
Huma-Num : Une infrastructure française pour les Sciences Humaines et Sociales. Stratégie, organisation et fonctionnement

Olivier Baude

olivier.baude@huma-num.fr

Centre National de la Recherche Scientifique, France

Adeline Joffres

adeline.joffres@huma-num.fr

Centre National de la Recherche Scientifique, France

Nicolas Larrousse

nicolas.larrousse@huma-num.fr

Centre National de la Recherche Scientifique, France

Stéphane Pouyllau

stephane.pouyllau@huma-num.fr

Centre National de la Recherche Scientifique, France

Introduction

La recherche en sciences humaines et sociales vit un tournant numérique qui provoque une évolution sensible des pratiques de recherche.

Aujourd'hui, les chercheurs produisent d'importants volumes de données numériques et utilisent des outils spécialisés pour y accéder, les manipuler, les visualiser et les diffuser. Cela nécessite d'associer ces grands volumes de données à des moyens technologiques qui se doivent d'être stables et conséquents. La maîtrise de ce nouvel environnement doit s'appuyer sur le développement de compétences diversifiées et en constante évolution.

La mise en œuvre de cet écosystème ne peut plus être réalisée avec les moyens dont disposent les chercheurs individuellement ou même à l'échelle d'une équipe de recherche. Une infrastructure, au sens élargi du terme, est nécessaire pour être à même de fournir et mutualiser différents services et d'accompagner les équipes de recherche tout au long de leurs projets numériques.

La TGIR Huma-Num est une infrastructure de recherche française au service de la recherche en sciences humaines et sociales. Elle est bâtie sur une structure originale destinée à proposer des services dédiés à la production et à la réutilisation de données numériques mais aussi à favoriser et accompagner l'émergence d'outils et de bonnes pratiques issus des communautés scientifiques des sciences humaines et sociales. Enfin, Huma-Num est étroitement associée à la construction d'infrastructures européennes permettant tout autant l'échange et la valorisation de services que le renforcement de réseaux d'expertises.

Notre présentation s'articule autour des trois grands axes développés par Huma-Num pour répondre aux besoins de la recherche en sciences humaines et sociales.

- La création et le soutien de consortiums, cercles d'expertise disciplinaires ou non, permettant une concertation collective par les communautés ;
- Le développement de services technologiques destinés à l'outillage des données de la recherche ;
- L'ouverture à l'international par la participation à des projets d'infrastructures Européennes mais aussi par des échanges bilatéraux plus ciblés.

Les consortiums réunissent des acteurs issus de structures et de projets de recherche divers, autour de thématiques, d'objets ou de méthodes communs. L'objectif est de favoriser l'appropriation des dispositifs numériques par ces communautés et ainsi de créer des synergies en lien avec les services de Huma-Num. Le spectre des activités de ces consortiums est large, allant, par exemple, de la création d'applications jusqu'à la définition de schéma de métadonnées, en passant par la production de guides de bonnes pratiques associées à des formations spécifiques. Actuellement, Huma-Num soutient les activités de huit consortiums, principalement disciplinaires, qui s'articulent avec le réseau des Maisons des Sciences de l'Homme (réseau géographiquement réparti sur les campus universitaires et du CNRS en France métropolitaine) et dont certains collaborent à des projets européens.

Les services gérés directement par l'équipe de la TGIR Huma-Num sont destinés à fournir des outils tout au long du cycle de vie des données : de la production de données brutes jusqu'à leur préservation après traitement. Pour pouvoir proposer des ressources bien dimensionnées et adaptées, Huma-Num n'a pas créé

son propre centre de données, mais s'appuie sur les compétences et les moyens de centres de calcul nationaux français.

L'un des objectifs principaux est de mettre en œuvre un cadre d'outils permettant aux chercheurs d'exploiter au mieux leurs données mais surtout de les amener à les partager et, in fine, à les préserver. Il s'agit de permettre l'appropriation par les chercheurs de méthodes et services numériques au cœur des processus de recherche scientifique. Les technologies d'interopérabilité, en particulier celles proposées par le Web Sémantique, sont au cœur de cette chaîne de traitement permettant des échanges fluides entre les différentes briques qui la composent. Le Web Sémantique permet aussi de décloisonner les données de la recherche par l'utilisation de référentiels et normes communs.

Ces outils appuient une démarche générale d'accompagnement des utilisateurs. Ainsi, lors d'une demande d'utilisation, Huma-Num effectue de manière systématique une sensibilisation à la pérennisation, à la curation, au partage des données et aux enjeux de la normalisation.

