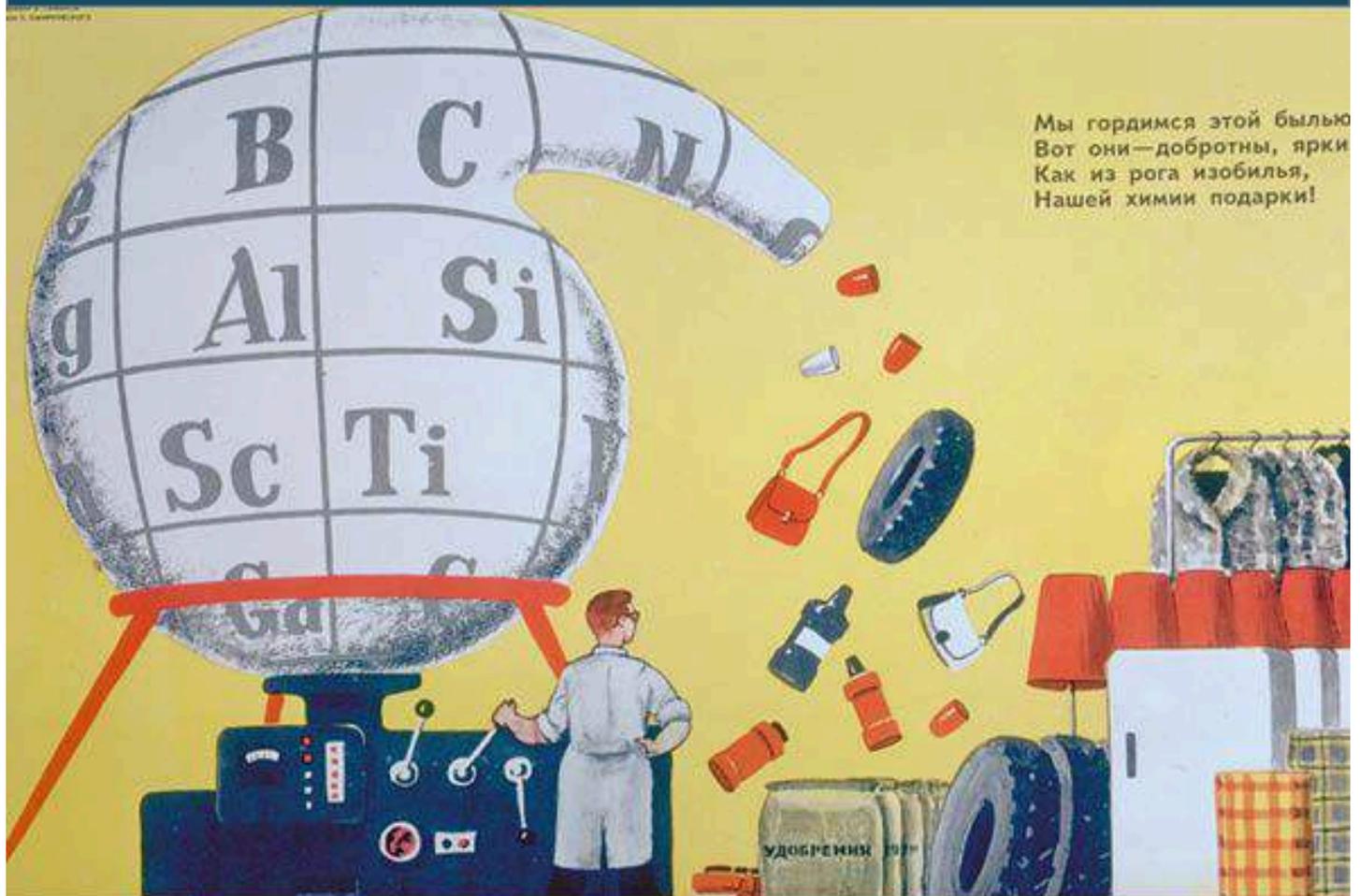


THATCamp Paris 2012

Non-actes de la non-conférence
des humanités numériques



non-collection
des Editions de la Maison
des sciences de l'homme
en libre accès sur
<http://editions.msh.revues.org/>

@œuvre collective
#tcp2012

THATCamp Paris 2012

Non-actes de la non-conférence des humanités numériques

DOI : 10.4000/books.editionsmsmh.278

Éditeur : Éditions de la Maison des sciences de l'homme

Lieu d'édition : Paris

Année d'édition : 2012

Date de mise en ligne : 1 octobre 2012

Collection : La Non-Collection

EAN électronique : 9782735115273



<https://books.openedition.org>

Référence électronique

. *THATCamp Paris 2012 : Non-actes de la non-conférence des humanités numériques*. Nouvelle édition [en ligne]. Paris : Éditions de la Maison des sciences de l'homme, 2012 (généré le 15 novembre 2023). Disponible sur Internet : <<http://books.openedition.org/editionsmsmh/278>>. ISBN : 9782735115273. DOI : <https://doi.org/10.4000/books.editionsmsmh.278>.

Légende de couverture

Travail et patrie - URSS 1970

Crédits de couverture

Éditions de la Maison des sciences de l'homme

Ce document a été généré automatiquement le 15 novembre 2023.



Le texte seul est utilisable sous licence CC BY-SA 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.

RÉSUMÉS

Les 25 et 26 septembre 2012, le Centre pour l'édition électronique ouverte a organisé son deuxième THATCamp à l'Ecole des Hautes Etudes en Sciences Sociales à Paris. Inventés par le Center for History and New Media de l'Université George Mason à Washington, les THATCamps (The Humanities and Technology Camp) sont des rencontres informelles entre acteurs des humanités numériques qui leur permettent de partager expériences et bonnes pratiques dans ce domaine de recherche en pleine expansion. Plus de cent cinquante personnes venant de toute la France et provenant des toutes les disciplines des sciences humaines et sociales ont participé à l'événement. Cette rencontre s'est déroulée près de deux ans après la première du genre en France, au cours de laquelle avait été rédigé le *Manifeste des digital humanities*. Réalisé en *booksprint* par une équipe de plus de dix personnes en deux jours, cet ouvrage fait le compte rendu des échanges qui ont eu lieu au cours de cette non-conférence.

NOTE DE L'ÉDITEUR

Les tweets échangés à propos du THATCamp Paris 2012 ont été compilés dans un tableur disponible en accès ouvert à l'adresse suivante : <https://archive.org/details/Tcp2012>

SOMMAIRE

Avant-propos

Introduction

Collectif

Les digital humanities sont-elles solubles dans le Steampunk ?

Conférence inaugurale

Paul Bertrand

Première couche : Une insoutenable tension dialectique : autour d'une esthétique du *steampunk*

2. De grands espoirs : la dissémination et l'appropriation

3. Deuxième couche : qu'est-ce qui coince ? L'idéologie du Steampunk

3. Deuxième couche : pourquoi ça coince ?

4. Sortir de l'idéologie du *steampunk*

Discussion

Notions, formation, compétences

L'historien programmeur ?

Proposé par : Frédéric Clavert, Aurélien Berra, Franziska Heimbürger

Collectif

Introduction

I. Pourquoi un chercheur se pose la question du code ?

II. La place du chercheur dans la chaîne de développement, de l'expression du besoin à l'aboutissement du projet

III. Interrogations

Conclusion

Formations et humanités numériques en France

Aurélien Berra et Florence Clavaud

Collectif

Présentation

Conclusions

Quelles compétences et littératies pour les humanités numériques ?

Proposé par Olivier Le Deuff et Aurélien Berra

Collectif

Un flux de questions

Une formation ou une méthode ?

Le point de vue des informaticiens

Question de gouvernance

Qui sont les Digital Humanists ?

Lieux et communautés

Quels espaces physiques pour des humanités numériques ?

Atelier proposé par Nicolas Catzaras & Johann Holland
Collectif

Introduction
Quelques développements à partir des mots clefs
Tour de table des intérêts
Synthèse des sujets abordés
Comment concevoir les espaces en fonction des usages ?
Éléments de conclusion/actions
Référence proposée

Les réseaux sociaux numériques de chercheurs en SHS

Proposé par Elifsu Sabuncu et Antoine Blanchard, animé par Nicolas de Lavergne et Olivier Le Deuff
Collectif

Des réseaux sociaux scientifiques
Les pratiques des chercheurs
Prédominance des besoins de veille
Besoins et activités
Périmètre

Sciences participatives

Proposé par Jean-Pierre Girard
Collectif

Introduction
Tour de table et discussions
Pour conclure : quelques pistes de réflexion

Plate-formes participatives

Proposé par Marie-Madeleine Mervant-Roux, ARIAS
Collectif

Présentation
Un cas
La présentation de l'outil
Comment le projet peut-il se transformer avec l'outil ?
L'ouverture de la plate-forme
La question du temps

Quelles coopérations d'acteurs dans une politique d'édition numérique pour un laboratoire en SHS ?

Dorian Ryser et Jean-Pierre Masse
Collectif

L'exemple d'un laboratoire en SHS
La communauté s'interroge
Pour conclure : une question de gouvernance

Comment écrire pour un carnet de recherche ?

Proposé par Mélodie Faury et Pierre Mounier
Collectif

Illustrations
Pourquoi la question se pose-t-elle ? Démêler quelques enjeux
Des fonctionnalités aux fonctions des carnets de recherche : de nouvelles formes d'écriture
Conclusion

La diffusion de la recherche

Comment commencer, poursuivre et pérenniser une revue en ligne ?

Proposé par Paul-Henri GIRAUD
Collectif

Pourquoi une revue en ligne ?

Complémentarité entre édition papier et édition numérique

Il est très difficile, à l'heure actuelle, de publier une revue en ligne en dehors d'une plate-forme.

L'édition

Éditeurs SHS et humanités numériques

Proposé par Marion Colas et Maria Vlachou
Collectif

Introduction

L'édition, différentes manières de faire

L'édition et les humanités numériques

De l'édition en sciences humaines et sociales sans éditeur

Quel(s) modèle(s) économique(s) pour l'édition électronique ?

L'impression à la demande, un service demandé et apprécié

Et le libraire dans tout cela ?

L'édition multi-supports et de la diffusion multi-canaux

Le métier d'éditeur perdurera

Évolution de l'archive ouverte HAL-SHS

Proposé par Laurent Capelli
Collectif

Intégration d'un moteur de recherche

Élargissement de la typologie de documents déposables aux vidéos

Le référentiel chercheur et la question de l'évaluation

Dépôt du texte intégral et de notices simples

Épisciences.org : création d'épi-revues

Conclusion

Open data en SHS

Proposé par Cynthia Pedroja, Elifsu Sabuncu, Anne-Laure Stérin
Collectif

Introduction

Quelle est la pertinence d'une diffusion des données SHS en open data ?

À qui appartiennent les données ?

Comprendre l'open data au-delà de la seule recherche académique

La nécessaire définition de protocoles et de formats pour la diffusion des données

Conclusion

La diffusion des sciences humaines et sociales vers un large public : panorama des outils de médiation numérique

Proposé par Isabelle Jouve et Lars Wieneke
Collectif

Introduction

Un exemple de médiation numérique : Le projet Goulag

Communication et médiation, quelle différence ?

L'interaction dans le numérique

Objectifs et motivations de la médiation numérique ?

Les compétences nécessaires

La pérennité

Instrumenter la recherche

Utiliser l'ontologie CIDOC CRM pour l'information relative au patrimoine culturel

Proposé par Anne-Violaine SZABADOS (équipe LIMC-ArScAn / Archéologies et Sciences de l'Antiquité - UMR7041, CNRS), Katell BRIATTE (Ministère de la Culture et de la Communication) et Rosemonde LETRICOT (Irht)
Collectif

Présentation

Introduction

Interopérabilité

Communauté en France

Discussion

Conclusions, perspectives

Utilisons RDFa pour nos corpus

Proposé par Stéphane Pouyllau
Collectif

Introduction de Stéphane Pouyllau

L'utilisation des syntaxes RDF et RDFa

Identifier les ontologies utiles aux SHS

Automatiser la sémantisation des corpus

Inciter à l'utilisation d'une structuration de l'information exprimée à l'aide de la syntaxe RDFa

L'historien proto-programmeur. Outils et méthodes pour (re)travailler ses données

Proposé par Sylvain Machefert
Collectif

L'autonomie dans la manipulation de ses données

La structuration des données

De la ligne de commande aux interfaces visuelles

Conclusion

Ressources

Sommes-nous en train de perdre la mémoire ? Mémoire et archivage du web

Proposé par Frédéric Clavert et Clément Oury
Collectif

Archiver le Web : la politique de la BNF

Des problèmes classiques qui se posent de façon plus aiguë

Repenser la notion de préservation

Conclusion

Rôle et enjeux du design graphique

Proposé par : Nicolas Thély, David-Olivier Lartigaud et Gilles Rouffineau
Collectif

Les enjeux du design graphique

Démonstration à partir d'un outil de visualisation

Le public visé par le design : le cas des chercheurs

Avant-propos

- 1 L'objet que vous avez devant les yeux est une expérimentation collective. Ce sont les « non-actes de la non-conférence » THATCamp qui rassembla à l'EHESS à Paris en septembre 2012, des personnes venant de tous les métiers de la recherche, de plusieurs pays européens, de toutes les disciplines des sciences humaines et sociales autour de la notion d'« humanités numériques ».
- 2 Un THATCamp est conçu comme une non-conférence, c'est-à-dire un rassemblement informel d'acteurs construisant collectivement leur propre programme et partageant expériences et idées au cours d'ateliers auto-gérés. Que peut-il rester d'une non-conférence, hormis le souvenir joyeux dans la mémoire des participants d'échanges inattendus et enrichissants ?

Atelier visualisation de réseaux



- 3 Nous avons voulu aller un peu plus loin cette fois-ci, en faisant le pari de publier les « non-actes » de cette rencontre. Le texte que vous vous apprêtez à lire a été rédigé et édité par une dizaine de personnes, en deux jours à la suite du THATCamp. « Booksprint » est le terme qui désigne cette méthode originale de publication inventée par Tomas Krag et utilisée avec succès pour la rédaction de manuels.
- 4 Un ouvrage rédigé et préparé en deux jours n'est pas comparable à celui qui aura suivi un long processus d'édition au cours de plusieurs mois, c'est évident. La forme en est moins parfaite. Des erreurs subsistent. Mais il bénéficie aussi de l'énergie et de la vitalité qui s'est dégagée du fait du travail très collectif qui l'a engendré juste avant et pendant sa préparation. Les textes en sont moins dilués ; ils vont à l'essentiel. Il permet surtout à ceux qui ont participé à la non-conférence, d'en obtenir immédiatement une trace écrite, avant que leur attention ne soit capturée par d'autres projets.
- 5 La réalisation de cet ouvrage n'aurait pas été possible sans l'investissement de toute l'équipe des Éditions de la Maison des sciences de l'homme, qui en a conçu l'idée, a accueilli l'équipe de rédaction dans ses bureaux, et s'est investie totalement dans la rédaction et la préparation éditoriale de l'ouvrage. Il s'agit de « non-actes » qui, comme la « non-conférence » repose sur les principes d'une intelligence collective, ou mieux, distribuée. Ne cherchez pas de chef d'orchestre, de directeur d'ouvrage ou de chaîne de commandement, car il n'y en a pas. Ne cherchez pas d'auteurs solitaires et géniaux, car il n'y en a pas. Cherchez plutôt des collectifs au sein desquels le travail a circulé de manière fluide en fonction du bon vouloir et des compétences de chacun.

Début du BookSprint



- 6 Cherchez une équipe de rédaction : Valérie Aubourg, Ouafa Boubia, Emmanuel Chateau, Frédéric Clavert, Emmanuelle Corne, Marin Dacos, Nicolas De Lavergne, Mélodie Faury, Nathalie Fourrier, Carmen Husti, Cécile Lochet, Pierre Mounier, Solène Roulet, Luc Tesnière, Astrid Thorn Hillig, Nahalie Vaillant
- 7 ... et surtout, cherchez les 150 THATCampeurs qui ont coproduit par leurs échanges, via leurs tweets et les pads qu'ils ont alimentés, un état des questions sur les humanités numériques dont les textes qui suivent rendent compte.

- 3 Ce qui compte dans un THATCamp, c'est la participation des *campers*. Qu'est-ce que cela veut dire ? Après la conférence inaugurale, va se tenir une assemblée générale au cours de laquelle le programme sera présenté par vous, les ateliers définis par le vote des participants.
- 4 Il y aura 5 salles à disposition, la qualité du programme ne dépend que de vous. Certains ont d'ores et déjà fait des propositions sur le Wiki. Ils auront 5 minutes pour les présenter et les proposer. D'autres propositions non présentées sur le BarCamp pourront également être faites.
- 5 La qualité du programme dépend donc de votre participation. Réfléchissez-y donc d'ici l'assemblée générale.
- 6 Une fois le programme déterminé, les organisateurs d'ateliers feront une introduction d'une dizaine de minutes, mais les ateliers eux-mêmes seront le produit des interventions des participants à chacun des ateliers. En règle générale, le but est de mêler sciences sociales et technologies. L'objet des ateliers devant plutôt être concret et viser à un transfert de savoir-faires.
- 7 **Pierre Mounier** : Quelques aspects pratiques : Il y a une équipe à l'organisation du THATCamp que vous avez peut-être déjà repérée : elle se compose de Nadine, Lisa, Céline, Annik, Emmanuelle, Solène. Vous pouvez vous tourner vers elles pour vos questions. Repérez-les bien, je les remercie du fond du coeur pour leur engagement.
- 8 Ce THATCamp n'aurait pas été possible sans l'appui d'un grand nombre de partenaires très divers qui nous ont soutenus par tous les moyens possibles dans l'organisation de cet événement, je souhaiterais leur donner la parole.
- 9 **Jean-Luc Pinol** : pour le TGE Adonis (et Nadine Dardenne également présente qui pourra donner toute une série d'informations). J'étais présent au THATCamp 2010 pour une session sur les données géographiques. Je savais alors déjà que je serai nommé à la direction du TGE. À l'époque le TGE c'était surtout des projets, aujourd'hui le TGE ce sont des réalisations : l'accessibilité aux données numériques au travers d'une grille de service, c'est aussi une plate-forme de recherche et d'enrichissement des données et métadonnées Isidore menée par Stéphane Pouyllau, le TGE est également tête de réseau pour DARIAH en France, représentée par Sophie David, ici présente aussi.
- 10 Il faut remercier également le soutien du Campus Condorcet qui a permis à trois jeunes chercheuses italiennes et espagnole de venir participer à nos travaux : **Françoise Plet-Servant** : ce projet a comme caractère de créer le premier campus innovant en sciences humaines et sociales (ses membres fondateurs sont l'EPHE, ENC, Paris I, Paris III, Paris XIII, la Fondation maison des sciences de l'homme, etc.). Deux sites sur le Nord Est parisien, l'un à Aubervilliers, et l'autre, plus petit et consacré aux Licences, Porte de la Chapelle.
- 11 **Johann Holland** chargé de mission numérique pour le campus : en plus d'un projet immobilier et scientifique, c'est un projet numérique. Cette dimension est très importante dans le projet, on veut en effet offrir un environnement totalement innervé par le numérique, espace qui permette un certain nombre d'usages innovants. Nous espérons également pouvoir nous inscrire en tant que corpus dans le paysage des humanités numériques, d'où notre participation. C'est au titre des ateliers Condorcet que nous avons pu soutenir cet événement.

