih.org* project has reached a mature stage<sup>1</sup>. With an active SPARQL endpoint, the system is now interoperable and interconnected<sup>2</sup>. It is housing 16 ongoing projects, including several on a European scale (France, Germany, Switzerland, Belgium); there are about 50 active users and it contains nearly 1,700,000 data rows. The research projects include, for example, the Siprojuris project<sup>3</sup>, a prosopographic data publishing site interconnected with IdRef, the authority data repository of the French higher education libraries' catalogue<sup>4</sup>; the Historical Atlas of Political Territories<sup>5</sup>, a project that is mapping the evolution of world political borders; Society Religion Science (SRS)<sup>6</sup>, an experimental digital intellectual history website providing documents encoded in XML following the recommendations of the *Text Encoding Initiative* (TEI) and annotated using the *symogih.org* ontology<sup>7</sup>.

The institution hosting the project is the *Laboratoire de recherche historique Rhône-Alpes*<sup>8</sup>. Its human and material resources being limited, to avoid a growth crisis and with a view to building an

1 *SyMoGIH: Système Modulaire de Gestion de l'Information Historique* - Modular Historical Information Management System. The project site: <http://www.symogih.org/>. Cf. Beretta, F., Vernus, P., and Hours, B. (2012).

2 <http://www.symogih.org/?q=rdf-publication> . Cf. Beretta, F., Ferhod, D., Gedzelman, S. and Vernus, P. (2014).

3 <http://siprojuris.symogih.org/>

4 <http://www.idref.fr/> .

5 <http://geo-larhra.ish-lyon.cnrs.fr/?q=atlas-historique-des-territoires-politiques> . Cf. Beretta, F. and Butez, C. (2014).

6 <http://srs.symogih.org/> .

7 [https://groupes.renater.fr/wiki/symogih/symogih\\_manuel/edition\\_de\\_textes\\_en\\_xml-tei](https://groupes.renater.fr/wiki/symogih/symogih_manuel/edition_de_textes_en_xml-tei) .

8 LARHRA CNRS UMR 5190 – Universités de Lyon et Grenoble: <http://larhra.ish-lyon.cnrs.fr/>.

international community of users, the evolution of two principle aspects of the architecture of the *symogih.org* project needs to be reconsidered and planned for from today: these are, its interface with other information systems, in particular with those put in place by heritage institutions; and the management of a growing community of users.

The purpose of our DH 2016 Conference poster is to share our prospects with other DH projects and to present a proposal for the evolution of the platform and operational information, for both the technological and project management domains, and to call for an international collaboration which could function as a consortium.

### **Towards a distributed and multi-instantiated database**

A possible technological answer to the challenge of the increasing number of projects hosted by the *symogih.org* platform, i.e. to provide the expected information centralization without creating a bottleneck and low performance, is to evolve its architecture by distributing the data between a primary node and several secondary nodes of a clustered database.

The primary node would contain all of the shared repositories (object authority records and instances of the collaborative ontology), as well as general-interest data. The secondary nodes, hosted in various research institutions, would contain the information produced by the local projects. To allow the collaborative management of a shared, good-quality ontology to be maintained, avoiding redundant and messy data, a user of any secondary node should be able to access all the data, whether it is held in their database instance or not, via a table held in the primary node, which would hold the keys of all the existing knowledge units and their location in the distributed system.

The whole information system will be interoperable with reference ontologies like CIDOC-CRM or FRBRoo and interlinked with authority files repositories (ISNI, VIAF, BNF, etc.). In addition, a growing portion of the information will be directly accessible in RDF format through a SPARQL endpoint, allowing the data to be queried at the same time as those from other linked data warehouses. The data will be available under a Creative Commons 4.0 international licence.

### **Establishing an international community of users**

This distributed platform management entails the creation of an international organized community of users in the form of a consortium. This community will need to be built around a governance committee, responsible for ensuring a collaborative approach at all levels, from managing the definitions of ontology instances through to data capture. The primary node could be managed, as at present, by the *Pôle histoire numérique* (Digital History Department) of the LARHRA, founding

institution of the *symogih.org* project; but it could be hosted in the data centre of an institutional structure such as, in France, the TGIR HumaNum<sup>9</sup>.

The other nodes would be managed by teams made up of researchers and IT technicians, ensuring not only management of the respective databases and software development for the whole project, but also running local scientific projects and training users in data modelling and capture, working closely with the governance committee. The evolution towards the participation of projects using different languages will imply the implementation of a multilingual version of the information system, both at the level of the interface and for drafting instances of the collaborative ontology.

By offering access to a collaboratively designed ontology and a robust technological solution, intended for an increasingly wide community of users, the *symogih.org* project will contribute to the evolution of practices in the domain of data production and data curation in historical research.

## References

Beretta, F. and Butez, C. (2014). SyMoGIH project and Geo-Larhra: A method and a collaborative platform for a digital historical atlas. *Digital Humanities 2014 Conference*, Lausanne, Switzerland, <http://dharchive.org/paper/DH2014/Paper-831.xml>.

Beretta, F., Ferhod, D., Gedzelman, S. and Vernus, P. (2014). The SyMoGIH project: publishing and sharing historical data on the semantic web. *Digital Humanities 2014 Conference*, Lausanne, Switzerland, <http://dharchive.org/paper/dh2014/poster-930.xml> ; <https://halshs.archives-ouvertes.fr/halshs-01097399v1>.

Beretta, F., Vernus, P., and Hours, B. (2012). The Système Modulaire de Gestion de l'Information Historique (SyMoGIH): a collaborative and cumulative platform for storage and use of geo-historical information. *Digital Humanities 2012 Conference*, Hamburg, Germany, <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/le-systeme-modulaire-de-gestion.1.html>.

---

9 Très Grande Infrastructure de Recherche (very large research infrastructure): <http://www.huma-num.fr/>.