Access to cultural heritage data: a challenge for the Digital Humanities

Anne Baillot
anne.baillot@gmail.com
Centre Marc Bloch, Germany

Marie Puren
marie.puren@inria.fr
INRIA, France

Charles Riondet
charles.riondet@inria.fr
INRIA, France

Laurent Romary
laurent.romary@inria.fr
INRIA, France

Access to Cultural Heritage data is a key issue in the future development of Digital Humanities (Murray-Rust, 2013). Cultural Heritage data encompass digitized resources (scanned artefacts) as well as the attached metadata, annotation or further enrichments, all of which are a necessary basis for reliable computational research. Access to high quality Cultural Heritage data and metadata is, in that sense, the condition for reliable, performing and verifiable research in many arts and humanities fields - and not just of interest to librarians and archivists.

One of the core challenges of giving access to high quality Cultural Heritage data is often the lack of institutional connection between local GLAM Institutions (e.g. Galleries, Libraries, Archives and Museums), infrastructures and research (University College London, 2010; Higgins, 2013). The initiative we want to present in this paper addresses this issue by bringing together several supra-national infrastructures. These research infrastructures develop a common online environment that allows all the relevant actors to connect and improve together access to Cultural Heritage data. This project is based on in-depth exchanges on recognized standards as well as on strategies of access, data curation and management, licensing, and an effort towards open data (Romary et al., 2016).

The "Cultural Heritage Data Reuse Charter" is currently being developed by several organisations and projects grouped together within a steering committee: DARIAH-EU, Europeana, Clarin, E-RIHS and APE together with the European projects Parthenos, HaS and IPERION-CH .

It offers a comprehensive framework regarding all aspects relevant to co-operations revolving around access to and reuse of Cultural Heritage data.

The Cultural Heritage Data Reuse Charter: an encompassing cooperation framework

The Cultural Heritage Reuse Charter is an online environment dedicated to all actors taking part in scholarly reuse of digital data generated by Cultural Heritage Institutions. It addresses five actors: Cultural Heritage Institutions, Cultural Heritage Labs, Researchers, Data Centers and Research Institutions.

- Cultural Heritage Institutions (GLAM) are considered in their function as curators of collections and objects in their physical form and as potential primary initiators of corresponding digital surrogates, from basic descriptions (catalogues of collections, metadata for specific objects) to more elaborate outputs (scans, 3D models, physical analyses, etc.) (Ray, 2014).
- Primary data can be hosted by CHIs or by Higher Education Institutions like universities, but they are in many cases curated by dedicated data centers. These centers play a key role in guaranteeing the stability, the visibility and the long time availability of the primary data. The engagement expected from them in the context of the Charter is of a more technical nature and should ensure a concrete implementation of the CHI-researcher relationship.
- Cultural Heritage laboratories have a high-level expertise in Cultural Heritage. They give essential insights into Cultural Heritage history, technologies, environment, and alteration.
- Researchers are invited to sign in person, independently from the institution for which they are working at the time they sign the Charter. However, academic institutions (departments, universities, research

institution or funding agencies) wishing to sign the Charter, or even make it a requirement for their members or the projects they fund or host, are welcome to do so as well.

The Charter environment allows all five actors to declare general principles (common work ethics), and more broadly all the relevant information needed to understand how a given dataset can be reused. It allows its users to get in contact with partners they would want to work with. Institutions can declare their collections; researchers their research interests and existing publications so that these are connected together. Doing so, all of them always have the possibility to define precisely which aspects of their profile information they wish to make public and which not.

By joining forces, and by sharing the information associated to Cultural Heritage collections, the Charter will help document the knowledge generation process and, consequently, increase the quality of data and metadata accessible to research.

Signing the Charter implies making a statement about the technical quality of the data to be reused, or the data derived by such a reuse. The implementation of appropriate standards is considered a key node for the stabilization of data access (Romary, 2011). More broadly, the Charter offers a concrete implementation framework for the FAIR principles (make the data findable, accessible, interoperable and reusable).

The online environment in practice

The principles described below are addressed by a series of components allowing to define the conditions of reuse for each type of data. The Charter environment offers a framework that can be either picked among a set of recommendations or formulated in a text field by the concerned institutions or actors according to their needs and wishes.