- 12 Autre soutien, le Centre virtuel de la connaissance de l'Europe, à Luxembourg, **Frédéric Clavert** : le CVCE est un institut scientifique sous tutelle du gouvernement luxembourgeois, un centre de recherche et de documentation sur l'histoire de l'intégration européenne. Nous avons 20 000 documents en ligne. Le rôle du Digital Humanities Lab dont j'ai la responsabilité est de faire le lien entre chercheurs et développeurs, réfléchir à l'évolution de nos méthodes de travail et de ces corpus de recherche qui sont en ligne, mais aussi de participer à diverses réunions pour y apporter notre contribution.
- 13 ALLC The European Association for Digital Humanities : **Marjorie Burghart** : vous connaissez sans doute l'association par le biais de l'organisation des conférences DH. Elle a été créée en 1973. Récemment elle a changé de nom pour devenir une association pour les *digital humanities*, avec la volonté d'ouvrir son champ d'action. Vous connaissez sans doute le journal de la société, LLC. Ou son action pour le soutien de petits projets ou workshop en DH. Il y a actuellement un appel à petits projets en cours (date limite 15 novembre 2012 pour des grants entre 800 et 2 000 euros), pour cet appel, la participation est ouverte aux non membres.
- 14 La Fondation Maison des sciences de l'homme : **Nicolas de Lavergne** : la FMSH soutient l'événement par le biais de l'organisation du *booksprint* par les Editions de la Maison des sciences de l'homme et par l'hébergement du *codesprint* dans ses locaux. La FMSH qui fêtera son 50e anniversaire l'année prochaine, est un acteur fort dans le numérique : vous connaissez sans doute les Archives audiovisuelles de la recherche, mais également le programme TIC-Migrations sur la modélisation et l'usage des TIC par les diasporas sur le Web. On devrait bientôt entendre parler de nouveaux projets numériques de la Fondation.
- 15 La Maison méditerranéenne des sciences de l'homme : **François Siino** : il membre d'un des laboratoires qui la compose. Comme d'autres MSH, c'est une fédération d'unités pluridisciplinaires. Pour ce qui concerne les activités liées au numérique, de nombreuses revues sont hébergées par Revues.org depuis longtemps, la MMSH travaille au développement d'une phonothèque, et héberge des blogs scientifiques avec un portail spécifique sur la plate-forme Hypothèses.
- 16 Le Centre pour la communication scientifique directe (CCSD), est présenté par **Laurent Capelli** : le CCSD est une unité de service du CNRS, et se compose d'informaticiens qui offrent des services à toute la communauté de la recherche. C'est le développement de l'archive ouverte HAL, d'une plate-forme ouverte d'organisation de conférence (Sciencesconf.org), et bien sûr la maîtrise d'œuvre, sous la direction du TGE Adonis, de la plateforme Isidore. En 2013, viendra la publication d'une nouvelle version de l'archive ouverte et la mise en place d'épi-revues, qui seront comme des revues au-dessus des archives ouvertes.
- 17 **Claire Clivaz** présente les prochains événements de l'Université de Lausanne : Ouverture du premier lab DH en août, le premier *DH summer school* aura lieu à Berne les 26 au 26 juin 2013, avec des ateliers de formation à la TEI, etc. L'université accueillera *DH 2014* (en juillet) à Lausanne, après celui qui aura lieu au Nebraska en 2013, sur le thème du *cultural empowerment*.
- 18 Et enfin l'Unité Régionale de Formation à l'Information Scientifique et Technique (URFIST) qui a un réseau d'unités partout en France et qui fait de la formation avancée à l'IST. Évidemment la notion d'humanités numériques pour les SHS est fortement

connectée à cette notion d'IST et cette conscience est en train de se développer au sein des URFIST.

- 19 Des remerciements également au Cléo et à ses tutelles, le CNRS, l'Université d'Aix-Marseille et l'Université d'Avignon pour l'organisation de cet événement. Un remerciement tout particulier à l'Ecole des Hautes Etudes en Sciences Sociales qui accueille cette rencontre dans ses locaux et assure un soutien logistique indispensable à son bon déroulement.
-

INDEX

Mots-clés : THATCamp, humanités numériques

Les *digital humanities* sont-elles solubles dans le Steampunk ?

Conférence inaugurale

Paul Bertrand

NOTE DE L'ÉDITEUR

Transcription de la conférence inaugurale. Celle-ci peut être écoutée via un lecteur audio : <http://archive.org/embed/LesDigitalHumanitiesSont-ellesSolublesDansLeSteampunk>

- 1 Le plus important pour moi d'abord ici, c'est de vous remercier et de remercier les organisateurs. Je remercie Pierre en particulier pour son invitation qui m'honore beaucoup et qui m'a fait me poser beaucoup de questions. Dans mon expérience des tous premiers THATCamp, je n'avais pas grand souvenir de conférences inaugurales. Je n'ai pas participé à celle de 2010 ici, mais pour ce que je connaissais, cela ne me disait rien. Je me suis donc dit deux choses : la première, soit cela ne m'avait fait absolument aucun effet, et donc je sais bien ce que je ne dois pas faire, ou plutôt je vais essayer de faire autre chose. Et la seconde, c'est que cela soit original ou pas, c'est un défi. Faisons donc une non-conférence inaugurale ! Ce qui somme toute, est, me connaissant, relativement faisable.

Paul Bertrand



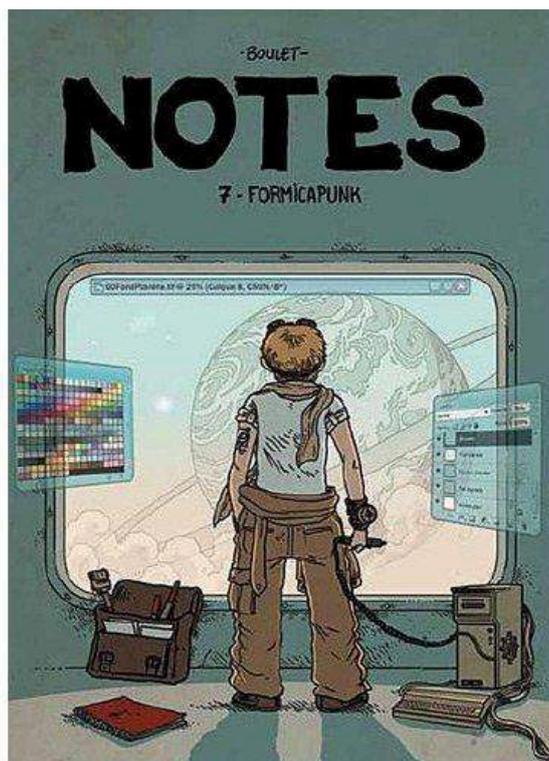
- 2 Dans la mesure où il faut partir d'un concept, le concept que j'ai choisi est celui du « *steampunk* ». Cela, pour deux raisons : la première pour rendre hommage à l'un de mes collègues de la blogosphère avec lequel nous avons commencé, il y a bien longtemps, et qui est le dessinateur Boulet dont le septième tome des *Notes* vient de sortir. Boulet est un grand adepte de cette notion. Pour ceux qui connaissent un peu le monde du blog, c'est un personnage ; et ici je lui rends hommage, vous pouvez le twitter, désormais je suis prêt à devenir son ami et à aller boire des verres avec lui ! La seconde chose, la plus importante, est que le *steampunk* est un mode de réflexion, plutôt une esthétique, très intéressante qui est en train de prendre de l'importance ces dernières années – depuis déjà au moins une petite vingtaine d'années, mais plus fortement ces dernières années. Au départ, c'est un genre littéraire, un sous-genre de la science-fiction dite « uchronique » – c'est une définition que j'ai récupérée à partir de Wikipédia, c'est pour vous dire si c'est sérieux ! – dont l'intitulé a été forgé par allusion au « cyberpunk ». Pour cette raison, il est plus approprié de parler de « rétro-futurisme ». L'expression *steampunk*, je continue la définition, est un terme qui signifie « punk à vapeur », au cas où vous auriez un souci de traduction, et peut être traduite par « futur à vapeur », terme inventé pour qualifier un genre de la littérature de science-fiction, né à la fin du *xx*^e siècle, dont l'action se déroule dans l'atmosphère de la société industrielle du *xix*^e siècle. Le terme fait référence à l'utilisation massive, je cite toujours, des machines à vapeur au début de la révolution industrielle et à l'époque victorienne.

aurelberra

RT @jpmasse : @openeditionsays en latin ?

c'est totalement *steampunk* non ? ;-) #tcp2012

Notes 7 : Formicapunk, de Boulet



- 3 À l'origine, c'est un genre littéraire. Mais le plus intéressant, c'est de voir que très rapidement c'est devenu un genre en soi qui s'est répandu dans la société, et est devenu un mode de représentation esthétique. Vous savez que de nos jours, l'uchronie a beaucoup de succès, et parmi les formes d'uchronies, le *steampunk* en est une très importante. Une des différences principales entre le *steampunk* et la science-fiction c'est la présence d'éléments anachroniques plus tardifs, comme les ordinateurs, les manipulations génétiques, ce qui permet d'y développer tout un humour spécifique... L'idée est donc de créer un décalage ; un décalage entre, d'une part un contexte dix-neuvième siècle, et de l'autre toute une technologie, ou toute une aspiration à un progrès technologique qui l'est plutôt avant-gardiste. Pour ceux qui ont quelques connaissances de la série *Star Trek* à la télévision, il y a des épisodes qui sont déjà clairement *steampunk*, où l'on voit des gens vivant façon dix-neuvième siècle (avec l'œil américain évidemment), mais qui à côté de cela arrivent à développer des armes de destruction massive avec pratiquement deux bouts de caoutchouc ou des bouts de bois. Il est donc très intéressant de voir que ce système esthétique est en train de se développer.
- 4 Ce qui est encore plus intéressant, c'est que ce style se rapproche très fortement, par certains côtés esthétiques, de la vision actuelle que l'on peut avoir du numérique et de son évolution. Je vais vous citer une partie de la définition que j'avais trouvée dans Wikipédia : « tandis que l'ère de l'électronique tend vers la miniaturisation – grand fantasme – l'élément esthétique fondamental du *steampunk* est son gigantisme » : demandez à l'IN2P3 et sa salle des serveurs ! « Cette technologie, délibérément bloquée au stade de la vapeur, produit des constructions démesurées, complexes, extravagantes de tuyauteries actionnées par des leviers et des claviers aussi nombreux que compliqués. Les véhicules à vapeur semblent aussi lourds et difficilement maniables, ils

nécessitent une très forte dépense d'énergie pour un rendement faible. Et surtout, enfin, le héros classique dans ce genre d'univers est le mécanicien de génie », on aurait pu dire le *geek*, cela aurait été aussi bien.

- 5 À côté de cela, continuons dans les définitions, les *digital humanities* désignent « une transdiscipline porteuse de méthodes, de dispositifs et de perspectives heuristiques liées au numérique dans le domaine des sciences humaines et sociales. [...] Nous appelons à l'intégration de la culture numérique dans la définition de la culture générale du XIX^e siècle. » Vous avez reconnu le Manifeste des *digital humanities* !
- 6 Je voulais poser ces deux éléments de définition, parce que maintenant je vais les faire se télescoper. Ce dont je vais parler ici, c'est de ma vision, totalement subjective qui est le résultat de mon expérience et de mes connaissances (quelles que soient leurs qualités) des *digital humanities* depuis maintenant presque une dizaine d'années, et même davantage si remonte aux débuts quand je commençais à bricoler avec des bases Access. Je vais donc mobiliser cette expérience, en me référant évidemment au *steampunk*, mais pas seulement, pour évoquer ma vision de l'évolution actuelle des *digital humanities*. C'est une vision qui paraîtra par certains aspects dure voire négative, mais je vais essayer de présenter les choses telles que je les ressens. L'objectif c'est de vous faire bouger, de vous faire réagir.

Première couche : Une insoutenable tension dialectique : autour d'une esthétique du *steampunk*

- 7 La première couche d'interprétation du *steampunk*, forcément, c'est celle qu'avait donné en son temps dans un congrès sur les Linked Open Data for Ancient Studies¹, Andrew Reinhard. Je ne suis donc pas le premier à avoir parlé de *steampunk* pour les *digital humanities*. Andrew Reinhard avait lancé un pavé dans la mare en avançant l'idée suivante : « [...] the current scholarly publishing enterprise is essentially *steampunk*, 21st century work with 19th century models [...] » Il y a une certaine violence dans cette déclaration ; et en même temps un certain réalisme.
- 8 Si je me réfère à mon expérience, les modèles éditoriaux que je retrouve en règle générale dans les équipes avec lesquelles j'ai travaillé jusqu'ici correspondent à l'état d'esprit décrit par Andrew Reinhard - avant de me retrouver à l'université de Louvain, j'ai été comme certains d'entre vous le savent peut-être, et je le suis toujours d'ailleurs en grande partie, rattaché à l'Institut de recherche et d'histoire des textes (IRHT) ; j'ai dirigé, et je dirige toujours, la partie scientifique du centre de ressources numériques Telma. On arrive avec une édition, souvent d'une grande qualité scientifique, que l'on veut à tout prix publier sur le web, mais en PDF. Et on la met en ligne en l'état parce que l'on n'a pas le choix, sous l'effet d'une pression qui vient de tous les niveaux de la recherche, et même si on peut avoir d'autres perspectives et d'autres points de vue sur cette publication.
- 9 Je ne donnerai aucun exemple particulier dans les quelques mots qui vont suivre parce que, vous me connaissez, j'ai horreur de dire du mal des gens ! Je me contenterai d'un commentaire long sur quelques éléments que j'ai pu retirer ici ou là, autour de points de changement nécessaires dans cette vision des *digital humanities* qui se développe actuellement.

- 10 Le premier point, est de savoir comment régler cette question de l'urgence de la publication, et ce divorce constant entre chercheur et ingénieur.

NuevoMundoMN

El divorcio entre el investigador y el ingeniero :

¿un punto clave de las humanidades digitales ? #tcp2012

- 11 Il y a pour moi une tension dialectique qui est perpétuée, que l'on ne retrouve pas seulement dans la publication d'un corpus quel qu'il soit. Je vous ai parlé des publications en PDF, mais je pourrais évoquer aussi des corpus publiés dans des formats encore plus frustrés ou encore plus problématiques, soit parce qu'ils sont trop anciens, soit parce qu'ils sont structurés selon des modèles qui nous font hérissier le poil... jusqu'à voir des éditions publiées sur le web en .doc au prétexte que le fonction CTRL-F est largement suffisante ! Et cette tension se reproduit également dans d'autres domaines, comme par exemple les bibliothèques.
- 12 Il est vrai que les bibliothèques ont dix longueurs d'avance par rapport au monde de la recherche, du point de vue de l'appropriation des nouvelles techniques du web sémantique, de l'appropriation des concepts de bonnes pratiques ou d'archivage pérenne, ou encore de l'utilisation des moteurs ou des ontologies. N'empêche que beaucoup de bibliothèques, surtout à des niveaux locaux, car je ne parle pas de machines de guerre comme la BnF, subissent une pression qui vient de leur tutelle, par les collectivités elles-mêmes, pour que les collections soient publiées de manière très rapide, très frustrée, éventuellement dans des formats qui sont comparables à ceux dont je parlais tout à l'heure, avec des métadonnées frustrées également : le Dublin Core du bas du tableau. Quand certaines bibliothèques ou dépôts d'archives ne disent pas tout simplement : « une partie seulement est visible sur le web, et pour le reste venez chez nous pour avoir accès à tout ». Donc, une sorte de libre accès à deux vitesses qui se trouve malheureusement être une constante, et qui est non pas commandé par les bibliothèques elles-mêmes ou les archives mais par les collectivités. Il y a donc un problème entre une demande d'un côté et, d'autre part, une capacité de production, des règles, un règlement, une organisation, qui eux se trouvent en porte-à-faux.

2. De grands espoirs : la dissémination et l'appropriation

- 13 Nous avons de grands espoirs pour régler le problème de l'utilisation de formats nouveaux, pour mettre en œuvre ces fameuses bonnes pratiques qu'on connaît bien maintenant, pour mettre en œuvre tout une série de procédures d'archivage pérenne, pour participer aux grandes procédures d'interopérabilité : on avait beaucoup d'espoirs que cela puisse décoller assez vite.



- 14 Voilà maintenant à peu près cinq années que l'on peut dire que les *digital humanities* ont pignon sur rue en France. Plus, si on compte les pratiques sur le terrain. Là encore, je prends l'exemple de ce que je connais bien : je me souviendrai je crois toute ma vie du moment où nous avons lancé la plate-forme Telma. Pendant des années cette plate-forme a peiné à démarrer parce que l'on n'avait pas de reconnaissance institutionnelle. C'était un partenariat École des Chartes / IRHT qui fonctionnait très bien du côté de l'École des Chartes tandis que de celui de l'IRHT, on ne comprenait pas très bien à quoi cela pouvait servir. Et c'est lorsque nous avons réussi à faire une démonstration, deux ans après la première publication, que l'on a réussi à convaincre *in fine* la direction (à partir du moment où elle partait, c'est cela qui est le plus intéressant !) de l'importance de l'objet. Parce qu'en parallèle se mettait en place toute une série d'autres projets similaires un peu partout en France, la direction se rendait compte de l'importance de l'initiative. Et tout à coup, toutes les portes nous ont été ouvertes. À tel point que la grosse verrue sur le dos de l'IRHT, si je puis dire, est devenue, tout d'un coup, un fleuron ; une arme importante pour la grandeur de l'institution.

Les *digital natives*

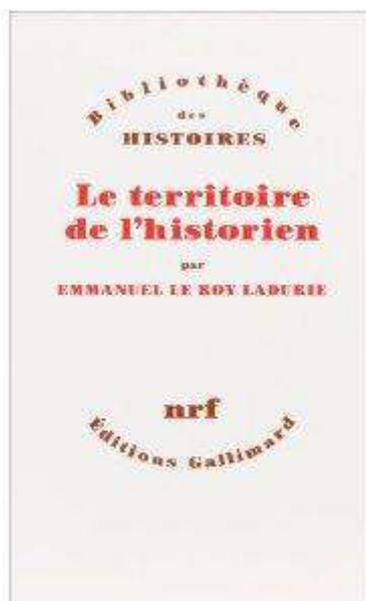
- 15 Il y a quatre ou cinq ans, on misait beaucoup sur les *digital natives*, ces jeunes qui sont nés dans le monde numérique et que l'on retrouve maintenant à l'université, que ce soit à Louvain ou ailleurs. Pour ma part, je suis revenu des *digital natives* qui sont plutôt des *Facebook natives*. De fait, on est très loin de l'appropriation du numérique en tant que tel. Lorsque l'on propose à des chercheurs, très jeunes, qui viennent d'arriver, qu'ils soient en thèse ou juste après, de se lancer, ils peuvent ouvrir un carnet de recherche – ce qui correspond bien en quelque sorte aux pratiques de réseau social qui sont en train de se développer et c'est une excellente chose. Mais lorsqu'il s'agit d'aller

plus loin dans l'appropriation des *digital humanities*, même de manière assez distante, il y a comme un blocage et une interrogation. Et l'appropriation est lente et difficile.

