



Huma-Num, 10 years of building an infrastructure in SSH: a revolution in practices or an evolution?

Nicolas Larrousse, Olivier Baude, Stéphane Pouyllau, Hélène Jouguet

► To cite this version:

Nicolas Larrousse, Olivier Baude, Stéphane Pouyllau, Hélène Jouguet. Huma-Num, 10 years of building an infrastructure in SSH: a revolution in practices or an evolution?. SciDataCon-IDW Salzburg 2023, CODATA, Oct 2023, Salzburg (AU), Austria. halshs-04333382

HAL Id: halshs-04333382

<https://shs.hal.science/halshs-04333382>

Submitted on 9 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC0 - Public Domain Dedication 4.0 International License

Huma-Num, 10 years of building an infrastructure in SSH: a revolution in practices or an evolution?

Authors:

Nicolas Larrousse : Huma-Num CNRS UAR 3598 - ORCID [0000-0002-4968-797X](https://orcid.org/0000-0002-4968-797X)

Olivier Baude : Huma-Num CNRS UAR 3598 - ORCID [0000-0002-8627-5229](https://orcid.org/0000-0002-8627-5229)

Stéphane Pouyllau : Huma-Num CNRS UAR 3598 - ORCID [0000-0002-9619-1002](https://orcid.org/0000-0002-9619-1002)

Hélène Jouguet : Huma-Num CNRS UAR 3598 - ORCID [0000-0001-6719-3765](https://orcid.org/0000-0001-6719-3765)

Happy birthday Huma-Num! After 10 years of building an infrastructure dedicated to SSH in France: it's time to look back and see how far we have come.

Birth

At the beginning, around 2010, a time when digital services were only in their infancy, two existing infrastructures handled their use: the TGE Adonis, which stands for "Très Grand Equipement", and the Research Infrastructure "Corpus".

The first one was dedicated to providing digital services and the second one was designed to support SSH communities in digitizing corpora. In 2013, Huma-Num was created as a national infrastructure by the ministry of research and the CNRS by merging these two structures. The initial idea was to develop technological services for digital research data in close collaboration with research communities.

But what exactly is an infrastructure for SSH? This was a concept that had to be invented.

From Childhood to Adolescence

At the time of its creation, the most urgent task was to try to save data that were at best on a hard disk and sometimes on a floppy disk. We quickly discovered that preserving raw data was not sufficient - what was the point of preserving data if it was impossible to understand them? We had already gained some experience of this issue through long term preservation projectsⁱ but, as SSH are made up of diverse disciplines, we needed the support of experts. This is where Huma-Num consortia, the second part of Huma-Num inherited from the RI "Corpus", took over. A Huma-Num consortium is a group of people funded by Huma-Num, often from different institutions, sometimes from different disciplines, working on the same scientific objects, methods and themes. They provide expertise about their specific scientific objects in order to define good practices, training and toolsⁱⁱ.

It is an illustration of the fact that an infrastructure cannot be reduced to providing technological services, but that the expertise of communities and their involvement in the construction are key to building an infrastructure.

As for the technology, in 10 years we were well positioned to monitor the rapid evolution of needs. SSH research projects nowadays use more sophisticated tools (e.g. Geographical Information System, Rstudio, Deep Learning software, etc.), new types of data (e.g. Social media, LIDAR, 3D objects, etc.) and above all the volume of data has evolved very rapidly.

Toward maturity

Huma-Num has reached a certain level of maturity from a technological point of view even if there is still room for improvement as the needs and technologies are evolving rapidly: from a quantitative point of view, Huma-Num hosts more than 1 petabyte of data, about 400 virtual machines and 800 web sites. More generally, we are able to provide services for each step in the research data lifecycle. In particular, two services have been developed by Huma-Num to process, store and display research data while making them FAIR: NAKALAⁱⁱⁱ to share data and ISIDORE^{iv} to disseminate them.

Furthermore, Huma-Num's consortia are constantly changing and evolving to take into account new needs, for example supporting a standard such as the Music Encoding Initiative for the "Musica consortium" or providing a toolbox for historicized data for the "Paris Time Machine consortium".

At this point, we were able to focus on addressing one of the main issues we face: the quality of data, both of the digital object itself and its metadata.

In order to address this issue, we can of course rely on the support provided by Huma-Num's consortia, but we need to have other complementary means to be able to support research projects throughout the whole data life cycle. Two years ago, we set up a dedicated service internally, currently made up of three people but which is expected to grow: they provide documentation, training and personalized support for research projects.

This is of course not enough to take into account all needs and Huma-Num has established partnerships with regional SSH infrastructures, the "Maisons des Sciences de l'Homme"^v, in order to provide additional support to local projects and thus fluidify the relationship with Huma-Num by clearing the way.

In the same vein, as a national infrastructure, it is important to establish contacts abroad to remain well-informed and also to be proactive in promoting services and standards developed by France. For the last 7 years, we have been establishing collaboration at the international level notably by participating in the construction of European Infrastructures such as DARIAH^{vi} and participating in European infrastructure projects. We also have developed more targeted cooperation with Canada, for instance on innovative tools.

What's next?

We now have the necessary building blocks to develop new services and partnerships. The main one, which started this year, aims to establish a collaboration to link data and publication with other national infrastructures dedicated to publications, OpenEdition^{vii} and METOPES^{viii}. This ambitious project, called COMMONS, will last 8 years for a total cost of around 30 M€.

During the past 10 years, about 10,000 users have used the infrastructure, so the following question can be asked: Has Huma-Num influenced the ways of doing research in SSH or has it simply accompanied the general evolution of Digital Humanities?

The Huma-Num lab was created in part to get an idea of the answers to these questions by investigating the evolution of current research practices in SSH through the prism of digital technology.

This paper will present in greater detail what we learned the last 10 years the perspectives for the coming years.

References:

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

Edmond, J. (2015). Collaboration and Infrastructure. In A New Companion to Digital Humanities (pp. 54–65). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118680605.ch4>

Laurent Romary, Jennifer Edmond. A Tangential View on Impact for the Arts and Humanities through the Lens of the DARIAH-ERIC. Bente Maegaard; Riccardo Pozzo. *Stay Tuned To The Future - Impact of the Research Infrastructures for Social Sciences and Humanities*, Leo S. Olschki Editore, 2019, 978 88 222 6643 9.

Olivier Baude, Adeline Joffres, Nicolas Larrousse, Stéphane Pouyllau. Interoperability: a new horizon for data sharing in Humanities and Social Sciences. The input of three digital services developed by Huma-Num. *Digital Humanities 2016.*, Jul 2016, Cracovie, Poland. <https://hal.archives-ouvertes.fr/hal-02153019>

ⁱ <https://hal.archives-ouvertes.fr/hal-02153003>

ⁱⁱ <https://documentation.huma-num.fr/humanum-en/#huma-nums-national-consortia>

ⁱⁱⁱ <https://www.nakala.fr>

^{iv} <https://isidore.science>

^v <https://www.msh-reseau.fr>

^{vi} <https://www.dariah.eu>

^{vii} <https://www.openedition.org>

^{viii} <http://www.metopes.fr>