



HAL
open science

Les sources numériques de la recherche dans les sciences humaines et sociales

Fatiha Idmhand, Richard Walter

► To cite this version:

Fatiha Idmhand, Richard Walter. Les sources numériques de la recherche dans les sciences humaines et sociales. Le numérique bouleverse la donne : vers de nouvelles démarches scientifiques ?, Colloque annuel de l'Académie Sciences Lettres Toulouse, Académie Sciences Lettres Toulouse, Oct 2014, Toulouse, France. halshs-01225484

HAL Id: halshs-01225484

<https://shs.hal.science/halshs-01225484>

Submitted on 6 Nov 2015

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 Attribution - NonCommercial - NoDerivatives 4.0
International License

Colloque annuel de l'Académie Sciences Lettres

Toulouse, 15 octobre 2014

« Le numérique bouleverse la donne : vers de nouvelles démarches scientifiques ? »¹

Les sources numériques de la recherche dans les sciences humaines et sociales

Fatiha Idmhand (ITEM, Université du Littoral Côte d'Opale)

Richard Walter (ITEM, CNRS)

L'arrivée de l'outil numérique au sein des Sciences Humaines et Sociales a bouleversé les métiers de la recherche et introduit de nouvelles formes de gestion, de diffusion, d'exploitation et d'archivage des sources. Alors que les objets de recherches sont, de plus en plus souvent, numériques, les chercheurs utilisent eux aussi ces nouvelles technologies pour réaliser leurs recherches et produire, à leur tour, des données numériques. Les sources qui circulent aujourd'hui sont donc, de plus en plus, nativement numériques, et chercher, lire ou exploiter des fichiers électroniques sont devenus des actions « naturelles ».

Les conditions d'analyse, d'édition et de sauvegarde de ces sources numériques de différentes natures constituent donc un nouvel enjeu de réflexion épistémologique et scientifique, et cela est d'autant plus vrai dans un laboratoire comme le nôtre, l'Institut des textes et manuscrits modernes (UMR 8132), où la source papier manuscrite et/ou imprimée, est encore au cœur des préoccupations du laboratoire et des projets de recherches.

En effet, les archives des écrivains contemporains représentent, par leur originalité, leur diversité et leur amplitude, un défi pour les chercheurs qui les étudient, et pour les institutions qui acceptent de les recevoir et d'assumer leur conservation, leur préservation et leur consultation. Ces patrimoines sont l'une des « raisons d'être des bibliothèques et des archives,

¹ <http://academie-sciences-lettres-toulouse.fr/colloque-annuel-de-lacademie-le-15-octobre>

publiques et privées [qui] conservent ces aventures de la pensée » (Pierre-Marc De Biasi²), aujourd'hui, ils incarnent aussi une réserve de données importantes pouvant être exploitées et/ou éditées autrement grâce à l'outil numérique. Mais à condition de respecter certaines règles.

Les recherches en SHS au sein du paysage numérique

Depuis quelques années, les projets d'éditions électroniques des sources se multiplient au sein de la recherche en SHS et nombreux sont les chercheurs qui veulent diffuser des matériaux, supports ou résultats sur le web de même que les bibliothèques et centres de documentation qui numérisent et republient, pour mieux les valoriser, leurs collections et patrimoines culturels et/ou scientifiques. Or ces (re)publications de sources, de natures et formats multiples, présentent quelques difficultés ; en effet, si la transition massive vers le numérique semble inéluctable, la perte de données l'est également si aucune structuration de l'information à archiver n'a été réalisée et si le respect des normes et standards, méthodes et coutumes, actuelles et futures n'a été anticipée. Le passage au numérique exige donc une préparation fine car chacun des choix éditoriaux peut avoir des conséquences irréversibles : c'est donc très en amont qu'il faut mesurer les impacts de ces choix, dès le montage du projet.