Les historiens programmeurs de Le Roy Ladurie ?

- 16 Deuxième espoir que nous avons, les historiens programmeurs, et je sais que cela a fait l'objet de beaucoup de discussions ces derniers temps. Sur ce concept aussi, je suis très perplexe. Je jette ma pierre dans le marigot de la discussion dans laquelle tout le monde se dépêtrera demain. Je pense que ce concept d'historien programmeur est dangereux et n'a pour l'instant pas donné de résultats. En tous les cas, comme le concevait Le Roy Ladurie, c'est-à-dire comme un historien tout à coup capable de jongler, qui avec du Python, qui avec du Java, voire qui avec du Perl. Car le problème est que beaucoup de ces historiens programmeurs, - on en connaît actuellement qui ont en effet appris la programmation - programment fort bien, mais qu'un historien a une vie d'historien déjà bien remplie en soi, une vie de recherche. Lorsqu'à un moment ou un autre il se lance en profondeur dans une démarche d'apprentissage technologique (un langage, une méthode de programmation), il s'y lance à fond, et applique cette technique à son objet de recherche. Mais, vient alors le reste de la vie. Et le reste de la vie, ce sont les charges d'enseignements, la recherche qui se développe, les projets d'ANR. Et il n'a plus la possibilité matérielle de développer ce savoir, de le faire évoluer, de mettre en place pour lui-même une veille et continuer à développer ce savoir faire technologique. Ce savoir se retrouve donc bloqué à un instant *t*.

Le Territoire de l'historien



- 17 Cela ne serait pas grave si l'historien lui-même acceptait l'idée que ses connaissances technologiques peuvent être dépassées. Or le problème est que beaucoup de ces historiens, pour l'instant en tous les cas, sont persuadés que le langage qu'ils connaissent et qu'ils maîtrisent parfaitement est le « couteau suisse » parfait. Il y a ainsi aujourd'hui des chercheurs qui bloquent des processus d'avancement dans une discipline parce que l'on ne peut pas arriver à les faire passer de cet outil à autre chose. J'ai un exemple récent concernant le passage d'une base de données codicologique et paléographique

sur un autre support PHP : « Oui mais tout de même, j'aimais bien mon ancien système. Ne peut-on pas le laisser à côté, pour tous les autres chercheurs qui y sont habitués ? » Il y a une difficulté terrible à évoluer et à passer à autre chose. Le risque est donc un blocage dans cette évolution, ce processus de construction des *digital humanities*, avec des chercheurs qui ne voudront pas tout simplement évoluer vers de nouvelles étapes de développement.

- 18 C'est la raison pour laquelle pour moi la notion d'historien programmeur reste dangereuse. Où alors il faut envisager un historien avec une teinture de *digital humanities* assez forte pour en comprendre les enjeux, et qui suive cette teinture en étant en contact direct avec des gens qui puissent lui instiller régulièrement, lui faire prendre conscience de l'évolution des choses et lui donner les moyens d'une veille rapide et efficace. En fait, ce dont les chercheurs ont besoin, c'est de se dépêtrer dans la masse de d'informations produites autour des *digital humanities*.

piotrr70

Besoin énorme pour les chercheurs de se repérer dans le foisonnement des DH #tcp2012

spouyllau

@piotrr70 il faut des boussoles en fait #tcp2012

- 19 Je suis ravi du nombre de colloques, du nombre de conférences, du nombre de journées d'étude, du nombre de symposium qui s'organisent. C'est parfait, c'est la preuve que quelque chose se passe. Mais un chercheur lambda, quel qu'il soit, ne peut pas se débrouiller dans tout cela. Comment voulez-vous qu'il arrive à trouver les quelques éléments qui puissent lui montrer vers quoi on va et ce qu'il faudrait faire ? Il est donc nécessaire de rationaliser l'information.

3. Deuxième couche : qu'est-ce qui coince ? L'idéologie du Steampunk

Un risque d'appauvrissement des données

- 20 Continuons avec autre chose. J'en mets une couche. Jusqu'à présent, on en est resté à un niveau esthétique. Il n'y a pas de problèmes techniques, on n'est pas confrontés réellement à un danger pour la recherche elle-même, mais plutôt à un ralentissement dû à une absence de dialogue. Le problème est que le *steampunk* est aussi une idéologie qui implique une façon de voir, ou de conceptualiser les données historiques, à les figer à un certain niveau. Je pense, vous allez comprendre où je veux en venir, qu'on est arrivé à un certain palier dans le développement des *digital humanities* et que grandit le risque d'un appauvrissement des données.
- 21 On court le risque d'un appauvrissement des données plutôt qu'à un enrichissement de ces données, au moment du passage au numérique. L'appauvrissement, c'est un mot très dur que je vais tenter d'expliquer.
- 22 En fait, je vais tout vous dire : pour préparer cette conférence, je me suis demandé quelle démarche heuristique mettre en oeuvre. Cela serait intéressant aussi pour le THATCamp de savoir comment mettre en place une heuristique *digital humanities* à

partir des données. Comment ai-je préparé ma conférence ? J'ai décidé de prendre trois semaines de tweets. J'ai pris tous les tweets que j'ai reçus depuis trois semaines et j'ai été voir chacun des liens vers lesquels on me renvoyait en me disant que cela me faisait un état, qui vaut ce qu'il vaut, mais un état de la situation de la réflexion de tous ceux qui travaillent dans les *digital humanities* que je connais et avec lesquels je travaille. Un certain nombre d'éléments m'ont sauté aux yeux et je vous les livre ici.

Vive le texte !

- 23 Le premier élément concerne la notion de « texte ». Je pense en particulier à une façon de voir les choses qui a été développée par un chercheur anglo-saxon du nom de Tim Hitchcock sur son blog (c'est un lien qui je crois m'avait été renvoyé par Frédéric Clavert). Cette note de blog disait clairement que l'on n'avait plus besoin du livre. En effet, lorsque le texte passe dans le monde des *digital humanities*, le concept de livre devient secondaire et on entre dans une autre perspective fondée sur les données. Cela signifie que le métier d'historien lui-même qui était basé sur le livre et son appropriation par le chercheur – je parle de l'historien mais on pourrait évoquer n'importe quel chercheur en sciences humaines –, le chercheur qui s'appuie sur concept fondamental du livre comme élément de référence, voit son métier déconstruit en quelque sorte par la disparition d'un objet emblématique pour lui, progressivement remplacé par le texte.
- 24 Le problème est que le texte n'existe pas. Il n'a jamais existé par lui-même. Le texte n'existe que parce qu'il a été conçu dans un cadre précis ; un livre, un article, une note de blog éventuellement. Mais le texte est conçu de manière complètement liée intellectuellement à l'objet matériel de départ, physique ou numérique. Donc, le livre n'est en quelque sorte pas mort, si l'on prend en considération l'objet de départ. Un texte est lié à son contexte culturel de lecture, de production, et on ne peut pas le dégager de ce contexte. Le risque que nous prenons tous avec les *digital humanities* telles qu'on les voit se développer aujourd'hui est que le texte soit détaché de son contexte de production.

Neuromancien

paul bertrand considère que le texte n'existe pas
indépendamment et qu'il est lié à l'objet matériel
de départ #tcp2012

Cclivaz

#tcp2012 Fahrenheit 451 est un roman paradigmatique
de la culture digitale : c'est l'humain qui porte la culture,
quels que soient les médias

Vive la donnée !

- 25 L'autre élément sur lequel je voudrais revenir est la donnée. Là encore, on constate des fantasmes. Tout le monde nous parle des *data*. La donnée est devenue un objet absolument essentiel, on est certain qu'elle est devenue l'élément premier de notre recherche, et qu'avec cette donnée nous allons pouvoir, en l'organisant, en

l'interrogeant, en l'enrichissant par des procédures du web sémantique, produire des résultats, des idées, de la nouveauté intellectuelle.

- 26 Or le problème est que la donnée n'existe pas. La donnée cela n'existe pas ! De quoi s'agit-il ? Est-ce un texte, une ligne de texte, un mot, un concept ? Est-ce que c'est un concept derrière un concept, est-ce que c'est une idée derrière un concept ? Je sais bien, je ratiocine, mais tout de même, je demande qu'on me donne une explication claire de ce qu'est une donnée.

Merryarchive

#tcp2012 P. Bertrand. Oh là là, la donnée n'existe pas non plus !

Inactinique

Lisez, sur les données, l'article de danah boyd
dans le ReadWriteBook 2 #tcp2012

- 27 Une base de données est un outil extrêmement utile, mais elle crée une information qui elle-même n'existait pas jusque-là. Prenons l'exemple d'une base de données autour de la codicologie. On part de deux manuscrits en tant que tels et on commence à prendre des informations codicologiques autour de chaque manuscrit ; on en fait une très belle base codicologique sur les reliures, ou bien sur les réglures ; pourquoi pas, chacun ses fantasmes ! Le problème est que la donnée codicologique est une construction pure, elle est récupérée de l'objet livre dans des conditions qu'il faudrait encore pouvoir définir, et l'objet livre n'est plus lui-même dans la donnée. Cela signifie que l'on va interpréter les résultats de recherches statistiques en les détachant de la réalité du livre lui-même, de la réalité du manuscrit. Et forcément, on ne pourra pas parler d'une donnée complète qui pointerait complètement vers toutes les réalités du manuscrit.
- 28 Or, la nouveauté en recherche, on le sait bien, ne vient pas de la mise en relation de données dont on se doute bien qu'en les mettant ensemble on arrive à des résultats, mais de la mise en relation de données à propos desquelles on n'avait aucune idée de leur possible mise en relation. Le risque, avec la création de cette donnée, c'est de n'avoir qu'une donnée partielle et souvent non définie. Qu'est-ce qu'elle représente ? Est-ce que c'est une source, une partie de la source, est-ce que c'est une métadonnée au-dessus de la source ? Lorsque l'on étudie un objet documentaire, on peut le découper en autant de données que notre imagination le permet et ce concept de données n'est pas posé de manière complète.

Vive les *big data* !

- 29 L'autre grand fantasme, ce sont les *big data*, les grandes masses de données qui font évidemment beaucoup rêver de l'autre côté de l'Atlantique mais qui ici aussi commencent. C'est tout le problème de l'histoire du chiffre. En sciences humaines, il y a 30 ans, on ne jurait que par la statistique et par le quantitatif. On a depuis jeté aux orties le quantitatif de manière très virulente, et via les *big data*, il semble revenir par la fenêtre ; Pourquoi pas après tout ? Le problème est que ces énormes masses documentaires aplatissent la qualité des données, ou en tous cas l'information de base. La critique a d'ailleurs été bien posée dans l'article publiée dans le Read/Write Book 2 :

les big data sont à la fois incontournable et en même temps très dangereuses². Il n'y a rien de plus difficile à analyser, pour moi, qu'une série de chiffres.

EmmaCorne
il y a 20 ans en shs on ne jurait que
par le quantitatif, on l'a viré et là il revient
par la fenêtre @PaulBertrand #tcp2012

Vive les plate-formes et les méta-moteurs !

- 30 Evoquons les métadonnées, très rapidement. Parce que là aussi c'est encore un fantasme. Les métadonnées sont souvent la seule clef d'accès, à l'heure actuelle, pour bien des documents et bien des textes. Le problème est que les métadonnées que l'on utilise actuellement sont souvent très faibles. La plupart du temps, et je parle d'expérience, j'ai vu les métadonnées qui sont utilisées dans les disciplines où je travaille, et il faut bien avouer que c'est un peu triste. Il est vrai qu'heureusement les plate-formes et les méta-moteurs existent, et on ne peut qu'en être satisfait, tout en même temps en restant extrêmement prudent car ils dépendent en grande partie de leur qualité de contenu, de la qualité du marquage. C'est vrai – et là je m'incline devant Stéphane Pouyllau qui est là – qu'Isidore³ résout déjà une partie du problème grâce à son enrichissement sémantique. Il n'empêche que – et là, je crois qu'il le reconnaîtra avec moi – la qualité d'Isidore reste encore très variable. Oui évidemment, je suis persuadé qu'on peut l'utiliser pour la recherche. Mais comment l'utiliser ? Et comment améliorer les métadonnées que ces plate-formes et méta-moteurs utilisent ?

Vive les ontologies !

- 31 On a enfin les ontologies, un vieux problème en discussion depuis de nombreuses années. Parce que hélas, elles sont extrêmement complexes et extrêmement lourdes à mettre en place, souvent au cas par cas. Moi je n'ai pour l'instant pas encore vu beaucoup d'ontologies fonctionner de manière ample dans nos disciplines. Je ne dis pas qu'ailleurs il n'y en a pas. Mais les quelques rares ingénieurs de *digital humanities* qui s'y collent me regardent souvent avec des yeux effarés quand je leur demande : « Alors, quand est-ce que l'on va pouvoir travailler avec vos ontologies, quand est-ce que vous nous aidez à les mettre en place ? »
- 32 Lorsqu'on parle d'interopérabilité par les ontologies, je me demande si on ne se paie pas de mots. Le risque que j'ai vu apparaître dans des laboratoires très proches, où l'on me dit qu'il faut développer l'interopérabilité – et je suis le premier à dire, bien sûr c'est une excellente idée – est qu'elle se résume à permettre un accès autour des noms d'auteurs, des cotes des manuscrits ou des livres. C'est-à-dire quelque chose que l'on faisait très bien avec des index traditionnels en papier. . On pourrait aller un peu plus loin, mais comment procéder ? La pression institutionnelle fait que le retour aux index traditionnels pointe à l'horizon.

piotr70
P. Bertrand ne se paie pas de mots :

il rend compte sincèrement d'interrogations
réelles que se pose un chercheur bien au courant #tcp2012

Vive les bonnes pratiques !

- 33 Enfin, et là je vais jeter un gros pavé dans la mare, il faut parler des bonnes pratiques. Parce que les bonnes pratiques sont peut-être une excellente chose, mais elles sont aussi un frein. Tout d'abord il faudrait voir exactement ce qu'il y a derrière, car c'est aussi un terme dont on se gargarise, il faut bien l'avouer. Par exemple, on se focalise sur l'archivage – je ne dis pas que ce n'est pas important –, on se focalise sur la structuration et la possibilité de mise en opérabilité, et on oublie (ou en tous cas on ne se positionne pas) sur le reste. On ne se positionne pas sur l'innovation conceptuelle par exemple, sur l'enrichissement sémantique, sur la granularité des outils d'interrogation, ou simplement sur la décomposition de ces fameuses données. J'ai peur qu'on se contente de bonnes pratiques aujourd'hui dans les *digital humanities* et que l'on renonce à un certain devoir d'audace.

NuevoMundoMN
Humanidades digitales : imposibles
sin »deber de audacia » #tcp2012

3. Deuxième couche : pourquoi ça coince ?

- 34 Je ne voudrais pas être trop pessimiste et chercher des raisons à ces blocages. J'en vois deux.

Le poids de l'institutionnalisation

- 35 La première c'est le poids de l'institution. Nous sommes dans une société où on se satisfait des bonnes pratiques parce que l'on n'a pas le temps de l'innovation. L'institution, les institutions au sens large du terme, pèsent de tout leur poids sur nos disciplines, nos disciplines de recherche et nos disciplines *digital humanities* avec une exigence de résultats rapides et efficaces. Je le vois, lorsque je travaille avec des développeurs pour les différents projets que l'on a mis en place, que cela soit à l'IRHT, à Telma ou ailleurs, à travers les demandes de chercheurs qui sont soumis aux pressions des financeurs et ce poids fini par retomber sur les épaules des développeurs et des spécialistes des *digital humanities*. Il faut aller vite pour publier quelque chose dans les temps. Alors, on fait du PDF et on réfléchit après, ou alors on fait un peu de TEI-Lite, et puis on publie. Mais c'est de la TEI *light*, et on ne pourra pas aller plus loin car on n'aura pas le temps d'approfondir. Le poids de l'institution est terrible ; il faudrait pouvoir y réfléchir et s'en dégager.

L'accélération du temps

- 36 Deuxième élément, l'accélération du temps. C'est un concept que je prends au sociologue Harmut Rosa qui a publié un superbe ouvrage il y a un an ou deux sur cette question⁴.
- 37 De la même façon que l'accélération du temps a pesé sur notre manière de voir les choses et sur notre inconscient, l'accélération du temps pèse sur nos méthodes de construction de projets, sur les projets eux-mêmes et la nécessité de résultats dont je parlais tout à l'heure. On peut parfaitement mettre en place un projet de recherche ANR sans jamais avoir jamais rencontré les personnes avec qui on travaille ; je le sais bien, je l'ai fait.
- 38 Donc, comme disait Frédéric Clavert sur Tweeter, on est dans du « post-post-modernisme », on ne reconceptualise plus la donnée parce que l'on n'en a plus le temps et plus les moyens. On revient à une situation critique positiviste, similaire de celle d'avant l'Ecole des Annales, utilisant les documents comme autant de carrières de faits, et figeant ces faits en les atomisant tout simplement. On n'a pas le temps d'aller faire autre chose. On est dans un positivisme si fort, que le chercheur parle encore moins qu'avant avec le développeur ; il n'en a plus le temps.
- 39 En ce sens, les *digital humanities* rejoignent donc non-seulement l'esthétique, mais aussi une idéologie *steampunk* : coincées dans un fatras technologique institutionnel, engoncées dans ces cols amidonnés victoriens où le mécanicien est roi et l'usine à vapeur, pour ne pas dire l'usine à gaz, refont surface. Le risque n'est pas tant de se fondre dans l'idéologie ou l'esthétique *steampunk* – qui est clairement une esthétique de crise, c'est clair – mais c'est surtout de fonder cette idéologie qui prend le pas et qui s'impose, comme naturellement, à beaucoup d'entre nous.

4. Sortir de l'idéologie du *steampunk*

- 40 Alors que faire, car je ne veux pas vous laisser là, tout triste (et surtout je ne veux pas que vous me jetiez trop de tomates quand j'aurai fini de parler). J'ai deux propositions.

jpmasse
#tcp2012 P Bertrand ne veut pas nous laisser tous tristes ;
non à la #tristitude <http://t.co/7ilphrsH>

Reconceptualiser le « donné »

- 41 La première proposition consiste à reconceptualiser le « donné » et pas seulement avec des balises, mais d'un point de vue intellectuel, sur pièces. Il faut une nouvelle critique car la critique sur laquelle nous nous basons date du XIX^e siècle, un moment tout à fait important de l'histoire mais qui n'est plus adaptée à nos besoins. J'ai la chance de donner un cours de critique historique à des étudiants de l'université en Belgique qui est peut-être le seul pays d'Europe où l'on enseigne la critique, en tant que telle, comme discipline. Et je me rends compte que celle que l'on enseigne à l'heure actuelle, suivant les vieux systèmes, Langlois et Seignobos⁵ – enfin on a un peu dépassé, mais pas beaucoup – est complètement à revoir.