This framework encompasses all questions related to the reuse of Cultural Heritage Data:

- Long-term and persistent access to metadata, texts, images (in the case of a manuscript for instance: archival metadata, scan of the manuscript, transcription, annotation)
- Licensing of the content (linking to relevant documentation allowing for instance researchers to gather information on licensing and citation practices they often lack)

- Formats and standards (also connecting to further information)
- Enrichments (connection of scholarly work and CHI work)
- Dissemination of both CHI information and research (visibility of the work of all stakeholders)
- Retro-provision (communicating enrichments based on CHI data to the CHI they originally emanate from)
- Quality control at all levels according to appropriate standards.

In practice, users of the Charter register in the online environment in their primary function as Cultural Heritage Institution, Researcher, Cultural Heritage Lab, Data Center or Research Institution. Identification of entities are realized on the basis of existing standards such as ORCID for researchers.

In the researcher profile, three main areas are to be defined by the registered user. First, he/she has to abide to the reuse principles defined by the Cultural Heritage Institutions regarding the collection he/she wants to work on; this is the "use of primary data" area. Second, he/she has to declare the dissemination principles he/she favours. In this "dissemination of secondary data" area, he/she can gather information on licences. The third area is that of the "cooperation ethics", in which the researcher declares that he/she will follow best practices in citing the other Charter partners involved in his/her endeavour. This threefold profile is the basis on which the researcher can reach out to institutions or collections he/she wishes to work with.

Timeframe

The Cultural Heritage Reuse Charter is currently under development. Workshops in which information will be gathered especially on the expectations of Cultural Heritage Institutions will take place in Berlin (November 2016), Paris (November 2016), Rome (January 2017) and Dublin (February 2017). Additional input from the other actors is gathered in parallel (interviews). A soft launch of the web interface is planned for the summer of 2017, so that the interface as well as the benefits for the first signatories can be demonstrated in Montreal .

This abstract is in English in order to reach the widest possible community. Presenting the paper in French or having the slides to the presentation in French would be possible as well.

Bibliography

DeNardis, L. (2011) *Opening Standards: The Global Politics of Interoperability*. Cambridge, Mass.: MIT Press.

Europeana (n.d.) The Europeana Licensing Framework. Accessed October 28, 2016. https://goo.gl/947T4z

Higgins, S. (2013). "Digital Curation. The Challenge Driving Convergence across Memory Institutions", *The Memory of the World in the Digital age: Digitization and Preservation*, UNESCO. Accessed October 30, 2016. https://goo.gl/JteZQe

Inist (n.d.) « Textes de références », *Libre accès à l'information scientifique et technique*. Accessed October 28, 2016. https://goo.gl/jXtmg4

Murray-Rust, P. (2013) "Open Data in Science", Serials Review, Vol 34, No 1. Accessed October 28, 2016. https://goo.gl/9ZqdiQ

Osswald, A., and Strathmann, S. (2012), The Role of Libraries in Curation and Preservation of Research Data in Germany: Findings of a survey, IFLA World Congress. Accessed October 30, 2016. https://goo.gl/GSOxRX

Ray, J. (2014), *Putting Museums in the Data Curation Picture*, Springer.

Romary, L. (2011) "Stabilizing knowledge through standards - A perspective for the humanities", *Going Digital: Evolutionary and Revolutionary Aspects of Digitization*, Science History Publications. <inria-00531019>

Romary, L., Mertens, M., Baillot, A (2016) "Data fluidity in DARIAH – pushing the agenda forward", BIBLIOTHEK Forschung und Praxis, De Gruyter, 2016, 39 (3), pp.350-357. <hal-01285917v2>

Sabharwal, A. (2015) Digital Curation in the Digital Humanities. *Preserving and Promoting Archival and Special Collections*, Chandos Publishing, 2015.

Suber, P. (2012), *Open access*, The MIT Press, Cambridge, London.

University College London (2010), Advancing Research and Practice in Digital Curation and Publishing. *Summary Report and Recommendations of the Workshop on Next Steps in Research, Education and Practice,* 2010. Accessed October 30, 2016. https://goo.gl/hqKuKQ