Le travail en « mode projet » est nouveau pour les SHS, c'est une nouvelle façon d'appréhender et de gérer le projet scientifique qui désespère encore le chercheur en SHS, notamment s'il a un projet « numérique » à développer. Si la démarche scientifique n'est pas remise en cause par de telles approches, gérer des sources numériques demande au chercheur d'affronter deux domaines qui ont, durant longtemps, été étrangers pour lui : celui de la technique et celui de la gestion du projet. Ce phénomène a d'ailleurs été largement accentué depuis la création de l'ANR (2005) et de l'AERES (2006) et depuis la généralisation, par le biais de ces instances, de l'évaluation de la recherche et d'une partie importante de son financement sur des appels à projets. L'instauration de ces nouvelles façons de financer la recherche, qui ont coïncidé avec la généralisation des usages du numérique et l'émergence de l'édition électronique et « ouverte »³, ont conduit les chercheurs en SHS, depuis 2005, à apprendre à mesurer le temps dans un domaine qui ne le métrait pas forcément, à évaluer des coûts en euros et « personne-mois », bref, à acquérir des compétences techniques et managériales comme un chef de projet.

² Pierre-Marc De Biasi, « Le patrimoine écrit », en ligne : <http://www.item.ens.fr>

³ Celle-ci, bien qu'elle parut moins chère dans un premier temps car la mise en ligne semblait effacer les coûts, se révèle coûteuse en temps, en infrastructures et en ressources humaines.

L'édition numérique de sources en SHS

Les projets d'éditions numériques se sont développés au sein des SHS en même temps que le bouleversement du contexte institutionnel et financier. Si les premiers projets financés ont malheureusement souffert du manque d'accompagnement « numérique », le recours aux prestataires privés pour la conception des sites web ou la mise en place des bases de données n'a, souvent, pas anticipé les mutations du web ou les problématiques de l'hébergement pérenne des données. Le secteur public, qui avait longtemps négligé ce domaine, a mis en place depuis 2008, des services destinés à accompagner et soutenir les projets numériques au sein des SHS. Dorénavant, il existe, comme dans les sciences dites « dures », de « Très Grandes Infrastructures »⁴ (TGIR/TGE) en SHS et des consortiums qui apportent, en amont, des conseils et des recommandations techniques aux porteurs de projets (appelées « bonnes pratiques ») avant la mise en œuvre d'actions dont les conséquences sont irréversibles.

Dans le même temps, le « champ numérique » des projets en SHS se caractérise aussi également par la nécessité de la formation aux outils et aux usages : la notion de « formation tout au long de la vie », régulièrement employée en France depuis la loi du 25 novembre 2009⁵ et associée à l'acquisition de compétences techniques et « professionnalisantes » commence à trouver un écho au sein des SHS. Depuis quelques années, des « formations aux outils numériques » sont régulièrement proposées par les Consortiums des TGIR, par des MSH ou à la suite d'initiatives locales issues des centres de recherches, pour aider les chercheurs et ingénieurs des SHS à comprendre l'environnement qui les entoure et à prendre conscience de l'importance de la constitution de l'équipe. .

Aujourd'hui, s'engager dans un projet numérique demande l'association de différentes compétences et ce, qu'il s'agisse de sources nativement numériques ou de sources qui ont été transformées en objets numériques (grâce à la numérisation notamment). Face au « tournant numérique », les conditions de production et de diffusion des savoirs ont évolué : on parle dorénavant d'« Humanités Numériques » pour définir l'espace de recherche interdisciplinaire et transversal, dans lequel s'inscrivent les projets « numériques » des SHS.

⁴ TGE Adonis, TGIR Huma-Num, Prodego...

⁵ Voir <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000021312490&categorieLien=id> et voir également : <http://www.education.gouv.fr/cid217/la-formation-tout-au-long-de-la-vie.html>

Les « Humanités numériques » ou les SHS à l'heure du numérique

Les « Humanités Numériques » désignent un domaine traversé par plusieurs disciplines et corps de métiers qui « collaborent »⁶ ensemble à la réalisation d'un même projet de recherche. En 2010, le « Manifeste des Humanités Numériques » a proposé la définition suivante :

Les *digital humanities* désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales.⁷

A l'époque, la définition des *ThatCampers* parlait des *Digital Humanities* car la notion ne s'était pas encore « francisée » mais depuis 2013, après le ThatCamp de Saint-Malo, les *ThatCampers* se sont mis d'accord pour utiliser l'expression « Humanités numériques » et pour constituer une « Association francophone des Humanités Numériques », à l'instar d'autres associations importantes du monde des *Digital Humanities*.⁸ La création récente d'« Humanistica »⁹, dont les statuts ont été votés en juillet 2014, la multiplication de rencontres ou de colloques qui tentent d'analyser le « phénomène numérique », l'accroissement d'écrits¹⁰ au sein des *Internet studies*¹¹ et de rapports d'activités montrent que le mouvement se structure mais qu'il est encore difficile, au sein des SHS en tous cas, de circonscrire le champ de ses applications et de ses actions.

