

XIth Heloise Workshop: Paris, 12-13th of May 2022

Paris : Conservatoire National des Arts et Métiers, 12th of May

Paris Panthéon-Sorbonne 13th of May

Digital Academic History and the Challenge of Uncertainty

The members of the Heloise network have now been working for more than ten years on the development of databases of different sizes and ambitions. They share the conviction that education is one of the most crucial component of any cultural system, and that in this field, universities offer an enormous deposit of sources which provide historians with exceptional data, especially for long-term or comparative history: to link and enrich the data with economic, social, or intellectual sources, it is necessary to formalize them, and this is certainly why the development of databases has become so important for us, since they are a requisite for this type of approach. However, databases are not without problems. We have dedicated our attention to what is still a main concern in historical research in general, the possibility to devise a common structure for the exchange of information and for a collective exploitation of individual databases. This remains a preoccupation, and we will be keen to learn about what has been going on in this field in a session devoted to this topic. But we would like in Paris to trod a new path, in the wake of the discussions we had during the Leipzig workshop in 2019, and to explore the problem of uncertain data, since we have the opportunity to benefit from the experience of the DAPHNE program, which has associated several groups of HELOISE with teams of computer scientists working in this field. In fact, uncertainty is everywhere. It may come from the sources, which can be incomplete, unclear, or ambiguous, or from the language of the sources, or from the language the historian himself uses to describe the content of his sources. We should like to make this the chief subject of our meeting, while another session will offer the possibility to present new databases or significant progresses made on existing databases or new data driven research results.

CALL FOR PAPERS

We therefore expect that papers could be devoted to three main topics in our next meeting :

- I. – The actuality of research on a common structure for the exchange of information.
- II. – New databases or significant progresses made on the existing ones or new data driven research results.
- III. – A more specialised workshop, conducted by Cédric Du Mouza and Stéphane Lamassé, directors of the DAPHNE program, centered upon the concept of uncertainty :

Uncertainty is generally considered as a subjective measure of the certainty of something. A state of uncertainty is created when it is impossible to determine whether the information is true or false. The U.S. National Research Council defines uncertainty as follows: "uncertainty is a general concept that reflects our lack of sureness about something or someone, ranging from just short of complete sureness to an almost complete lack of conviction about an outcome"¹ [NRC00]. In constructing records of individuals, historians are confronted with data of uneven quality, often unclear. Input data may be ambiguous or unclear, incomplete, imprecise, contradictory. Knowledge of

¹ [NRC00] National Research Council, Risk Analysis and Uncertainty in Flood Damage Reduction Studies. National Academy Press, 2000.

events can be questionable and can take the form of plausibility, probability and/or possibility. In this workshop, we intend to confront the relationship historians have developed with these differences in data quality and uncertainty and to observe how their consideration can become an opportunity for prosopographical studies. Some representative topics of interest that participants may wish to explore include, but are not limited to:

- a. reflection about issues related to:

- The reliability and quality of the original documentation (demographic, economic, administrative, religious, family, etc.), the authors of the documentation, their credibility...
- Does the failure to take uncertainty into account raises problems for systems which handle prosopographical information today?
- The heterogeneity of records in terms of their size and the nature of the descriptive elements they contain. This can have consequences on the processing of the information and the results that result from it (statistical, geographical ...)
- a more epistemological reflection on the different natures of uncertainty in the experimental spaces defined by the databases.

b. practices and solutions used :

- The rules that historians use to reduce, to circumvent, the uncertainty during their prosopographical investigations.
- The capture of quality and certainty of information
- Machine learning methods as tools for reducing uncertainty
- The use of repositories to decrease or deal with uncertainty (Viaf, IDref..)
- How to take into account the uncertainty dimension when aligning and/or merging data?
- The modeling of time and uncertainty dimensions related to people, places, factoids and sources, which are real difficulties.
- Is it possible to take uncertainty into account in the design of a database (MCD, ontology,...)?

Proposals for papers for all sessions should be sent to the following addresses by 21 March 2022:

dumouza@cnam.fr

stephane.lamasse@univ-paris1.fr

Cedric du MOUZA, Laboratoire CEDRIC, CNAM Paris, 292 rue Saint-Martin 75141 Paris Cedex 03

Stéphane LAMASSÉ, LAMOP, UFR d'Histoire, Université Paris 1 Panthéon-Sorbonne, 17 rue de la Sorbonne, 75005, Paris