

Huma-Num, les humanités numériques en réseau

Huma-Num, très grande infrastructure de recherche (TGIR), vise à faciliter le tournant numérique de la recherche en sciences humaines et sociales. Stéphane Pouyllau, son directeur technique, revient ici sur les raisons de sa mise en place et ses enjeux...

Depuis 20 ans, la recherche en sciences humaines et sociales (SHS) s'est progressivement tournée vers l'édition, la production et l'utilisation des documents et des données numériques. Au-delà de la création et de l'utilisation de bases de données bibliographiques, pratiques anciennes dans les SHS, les chercheurs sont aujourd'hui pleinement entrés dans « le monde numérique », décrit en 2008 par Gérard Berry lors de sa leçon inaugurale de la Chaire d'Innovation technologique Liliane Bettencourt au Collège de France¹.

Le monde de la recherche en SHS devient lui aussi numérique, cela touche autant les outils que les documents et données utilisés par les enseignants-chercheurs dans leurs travaux de recherche. Le web a favorisé dans un premier temps la diffusion des métadonnées, puis des documents numériques eux-mêmes, jusqu'à devenir depuis le début des années 2000, l'un des principaux cadres du développement d'outils de traitement de ces documents. Les SHS ne sont pas restées à l'écart de ce mouvement. En effet, leurs disciplines ont – à des rythmes différents – bousculé et fait évoluer leurs méthodes afin de pouvoir utiliser les documents et les données au travers du web. Ces évolutions, nécessaires à l'heure où certains supports de données devenaient exclusivement numériques (photographies, enregistrements audio/visuels), ont aussi largement modifié les métiers et les façons de « faire de la recherche » dans les laboratoires. En effet, au moment où les SHS découvraient aussi la recherche par projets, les pratiques de collecte, de structuration et de documentation des données ont changé pour devenir de plus en plus collectives. À l'heure du web sémantique et du partage d'ontologies et de thésaurus, la fabrication de métadonnées s'envisage désormais de façon « collaborative » et concertée au travers des communautés scientifiques et *via* le web. Dans certaines disciplines, comme l'histoire des sciences, la participation « d'amateurs » à la documentation de corpus de données numériques partagées en ligne a été proposée. Mais au-delà, ce sont aussi les pratiques liées à l'édition scientifique, à la diffusion des corpus et à leur conservation que le numérique et le web sont venus modifier. Il est possible aujourd'hui de

collecter des données, de les documenter, de les classer, de les éditer, de les relier à d'autres données, en se servant de plusieurs outils de traitement ou d'édition n'utilisant comme support que les technologies du web.

ACCOMPAGNER LE CHANGEMENT...

Ces évolutions profondes ont provoqué une période d'interrogation méthodologique et professionnelle et une phase d'adaptation nécessaire aux acteurs (chercheurs, professionnels de l'IST, informaticiens) afin de s'habituer aux pratiques nouvelles. C'est ce moment que nous vivons et qui prend la forme d'un mouvement d'échanges et d'examen collectif des méthodes et pratiques dans les SHS que l'on nomme parfois « humanités numériques ». Comment assister les chercheurs et les équipes dans cette évolution rapide où les savoir-faire s'hybrident et où le recours à l'utilisation d'outils web nécessite tout de même de savoir préparer ses données ? Comment partager les outils, ou des parties d'outils ? Comment bien les choisir sans hypothéquer l'intégrité de ses données en s'interdisant du coup des réutilisations futures ?

À ces questions, le ministère de la Recherche et de l'Enseignement supérieur a proposé la mise en place d'une très grande infrastructure de recherche portée par le CNRS, Aix-Marseille Université et l'établissement du Campus Condorcet. Huma-Num, née de la fusion du TGE Adonis et de l'infrastructure de recherche Corpus, propose une organisation originale qui repose sur la mise en œuvre d'un dispositif humain impliquant les communautés scientifiques elles-mêmes et un ensemble de services technologiques (stockage, archivage, exposition et enrichissement de données, par exemple Isidore²) à l'échelle nationale et européenne (infrastructures Dariah et Clarin).

... EN CONCERTATION AVEC LES COMMUNAUTÉS CONCERNÉES

La principale mission de la TGIR Huma-Num est d'intervenir, par l'intermédiaire de consortiums scientifiques³ regroupant des acteurs des communautés scientifiques, sur la production de corpus de données et documents de la recherche⁴ et de mettre en place un dispositif technologique per-

[1] www.college-de-france.fr/site/gerard-berry/inaugural-lecture-2008-01-17-18h00.htm

[2] www.rechercheisidore.fr

[3] www.huma-num.fr/service/consortium, voir aussi encadré p. 15.