- 42 Il faut la revoir de manière ouverte, et en mouvement. Car ce qui est intéressant, c'est de repenser le donné (je dirais, plutôt que les données), à partir de sa situation. Replacer la situation de chaque donnée dans son contexte, que ce soit du mot jusqu'au corpus, puis après cela à toutes les déclinaisons du corpus puisqu'une déclinaison du corpus est un nouveau document en soi. Une base de données crée un nouveau donné. Il y a création de quelque chose de neuf.
- 43 Je propose de développer une réflexion un peu similaire à celle que j'avais essayée d'avoir avec Christophe Jacobs, il y a de cela deux ans, pour l'ANR mis en place par Jean-Philippe Jeunet autour du numérique. Une tentative de mouvement critique qui est encore à continuer et à approfondir franchement, notamment parce que le donné ne peut plus seulement être considéré comme un élément que l'on replace dans un contexte, avec un avant, un après, un contenu, un contenant. Mais aussi parce que le donné est en mouvement, en relation. C'est un donné que l'on doit mettre en relation avec d'autres, et soumettre à la critique le type de relation ainsi. Il y a une nouvelle critique documentaire à avec les spécialistes des *digital humanities* qui peut passer par une sémantisation des données, mais qui ne pourra surtout pas faire l'économie d'une reconceptualisation critique.
- 44 Je propose par ailleurs de créer un espace hors de la pression du temps et des institutions, dédié à l'innovation, presque pour le plaisir. Il s'agit de réfléchir avec ceux qui le veulent autour de ce nouveau guide critique du donné tel quel le chercheur en *digital humanities* le voit apparaître et tel que le chercheur lui, de son côté, tente de le mettre en place avec ses propres moyens. C'est un espace de dialogue qui s'établit au niveau de la critique qui peut nous permettre d'échapper à cette idéologie potentielle, et de toute façon certainement de cette esthétique de crise qu'est le *steampunk*.
- 45 Voilà, j'ai beaucoup parlé, maintenant à vous de me jeter les tomates ! Merci.

MajBen
comment faire parler une salle
qui ne f/sait que tweeter #tcp2012

cbarthonnat
 Et maintenant, Paul Bertrand (@medieviz)
 va analyser tous les tweets relatifs
 à sa non-conférence :-D #steampunk #tcp2012

Discussion

- 46 Lou Burnard : Merci beaucoup Paul pour cette intervention très provocatrice. Il me semble que (je ne suis pas historien, donc je me trompe peut-être) mais je crois que la

plupart des institutions actuelles, quelques-unes c'est sûr, ont été fondées par le passé en réaction à ce que tu viens juste d'évoquer : un mécontentement avec les institutions du moment. N'échappera-t-on donc jamais à ces cycles de création d'institutions ?

- 47 Paul Bertrand : Oui c'est certain. Jusqu'à un certain point, le THATCamp commence à être une institution.
- 48 Pierre Mounier : Pas pour le moment car les THATCamps sont justement porteurs d'une rhétorique anti-institutionnelle et se définissent un peu comme des zones d'autonomie temporaire.
- 49 Paul Bertrand : Certes, mais c'est une initiative qui a réussi à se positionner dans le temps avec des structures, avec un cadre, avec un logo. Il y a toujours de l'institutionnalisation. C'est pour cette raison qu'il faut créer un espace en dehors du temps qui pourrait même être virtuel, et même s'il y aura une institutionnalisation inévitable avec la production d'un livre ou le partage d'un bon moment. Le plus important pour moi, et c'est seulement un avis, c'est de pouvoir se dégager du poids des exigences d'efficacité et de rapidité de ces institutions qui nous gouvernent, si je puis dire, dans lesquelles nous tentons de travailler et nous tentons d'avancer. Si nous nous débarrassons de cet idée que l'on veut à tout prix avoir vite des résultats, et être très rapidement efficace, alors, on risque de réussir.
- 50 C'est très idéaliste ce que je propose, mais si on n'est pas idéaliste, alors on ne sert plus à rien.
- 51 Jean-Luc Pinol : Je poserai quelques questions pas du tout comme directeur du TGE Adonis, mais comme historien.
- 52 Globalement, je suis historien des villes. La ville est une institution qui se recompose en permanence. Si elle ne se recompose pas elle meurt. C'est un peu la même chose pour les institutions ; il faut qu'elles se reconstituent, se recomposent en permanence. Je ne ferai qu'une petite parenthèse, si l'on regarde le devenir des grandes infrastructures de recherche et la manière dont on repense, au-dessus d'elles, leurs missions, on est en plein dans ces questions.
- 53 La deuxième question est plus étroite et, d'une certaine manière plus corporatiste. Tu as évoqué les chiffres, et tu as dit qu'avec les chiffres on n'a jamais rien démontré. Je suis d'un avis exactement opposé. Précisément, le numérique nous permet avec des méthodes de visualisation de faire des choses totalement différentes et nous oblige à repenser nos données. J'ai vu, comme directeur de laboratoire, des gens qui n'avaient pas quitté le XIX^e siècle. Je suis tout à fait d'accord avec le fait que les étudiants sont bien souvent plus des *facebook natives* que des *digital natives*. La plupart, même s'ils ont 20 ou 25 ans de moins que moi, pensent qu'avec Word on fait tout. Mais c'est vrai qu'il faut réfléchir aux nouvelles manières de penser, en l'occurrence ici, la recherche historique, en utilisant ces nouveaux outils.
- 54 Paul Bertrand : Pour répondre rapidement, sur l'institutionnalisation et sur le fait que l'institution se refait, c'est très bien, et même au contraire c'est une excellente chose. C'est d'ailleurs la raison pour laquelle, la refonte de l'institution vient souvent d'une critique externe à l'institution, voire au sein de l'institution même.
- 55 Sur la question du chiffre je suis d'accord. On peut faire des choses avec les données numériques. Je sais qu'il y a beaucoup d'outils extrêmement efficaces mais le risque c'est de s'en contenter. Ce que je mets en avant c'est le fantasme des big data.

- 56 Jean-Luc Pinol : Là on est d'accord. Mais simplement, je pensais au dernier ouvrage de Franco Moretti sur l'histoire littéraire où il dit, de façon très provocatrice, arrêtez de lire, comptez ! Son livre, c'est en français *Graphes, cartes et arbres*⁶. Il montre bien, d'ailleurs avec des méthodes que pour autant je peux considérer, en termes de cartographie, assez frustrées, comment on est obligés de repenser la manière de faire de la recherche.
- 57 Marjorie Burghart : Je voulais revenir sur le côté *steampunk*, qui m'a beaucoup plu. Sur une de tes premières diapositives, tu citais ce passage sur les *Current scholarly publishing*, et je me demandais où était le « steam » et où était le « punk » dans tout cela. Quand on entend cela, on imagine des gens qui amènent leurs trucs tout poussiéreux et qui les donnent à des gens très pointus que l'on voit presque avec des costumes de la NASA en train de se demander ce qu'ils vont en faire Est-ce que l'on n'est pas en fait dans la situation tout à fait inverse, une situation où des personnes sont en train de faire leur révolution industrielle, des gens tout à fait à la pointe au niveau intellectuel, et se confrontent à quelque chose de très vapeur, à quelque chose de très victorien qui rend les projets très lourds à mener. Voilà, il y a un anachronisme c'est sûr. Mais, je me demandais de quel côté sont les humanités, et de quel côté est le digital dans cet anachronisme.
- 58 Paul Bertrand : Je suis tout à fait d'accord. On peut poser la question dans les deux sens. La façon dont je l'avais entrevue, comme Andrew Reinhard l'avait entrevue, c'était dans ce premier sens. Mais on pourrait tout à fait envisager les choses dans ce sens-là aussi. Le *steampunk* est dans les deux sens, tout à fait.
- 59 Paul Girard : Je vous remercie pour cette intervention. Je vais plutôt donner le point de vue d'un ingénieur, du coup. J'ai entendu beaucoup de questions et de problématiques mais en fait peu de réponses concrètes sur la question des données. Ne serait-ce pas justement une bonne pratique que de définir ce qui devrait être la « bonne donnée » pour les *digital humanities* ? Il y a en tout cas probablement un juste milieu entre le quanti et le quali qui est celui que vous proposez à la fin sous la forme de laboratoires comme ceux qui existent à Sciences-Po ou à l'ENS Lyon. Il y a probablement un juste milieu entre le PDF et la TEI. On peut avoir un texte qui ne soit pas un texte brut et qui pour autant soit traitable par des informaticiens, et qui ne soit pas bloqué comme le PDF qui malheureusement est ingérable pour les ingénieurs. N'y a-t-il pas à définir, à défaut de bonnes pratiques, de bons principes pour l'interopérabilité entre les chercheurs ?
- 60 Paul Bertrand : Je pense qu'on peut même aller encore plus loin. Je pense, que ce qu'il faudrait vraiment – au-delà des formats, au-delà des problèmes que j'ai soulevé – c'est se poser la question de savoir ce que c'est que ce donné que nous avons. Avant même d'essayer de le formater, ou de le mettre dans un format, je voudrais simplement que l'on réfléchisse. C'est quelque chose que tout le monde attend et qui permettrait à la communauté des *digital humanities* de savoir jusqu'où aller plus loin, et peut-être de choisir ces bonnes voies. Mais avant de choisir ces bonnes voies, il faut déjà savoir sur quoi on travaille. Et pour l'instant, moi, cela fait des années que je ne vois pas cette réflexion se mettre en place. Non parce qu'il n'y a pas de volonté mais parce que l'on n'a pas le temps de le faire, parce que l'on a une pression énorme qui nous empêche de mener cette réflexion. Ce que je propose c'est de faire un non-laboratoire en quelque sorte ; un peu dans l'idée du THATCamp, l'idée d'un espace hors du temps, hors du

temps de l'institution, même si c'est un fantasme, dans lequel on tente de réfléchir à ce donné.

- 61 Stéphane Pouyllau : Je suis principalement d'accord avec tout ce que tu as dit tout au long de ton exposé, cela n'est pas vraiment une surprise. Sur cette problématique autour de la donnée, je suis assez d'accord avec l'idée du non-laboratoire, l'idée d'avoir un espace de réflexion et de discussion qui se réunisse à intervalle régulier. Le problème, c'est que dès qu'on va le réunir à intervalle régulier, on va l'institutionnaliser. On sera toujours, je pense, obligés de passer par ce genre de choses. Mais je pense que c'est essentiel parce que la question des données et de la capacité qu'ont les chercheurs à les structurer, à les malaxer pour les réutiliser, est liée à un besoin qu'il faut qu'on arrive à faire émerger, et à carrosser sur le plan de la formation pour qu'on ait à tous points du territoire, dans tous les laboratoires, à un moment donné quelqu'un qui dise « attention, je dois me préoccuper de la réflexion sur les données ». Et là où je te rejoins, c'est que la pression institutionnelle l'empêche et agit comme un couvercle qui tombe par-dessus ces réflexions. Il est très important de relever ce couvercle, pour donner de l'air et de retrouver du temps pour savoir ce que l'on est en train de faire avec ces données et comment on va faire avec les données que l'on a à notre disposition. Cela me semble essentiel, et je pense qu'il faut arriver à trouver une solution ; je souscris à cette idée de non-laboratoire, ou de laboratoire virtuel qui permettrait de prendre ce temps-là.
- 62 Par ailleurs, je pense qu'il y a quelque chose à inventer pour que les collègues qui utilisent les bases de données, quel que soit le format (je ne me positionne pas en termes de formats ou de logiciels utilisés) arrivent à se mettre dans la situation où ils sont toujours dans une position de chef de projet, en assistance à la maîtrise d'ouvrage. Il faut qu'ils soient capables de mobiliser rapidement leurs compétences pour produire une base de données mais qu'ils aient conscience qu'à un moment donné il faut qu'ils fassent appel à d'autres personnes pour faire le travail parce que celui-ci sort de leurs compétences. C'est la notion de relation : savoir à partir de quel moment on va devoir passer le témoin à des gens qui ne sont pas issus du monde académique, qui ne connaissent pas le terrain, mais que l'on va aller chercher parce qu'ils savent faire, ou qu'ils savent faire autre chose que l'on ne sait pas faire. Et de la tension qui va s'installer, de la discussion qui va s'installer entre eux et le chercheur, c'est de là que va naître en fait l'innovation. Il y a un travail à faire auprès des collègues historiens, géographes, ou du monde académique en sciences humaines et sociales, pour qu'il se comportent en chefs de projets. Qu'ils sachent faire une partie du travail, très bien, mais à un moment donné aient conscience qu'il faut chercher d'autres compétences ailleurs.
- 63 Paul Bertrand : Je ne peux qu'être d'accord avec toi, c'est évident. Pour le non-espace, ce que je veux vraiment dire c'est que l'on relance une démarche d'innovation – je sais bien que le terme est très connoté, très politiquement correct et en même temps pas très intéressant. Mais derrière le concept d'innovation, il y a la recherche. La recherche qui, par nature, tend toujours à aller plus loin. L'idée c'est de relancer une démarche un peu audacieuse parce que je pense que l'on en manque assez ces derniers temps.
- 64 On est souvent confronté au problème très concret du contact entre les métiers. Faire en sorte que le chercheur comprenne qu'un ingénieur *digital humanities* n'est pas là pour publier son corpus, mais qu'il est là pour accompagner son corpus du début à la

fin. L'idée c'est d'accompagner. Et quand les problèmes d'accompagnement seront passés dans les mœurs alors la question sera résolue ; presque.

- 65 Frédéric Clavert : Ma question a été un peu vidée de son sens par celle de Stéphane. Mais peut-être une simple question, concrètement, ce non-laboratoire, qu'est-ce que c'est exactement ? Comment l'organise-t-on, enfin on ne l'organise pas, c'est un non-laboratoire : mais qu'est-ce que l'on fait ?
- 66 Paul Bertrand : Sur comment l'on fait, je n'ai aucune idée. Sur ce qu'on y ferait, là j'ai une petite idée. L'idée que j'ai c'est de se retrouver d'abord simplement autour de nos données ; sur ce que nous entendons chacun par « données ». Puis d'aller plus loin, en mettant à plat, chacun en essayant de décortiquer la réalité face à laquelle nous nous trouvons ; tout en essayant de le raccrocher avec les projets dans lesquels nous sommes, essayer de trouver des points de convergences, des points de ralliement. C'est une démarche très philosophique, très intellectualisante que je propose là. Mais vous savez bien que les *digital humanities* c'est une approche très philosophique de la donnée, ce n'est pas beaucoup autre chose. Il faudrait donc travailler sur les données que nous rencontrons et essayer de les classer, de les répertorier, non pas à la façon de Linée, mais dans une perspective de structuration sémantique, matérielle éventuellement, documentaire, etc. C'est à décliner, je lance les choses, mais tout est à construire.
- 67 Frédéric Clavert : Un atelier demain là-dessus ?
- 68 Paul Bertrand : Pourquoi pas. C'est un peu l'idée que j'avais en tête. Mais j'attendais les tomates ou pas.



- 69 Pierre Mounier : Je suis content de voir que cette idée d'une démarche beaucoup plus réflexive recoupe beaucoup les textes qui ont été choisis pour le *Read / Write Book 2*. Vous verrez, beaucoup de textes ont été choisis pour cette dimension là, à la fois réflexive et critique sur ce que deviennent, ou sont en train de devenir, sous un certain

nombre de pressions, les *digital humanities*. Je pense que c'est une démarche vraiment importante et je suis ravi qu'on la partage.

- 70 Deuxième chose, je peux annoncer que si les acteurs des *digital humanities* ont découvert le *steampunk* grâce à Boulet, il faut savoir aussi qu'il y a une demi-heure, Boulet a découvert les *digital humanities* sur twitter, en particulier grâce à Rémi Mathis qui l'a tweeté, et Boulet a répondu !

- 71 Ce média ne peut être affiché ici. Veuillez vous reporter à l'édition en ligne <http://books.openedition.org/editionsmsh/327>

NOTES

1. <http://classicslibrarian.wordpress.com/2012/06/04/lawdi-conference-on-linked-open-data-for-ancient-studies/>
2. *Read/Write Book 2, Une introduction aux humanités numériques*, Pierre Mounier (dir.), Openedition press, 2012, <http://press.openedition.org/226>, page consultée le 27 septembre 2012.
3. <http://www.rechercheisidore.fr>
4. Harmut Rosa, *Accélération, Une critique sociale du temps*, coll. Théorie Critique, Paris : Éditions La Découverte, 2010.
5. Langlois et Seignobos, *Introduction aux études historiques*, 1898.
6. Franco Moretti, *Graphes, cartes et arbres, Modèles abstraits pour une autre histoire de la littérature*, Paris : Éditions Les prairies ordinaires, 2008.

INDEX

Mots-clés : steampunk, digital humanities, cyberpunk, recherche, donnée, métadonnée

Notions, formation, compétences

L'historien programmeur ?

Proposé par : Frédéric Clavert, Aurélien Berra, Franziska Heimburger

Collectif

Introduction

- 1 Entendons ici, le terme d'historien au sens large. En tant que chercheurs en sciences humaines et sociales, nous manions tous des documents, des sources primaires. Il ne s'agit pas de poser la question de la programmation par les chercheurs eux-mêmes, mais celle de la relation entre le programmeur et l'historien en partant d'exemples concrets pour ensuite en discuter¹. Nous avons pris, comme point de départ, les deux éditions de *Programming Historian*².
- 2 Trois éléments doivent être abordés : pourquoi sommes-nous amenés à nous poser la question du code ? Cette interrogation débouche sur une autre : quelle est la place du chercheur dans la chaîne d'un projet scientifique impliquant du développement ?
- 3 Dans quels cas est-il nécessaire de se lancer dans un peu « ou beaucoup » de programmation ?
- 4 Y a-t-il des exemples concrets où une programmation par un historien a été décisive dans sa recherche ?

I. Pourquoi un chercheur se pose la question du code ?

3

- 5 Pour la première édition de *The Programming Historian*, les deux auteurs avaient des profils différents : un informaticien converti à l'histoire avec 30 ans d'expérience du code informatique (William J. Turkel) et un historien ayant appris récemment à développer pour les besoins d'un projet précis (Alan McEachern). Pour la seconde édition, l'équipe, plus large et d'autant plus hétérogène, donne un très bon exemple de l'imbrication de nombreux corps de métiers et savoir-faire articulés autour d'une plate-

forme collaborative et typique des humanités numériques. Cette édition est plus variée dans les sujets qu'elle aborde et les outils et langages de programmations utilisés⁴.