Le rapport qui a été réalisé par Marin Dacos et Pierre Mounier à la demande de l'Institut Français, *Humanités numériques : état des lieux et perspectives*¹² en 2013 est l'état des lieux le plus récent sur les « Humanités Numériques » en France. Ce rapport propose une carte des principaux projets en « Humanités Numériques » en France, en Europe et dans le monde, il confirme l'essor du mouvement¹³ et offre un historique des « Humanités Numériques » depuis

⁶« Digital approaches are conspicuously collaborative and generative » in Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, Jeffrey Schnapp, *Digital Humanities*, Massachusetts, The MIT Press, 2012, p.3.

⁷<http://tcp.hypotheses.org/318>

⁸Par exemple : Association for Computers and humanities ou Humanidades Digitales...

⁹<http://www.humanisti.ca>

¹⁰ Voir la bibliographie sur le sujet publiée que nous avons élaborée à l'occasion du colloque [DH] <http://dhnord2014.meshs.fr/ressources-multimedia/bibliographie-1>, ou celle d'autres sites comme celui d'Aurélien Berra <http://philologia.hypotheses.org/author/philologia> ou de l'Association for Computers and humanities : <http://digitalhumanities.org/answers/topic/lectures-de-base-en-francais-pour-les-digital-humanities>

¹¹Michel Wieviorka, *L'Impératif Numérique ou La nouvelle ère des sciences humaines et sociales ?*, CNRS éditions. 2013, p. 7-9.

¹² Marin Dacos, Pierre Mounier, *Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international*, Rapport commandé par l'Institut français, Ministère des Affaires étrangères pour l'action culturelle extérieure de la France, en ligne sur : <http://www.institutfrancais.com>, 2014.

¹³ Le rapport souligne également l'écueil du soutien institutionnel.

leur émergence dans les années 1950 aux États-Unis¹⁴, depuis les *Humanities computing/Literary and linguistic computing*¹⁵ qui seraient à l'origine des « Humanités Numériques » et des premiers projets associant les SHS et l'informatique. En réalité, l'expression « Digital Humanities » n'a été forgée que très récemment, en 2006, après la publication de l'ouvrage collectif *A Companion to Digital Humanities* par Susan Schreibman, Ray Siemens, John Unsworth.¹⁶ Mais si l'expression « Humanités Numériques » renvoie l'imaginaire vers la dimension « numérique » des « humanités », il faut rappeler qu'au cœur de ces domaines il y a toujours et en premier lieu, un projet scientifique : celui d'un ou de plusieurs spécialistes d'un objet, qui ont une expertise importante de celui-ci, laquelle est généralement issue d'une formation disciplinaire poussée, et qui désire(nt) enrichir l'examen de cet objet à l'aune des nouvelles propositions et des nouveaux outils : ceux du numérique en particulier. La prise en compte de ces outils, parmi les instruments d'analyse traditionnels, est une conséquence de la « grande conversion numérique » de notre société, une métaphore religieuse que Milad Doueïhi¹⁷ utilise pour signaler la « tendance universaliste »¹⁸ d'un phénomène qui transforme en profondeur les fondements de nos cultures. Steve Jones, lui, passe par une métaphore médicale pour décrire ce changement de culture qui rompt avec le passé, en parlant d'« éversion »¹⁹, il évoque le passage de la culture imprimée et papier, à la culture numérique.

