



HAL
open science

A Techno-Human Mesh for Humanities in France: Dealing with preservation complexity

Nicolas Larrousse, Joël Marchand

► **To cite this version:**

Nicolas Larrousse, Joël Marchand. A Techno-Human Mesh for Humanities in France: Dealing with preservation complexity. DH 2019, Jul 2019, Utrecht, Netherlands. hal-02153016

HAL Id: hal-02153016

<https://hal.science/hal-02153016>

Submitted on 11 Jun 2019

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.

A Techno-Human Mesh for Humanities in France: Dealing with preservation complexity

Nicolas Larrousse (nicolas.larrousse@huma-num.fr), Huma-Num CNRS (Centre National de la Recherche Scientifique), France

Joel Marchand (Joel.Marchand@huma-num.fr), Huma-Num CNRS (Centre National de la Recherche Scientifique), France

Nowadays, as the use of digital data for research in Humanities has become the norm, researchers are dealing with a huge amount of data. As a consequence, the risk of data loss is increasing. Another difficulty is to provide full access to this flood of data to users often located in distant areas. These problems can no longer be addressed individually by researchers or even at a laboratory level: it is therefore necessary to use a technical infrastructure with specific skills to provide stable preservation services.

Huma-Num¹, the French national infrastructure dedicated to Digital Humanities, was looking for a way to address these challenges. The main goal was to deploy a technology that would be readily usable and transparent for average users. Scalability was mandatory considering the rapid evolution of the mass of data, and the system should be, ideally, distributed to ensure better security.

Besides these purely technological requirements, we also had some political and organizational concerns. We wanted to delegate the close relationship with users and local administration to an existing robust network of regional centres. By doing so, we expected a better appropriation of the proposed solution and also an enhanced capacity to respond to their specific needs.

This paper will present the implementation of a preservation system in France, branded “Huma-Num-Box”, which aims to address all the above-mentioned goals. Then, we will give some feedback about this experiment and actions for the near future.