- 6 Arrêtons-nous, un instant, sur la première version.
- 7 Le but des deux auteurs était de montrer comment récupérer de l'information sur le web, comment la stocker dans Zotero, pour ensuite exploiter les données récupérées. C'est ce que l'on appelle du *data mining*. Cette méthode a été appliquée aux archives en ligne *The Old Bailey on-line*⁵ et résumée dans le très bon rapport « Data mining with criminal intent⁶ ». L'origine de ce livre repose sur la conviction des deux auteurs que l'histoire numérique est le futur de la discipline historique. L'histoire numérique suppose-t-elle toutefois d'acquérir les compétences nécessaires au développement informatique ?
- 8 Frédéric Clavert explique pourquoi il en est venu à s'intéresser à ces questions. *Il a eu besoin de trouver* sur le web, de stocker dans Zotero et de manipuler des données sur l'Union économique et monétaire. Il s'agit ici d'une programmation très légère, quelque chose que l'on peut qualifier de « script » plutôt que de « développement » à proprement parler.
- 9 Toutefois, les humanités numériques sont très orientées vers la pratique et la méthodologie, plus que vers la théorie. Comment appréhender ce qu'implique cette pratique sans avoir des notions de code ? Il s'agit là d'un autre aspect de l'historien programmeur : non apprendre à développer pour son petit projet, mais apprendre à développer pour comprendre les grands projets de recherche impliquant plusieurs corps de métiers, dont des programmeurs. Par exemple, il est difficile de saisir pleinement le cycle de développement au sein d'un projet si l'on ne sait pas écrire un peu de code. Ainsi peut-on réaliser la complexité du développement, qui doit prendre en compte tous les comportements possibles de l'utilisateur pour lequel on conçoit un programme.
- 10 On peut avoir des expériences inverses : mettre un corpus de documents variés en ligne et proposer des outils d'exploitation de ce corpus. Le chercheur ne rencontre pas d'obstacles sur certaines opérations, comme modifier des textes formatés en HTML. Dans ces cas, l'intervention directe du chercheur permet même un certain gain de temps. Mais il y a un problème de veille sur les nouveaux outils, comparaison des diverses applications disponibles, analyse des avantages/faiblesses des diverses possibilités en fonction des données spécifiques.
- 11 Le conseil et l'orientation sont cruciaux pour les choix de l'historien. Dans le secteur privé, on fait appel à des sociétés de consultants spécialisés dans divers domaines. Ces consultants travaillent sur des projets similaires et mutualisent les expériences et les connaissances. Les chercheurs en sciences humaines et sociales (SHS) manquent de ce genre de structure où des « consultants », informaticiens spécialisés en SHS, pourraient intervenir à des degrés et avec des implications diverses.
- 12 Une autre expérience a été partagée avec le langage d'interrogation de bases de données SQL. Comment traiter et stocker des données qu'on produit ? La prise de note doit-elle être faite avec Word ? En XML TEI ? La prise de notes étant collective, il est nécessaire de modéliser les données et il devient important de faire comprendre à une équipe complète que produire des données revient à produire un modèle de données. La modélisation des données est, en fait, la première dimension scientifique de l'entreprise. La seconde touche aux partages des données : chaque chercheur

produisant des données dans un contexte de recherche précis doit apprendre à distinguer les questionnements scientifiques des données brutes. Passé la modélisation, on touche aux limites de ce que l'historien peut faire (gestion du serveur, choix des logiciels, etc.). Mais la question des compétences techniques revient au moment de l'exploitation des données. Ensuite vient le point crucial : les chercheurs doivent connaître suffisamment le SQL pour pouvoir utiliser les données et produire de nouvelles connaissances.

- 13 Ainsi, besoin de différencier les technologies choisies du contenu et de son traitement. La pratique historienne de l'informatique n'est pas la gestion du serveur, la sécurisation du système, le choix des technologies utilisées, mais il lui faut les compétences nécessaires pour exploiter la richesse des documents disponibles⁷.
- 14 Ainsi se pose la question de la place et du rôle du chercheur lorsqu'il élabore un projet nécessitant un développement informatique.

II. La place du chercheur dans la chaîne de développement, de l'expression du besoin à l'aboutissement du projet

- 15 La discussion entre l'ensemble des participants de l'atelier s'est clairement orientée vers le rôle du chercheur au sein d'un projet de recherche impliquant du développement d'une part, et vers la nécessité de formation et de culture numériques des chercheurs d'autre part.

A. La place du chercheur au sein d'un projet numérique

- 16 Évoquant un projet d'histoire médiévale, une participante a formulé un besoin spécifique. Elle s'est adressée à un informaticien qui a développé un plug-in qui répondait à ses besoins. L'expérience – une discussion entre historienne et développeur – a été riche. Avec surprise, elle a rapidement constaté que l'informaticien lui faisait part de nouvelles perspectives sur l'interprétation des manuscrits médiévaux. Elle a par la suite suivi une formation en Java pour mieux communiquer avec son informaticien⁸.
- 17 Dans ce dialogue entre programmeur et chercheur, il faut insister sur la modélisation des données, dès le début du projet. Ainsi, faut-il revenir sur la question de la modélisation et du risque à prendre lorsqu'au début d'un projet des choix sont faits, impliquant à titre principal le chercheur. On ne peut revenir sur ces décisions, très difficiles à prendre.
- 18 Le rôle du chercheur dans la modélisation des données repose sur l'explicitation de ce qu'il veut faire. Il n'est pas sain de programmer tout seul et de faire, ensuite, l'exploitation des données. L'autre rôle, très important, du chercheur est de faire comprendre aux développeurs ce qu'est le monde des sciences humaines et sociales. Dans le cadre d'un travail impliquant informaticiens et paléographes, ces derniers ont pris deux jours à expliquer que la paléographie ne fonctionne pas avec des réponses vrai/faux ou des pourcentages d'exactitude.
- 19 Le rôle du chercheur dans un projet numérique change la temporalité de la recherche. Pour un projet européen, les sociologues impliqués ont voulu donner accès à l'ensemble

des interviews réalisées, plutôt que de laisser chaque groupe national collecter de son côté les données. Il a ainsi fallu un énorme travail en amont, avec les risques que cela comporte, pour modéliser et gérer les données. Pour être sûr de mener cet effort à bien, les chercheurs ont été obligés d'explicitier leurs buts, d'harmoniser le modèle de données, ce qui implique des choix techniques, mais également scientifiques. Or, s'il faut rendre la recherche explicite dès le début d'un projet, sa temporalité est modifiée et la recherche, plus risquée.

- 20 La modification de la temporalité de la recherche pousse à se pencher sur la question de l'économie du temps. Selon l'un des historiens présents, la recherche est un processus long, qui part de documents, de leur analyse, engendrant un besoin de traitement informatique. À ce moment-là, intervient le dialogue entre informaticien et historien, puis le développement propre par l'informaticien. Enfin, l'historien doit développer la maîtrise de son nouvel outil de travail, s'y familiariser. Dans certains cas, on doit envisager qu'il soit plus rapide, plus efficace que l'historien développe lui-même son outil, dont la modélisation se fait dès la lecture de la source par l'historien. Jean-Philippe Genet parle à ce propos d'une « méta source »⁹. Seul l'historien est capable de construire cette méta source – il ne peut pas déléguer à l'informaticien cette tâche dans la mesure où l'enjeu épistémologique est majeur : c'est sa connaissance de la période et son travail d'analyse des sources qui vont lui permettre de concevoir un modèle de données pertinentes.
- 21 Programmer ou savoir utiliser les logiciels ? Au cours d'une expérience d'une base de données collaborative associée à un projet d'analyse d'images, les interfaces et la base de données ont été créées par l'informaticien, avec des contraintes d'ergonomie et après le conseil des historiens. Cela a bien fonctionné. La conclusion à en tirer est que les chercheurs doivent d'abord connaître les logiciels développés pour eux, pas nécessairement qu'ils sachent programmer. Des logiciels puissants existent, il n'est pas toujours nécessaire de réinventer la roue. L'ignorance informatique des historiens doit être évitée, mais il est nécessaire de disposer d'une bonne culture informatique, autorisant un bon apprentissage des logiciels.

B. Formation et culture numériques des chercheurs

- 22 Puis vient la question de la formation et de la culture numériques des historiens. Ce qui semble important est qu'une culture informatique comprenant les bases de la programmation – mais pas nécessairement une pratique suivie et autonome – permette aux historiens – sociologues, anthropologues, – d'acquérir un vocabulaire les autorisant à dialoguer avec des informaticiens et à faire comprendre leurs besoins et leurs idées.
- 23 Quelle place doivent prendre ces éléments dans la formation des jeunes historiens ? Comment définit-on cette culture numérique ?¹⁰
- 24 Dans certaines disciplines, et notamment l'archéologie, la problématique des relations entre historiens et informaticiens est bien présente depuis trente ans. Les difficultés rencontrées sont souvent liées à la formation initiale des chercheurs, qui n'ont pas appris à programmer et doivent, quand le besoin s'en fait sentir, s'y auto-former. Mais alors se pose la question de la définition de la limite jusqu'à laquelle l'historien doit s'investir dans la programmation, du dialogue avec les informaticiens, de la disponibilité durable dans un projet ou une équipe de recherche des compétences en programmation et de la façon dont ces compétences sont mobilisées. Mais la conviction

que le chercheur est le seul à pouvoir comprendre le questionnement et la crainte de l'appauvrissement par l'explicitation l'emporte.

III. Interrogations

- 25 En premier lieu, le monde des informaticiens et le monde des sciences humaines et sociales ont-ils quelque chose en commun ? La question de la culture générale et de la connaissance de l'autre est à double sens. L'informaticien doit, autant que l'historien, chercher à comprendre le monde de l'autre. Un espace commun doit être créé.
- 26 On peut rappeler que l'adoption de Sparql, XML et R reste une exception chez les historiens. Apprendre à des doctorants à utiliser des outils basiques (traitement de texte, Zotero, etc.) est déjà difficile pour beaucoup de chercheurs. Peut-être faudrait-il qu'un historien-bidouilleur émerge, plutôt qu'un historien-programmeur. Le chercheur pourra ainsi être autonome (et rapide) sur des petites tâches de recherche, tout en évitant l'écueil, pour les grands projets, de faire de mauvais choix techniques, qui, plus tard, risquent d'entraver le partage et la diffusion de ces bases.
- 27 Une historienne estime que, si elle fait de l'informatique, c'est pour résoudre des problèmes qui se posent à elle. Elle travaille alors avec des informaticiens à l'INRIA. La moindre des choses est qu'il y ait une culture partagée : les informaticiens doivent connaître les sciences humaines et sociales, la culture des chercheurs avec qui elle travaille. Quant à la culture informatique des chercheurs, elle doit venir de loin, dès la maternelle. Que veut dire programmer ? Il faut, avant tout, collaborer, coopérer et travailler !

Conclusion

- 28 Le déficit de formation en informatique des chercheurs en sciences humaines et sociales est flagrant. Leur formation est nettement insuffisante. Elle peut inclure de la programmation, pas tant pour coder que pour apprendre les logiques informatiques.
- 29 Sur les modalités de cette formation, une interrogation subsiste ; l'existence d'une interface évolutive et accessible semble indispensable. Une formation spécifique risque de s'institutionnaliser et donc de se figer dans des certitudes qui ne sont pas du domaine des humanités numériques.
- 30 Se trouve également posée la question de la définition des humanités numériques : discipline ou ensemble de méthodes portant sur des usages nouveaux à la fois de sources et d'outils ?
- 31 Des rencontres régulières à la façon des universités d'été seraient peut-être la meilleure formule pour répondre à ce déficit de formation.
- 32 Les métiers d'historiens et d'informaticiens restent bien distincts. Au lieu de chercher à faire le travail de l'autre, émerge plutôt la nécessité de travailler ensemble, ce qui nécessite une culture commune.
- 33 Les données que produisent les chercheurs dictent, d'une certaine manière, le rôle de l'historien dans la chaîne de développement qui doit expliciter ses méthodes, dès le départ, identifier les risques de cette explicitation liés à la difficulté de « tout » expliciter en amont.

- 34 Il faut, aujourd'hui, pour comprendre la distinction entre le travail du chercheur et celui de l'informaticien, faire la différence entre la structuration des données – que les historiens doivent apprendre – et leur traitement.
- 35 De plus, avant de s'engager dans le développement d'un outil informatique, pour répondre au besoin de traitement de données historiques (ou autres), il est nécessaire de faire l'inventaire de ce qui existe dans son labo, sur le Web, etc., afin d'identifier les outils génériques pouvant être utilisés pour d'autres projets que celui pour lequel ils ont été développés (modèle de données non-spécifiques, usage de normes), d'identifier les outils adaptables pouvant être aisément modifiés (ex : code ouvert, licence de distribution, accès au programmeur) et afin de garantir un bon accès aux données.
- 36 La question de l'historien programmeur a finalement soulevé celle de la formation et de l'articulation du travail entre historien et informaticien.
-

NOTES

1. Outre le fameux article d'Emmanuel Le Roy Ladurie, « La fin des érudits ». *Le Nouvel Observateur*, 8 mai 1968, qui contient la fameuse citation « L'historien de demain sera programmeur ou ne sera plus », vous pourrez lire les éléments suivants: <http://lists.digitalhumanities.org/pipermail/humanist/2010-October/001666.html>, <http://www.clavert.net/wordpress/?p=385> et <http://www.boiteaoutils.info/2011/09/les-historiens-seront-ils-finalement.html>

2. Pour la première édition : William J. Turkel and Alan MacEachern, *The Programming Historian*, 1st ed. NiCHE: Network in Canadian History & Environment (2007-11), <http://niche-canada.org/programming-historian> ; pour la seconde édition : William J. Turkel (dir.), *The Programming Historian*, <http://programminghistorian.org/>, 2nd ed. NiCHE: Network in Canadian History & Environment (2012-07).

3. Frédéric Clavert fait un essai pour appliquer cette méthode à un domaine précis : en parcourant les sites web de l'Union Européenne (europa.eu / ecb.eu) et de la banque des règlements internationaux (bis.org) avec des mots clés bien réfléchis, le but est de récupérer tout document concernant l'Union économique et monétaire, puis de les stocker dans Zotero et, enfin, de commencer l'analyse avec l'extension (encore instable) pour Zotero appelée *Paper Machines* (<https://github.com/chrisjr/papermachines>). *Paper Machines* autorise une première visualisation, sans avoir besoin de développement, justement.

4. *The Programming Historian* est un guide d'autoformation avec des cours dont la durée varie entre une demi-heure et une heure. Les premiers articles sont des versions remaniées de l'édition de 2007, de nouvelles leçons ont été ajoutées (exemple, la commande Unix *wget*) d'autres sont en préparation (notamment sur Zotero).

5. <http://www.oldbaileyonline.org/>

6. <http://criminalintent.org/>

7. Intervention de F. Beretta et P. Vernus à TCP 2010 sur la modélisation des données en histoire : <http://tcp.hypotheses.org/341> - SyMoGih - http://larhra.ish-lyon.cnrs.fr/Pole_Methodes/SyMoGIH_fr.php

8. Graphoskop : <http://www.palaeographia.org/graphoskop/index.htm>

9. Jean-Philippe Genet, « Histoire, Informatique, Mesure », *Histoire & Mesure* 1 (1986) : 7-18.

10. Voir les autres chapitres sur la Formations et humanités numériques en France et Quelles compétences et littératies pour les humanités numériques ?

RÉSUMÉS

À partir du projet de site internet et de livre qui sont devenus collaboratifs pour la préparation d'une deuxième édition - et sur la base d'exemples concrets, cet atelier se propose de discuter de l'intérêt, pour des chercheurs en sciences humaines et sociales, d'apprendre les bases de la programmation.

INDEX

Mots-clés : histoire, programmation, livre collaboratif, sciences humaines

Formations et humanités numériques en France

Aurélien Berra et Florence Clavaud

Collectif

Présentation

- 1 Le repérage des formations en humanités numériques qui existent en est à ses débuts. On peut en fixer le point de départ avec l'état des lieux international des cursus en humanités numériques présenté à DH 2012 dans la table ronde "Digital Humanities as a university degree », organisée par M. Thaller et P. Sahle. Pour la France, c'est Florence Clavaud, responsable du Master « Technologies numériques appliquées à l'histoire » à l'École nationale des chartes, qui a été contactée. Cette question est centrale car c'est par la mise en place de véritables cursus que l'on pourra faire passer les humanités numériques de ce qui peut être perçu comme une « marotte individuelle » à une position plus établie dans le paysage institutionnel.
- 2 Florence Clavaud a récemment publié une synthèse des premiers résultats de son enquête. À l'université Paris 1, les membres du Pôle Informatique de Recherche et d'Enseignement en Histoire se sont attaché à établir une cartographie des centres et enseignement en humanités numériques francophones. Cette initiative se veut contributive et ceux qui y verraient des lacunes sont invités à y apporter les informations dont ils disposent.



- 3 En France, de rares formations de niveau master sont déjà bien repérées :
 - Le master « Technologies numériques appliquées à l'histoire », de l'École des Chartes
 - Un master à Caen
 - Le nouveau master « Architecture de l'information » qui vient de s'ouvrir à l'ENS de Lyon, comporte une spécialité « humanités numériques ».
- 4 Pour Florence Clavaud, l'insertion professionnelle à la sortie de ces formations est bonne, car elles produisent des profils mixtes qui sont appréciés. Pour autant, l'enquête reste à compléter. Un certain nombre d'éléments de formation existent, mais ils sont disséminés dans les cursus disciplinaires et de ce fait, échappent au radar.
- 5 C'est d'ailleurs un débat récurrent : faut-il établir des cursus « humanités numériques » spécifiques, comme le réclame le « Manifeste des Digital Humanities », ou instiller des éléments dans tous les cursus en sciences humaines et sociales ? À titre d'exemple, Frédéric Kaplan propose à l'EPFL de Lausanne un cours sur le sujet à destination d'étudiants de sciences expérimentales. Dans un autre registre, le King's College a fait le choix de renommer un certain nombre de ses cursus disciplinaires en y ajoutant une épithète « digital », comme, par exemple, les « digital historical studies ». À Lausanne, après avoir hésité à créer un doctorat « Digital Humanities », l'université semble se diriger plutôt vers des mentions « digital humanities » associées aux doctorats classiques.
- 6 De manière générale, un débat sur la question des formations soulève d'autres questions antérieures, plus profondes, qui relèvent des relations difficiles entre sciences humaines et technologies de l'information. Ces dernières pâtissent d'un manque de connaissance et de perceptions déformées chez les enseignants et chercheurs de ces disciplines. Ainsi, souvent, ceux-ci réduisent « l'informatique » à des compétences très sommaires. Tel étudiant viendra dans un cours sur les usages de l'informatique en histoire dans l'espoir de débloquer son antivirus. À l'extrême inverse, c'est presque en cachette que certains étudiants investis dans des cursus de sciences humaines, iront chercher ailleurs des cours de programmation qui ne sont pas intégrés dans celui auquel ils sont inscrits. Un historien s'intéressant aux SIG devra changer de discipline et se rallier à la géographie et aura du mal, par la suite, à obtenir une reconnaissance au sein du laboratoire sur cette double compétence.
- 7 Dans la gestion de certains projets, cela conduit à des situations absurdes : l'informaticien et le chercheur s'occupent du développement d'un site ou d'une plateforme en dehors de leur temps de travail habituel car il n'y a de reconnaissance ni

d'un côté ni de l'autre. Et lorsqu'il est possible de recruter un poste pour le projet, il est très difficile de trouver le profil adéquat.