[4] Cela inclut aussi les archives scientifiques des laboratoires et programmes de recherche.



➤ Exemple de changements épistémologiques induits par les humanités numériques : une organisation archivistique par l'analyse et la visualisation de réseaux de documents. Archives de la Société des Nations (ONU Genève).

mettant leur traitement, leur conservation, leur accès, en favorisant leur interopérabilité dans une perspective européenne. À cette mission principale s'ajoute une mission de prospective sur l'évolution des besoins documentaires et technologiques et des activités de recherche et de développement visant à améliorer ce dispositif technologique (particulièrement en liaison avec le domaine de l'information scientifique et technique). Dans le cadre de cette mission, Huma-Num prospecte dans le domaine stratégique de l'utilisation/réutilisation des données de la recherche, dans et pour les publications scientifiques, en favorisant des interactions avec les acteurs actifs dans ce domaine (plateformes d'éditions électroniques, archives ouvertes en particulier, en France et au niveau européen). Huma-Num propose en outre – et c'est là sans doute l'un des aspects les plus importants dans ce type d'infrastructure – des guides de pratiques technologiques à destination des chercheurs. Ces guides, qui sont nourris à la fois par les interactions au sein des consortiums, mais aussi entre les consortiums et avec nos partenaires internationaux, offrent une base d'information pour les projets de recherche qui ont à documenter, traiter, diffuser et archiver des données numériques. Ainsi, il ne s'agit pas d'offrir un « guichet » où des ingénieurs et chercheurs pourraient donner des conseils et apporter une « bonne parole technologique », mais plutôt de créer les conditions d'une concertation collective dans et avec les communautés en reliant leurs préoccupations à des solutions adaptées à l'état de l'art international (web sémantique, conservation des données, etc.).

DES CONSORTIUMS DISCIPLINAIRES LABELLISÉS

La concertation collective est assurée par les relations de la TGIR Huma-Num avec neuf consortiums disciplinaires, mais aussi avec les Maisons des

sciences de l'Homme, les laboratoires et équipes de recherche. Cette concertation permet d'identifier et de guider la mise en œuvre d'outils et services qui ont sens à être déployés et mutualisés au niveau national, voire le cas échéant au niveau européen. La TGIR s'appuie sur l'activité de consortiums disciplinaires ou multidisciplinaires, qu'elle labellise et finance pour faire évoluer son dispositif technologique. Ce mode de fonctionnement, croisé avec une veille technologique importante et une implication dans le domaine de l'IST, offre les conditions pour anticiper les évolutions du monde numérique dans le contexte SHS.

Ainsi les consortiums ont vocation :

- à mutualiser outils et méthodes relatifs à la constitution de corpus numériques pour la recherche en SHS (corpus de données dit « qualitatif »). Ils fédèrent les initiatives, aident à la mise en commun des sources primaires ou secondaires et facilitent leur diffusion par une sensibilisation à l'indispensable éditorialisation et mise en contexte de leurs métadonnées. À ce stade, Huma-Num encourage l'utilisation d'outils communs avec le monde des bibliothèques ;
- à produire des guides de bonnes pratiques disciplinaires, en s'appuyant notamment sur les échanges qu'ils entretiennent avec les pratiques « métiers » des domaines de l'information scientifique et technique, de l'informatique et de l'édition électronique⁵. Huma-Num relie dans ce cas les propositions interdisciplinaires afin de préfigurer des outils communs ;
- à accroître la visibilité des différents projets dont ils sont porteurs en les inscrivant dans les initiatives internationales et européennes telles que, par exemple, Dariah-Eu⁶. Il s'agit ici principalement de construire une pratique commune d'utilisation de normes adaptées (scientifiques, communautaires, techniques) dans la constitution et la réutilisation des corpus.

En conclusion, Huma-Num se propose d'accompagner les changements de méthodes, d'outils et de pratiques, liés au numérique en fournissant les conditions d'une concertation collective permettant aux communautés scientifiques de progresser, à leurs rythmes, dans la définition et la maîtrise des techniques de traitement, d'enrichissement, de signalement et d'archivage des données numériques.