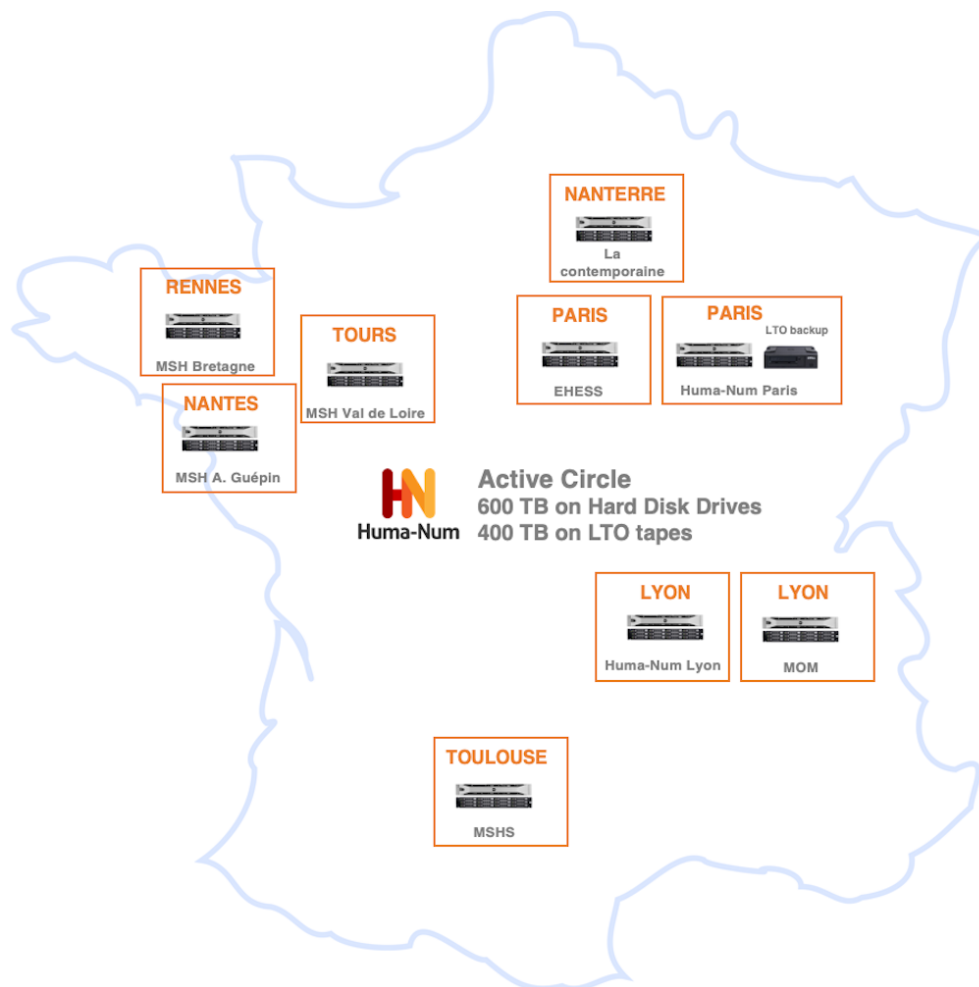
1. A technical choice

We all know that researchers do not really take preservation into account during the research data life cycle, especially at the early stage, but are more accustomed to making copies on a local device. Accordingly, the first goal was to provide a device as simple to use as a local hard drive. Moreover, we operate in a classical server landscape and we need to be able to access these data on servers using different technologies. We had substantial experience with “iRods software²” which was very efficient but not really user-friendly to say the least and not totally tailored to some servers’ technologies.

After some research, we decided to go for “Active circle³”, a software edited by a French company. Here are some reasons why we made this choice:

- It uses standard hardware which can be recycled and new hardware can be added with no re-replication
- The file system is natively distributed and provides all the protocols we desired (CIFS, NFS, FTP) for users and servers
- You can delegate user management to each node relying on a classical LDAP system
- You can mount a share as local and you can share data between nodes
- For each set of data, you can easily decide on the policy you want to apply (versioning, numbers and localization of the copies etc.)
- All the steps of the data life cycle are integrated in a single tool (storage, replication, daily checks of integrity, hash verification etc.)
- We knew that the support was very reactive.

This software seems to be perfectly adequate and even offers more than we expected but the choice of a commercial software was not an easy one. One downside is the cost: the cost of the licence is far from being marginal and is related to the data size.



2. A Human mesh

Huma-Num provides services at a national level, but relies on a network of 23 regional centres, called “Réseau des MSH (Maisons des Sciences de l’Homme)”⁴, to pass on information about its services and in return get some feedback. This network has been around for 20 years and each centre incubates research projects and provides local services.

It therefore seemed logical to set up our nodes in some MSH. Each node is associated with a technical correspondent who manages local accounts and shares under the supervision of Huma-Num. His/her role is also to ensure links with local system administrators and to perform the administrative tasks.

3. Feedback and the near future

After two years, we have a mesh of nine nodes geographically distributed all over France used by 500 users with 100 data shares. We now host around 500 TBs of data, which was quite unexpected so we were forced to expand the system. We were able to save endangered archaeological data located in remote “Ecoles Françaises à l’Etranger” (mostly Greece and Egypt). We also discovered a set of data (60 TBs) of very important contemporary historical archives which did not have a single backup because of the lack of local resources. So, we can say that it’s a success. A nice side effect is that we built a logical network above the national network⁵ to connect our nodes: this “private network for SSH” is ready to be used for future services. We now consider that this service is mature enough to make it available at a European level via the ERIC DARIAH⁶ infrastructure.

However, the use of the mesh is very uneven. Some nodes are quite empty, and we decided to use them as backups for other nodes. This means that there is probably more work to do to convince users of the benefits of using it. In order to address this, we organize meetings, called “Huma-Num-Bar”, to inform communities about our services: these meetings are broadcasted and archived. We also do a “MSH Tour” to interact directly with potential users.

The purpose of this paper is to demonstrate that providing a technology alone is useless: the key to success is user uptake. You absolutely need to rely on a network of expertise to explain the project and to make it work on a daily basis. We also learned that it takes time: from installing an “intrusive” machine inside an existing network until its use by researchers and engineers, every step comes with different difficulties. The main one is the real complexity of coordinating such different categories of actors involved in this project.

¹ <https://www.huma-num.fr/about-us>

² <https://irods.org/>

³ <https://www.oodrive.com/products/save/active-circle-storage-archiving-solution/>

⁴ <http://www.msh-reseau.fr/les-msh>

⁵ <https://www.renater.fr/?lang=en>

⁶ <https://www.dariah.eu/>