- 8 En France en particulier, le débat est surdéterminé par une opposition très forte et statutaire entre chercheurs et ingénieurs. Trop souvent, manifester un intérêt pour ce qui est perçu comme « technique » fait basculer du côté de l'ingénieur. Cette opposition est inconnue ou beaucoup plus faible dans certains pays, comme l'Allemagne ou le Luxembourg par exemple. Aux États-Unis, c'est la notion d'« alternative academic carrers » (#alt-ac) qui a récemment émergé sur la base d'une reconnaissance de double compétence.
- 9 Nombre de formations en sciences humaines incluent aujourd'hui ce qui n'est qu'un vernis numérique. Il est difficile d'aller plus loin car les savoirs-faire de base sont souvent manquants et l'intérêt des étudiants de sciences humaines pas toujours au rendez-vous. L'institution elle-même envoie d'ailleurs des signaux négatifs : les salles de formation ne sont pas toujours bien équipées, les matériels pas toujours bien entretenus. Tel cours sur les « ressources numériques » est systématiquement nommé « Informatique » sans plus de précision dans les maquettes de formation.
- 10 Il est possible de montrer l'intérêt des humanités numériques dans les formations, en montrant des outils, des réalisations ou en les reliant à la culture numérique. Par exemple, ce qu'a fait Harvard à propos de la controverse sur « l'évangile de la femme de Jésus ». Pour une part toutefois, cette dernière initiative peut être considérée comme relevant d'une opération de communication et non véritablement des humanités numériques.

Claire Clivaz - Un blog sur les Humanités Digitales - a Digital Humanities Blog



« ARTICLES RÉCENTS

■ L'«Évangile de la femme de Jésus» vu depuis la culture digitale: la Divinity School de Harvard maîtrise la communication digitale, à un point près.

■ Harvard Divinity School's Digital Control and Omission: The

23
Dimanche
SEPT 2012

L'«Évangile de la femme de Jésus» vu depuis la culture digitale: la Divinity School de Harvard maîtrise la communication digitale, à un point près.

POSTED BY CLAIRE CLIVAZ IN UNCATEGORIZED

≈ 2 COMMENTAIRES

- 11 Pour tenter de dépasser une situation peu favorable, deux voies semblent donc se dessiner.
- 12 – Utiliser la méthodologie comme cheval de Troie. Des cours de méthodologie spécialisés sur les technologies numériques peuvent être déclinés par discipline et dans ce cas, on constate souvent qu'ils sont davantage appréciés. Certains pensent que cette

offre pourrait être labellisée par l'intermédiaire de niveaux avancés de C2i. Et cela doit être poussé au niveau du doctorat où des séminaires transversaux sont souvent organisés au sein des écoles doctorales. Au Luxembourg par exemple, c'est le cas. Un grand nombre de ces séminaires portent sur différents aspects relevant des humanités numériques avec des intervenants invités. À l'EHESS, la formation « outils informatiques pour les historiens » relève aussi de ce cas de figure.

Un point est important : lorsqu'on évoque les questions de formation, on pense toujours à l'introduction de compétences informatiques dans les cursus de sciences humaines ; mais il serait sans doute intéressant de penser l'inverse. Il serait fructueux que des informaticiens puissent bénéficier d'une solide formation en sciences humaines. Il s'agit bien d'une compétence différente et très distincte de celle du chercheur. S'il est vrai que, le plus souvent les informaticiens ne font pas des études d'informatique pour aller dialoguer avec des chercheurs en SHS, certains conçoivent leur métier comme appartenant aux « sciences de l'information » et seraient prêts à faire cette démarche.

- 13 – Construire des cursus complètement hybrides, à cheval entre sciences humaines et sciences de l'information. C'est d'ailleurs l'un des intérêts du label « Digital Humanities » que de permettre d'imaginer ce type de parcours qui permet de produire des profils eux aussi hybrides, comprenant les enjeux aussi bien scientifiques que techniques. Le problème est qu'un affichage aussi visible produit quelquefois des réactions de rejet, ici chez les historiens, là chez les historiens de l'art.

Évidemment, la deuxième option suppose aussi que les humanités numériques s'imposent comme discipline ou « transdiscipline » comme le revendique le Manifeste. Prudence pourtant : les exemples de transdiscipline ne sont guère reluisants. Les profils « transdisciplinaires » sont merveilleux sur le papier mais non reconnus dans les commissions de sélection. C'est en réalité autour d'objets de recherche que les approches transdisciplinaires ont plutôt réussi. Mais alors, quel serait l'objet d'étude des humanités numériques ? On constate d'ailleurs que son autonomie est incertaine. Des postes jusqu'ici appelés « *digital humanities* » commencent à être requalifiés en « *big data* ».

Conclusions

- Deux aspects sont à prendre en compte : un tronc commun de connaissances et des spécialisations possibles.
- Le numérique peut-il être un vecteur de concepts d'une discipline à l'autre ?
- Les humanités numériques ont déjà une histoire et il est important que cette histoire fasse partie du tronc commun précisément pour concerner à la fois les ingénieurs et les chercheurs.
- Qui doit faire quoi ? On pense les formations d'après le cadre de formation français classique. Mais on pourrait penser à la création d'un centre de formation qui produise des programmes de formation adaptés aux champs disciplinaires.
- Les « écoles thématiques » sont un bon vecteur de formation, car elles réunissent des acteurs issus de nombreuses disciplines.
- La mise en place d'une vraie politique de formation passe par une activité de *lobbying* en faveur des humanités numériques, en particulier au niveau de la politique de la recherche

RÉSUMÉS

Dans cet atelier, il s'agit de faire avancer un projet d'enquête collective sur les formations intégrant des humanités numériques en France. Le point de départ sera l'état des lieux international des cursus en humanités numériques présenté à DH 2012 dans la table ronde « Digital Humanities as a university degree », organisée par M. Thaller et P. Sahle (18.07.2012). Pour la France, c'est Florence Clavaud, responsable du Master « Technologies numériques appliquées à l'histoire » à l'École nationale des chartes, qui a été contactée. L'atelier peut être l'occasion de contribuer à l'élaboration d'un questionnaire, de réfléchir aux modalités de sa diffusion et de comparer la situation française avec celle d'autres pays. Toutes les idées seront bienvenues.

INDEX

Mots-clés : humanités numériques, université, recherche, formation, enseignement supérieur

Quelles compétences et littératies pour les humanités numériques ?

Proposé par Olivier Le Deuff et Aurélien Berra

Collectif

NOTE DE L'ÉDITEUR

Cet atelier est le prolongement de l'atelier Formations et humanités numériques en France

Un flux de questions

- 1 Doit-on former des chercheurs qui comprennent la technique pour obtenir une autonomie du chercheur ? Cela devrait-il avoir lieu dans les écoles doctorales ou ailleurs ? Doit-on savoir coder avant de se lancer dans des projets d'humanités numériques ? Quelle est l'étendue de cette formation, dans quel domaine, avec quels outils ?

Une formation ou une méthode ?

- 2 Il est évident qu'une métamorphose des chercheurs (et de toute la science) a eu lieu grâce ou à cause du numérique. Cependant, pour répondre à ces questions, le recours à la notion de littératie est nécessaire. Il s'agit d'une notion d'origine anglosaxonne, qui a émergé chez nos collègues québécois, dont le champ sémantique est composé de plusieurs sens :
 - Premier sens : capacité à lire et écrire (*a simple ability to read and write*)
 - Deuxième sens : posséder des compétences et une certaine habileté (*having some skill and competence*)
 - Troisième sens : élément d'apprentissage (*element of learning*)

- 3 Cette extension de compétences portée par les littératies va de pair avec l'évolution de la place des outils numériques dans l'environnement des chercheurs. Face à ce constat, il s'agit de s'interroger sur les stratégies et les méthodes à mettre en place dans le domaine de la formation. Devrait-on parler plutôt d'une formation indépendante, transdisciplinaire et transversale ou bien d'une méthodologie ?
- 4 La diversité des métiers – et des champs de recherche impactés par les humanités numériques est telle que, trouver un terrain commun à toutes ces champs peut s'avérer très difficile. Cela pourrait plutôt passer par l'organisation d'ateliers communs.
- 5 Il est dangereux d'envisager une formation indépendante dans le domaine des humanités numériques : cela supposerait sans doute d'avoir des compétences illimitées dans des domaines très variés. Or l'essentiel concernant les humanités numériques relève des compétences collaboratives et de participation. Il faut communiquer ces nouvelles pratiques (nouveaux types d'écriture et nouvelles manières d'aborder les objets de recherche).
- 6 Actuellement il n'existe pas encore en France de formation dédiée exclusivement aux humanités numériques ; l'institution est encore absente de ce champ disciplinaire en construction, même si des initiatives récentes attestent toutefois d'une évolution de la situation.
- 7 Par ailleurs, les jeunes chercheurs qui investissent le champ des humanités numériques peinent à faire reconnaître leurs travaux par l'institution. Cela ressort pour partie d'une certaine incompatibilité hiérarchique entre jeunes chercheurs et institutions reconnues et pour partie de la difficulté rencontrée par les chercheurs de champs disciplinaires en cours d'institutionnalisation de bénéficier de la reconnaissance nécessaire pour avancer dans leur travail.
- 8 Par ailleurs, pour ces chercheurs en humanités numériques, les limites entre le travail (les recherches) et les loisirs s'effacent. Les frontières entre le lieu de travail et le lieu des loisirs deviennent étanches, mais le chercheur conserve malgré tout une bonne communication avec sa communauté de pairs.
- 9 Face à ces constats, on peut se demander quel est le profil des chercheurs des humanités numériques : le chercheur en humanités numériques serait-il un *alien* ? Le chercheur 2.0. par excellence ? Il serait très intéressant de mener une enquête sur cet objet de recherche.
- 10 De même, il est utile de savoir quelles doivent être les compétences de l'acteur des humanités numériques ? Le chercheur doit-il savoir coder ? L'ingénieur doit-il être au fait des thématiques abordées par les chercheurs ? Ou chacun doit-il garder ses compétences respectives et apprendre à mieux communiquer. Quelle est donc la distance à adopter par rapport aux données et comment appréhender cette distance ?
- 11 Quelques pistes possibles :
- 12 Les acteurs des humanités numériques – chercheurs et ingénieurs – doivent disposer d'un socle commun, d'une base commune, permettant l'échange professionnel.
- 13 Alors que pour quelques-uns, il est seulement nécessaire de connaître les bases du codage pour savoir exactement ce que l'on veut, être en mesure de l'exprimer en employant le vocabulaire adéquat et dialoguer, ainsi, avec l'informaticien. D'autres, au contraire, pensent que le chercheur, à l'instar de Dan Cohen, doit nécessairement

savoir coder, afin de disposer de compétences informatiques lui permettant d'être autonome et de ne pas dépendre d'un informaticien.

- 14 Mais, s'il est relativement aisé pour les chercheurs d'acquérir ces compétences techniques, recourir à ces compétences pour modifier les choses eux-mêmes, l'est beaucoup moins. On sait en effet à quel point il peut être complexe d'intervenir dans un code fait par quelqu'un d'autre et jusqu'où intervenir pour que le travail de l'informaticien soit préservé. Ainsi, dans les faits, pouvoir intervenir dans du code permet surtout d'évaluer les besoins au début d'un projet éventuel ; pour mieux savoir ce que l'on ne sait pas faire, et ce que l'on n'aura pas le temps de faire.
- 15 Cette question de la compétence peut être renversée : l'informaticien doit-il devenir historien par exemple ? Quel peut être l'étendue du domaine d'intervention de chacun ? Connaître le point de vue des informaticiens à ce sujet serait intéressant.

Le point de vue des informaticiens

- 16 Il serait souhaitable que les informaticiens aient eux aussi des formations en sciences humaines et sociales. Cependant, nous pouvons nous interroger sur ce qu'est le métier d'« informaticien ». Il est difficile de définir le champ d'intervention, de connaissance et de collaboration de ces acteurs. Il faut se garder de cloisonner le terme d'informaticien, car plusieurs personnes ont déjà suivi deux cursus différents, même si, pour la plupart des personnes, ces deux domaines sont souvent distincts.
- 17 Il est certain que tout le monde doit avoir une culture numérique, mais avoir une culture numérique ne signifie pas savoir programmer. Et quand bien même on souhaiterait savoir programmer, on pourrait aisément recourir à des outils d'aide à la programmation.
- 18 Dans certaines branches de la recherche, comme dans le domaine de l'archéologie, la connaissance informatique est plus « naturelle », bien qu'elle nécessite toutefois une mise à jour régulière. En effet, pour rester compétent en informatique et savoir utiliser les nouveaux outils, il est nécessaire de se former en continu.
- 19 Face au constat de cette évolution rapide, on remarque les limites de l'apprentissage informatique. Il s'agit de prendre en compte la question du temps de la formation, plutôt que celle de différence générationnelle.
- 20 Il y a la nécessité d'un apprentissage précoce, d'une culture numérique générale. Tout le monde devrait savoir utiliser les ordinateurs et des programmes comme Zotero. L'éducation nationale (C2i) prévoit des cours d'informatique (au collège et à l'université). Mais on y apprend des choses limitées. Des différences existent entre un citoyen né à l'ère du numérique et un chercheur du numérique. L'éducation est certes importante mais pas suffisante, car de manière générale il n'y a pas de culture numérique globale suffisante, même si cette culture est essentielle au chercheur pour qu'il puisse dialoguer avec ses interlocuteurs.
- 21 Par exemple, lorsque l'informaticien évoque le code, voire l'XML, il est déjà plus difficile de communiquer avec les historiens par exemple. On se demande alors où s'arrête la culture de base.
- 22 La principale difficulté en informatique est sa continuelle évolution. Pour la formation, cela devient problématique. Il faut donc essayer de trouver des fondements plus

durables tout en privilégiant l'utilité immédiate. Il faut former aux savoirs et à un savoir immédiat, dès le lycée, ce qui permettrait de mieux comprendre les processus et d'appréhender les enjeux.

- 23 Les humanités numériques permettent-elles d'en finir avec le cloisonnement entre intellectuels et techniciens ou bien accentuent-elles justement cette séparation ? Il faut garder à l'esprit la nécessité d'une distance critique nécessaire lorsqu'on est acteurs des humanités numériques, car le risque est que la donnée devienne un résultat en soi, tandis que la capacité d'interprétation des données demeure centrale pour les humanités numériques.

Question de gouvernance

- 24 Il est nécessaire de se demander à qui revient la tâche de définir les compétences, les cursus et les manuels. Par ailleurs, s'il y a un socle de compétences communes à conquérir, quels peuvent être les acteurs, à partir de quel moment commence-t-on la formation, à partir du doctorat ? Ou bien avant ?
- 25 Il est important de passer du discours à l'action et de valoriser les acquis de la réflexion. En tant qu'acteurs des humanités numériques, il faut convaincre nos pairs et insuffler l'envie à tout un chacun de poursuivre dans ce champ disciplinaire.
- 26 Cependant, la question de la gestion de la carrière des chercheurs se pose, car le monde du numérique ne bénéficie pas encore de la reconnaissance nécessaire. Bien au contraire, s'investir dans le champ des humanités numériques et jouer le jeu des humanités numériques, notamment en publiant ses recherches sur le web peut représenter un risque pour les chercheurs.
- 27 Pour les étudiants, même si la pratique proprement dite du logiciel est *a priori* plus utile et intéressante, il est plus important, notamment pour leurs CV, d'avoir une vue d'ensemble des logiciels utilisés.
- 28 Les compétences dans le domaine des humanités numériques sont rares donc et représentent une certaine valeur ajoutée, mais, malheureusement, elles ne sont pas toujours reconnues. Par ailleurs, la précarité relative qui touche les chercheurs en humanités numériques rend difficile la création de projets.
- 29 Il est à noter qu'il existe un domaine où le questionnement sur les compétences numériques semble évoluer : celui de l'archivage. De plus en plus de données numériques sont produites, sans toutefois que soit prise en compte la portée archivistique des travaux (on ne conserve pas toujours des copies). Or, il paraît essentiel d'assurer la pérennité de l'archivage des travaux, tout en se posant de « bonnes » questions par rapport aux formats, à l'évolution des systèmes afin d'adopter de « bonnes » pratiques. Pour répondre à toutes ces questions, les rôles et les compétences de chacun restent à définir, de même que l'étendue d'une culture numérique commune. Ainsi, la connaissance du langage HTML paraît aujourd'hui tellement évidente qu'on ne devrait plus avoir besoin de la mentionner sur notre *curriculum vitae*.

Qui sont les Digital Humanists ?

- 30 En guise de conclusion et de réponse générale aux différentes questions posées lors de ce débat, les participants tentent de définir ce qu'est le *digital humanist*.
- 31 Une enquête a récemment (mars 2012) été réalisée à ce sujet : Who are you Digital Humanists?. De dimension internationale, cette enquête pose la question du statut du *digital humanist*. Les résultats de l'enquête sont attendus avec impatience par la communauté des *digital humanists*.
- 32 La tentative d'identification du *digital humanist* passe par une identification des compétences de base, qui seraient notamment la maîtrise des langages HTML, TEI et Javascript. Ces compétences, pratiques, sont censées servir les recherches menées dans des disciplines différentes, dans un objectif de transversalité.
- 33 Mais le répertoire de compétences du *digital humanist* dépasse le cadre de ces compétences techniques. Il est donc, de fait, difficile de trouver une base commune à l'ensemble des compétences des *digital humanists*. L'enquête citée précédemment devrait donc permettre de définir avec plus de précision l'identité et le champ d'action du *digital humanist*. Mais cela sera-t-il suffisant ? Une enquête plus poussée sur les compétences nécessaires aux acteurs des humanités numériques permettrait de définir avec plus de précision l'identité et le champ d'action du *digital humanist*. Des financements existeraient pour cela.
- 34 Ce qui est essentiel et qui rallie les points de vue, est la nécessité du dialogue. Il ne faut pas forcément tout comprendre, seulement connaître le fonctionnement.
- 35 Il est également important de sortir d'une spécialisation trop forte, viser une multidisciplinarité et faire du *lobbying* à l'extérieur pour parler de la nécessité de formation et d'élargissement des compétences.
- 36 Le maître mot de la fin est : « Il faut être militant ! »

RÉSUMÉS

Un atelier qui concerne également l'identification de compétences. Question qui se pose souvent dans le cadre de projet. L'idée serait d'identifier ces compétences afin de contribuer à la définition de formations pour pouvoir disposer des personnes qui aient cette somme de compétences nécessaire pour monter des projets ambitieux. Sans doute faire un début de cartographie avec du mindmapping pour identifier des compétences clés à développer dans les formations. L'objectif est de débattre sur les compétences et littératies nécessaires pour s'investir dans les humanités numériques. Quelles sont les compétences numériques, informatiques et intellectuelles pour y parvenir. Faut-il faire évoluer les formations actuelles ? L'objectif est de tenter de dessiner une matrice des compétences minimales et d'envisager celles qui seront nécessaires à l'avenir pour accueillir et former les futurs chercheurs notamment en SHS.