STÉPHANE POUYLLAU

Directeur technique d'Huma-Num
stephane.pouyllau@huma-num.fr



POUR EN SAVOIR PLUS

SITE DE HUMA-NUM :

www.huma-num.fr

CARNET D'HUMA-NUM :

<http://humanum.hypotheses.org>

[5] Dernière parution : Guide méthodologique : Les outils de conversion vers le format PDF (2) : Traitement de texte, dessins techniques, édition scientifique, janvier 2014, www.huma-num.fr/ressources/guides

[6] www.dariah.eu



● ● ● LES 9 CONSORTIUMS LABELLISÉS Huma-Num

Consortium Mémoires des archéologues et des sites archéologiques (Masa)

Il réunit des acteurs nationaux déjà rassemblés dans le réseau national des Maisons des sciences de l'Homme traitant d'archéologie et le Musée d'archéologie nationale. En vue de favoriser toute recherche nationale ou internationale, il propose la mise en place d'un accès unifié à des corpus numérisés avec des procédures et des outils documentaires et archivistiques communs.

Consortium Sources médiévales (Cosme)

Pluridisciplinaire, ce consortium s'appuie sur une communauté scientifique largement ouverte à toutes les disciplines, rassemblant archéologues, historiens, historiens de l'art, philosophes, linguistes, littéraires..., réunie autour d'une pratique commune : la recherche sur le passé médiéval par le biais des sources les plus diverses. Ces sources sont au cœur des préoccupations du consortium : projets d'édition ou de transcription, de reproduction numérique ou d'un répertoire les recensant.

Consortium Cartes et photographies pour les géographes

Son objectif est de rendre accessibles, consultables et mobilisables des données cartographiques et photographiques nombreuses et éparses en géographie, qui constituent des fonds de laboratoires de recherche, de bibliothèques remarquables ou des fonds de chercheurs, ressources en général sous-exploitées faute d'accès en ligne ou de documentation suffisante.

Consortium Archives des sciences sociales du politique (Archipolis)

Il a pour mission d'archiver les recherches engagées à l'interface des sciences sociales et portant sur les différentes composantes de l'objet politique – des plus spécifiques telles que les élections, les partis, les syndicats et autres organisations politiques, le métier politique, l'action publique, les institutions, la citoyenneté, mais aussi plus largement toutes formes collectives d'appartenances et de comportements mettant en jeu les rapports de pouvoir.

Consortium Archives des mondes contemporains

Il a pour objet la création d'un instrument de recherche collectif qui permette de répondre aux problèmes qui découlent de la nature même des archives de la période contemporaine (XIX^e-XX^e siècles), caractérisées tout à la fois par leur masse, l'hétérogénéité des supports

(du papier au microsillon en passant par la photographie ou les disquettes informatiques, cédéroms et supports vidéo anciens et récents), leur fragilité, la diversité de leur statut juridique (public ou privé) et la dispersion des lieux où elles sont conservées.

Consortium Cahier - Corpus d'auteurs pour les Humanités : informatisation, édition, recherche

Transversal et interdisciplinaire, il fédère les différentes initiatives existantes ou en projet en France dans les domaines des « corpus d'auteurs », qu'ils relèvent de la littérature, de la philosophie ou d'une thématique liée à une école ou à une pratique, afin d'apporter coordination, partage d'expériences et de favoriser l'accès aux données. Les corpus pris en compte dans ce consortium sont fortement associés à une activité éditoriale, qu'elle soit seulement numérique ou qu'elle ait un double support (sur papier et en ligne).

Consortium Corpus oraux et multimodaux (Ircom)

Il a vocation à fédérer les équipes, laboratoires, chercheurs et enseignants-chercheurs engagés dans la constitution de corpus oraux et multimodaux, afin de faire converger les pratiques et de les rendre conformes aux standards internationaux.

Consortium Corpus écrits

Il fédère les équipes et laboratoires, les chercheurs, enseignants-chercheurs, ou ingénieurs engagés dans la production de corpus numériques écrits, quels que soient la langue et l'alphabet considérés, afin d'accompagner le développement des corpus écrits, d'en faire converger les pratiques et les besoins, et de financer des actions répondant à ses missions.

Consortium Archives des ethnologues

Il répond aux questions spécifiques posées aux ethnologues devant la singularité de leurs corpus, l'importance théorique et patrimoniale de leurs travaux, la richesse et la diversité humaine qu'ils représentent. Il s'agit, dans un premier temps, d'assurer la conservation et la pérennisation des données collectées par les ethnologues sur le terrain ainsi que de tous documents pouvant aider à leur contextualisation. Dans un deuxième temps, son objectif est de mettre à disposition ces matériaux dans le respect de la propriété intellectuelle et des normes éthiques.

Source : www.huma-num.fr/service/consortium