Linking Data & Publications in Social Science and Humanities: the role of infrastructures in the French national context

Tuesday 14 October 2025 11:30 (11 minutes)

In SSH (Social Science and Humanities) the link between data and publication can be seen from different angles depending on its potential use. The first use that comes to mind is to cite a dataset in a publication for the purposes of scientific verification. It can be done in a number of ways, from a simple text citation, both in the publication or in the description of a dataset, to a PID (Persistent IDentifier) in a specific metadata. Another possible type of link would be to show multimedia material (e.g. illustration, table, soundtrack, etc.) in a publication. Finally, in another vein, data papers can be considered as a case of linking data and publications. At a national level, an ecosystem of repositories and publication platforms are involved in the process of creating links between data and publications. More specifically, as part of two national projects (HALiance and COMMONS), four infrastructures are working together on SSH resources: Huma-Num (Data repository NAKALA), CCSD (Open Archive for research papers - HAL-SHS), OpenEdition (Publication Platforms for books and journals) and Métopes (Publishing Process). A study of the content of the different platforms carried out at the beginning of these projects showed that link-building practices already existed. Unsurprisingly, the study illustrated the diversity of the means used to create the above-mentioned links, and the fact that they are generally unidirectional.

What role can infrastructures play in this area? First of all, there is a need to simplify things for users by building bridges between platforms. For instance, when a research paper referring to a dataset in NAKALA is deposited in HAL-SHS, it would be convenient to have direct access to the dataset from the HAL environment. Similarly, to include multimedia material in a book from OpenEditionBook may require access to the files in NAKALA, which represents a different level of granularity. This means that it is necessary to work on seamless integration between the platforms; in particular by working on the consistency of information systems, developing specific APIs, and then finally by adapting interfaces.

Another crucial role of infrastructures is to guarantee the consistency of link information and to disseminate it in a standardised way. This requires automated communication between platforms to be able, for example, to create reciprocal links and maintain them over time: for instance, what happens to the link if the data disappears from the repository or if a new version is created?

This process which involves adaptations at various levels, both technical and organisational, did not start from scratch. A POC (Proof of Concept) of creating links between NAKALA and HAL-SHS was performed as part of the European project EOSC-Pillar. This enabled us to determine important issues to be resolved. It also helped us to identify standards that are emerging on this topic: in this regard, it was decided to use the COAR Notify protocol for communication between the platforms and the SCHOLIX standard to disseminate data-publication links which is supported by DataCite and Crossref and used in the European context by OpenAire. This work is well advanced and the link between the HAL-SHS archive and the NAKALA repository is already operational on the development platforms and is about to be put into production. This work will serve as the basis for the link between the NAKALA repository and the OpenEdition platforms, which will also use a structured standard developed by Métopes (COMMONS TEI-Publishing) for integrating data directly into a publication.

At this stage of the projects HALiance and COMMONS, communication between the various platforms via the COAR-notify protocol has been standardised and stabilised. This makes it possible to consider communication with other platforms in the national ecosystem, such as the national repository RDG (Recherche Data Gouv), which uses the same protocol. Technically, there is still work to be done: especially a better version management for publications and datasets and a better implementation of the SCHOLIX standard. Finally, it will be necessary to improve the integration between platforms through a potential common authentication,

for instance. However, the key to successfully implementing the link between data and publications lies in informing and training future users, which is an important part of our projects.

Bibliography

McGillivray B, Marongiu P, Pedrazzini N, Ribary M, Wigdorowitz M, Zordan. (2022). E. Deep Impact: A Study on the Impact of Data Papers and Datasets in the Humanities and Social Sciences. Publications. 10(4):39. https://doi.org/10.3390/publications10040039

Burton, Adrian, & Koers, Hylke. (2016). ICSU-WDS & RDA Publishing Data Services WG Interoperability Framework Recommendations (1.0). https://doi.org/10.15497/RDA00002

Edmond, J. (Ed.). (2020). Digital Technology and the Practices of Humanities Research. Open Book Publishers. https://doi.org/10.11647/obp.0192

Gassama, M., Szabo, D., Tang, C., & Bravo, S. (2024). Analyse de l'enquête sur les pratiques des scientifiques en matière de publication de data paper [Report, INRAE]. https://doi.org/10.17180/vrh7-r606

Harper, L. M. (2023). Data reuse among digital humanities scholars : A qualitative study of practices, challenges and opportunities [Université d'Ottawa / University of Ottawa]. http://hdl.handle.net/10393/45445

Arnold, T., Scagliola, S., Tilton, L., & Gorp, J. V. (2021). Introduction : Special issue on audiovisual data in dh. Digital Humanities Quarterly, 15(1). https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html

Kinnaman, A., & Guimont, C. (2023). Dh as data : Establishing greater access through sustainability. Digital Humanities Quarterly, 17(3). https://www.digitalhumanities.org/dhq/vol/17/3/000715/000715.html

Primary author: Mr LARROUSSE, Nicolas (Huma-Num CNRS - France)

Co-authors: Mrs KUNTZIGER, Bénedicte (CCSD CNRS - France); Mr BOURDOT, Charles (Métopes - Université de Caen Normandie - France); Mr ROUX, Dominique (Métopes - Université de Caen Normandie - France); Mrs JOUGUET, Hélène (Huma-Num CNRS - France); Mrs VERLEYEN, Julie (Huma-Num CNRS - France); Mrs GUIGONIS, Sandra (OpenEdition Aix-Marseille University - France); Mr BARBORINI, Yannick (CCSD CNRS - France)

Presenter: Mr LARROUSSE, Nicolas (Huma-Num CNRS - France)

Session Classification: Presentations Session 4: Data Stewardship

Track Classification: SciDataCon Persistent Themes: Data Stewardship