INDEX

Mots-clés : compétences, littératie, formation, méthode

Lieux et communautés

Quels espaces physiques pour des humanités numériques ?

Atelier proposé par Nicolas Catzaras & Johann Holland

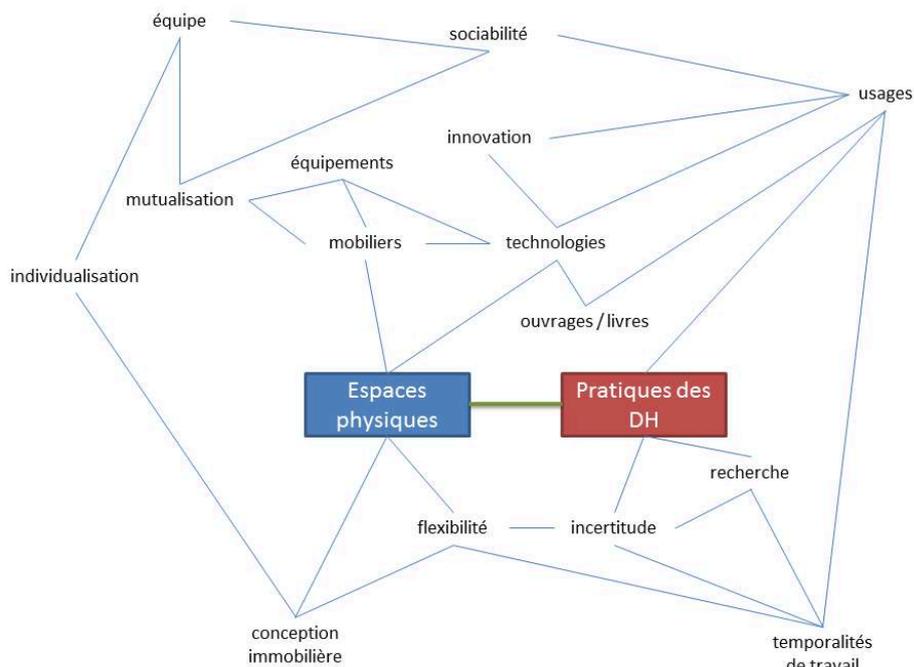
Collectif

Introduction

- 1 Un premier travail de réflexion mené sur le sujet a permis de dégager deux grands pôles : celui des espaces et celui des pratiques innovantes. L'un pouvant déterminer l'autre ou l'influencer, et *vice versa*. Si la question est de savoir quelle approche est aujourd'hui la meilleure afin d'équiper intelligemment l'espace physique pour instrumenter la recherche en SHS, il faut ajouter une dimension supplémentaire qui vient compliquer les choses. On ne peut pas mener cette réflexion sans tenir compte du facteur temps. Les technologies évoluent et les pratiques aussi, dans des rythmes plus ou moins synchronisés. Or notre questionnement porte aussi sur la réinvention permanente des pratiques, des équipements et des institutions.
- 2 Comment rester dynamique dans le temps ? Comment ces espaces peuvent rester suffisamment flexibles et dynamiques pour permettre la réinvention continue des pratiques et des approches heuristiques ? Ces questions se compliquent du fait de leur contexte où les temporalités ne sont pas les mêmes. D'un côté, le numérique nécessite des investissements importants – pour que ça fonctionne sans soucis techniques... – mais avec des évolutions et des innovations ultrarapides des technologies et des usages. De l'autre l'immobilier nécessite des investissements très lourds, pensés longtemps à l'avance en fonction de technologies stables et assez peu flexibles : les choix faits en matière d'immobilier conditionnent lourdement l'utilisation de l'espace et son lien avec les utilisateurs, en particulier dans le contexte numérique.
- 3 L'un des enjeux de la conception de tels espaces de recherche propices aux pratiques des humanités numériques et à leur évolution est de faire en sorte que tout fonctionne sans que l'utilisateur ait même à y réfléchir... Mais cela peut se faire au risque peut-être de passer sous silence la dimension réflexive de la pratique du chercheur telle qu'elle

s'inscrit dans l'espace physique et numérique. On parle ici du risque de la boîte noire dans laquelle on place sa confiance aveuglément sans s'intéresser à sa mécanique propre.

- 4 À partir de cet état de questions initiales, nous proposons pour lancer les discussions une carte de mots-clés et de relations indicatives. Cette carte peut servir à ouvrir un certain nombre de problématiques connexes. Le schéma proposé ci-dessous n'a pas d'autre ambition que de susciter des interrogations en faisceau.



Quelques développements à partir des mots clefs

Usages / Innovation

- 5 Le lien est très marqué entre usages et innovation, car l'innovation doit être entendue comme la marque de l'apparition d'un usage conséquent à l'échelle des pratiques considérées (ici de toutes les personnes dont les compétences servent à la recherche en SHS).

Équipements

- 6 Pour les équipements, on peut entendre des matériels déjà très connus ou « démocratisés » (écran, vidéoprojecteurs), mais aussi des dispositifs plus ambitieux et peu développés aujourd'hui (espaces d'holographie).

Dimension humaine

- 7 Dans la mesure où notre questionnement porte sur la réinscription du corps dans l'espace physique, une dimension plus humaine est à prendre en considération : la sociabilité, le travail en équipe, l'inscription de la personne dans l'espace.

Ressources

- 8 Tout cela est lié à la notion de ressources, qu'on entende par là les livres, revues, etc., sur supports matériels, mais aussi les ressources numériques ou numérisées auxquelles l'utilisateur doit pouvoir être lié. Pour des logiques d'accès aux sources et aux savoirs, mais aussi pour gloser et produire de nouveaux savoirs.

Lieu

- 9 La notion de lieu est également structurante. On parle ici du lieu en tant qu'espace privé dans l'espace de travail. La BnF pour cela offre un bon exemple. Dans les étages des tours, les personnels se sont réapproprié un espace défini comme neutre et interchangeable en le personnalisant, en s'y individualisant. À partir de la notion de lieu apparaît la notion de second lieu : au sein de l'espace de travail revient la notion d'espace privé. La notion de second lieu est essentielle. Elle a jusque-là été réservée aux bibliothèques (bibliothèques municipales, médiathèques, *learning centers*) mais assez peu aux bibliothèques de recherche. La question devient la suivante : faut-il étendre cette réflexion du second lieu au bureau du chercheur, un second chez soi, et quelles en sont les conséquences sur l'équipement numérique ? Cela touche aux infrastructures, bien sûr, mais aussi et de plus en plus à l'équipement personnel – ou acquis sur contrat – et apporté ou utilisé dans son labo, dans son bureau, à la bibliothèque... Cet état de fait du second lieu pose des questions d'interopérabilité, des questions de sécurité, mais surtout des questions d'usage.

Tour de table des intérêts

- 10 En faisant le tour de table des participants à l'atelier, nous avons pu identifier plusieurs grands types d'intérêts à la problématique proposée :
- La motivation institutionnelle pour l'immobilier (à construire ou à réhabiliter) conçu spécifiquement pour les pratiques innovantes de type humanités numériques, aussi bien dans la recherche et formation à la recherche que dans l'enseignement supérieur, à plus ou moins long terme (trois exemples potentiels de réalisation entre 2014 et 2018).
 - L'importance d'avoir des espaces physiques de collaborations entre des profils de compétences très divers (juristes, documentalistes, chercheurs, programmeurs, designers, éditeurs, populations tests...), pour sortir du tout virtuel.
 - L'anticipation des conditions de travail possibles dans le monde de la recherche, en vue de réinventer ou adapter sa pratique métier (par exemple dans le domaine de l'édition scientifique) aux nouveaux usages et outils.
 - L'étude historique de l'architecture, en particulier quand cela mobilise les notions d'espaces virtuels, et la problématique de l'inscription du corps ou des pratiques dans le numérique. Les architectes ont été assez peu prompts à réfléchir à l'espace numérique ou à s'y investir.

- L'évaluation des effets potentiels de la réinscription du corps dans des pratiques innovantes sur les formes de savoirs produites, réalisées (par exemple, dans l'écriture). La problématique semble assez peu étudiée pour le moment.
- La possibilité d'actionner le levier de l'espace physique pour que se développent vraiment dans le monde de la recherche les pratiques innovantes propres aux humanités numériques. L'articulation entre les espaces physiques et les espaces virtuels doit profiter aux seconds aussi.

Synthèse des sujets abordés

L'appétence pour les technologies

- 11 Actuellement, on peut se poser des questions quant à l'attitude générale des chercheurs vis-à-vis des nouvelles technologies. On peut même être assez pessimiste et constater un réel désintérêt pour les moyens qu'offrent ces nouvelles conditions de travail. Cependant, un tel constat n'est vrai qu'en partie. L'attitude des nouvelles générations de chercheurs qui utilisent massivement les outils numériques pour des usages « personnels » et professionnels change la donne à un horizon très court.

La sociabilité dans les espaces

- 12 On évoquait plus haut la dimension humaine de toute cette problématique. Dès lors qu'on peut intervenir sur un espace physique, on a une réelle opportunité d'agir sur les conditions de sociabilités des utilisateurs d'un équipement immobilier et technologique. Et cette sociabilité ne concerne pas seulement les chercheurs entre eux. Il est important notamment de travailler dès maintenant à une resocialisation entre les chercheurs et les techniciens par exemple. De changer de paradigme social dans le monde académique en dépassant le clivage habituel entre les chercheurs (c'est-à-dire les personnes jouant officiellement le rôle de chercheurs) et tous ceux qui sont trop fréquemment vus comme des prestataires de service.
- 13 Quand on a la chance, comme par exemple au Campus Condorcet, de pouvoir construire des espaces à partir de rien, on a un devoir de faire plus que de saisir une opportunité. On se doit d'avoir une nouvelle ambition sociétale. À savoir essayer de construire une nouvelle sociabilité au profit de la recherche pour que par hasard, des personnes de différentes disciplines et compétences puissent à un moment donné se retrouver comme par accident sur des mêmes problématiques, ou des croisements, par le biais du badinage, de la rencontre fortuite. Susciter la mise en relation de chercheurs, de documentalistes, qui se retrouveront dans un même espace et pourront « bavarder » dans des lieux de passages qui ne sont pas *a priori* des lieux de rencontres. On pointe ici le cas des chercheurs et des documentalistes – dont les métiers évoluent rapidement –, ou encore des éditeurs. Mais il est également important de tenir compte de tous ceux qui offrent des services mal valorisés à la communauté de la recherche (les « administratifs », qui par exemple gèrent les contrats, ou encore le gérant d'une conciergerie). La réflexion sera plus efficace si elle est systémique, embrassant un écosystème complet. Il faut trouver des moyens de connecter ces métiers et ces personnes pour former une communauté. Cela passe à la fois par du physique et du

virtuel, on ne pourra pas faire l'économie d'une double réflexion, surtout si on souhaite développer de nouvelles formes de sociabilité.

- 14 En ce qui concerne les possibilités de sociabilité renouvelée par l'espace immobilier et les technologies numériques, des lieux d'expérimentation commencent à émerger comme à Clermont-Ferrand (projet Epicentre : <http://culturestrafic.com/ct/epicentre-cowork/>). Il s'agit en l'occurrence d'espaces de *coworking* où des profils très divers se réunissent autour d'un même projet. Il semble également que des acteurs comme l'ENSCI se penchent sur ces questions (au travers d'un atelier sur les *digital studies*).

Comment concevoir les espaces en fonction des usages ?

- 15 La question n'est pas propre aux humanités numériques, elle se pose largement par ailleurs.
- 16 Une approche très générale et dichotomique peut nous tromper. On peut manquer le sujet en opposant deux « conceptions » extrêmes :
- On considère qu'un super espace est capable d'accueillir tous les usages et toutes les formes de sociabilité possibles.
 - On considère qu'on doit multiplier des espaces physiques hyper spécialisés pour favoriser des types de collaborations en particulier (ex : entre un éditeur, un juriste et un chercheur).
- 17 Le sujet peut être pris par un autre bout. On peut par exemple partir des types d'espaces physiques : espaces de travail isolé, espaces de travail ouvert et collaboratif, espaces de circulation, espaces de convivialité... Et réfléchir à leur articulation entre eux, notamment grâce au numérique. On reparle ici en fait de l'articulation entre l'espace numérique et l'espace matériel (physique). Car il est nécessaire de passer de la sociabilité dans le couloir, à la sociabilité dans le bureau. Et ce passage peut se faire en traversant des espaces virtuels qui nous baignent.
- 18 La circulation entre les espaces, reflet de la « circulation sociale » entre les individus / rôles / compétences dépasse d'ailleurs le microcosme des personnels qui font le cœur de la machine. Il est question, à travers un bâtiment équipé, de la relation avec l'environnement. C'est la question de la relation du monde de la recherche avec les invités extérieurs et plus largement le public, le bâtiment équipé pouvant jouer le rôle d'une interface ou d'une membrane perméable : je viens dans un bâtiment, j'y suis reconnu, je suis guidé à l'intérieur du bâtiment. La question est intéressante aussi à aborder sous l'angle de la compensation du handicap. Plus le bâtiment est pensé pour son usage par rapport à ces questions de compensation, plus le bâtiment est facile à vivre. Le numérique joue un rôle absolument central dans cette compensation.
- 19 On peut penser que l'espace de discussion numérique ne pourra se mettre en place qu'à partir du moment où on aura déjà des espaces physiques efficaces à disposition. Mais il y a un risque de retomber dans l'idéologie *steampunk* en s'attachant à des formes de sociabilité traditionnelles pour lesquelles le numérique ne fait que mimer les fonctions de l'immobilier (c'est le risque propre à tout « effet diligence » dont la logique est pourtant bien utile dans le cadre de l'accompagnement du changement). On peut espérer une sorte d'« agrégation naturelle » des espaces physiques et virtuels. La question alors peut devenir : comment passer du couloir (ou du bureau), à l'écran / interface ? Que vient faire là l'écran ? Et si l'écran n'est plus un écran mais une paroi ?

- 20 Certaines expérimentations existantes peuvent aider à imaginer ces articulations d'espaces. Par exemple celle d'un colloque où chaque participant est « tracé » grâce à une puce RFID qu'il porte. De plus, les relations entre participants sont enregistrées et représentées là aussi grâce aux interactions entre les puces RFID. Grâce à la puce, chaque participant peut projeter sur l'écran le plus proche son « environnement personnel » afin de retrouver des documents en lien avec sa conversation professionnelle. Et à l'issue du colloque, un site lui permet de retrouver l'ensemble des interactions qu'il a pu avoir pendant toute la durée de l'événement. Malgré son caractère angoissant (« Big Brother »), cet exemple est intéressant car il montre de la réinscription de l'espace physique dans l'espace numérique virtuel. Et inversement.
- 21 L'exemple qui vient d'être présenté est conforme à une tendance qui avait été identifiée il y a quelques années et qui a du mal à se confirmer. Celle prophétisant que l'individu ne se déplacera plus avec son matériel personnel, mais seulement avec son environnement personnel qu'il pourrait retrouver partout et sur n'importe quel poste connecté. Le développement important de l'informatique en nuage (cloud computing) laisse à penser que cette tendance est toujours d'actualité. Par exemple avec l'émergence du cloud, il semble que l'espace de stockage "situé" sur sa machine personnelle est amené à disparaître. Idem des applications logicielles, etc. Cette tendance qu'on peut appeler de manière un peu simpliste « cloud trend » est à opposer à l'habitus bien connu maintenant du « bring your own device ».
- 22 A ce stade on peut alors s'interroger sur ces deux tendances comme alternatives. L'avenir est-il aux usages de type « je viens avec mon environnement et sans son équipement » ? Ou au contraire à ceux où l'individu cherche à se réinscrire dans l'espace physique en venant avec son équipement ? En réalité l'avenir est sans doute à un meilleur équilibre des deux. On peut faire un mixte de tout cela, en offrant des espaces physiques équipés permettant de venir avec ou sans son matériel personnel, avec ou sans environnement personnel. L'essentiel dans tout cela est d'avoir une réflexion qui soit centrée sur l'utilisateur avant de se projeter sur les outils, et c'est toute la difficulté de l'anticipation que cela représente.
- 23 Cette idée de mixte est particulièrement pertinente dans les SHS. Car le chercheur y est encore très attaché à l'idée de travailler avec un matériel dont il a la propriété. Le jugement par les pairs étant très fort, il y a une sorte de forteresse que l'on construit autour de soi. Il paraît difficile, tout au moins à court ou moyen terme, de voir le chercheur arriver avec ses données dans un espace « neutre et interchangeable » même s'il pourra retrouver toujours son environnement de travail. L'enjeu pour lui est de ne pas perdre la main sur le fait de rester visible. Il s'agit d'être jugé par les pairs, donc d'être vu. Là perdure le fantasme du bureau, de l'ordinateur portable (le sien, c'est mieux). Un desktop collectif perturbe le chercheur. Actuellement, le climat n'est pas à la suppression des équipements, des espaces de travail physiques.
- 24 La recherche en SHS se trouve un peu coincée entre une perspective très globalisante et des difficultés à travailler à plusieurs. Comment faire en sorte que cette réalité ne plombe pas une dynamique collective ? On peut peut-être faire en sorte qu'il y ait une intégration virtuelle qui se fasse par compromis, progressivement. Là, le mouvement des DH a sans doute un rôle à jouer. Ne constituent-elles pas, justement, un levier pour accélérer un passage vers une recherche qui prend mieux en compte les dimensions collectives ?

- 25 Toutefois ce levier seul ne suffit pas. Des choix structurels sont à faire afin de permettre ou d'accélérer le changement. En ce sens l'ordinateur portable personnel est un intermédiaire. Il est un moyen terme qui anticipe l'espace de travail virtuel disponible partout et tout le temps, tout en présentant un caractère d'appropriation de l'espace physique plus aisée. Se rendre dans un espace de recherche neutre et impersonnel mais avec son ordinateur portable personnel serait ainsi un compromis entre le « mon bureau, mon ordinateur, mon inscription » et le tout virtualisé dans l'espace numérique délocalisé.
- 26 Le champ terminologique utilisé dans nos discussions est révélateur. On a parlé de « flottement » d'une part, et « d'ancrage » d'autre part. Il semble qu'on puisse filer une métaphore navale pour évoquer les attitudes possibles quant au rapport du corps aux espaces physiques et virtuels. Avec le flottement, l'utilisateur se retrouverait dans le collectif et des espaces (physiques, virtuels, disciplinaires aussi) qui sont poreux. Dès qu'il s'ancre, il obéit à un besoin plus institutionnel de reconnaissance, pour aboutir à une inscription (individuelle) plus forte, avec son matériel notamment mais surtout son travail de recherche solitaire qu'il se sera approprié et qu'il aura marqué du sceau de sa singularité. Mais pour filer la métaphore, une troisième attitude alors est possible : la navigation, le déplacement sur les flots quand on a un objectif précis. Le flottement est à rattacher à la notion de sérendipité : se laisser porter par ce qui se fait, ce qui se passe.
- 27 Pour en revenir à une des questions de départ, à savoir l'articulation des espaces physiques et virtuels, on est amené à affirmer que le laboratoire virtuel se double d'un laboratoire physique. Et comme dans d'autres domaines du numérique, on en arrive à une conclusion importante : l'espace virtuel, l'équipement numérique ne remplaceront pas l'espace physique mais viendront compléter le dispositif. Ce n'est pas une alternative. Il y a là aussi à dépasser les approches clivantes. Enfin, on peut créer un nouveau laboratoire physique, mais il doit être accompagné de laboratoires virtuels qui bougent. Un mouvement d'ensemble est à trouver, dans une conception plus globale des possibilités.