Deux corps professionnels sont donc particulièrement visés par ce domaine, les métiers de l'informatique et du numérique d'une part et ceux de la recherche en sciences humaines et sociales de l'autre. L'*Homo academicus*²⁰ de demain (ou d'aujourd'hui ?) doit donc associer des techniques et des expertises de plus en plus pointues, dans des domaines très spécifiques, pour accompagner la nouvelle « culture numérique » :

Qu'est-ce que la culture numérique, aujourd'hui si omniprésente, envahissante et familière ? Un ensemble de technologies conjuguées, ayant produit et continuant de produire des pratiques sociales (...) faite de modes de communication et d'échange d'informations qui déplacent, redéfinissent et remodelent le savoir dans des formes et formats nouveaux, et de méthodes pour l'acquérir et le transmettre. (...) elle exige des formes nouvelles et toujours changeantes de savoir-lire, de savoir-faire - une *compétence numérique* loin de se limiter au simple maniement des outils disponibles²¹.

¹⁴ Il est intéressant de remarquer qu'à cette même époque, dans d'autres parties du monde, la généralisation de l'usage de la machine à écrire donnait lieu à une première « mécanisation » de l'écriture.

¹⁵ Voir la définition proposée sur wikipédia : <http://fr.wikipedia.org>

¹⁶ Susan Schreibman, Ray Siemens, John Unsworth, *A Companion to Digital Humanities*, Oxford, Blackwell, 2004. En ligne <http://www.digitalhumanities.org/companion/>

¹⁷ Milad Doueïhi, *La grande conversion numérique*, Paris, Seuil, 2008, p. 23.

¹⁸ Milad Doueïhi, *La grande conversion numérique*, op. cit., p. 23.

¹⁹ Steven E. Jones, *The Emergence of the Digital Humanities*, New York, Routledge, 2013, p. 18.

²⁰ Milad Doueïhi, *La grande conversion numérique*, op. cit., p.195.

²¹ *Ibid.*, p. 37.

Ces qualités, qui sont assez difficiles à réunir en un seul être (surtout au sein des SHS), demandent donc la constitution d'équipes transdisciplinaire, dans un domaine traditionnellement assez cloisonné et associé au travail en solitaire. En effet, et contrairement aux sciences dites « dures », aller travailler au laboratoire a longtemps eu peu de sens au sein des SHS où la recherche est associée à la pensée, à l'isolement dans le bureau personnel, à la bibliothèque, ou à de longues enquêtes de terrain (linguistique, archéologie). Les projets « numériques » exigent dorénavant un travail et une gestion en « mode projet », des réunions d'équipe, des recrutements, la rédaction de cahiers des charges, la répartition des tâches, la définition d'un calendrier rigoureux, une fine coordination d'équipe et une gestion importante de la communication. Si la démarche scientifique reste inchangée, le choix des méthodologies et des outils pour réaliser les analyses ou exploiter les résultats est plus large qu'autrefois, d'où la nécessité de mettre en place des médiations entre l'outil informatique et les chercheurs et entre les utilisateurs et l'outil informatique. Le projet en Humanités Numériques intègre donc un « traducteur »/une « interface » pour transformer un contenu en codage, pour associer les exigences des auteurs/chercheurs avec les attentes des utilisateurs et avec le langage et la structure informatiques choisies (ou imposées). L'édition, traditionnellement située à l'issue du parcours de recherche, doit être, avec l'usage et la généralisation du numérique, préparée en amont.

L'édition numérique de sources en quelques étapes. Des « Humanités numériques » pour « archiver l'avenir »²²

Le numérique bouleverse les projets éditoriaux en permettant l'édition et la gestion de données hétérogènes dans un même espace virtuel, la publication d'une même source de différentes façons (littérale, diplomatique, éditorialisée, critique, etc.) et en supprimant la limitation du nombre de caractères ou de pages : deux éléments qui ont, pendant longtemps, représenté un frein quantitatif et financier. Les technologies numériques ont donc ouvert l'édition et l'analyse vers l'infini et permettent par ailleurs la comparaison d'éléments issus de médias différents : des sources sonores avec des sources manuscrites, iconographiques... Dans le cadre de travaux portant sur les processus de création, le numérique permet de représenter différents niveaux de relations et offre la possibilité de publier, sans limites, toutes les variantes

²² Voir Milad Douehi, *La grande conversion numérique*, Paris, Seuil, 2008, p. 205-248, le chapitre intitulé « Archiver l'avenir ».

et tous les états génétiques d'un même texte avec leurs transcriptions accompagnées de commentaires et d'analyses.