Ce point est crucial dans le fonctionnement de cette grille de services qui doit aussi permettre la montée en compétence des communautés.

Les activités internationales de Huma-Num sont de différente nature mais elles visent essentiellement à valoriser les dispositifs français d'infrastructure et de la recherche française et à construire un dialogue scientifique et des dispositifs dépassant les frontières nationales. Huma-Num porte ainsi la participation française à des ERICs (European Research Infrastructure Consortium), disciplinaires comme CLARIN depuis 2017, ou à vocation plus large pour les Sciences Humaines en incluant le domaine culturel comme DARIAH depuis sa création en 2014. Ces ERICs, établis sur une longue durée, ne sont pas financés directement par la Commission Européenne mais par les États participants. L'originalité du budget de l'ERIC DARIAH est que 90% de celui-ci est constituée de contributions « en nature » : l'objectif étant de ne pas dupliquer à l'échelle européenne des services qui existeraient déjà au niveau national. Le rôle d'Huma-Num est donc d'identifier les services et expertises nationaux, en particulier ceux en provenance des consortiums et des grands opérateurs français de l'information scientifique et technique et du patrimoine culturel, et d'en faire un ensemble cohérent capable d'intégrer et de compléter l'offre européenne constituée par les partenaires. En parallèle, Huma-Num participe à des projets à plus

court terme destinés à soutenir la construction de ces grandes infrastructures et qui sont, eux, financés par la Commission Européenne (dans le cadre des programmes H2020).

Enfin, hors Europe, Huma-Num maintient des liens étroits avec les mondes francophones (e.g. Québec) et hispanophones (e.g. Amérique du Sud). Là aussi, l'objectif est de valoriser les services d'Huma-Num ainsi que ceux issus des communautés nationales et de bénéficier en retour d'échanges d'expertises, notamment dans les domaines de la préservation et de la curation de données issues de la recherche en Sciences Humaines et Sociales.

La TGIR Huma-Num a élaboré sa définition de services technologiques autour du cycle de vie des données. Cette présentation vise à présenter, à partir d'une démarche réflexive, les services d'Huma-Num et la pertinence de leur articulation avec les autres composantes de l'infrastructure que sont les consortiums ainsi que leur projection au niveau international.

En effet, Huma-Num propose des outils adaptés à chaque étape du cycle des données de la recherche :

- Du simple stockage de sauvegarde au début du projet ; Suivant le taux prévisible d'utilisation des données, il est proposé plusieurs solutions technologiques : les données « froides » (i.e. peu utilisées) sont gérées à moindre coûts sur des stockages distribués alors que les données « chaudes » sont mises à disposition via des stockages performants de type NAS ;
- Des outils de traitement des données. Ceux-ci sont mutualisés comme par exemple des logiciels de SIG (Système d'Information Géographique) qui seraient trop coûteux à acheter pour une utilisation ponctuelle par un projet. Pour ces opérations de traitements, la TGIR met à disposition une grande puissance de calcul ;
- De l'hébergement Web pour présenter les données et permettre leur partage et leur accès ainsi qu'un travail collaboratif. Pour des besoins plus spécifiques des Machines Virtuelles sont mises à disposition afin d'offrir une grande souplesse d'exploitation ;
- Enfin, et c'est là l'originalité de l'infrastructure, il est proposé un ensemble de services qui permet la diffusion, la citation, la pérennisation et surtout la promotion de la réutilisation de ces données:

- NAKALA pour stocker des données documentées et les partager, associé à NAKALONA pour les éditorialiser
- ISIDORE pour diffuser ces métadonnées en les enrichissant, les classifiant et les positionnant dans le LOD (Linked Open Data)

Pour les données qui le nécessitent (e.g. des données de type patrimonial), un service de pérennisation à long terme (plus de 20 ans) en partenariat avec le CINES (<http://cines.fr>) dont c'est la mission, complète l'offre.

La pertinence des demandes d'ouverture de services est évaluée par un comité interne à HumNum qui s'appuie également sur les recommandations de son conseil scientifique.

Cet ensemble de services peut être visualisé de manière synthétique à cette adresse: <http://www.huma-num.fr/services-et-outils>

Par ses différentes composantes, la TGIR HumNum s'emploie à développer des réponses aux besoins nouveaux des différents acteurs de la recherche provoqués par l'utilisation du numérique. Le fil rouge qui les relie est de pouvoir rendre possibles de nouvelles recherches dans le domaine des sciences humaines et sociales et, au-delà, de constituer de manière dynamique un savoir partagé.