L'éditeur de ces sources ne se contente plus seulement de publier, il anticipe et/ou provoque les multiples usages de telles éditions. Prenons l'exemple d'une source comme le manuscrit : c'est un objet fragile, unique qu'il faut préserver pour pouvoir l'exploiter. En tant que document à traiter par l'archiviste, le problème de la définition des « unités documentaires » à traiter et du niveau de description de chacune de ces unités se pose lorsqu'il s'agit d'harmoniser l'approche scientifique et génétique du chercheur avec la cohérence et les principes de l'archivistique. Depuis la fin du dix-neuvième siècle, notamment dans le cas des écrivains, le manuscrit est également un objet sériel qui offre plusieurs « états génétiques » dont les contenus varient en fonction des modifications qui ont été apportées par l'auteur du manuscrit. Numériser et éditer un tel objet suppose donc de prendre en compte ces différentes dimensions de l'objet (documentaire, archivistique, génétique) dont le contenu varie selon les époques et comporte des informations de multiples natures qui nécessiteront d'être structurées (usage de descripteurs, de normes et de règles différentes). L'association entre le chercheur et l'ingénieur permet alors de définir, selon le projet scientifique (travaux sur l'objet, le texte, la langue, la phrase, l'expression, le mot, l'image, le contexte...), l'utilisation d'une technologie ou d'une autre, de processus de normalisations (point, orthographe, grammaire...) ou non mais surtout de principes de description de l'objet dont la bonne granularité est un bon gage de viabilité du projet.

Le chercheur éditeur-numérique est donc confronté, à chaque étape de son projet, à des choix de différents natures selon les niveaux de l'édition qu'il envisage : d'un côté l'édition scientifique, de l'autre les impératifs techniques ; d'un côté l'interopérabilité nécessaire en vue de la mutualisation et, de l'autre, des choix technologiques spécifiques et en fonction de besoins personnels... A chaque étape, une discussion et une médiation sont donc nécessaires car si l'édition critique numérique obéit aux mêmes principes d'exigence et de qualité que l'édition critique papier, celle-là a quelques impératifs techniques supplémentaires comme le respect des normes du domaine numérique mais permet, contrairement à la version papier, des mises à jour, des expansions et une profondeur d'analyse qui s'affranchit du coût du papier. La réussite d'un projet éditorial numérique dépend donc du respect de ces règles qui garantissent la compréhension de demain.

L'encodage est donc nécessaire pour diffuser, archiver, analyser, exploiter et préserver l'intégrité et les droits des documents publiés en ligne. Si ces opérations peuvent sembler extrêmement techniques, elles sont surtout scientifiques car l'encodage doit être compris

comme un processus destiné à faciliter ou à enrichir la compréhension et non à réduire le sens du document édité. Cette opération scientifique implique, *au premier chef*, le chercheur dont le rôle est à la fois de rendre cet objet accessible aux autres (chercheurs, public) mais également d'interpréter les données recueillies afin qu'elles soient diffusées, exploitées par d'autres, et, surtout, archivées. Ces opérations correspondent par ailleurs à des « temporalités scientifiques » différentes, indissociables et sensibles car la plupart d'entre elles sont « one shot » (à usage unique). Face à la masse croissante de contenus numériques produits, devant la variabilité et les contraintes du virage vers le *tout numérique*, les problèmes posés par la gestion de telles sources au sein des SHS sont, indéniablement, plus importants qu'autrefois.

Il n'existe aucune méthode ni solution « clé en main » pour l'instant et tous les problèmes se posent en même temps, les SHS doivent prendre en compte :

- le problème spécifique de la détérioration des supports avec le temps,
 - l'apparition de sources « nativement » numériques,
 - l'hétérogénéité des formats d'encodage,
 - la question de l'archivage à long terme de l'information numérique et de l'urgence à préserver cet « avenir fragile »²³,
 - le respect des normes internationales (Dublin Core, XML, TEI, Métadonnées...),
- et la gestion des compétences « novices » et enthousiastes (trop ?) des SHS dans ces domaines.

²³ Milad Douehi, *La grande conversion numérique*, op .cit., p. 205-248.