Les dynamiques de changement

- 28 Si le dynamisme du mouvement des DH est un vecteur de changement, il n'est pas le seul. Du côté institutionnel, des possibilités d'action sont possibles au moment de la conception immobilière. Mais comment aujourd'hui des projets comme celui du Campus Condorcet peuvent-ils faire sauter les distinctions traditionnelles ? On redoute que le projet immobilier de l'architecte soit évalué uniquement selon des critères du type « chacun son bureau ». C'est en amont, à la phase de conception immobilière proprement dite, au niveau institutionnel même, que les choix « qualitatifs » doivent être faits.
- 29 Et on peut même aller plus loin. Prenons le cas d'un des équipements de type *learning center* conçus ces dernières années. Il est arrivé que l'architecte n'ait pas tenu compte strictement du projet institutionnel en lui-même mais plutôt privilégié le geste architectural d'inscription du *learning center* dans l'espace. Ce qui n'est pas toujours un mal. Pour que le bâtiment livré corresponde effectivement aux attentes des utilisateurs dans leur diversité, il est important de ne pas concentrer la responsabilité au seul niveau institutionnel. Doivent être précisément définis les rôles des uns et des autres,

et en particulier des groupes d'utilisateurs ou des décideurs de projet qui tiennent compte de tout un système de contraintes.

- 30 Un contre-exemple bien connu peut être cité ici. Celui de la BnF François Mitterrand. Perrault a transcrit la description administrativo-physique dans son projet de bâtiment. Aux 4 départements étanches de la BN répondent les 4 tours éloignées. Or, c'est cette inscription physique qui a été contestée une fois la BN construite. Donc dans ce débat, il ne faut pas cantonner la réflexion au seul cas des chercheurs. Les autres acteurs doivent absolument être pris en compte, pour réfléchir à ce lien entre bâtiment physique et espace virtuel. Plus loin, il ne faut pas oublier que les structurations « corporatistes » ou administratives sont plus ou moins fortes selon les métiers. Le monde des bibliothécaires et celui des chercheurs n'est pas le même à ce sujet par exemple. Mais c'est justement un des obstacles à dépasser. Commencer par mettre les gens dans un même lieu (un *open space* pourquoi pas) serait un premier pas, toutes compétences confondues. Pas à temps plein sans doute, mais un jour par semaine par exemple.
- 31 Un exemple plus enthousiasmant peut être maintenant évoqué *a contrario*, histoire de positiver.
- 32 Il s'agit du projet de Centre Georges Pompidou virtuel. Ce projet marque un vrai changement dans le rapport espaces physiques / espaces virtuels. Et ce changement a été rendu possible parce qu'il y a eu une décision forte de la part du président du centre. Ce qui nous intéresse particulièrement ici, c'est qu'il y a eu une réinvention institutionnelle au travers de la redécouverte du projet initial du Centre Georges Pompidou qui est censé agréger des centres. Il est donc possible de faire bouger les lignes, même si l'institution est toujours en retard, par principe, on pourrait dire. Elle ne s'adapte que sous la contrainte des usages, des mœurs, etc. Il est pourtant important de réaffirmer que l'institution peut être une bonne chose dès lors qu'on parvient à en sortir régulièrement, à jouer avec ses vertus et ses vices. Si on comprend l'institutionnalisation comme un besoin anthropologique ou symbolique d'ancrage, on lui accorde qu'elle rend possible l'attitude d'inscription du corps dans un territoire qui n'est pas que physique (reconnaissance de la société, des pairs...). À nous d'être plus intelligents pour dépasser les oppositions traditionnelles.
- 33 Encore un exemple. Il s'agit cette fois-ci d'un projet qui débute tout juste. C'est un Equipex et il s'appelle Biblissima (<http://www.libraria.fr/fr/blog/un-%C3%A9quipement-dexcellence-pour-le-patrimoine-%C3%A9crit-biblissima>). Son objectif est de clusteriser les fonds d'images et de textes médiévistes issus de différents entrepôts en France (voire en Europe). Pour l'instant, il s'agit simplement d'un projet d'espaces virtuels accessibles à la navigation et l'appropriation depuis un seul portail ou méta-moteur. Mais qui peut dire, quand beaucoup de médiévistes l'utiliseront, s'il n'y aura pas du mouvement, des recompositions institutionnelles, rien que sur plan symbolique dans un premier temps ? On peut l'espérer en tout cas.
- 34 En règle générale, aujourd'hui, si les jeunes chercheurs, les contractuels, ou encore les non-statutaires sont beaucoup plus présents dans l'espace numérique que les chercheurs eux-mêmes, c'est parce qu'ils n'ont pas de place dans l'espace physique qui leur soit vraiment accordée. Il y a là une sorte de compensation logique. On peut dire à ce sujet que s'inscrire dans un espace virtuel et s'y fabriquer une identité peut aussi correspondre à une logique d'ancrage du même ordre que celle de l'appropriation d'un bureau physique. Le cas de la délocalisation à Orléans de certaines équipes techniques

du CNRS, qui de fait se retrouvent séparés des chercheurs restés à Paris, est parlant quant à l'importance de l'inscription géographique des personnes, au risque dû à la « séparation des corps » qui impactent les dynamiques collectives. Il y a eu par réaction, à Orléans, une sorte de mise en place naturelle des *digital humanities*, du fait de la délocalisation physique « imposée » ou en tout cas mise en place. Un rééquilibrage des espaces de travail personnalisés ou collectifs entre les différentes catégories de personnels peut faire bouger les lignes à l'avenir. Ainsi que leur distribution sur le territoire.

- 35 D'autres exemples peuvent être donnés pour continuer de caractériser les difficultés actuelles de circulation symbolique et disciplinaire, mais aussi pour dessiner un horizon plus ouvert qu'on ne le croit. À l'INHA galerie Colbert, on a copié le feuilletage de la rue Michelet (Paris 1, Paris 4, etc.). Un étage par institution. On aurait tout aussi bien pu regrouper les départements ayant une proximité disciplinaire. La possibilité d'accéder à un étage est déterminée par le badge détenu (laissant accès à un étage, pas à d'autres). La cafétéria (espace de rencontres par définition) n'est pas accessible sans badge. La difficulté ici n'est pas qu'institutionnelle, elle est aussi technique et administrative. Chaque badge délivré par une de ces institutions repose sur une base de données, un système d'informations pas toujours interopérable avec d'autres systèmes d'informations. De plus la question de la sécurisation des espaces est centrale pour les gestionnaires de bâtiments, notamment lorsqu'il s'agit d'ERP (établissements recevant du public). On voit bien que ces contraintes techniques et administratives heurtent de plein fouet la logique de convivialité, d'échange, de « rencontre inattendue » qui devrait présider dans un lieu ouvert à la recherche. Ici le numérique peut peut-être constituer une réponse pour un système qui soit d'une part respectueux de la sécurité des biens et des personnes ET qui réponde à ce besoin d'ouverture. Très concrètement, au France, bâtiment dans lequel se trouve l'EHESS, l'EPHE, des éléments de Paris IV et du CNRS et la FMSH, le fait que l'espace de cafétéria ne soit pas un espace ERP – et que les coûts pour le transformer soient extrêmement élevés – a abouti à la même conclusion : l'espace est théoriquement fermé aux publics extérieurs, ce qui revient à construire une barrière publics internes / publics externes très peu souple et très peu respectueuse des besoins de la recherche.
- 36 Enfin, pour clore cette partie sur les dynamiques de changement, une manière de saisir la question peut consister à voir dans un premier temps que la question semble être pilotée par celle des ressources humaines. Que l'on mette des personnes avec des compétences différentes (et affamés de collaborations interdisciplinaires) dans un même lieu physique, et des pratiques innovantes et intéressantes suivront. Ensuite la question des infrastructures (équipements d'Input/Output, serveurs, espaces de collaboration) peut être résolue avec du budget. Elle n'est peut-être pas prioritaire. Enfin la question des droits d'attributions soulevée par les collaborations et le partage de matériel commun peut être résolue avec des licences et du personnel juridique.

L'urbanisation numérique et le partage des données

- 37 En écho à la question de départ de l'atelier : « quels espaces physiques pour les humanités numériques ? » peut être posée la suivante : « quelle urbanisation des espaces virtuels en intelligence avec les espaces physiques ? ». L'idée ici est de discuter d'urbanisation numérique dont le principe est de concevoir un écosystème avec une

approche d'urbaniste-architecte. Pour dépasser la contrainte que l'immobilier impose aux équipements numériques quand il faut repenser les espaces virtuels à partir de l'existant physique. On peut par exemple essayer d'établir un continuum entre les espaces, qu'ils soient physiques ou virtuels. Pour que des espaces physiques puissent être augmentés par le numérique, il faut aussi réussir à penser des espaces virtuels à partir de l'existant physique. Il y a là une contrainte des lieux physiques. L'ambition est de co-concevoir l'espace numérique et l'espace immobilier/architectural pour dépasser les jeux de contraintes mutuelles, dans un seul geste cohérent. Par exemple, n'y a-t-il pas une traduction directe de la distinction entre « espaces privés ou espaces publics » du monde physique, dans le numérique ?

- 38 Cette distinction entre espaces privés et espaces publics conduit à interroger les possibilités de partage des savoirs, des savoir-faire, des ressources, et tout particulièrement quand ces dernières sont sous forme numérique. Des porosités sont nécessaires, soit à tout moment, soit ponctuellement. À la BnF par exemple, à l'occasion du « records management » (à la fois une contrainte et une mission de la BnF), on a imposé alors une telle « porosité » pour faire sauter le bureau personnel. L'enjeu principal était surtout d'éviter la perte de mémoire, de savoir, quand un conservateur/trice part avec son espace de travail (sur son ordinateur personnel) qui contient tout son travail. Il fallait faire entrer l'espace privé dans le commun de l'institution.
- 39 Suivant les disciplines universitaires considérées, la limite entre données privées et données publiques ou collectives n'est pas la même. En recherche médicale, les carnets de recherche sont considérés comme appartenant au laboratoire. Ce n'est pas le cas en SHS, à l'instar de ceux disponibles sur Hypothèses.org. Ensuite, le travail de collecte de ce qui est produit par les chercheurs pourrait être mutualisé, beaucoup plus que ce qui se fait actuellement. La Bn de son côté anticipe la durée de vie des documents, et organise donc à l'avance les conditions d'archivage. Ce qui est, bien sûr, une façon de déterminer à l'avance ce qui va rester. Mais qu'on fasse ici de l'historiographie de la recherche, ou qu'on cherche à définir l'archive, la question reste la même : à qui appartient ce qui est produit ? Au labo ou au chercheur ? Tout cela est en plus à considérer avec un œil juridique. La propriété des données de la recherche est réglementée parfois, notamment lorsque la recherche est financée sur fonds publics. Cela conditionne donc son mode de diffusion, d'archivage, etc. On peut se référer ici aux réflexions nationales et européennes sur ces enjeux. Mais l'enjeu à ce sujet est bien de déterminer, à l'avance, dans les grandes infrastructures, ce qui va être partagé, qui va pouvoir ensuite être disponible dans des espaces collectifs (physiques ou virtuels).
- 40 Le questionnement porte donc à la fois sur les archives de la recherche mais aussi sur le travail des sources du chercheur. On peut se demander si le travail d'archivage est si différent avec ou sans le numérique dans son principe. La FMSH par exemple reçoit des dons, avec des archives très intéressantes, au sein desquelles on voit la recherche en train de se faire. Le numérique systématise et accélère le processus. Mais ne modifie pas fondamentalement le métier.
- 41 Si cet objectif de pérennisation de la recherche est un réel enjeu, il fait d'un système sans doute plus vaste dans lequel d'autres enjeux apparaissent. Par exemple dans le cas de Condorcet, aux services pour la recherche et pour la pérennisation des données est corrélée une gamme de services dédiés à la vie de campus. On désigne par là l'ensemble des usages venant enrichir le quotidien de la personne, sans relever de l'activité de la

recherche proprement dite. Là aussi la porosité est importante entre des usages souvent pensés séparément. Une urbanisation des usages est peut-être à penser.

- 42 Dans l'espace physique, on parle parfois d'espaces banalisés. Il serait bon de les envisager aussi dans l'écosystème virtuel. De tels espaces ne seraient pas pour autant une *agora*, plutôt un espace « libre pour les hackers ». Il y a un besoin de ne pas y être contraint par un programme prédéfini, ce que l'espace d'*agora* va nécessairement induire. Revues.org et Hypotheses.org sont dans cette logique d'*agora*, et ces plateformes ont bien pris justement parce qu'elles reposent sur des outils dont le fonctionnement est reconnu. Cela est nécessaire, mais cela contraint aussi. En complément, il faut aussi laisser possible un espace permettant de faire du *trash*. On peut parler de « bacs à sable » éventuellement, articulés autour d'un espace d'accès et de rencontre où l'on relaie de l'information, où l'on fait des remontées éditoriales (logique de portail éditorialisé).

Éléments de conclusion/actions

- 43 Il est proposé de continuer la discussion après le THATCamp Paris 2012. Les modalités pratiques restent à définir, mais ce qui est stimulant c'est qu'on part d'encore moins loin !
- 44 En amont d'une prochaine discussion, on peut commencer à recenser ici les projets de construction ou de réhabilitation d'espaces dédiés aux pratiques innovantes de type DH.
- 45 En plus du projet de Campus Condorcet, et du bâtiment du 54 Bd Raspail, il y a un projet à l'Université de Luxembourg.
- 46 À côté de ça, on a identifié au moins deux réseaux d'interlocuteurs possibles et qu'on pourrait essayer de mobiliser sur ce type de questions. Il s'agit d'enseignants à l'ENSCI et des organisateurs d'espaces de coworking à Clermont-Ferrand.
- 47 Enfin, pour aider à avoir un débat encore plus précis, il sera important de mieux définir les mots, les notions entendues. L'enjeu notamment est de mieux identifier les espaces, et les usages qui s'y jouent.
- 48 **Adresses des personnes intéressées pour poursuivre la discussion :**
- 49 emchateau@laposte.net
- 50 alsterin@free.fr
- 51 benoit.majerus@uni.lu
- 52 nicolas.catzaras@msh-paris.fr
- 53 clementvallois@gmail.com
- 54 sur twitter #tcp2012 #espaces-dh

Référence proposée

- 55 Sur DH Questions and Answers, vous trouverez peut-être quelques indices pour votre atelier, à l'adresse : <http://digitalhumanities.org/answers/topic/what-hardware-and-software-would-you-put-in-a-dhmultimedia-lab>

RÉSUMÉS

Notre proposition d'atelier est motivée par un besoin qu'on pourrait dire un peu plus institutionnel que pratique. Nous avons tous les deux une problématique commune qui est celle de construire ou réhabiliter des espaces physiques en fonction des nouvelles pratiques des SHS. Nous souhaitons aborder un certain nombre de problématiques en articulant les pratiques innovantes, les espaces physiques, et l'évolution des technologies. Et d'une certaine manière nous nous posons la question générale suivante : quel est l'immobilier nécessaire dans un laboratoire des humanités digitales ? Mais en première approche, nous pouvons dire que nous nous intéressons en particulier à une manière de réengager le corps tel qu'il s'inscrit dans les pratiques de recherches. Avec un peu plus de recul, la question de l'inscription du corps dans l'espace est peut-être l'élément le plus déterminant de cette problématique.

Plus précisément, la proposition d'atelier porte sur le point de rencontre entre deux mouvements actuels et l'impact produit par cette rencontre sur l'espace immobilier. Ces deux mouvements sont d'une part le développement innovant des humanités numériques et tout particulièrement de ses pratiques instrumentalisées, et d'autre part les tendances technologiques en cours et toute leur (im)prévisibilité. Comment peut-on concevoir et réhabiliter aujourd'hui des espaces de recherche scientifique en fonction de ces pratiques et de ces tendances technologiques ? Comment en retour cette conception (ou cette réhabilitation) informe ces pratiques à travers les possibilités de dialogue et de sociabilité, de manipulation et de traitement de données de recherche ? L'atelier pourra réunir des chercheurs, des architectes-designers, des informaticiens, des sociologues, etc. pour tenter de faire émerger des pistes de réflexion et d'action en lien avec cette thématique.

INDEX

Mots-clés : architecture, réseau, design, espace, institution

Les réseaux sociaux numériques de chercheurs en SHS

Proposé par Elifsu Sabuncu et Antoine Blanchard, animé par Nicolas de Lavergne et Olivier Le Deuff

Collectif

NOTE DE L'ÉDITEUR

Olivier Le Deuff est maître de conférence en sciences de l'information et de la communication, à l'université Bordeaux III et responsable de l'axe « réseau social » à l'Institut des humanités numériques.

Nicolas de Lavergne est responsable « Communication et innovation numérique » à la Fondation Maison des sciences de l'homme.

Cet atelier a finalement davantage tourné autour de questionnements, d'interrogations sur les motifs, les contours, les fonctions d'un éventuel réseau social SHS.

- 1 Les deux animateurs indiquent qu'un projet de réseau social scientifique pour les SHS est en cours d'élaboration au sein de l'Institut des humanités numériques de Bordeaux, dans un état déjà avancé (préparation d'un budget, d'un plan de développement, d'un modèle économique), cependant qu'à la Fondation Maison des sciences de l'homme, au cœur d'un réseau national et international de chercheurs en SHS, la réflexion est bien engagée sur la création d'un réseau équivalent, avec une dimension francophone dans un premier temps, mais qui devra s'ouvrir ensuite. L'atelier voudrait donc s'orienter vers l'explicitation des pratiques et des expériences concrètes des chercheurs sur les réseaux sociaux, en particulier les réseaux sociaux scientifiques, ainsi que sur les fonctionnalités nécessaires ou désirées dans un nouveau réseau social spécifiquement dédié aux SHS.
- 2 On peut partir du constat d'une profusion de réseaux sociaux numériques, et avoir pour première réaction ne pas en ajouter un de plus. Mais en analysant les inconvénients des réseaux existants (massivement anglophones, très orientés sciences dures, peu actifs comme academia, développés par des entreprises privées et avec quel modèle

économique ?), on peut se dire que, étant donné ladite profusion de réseaux sociaux numériques, on peut sans vergogne en ajouter un – ça ne fera pas une grosse différence au niveau global – qui obéit à des exigences différentes (public, gratuit, francophone – même si pas seulement –, orienté SHS)

Des réseaux sociaux scientifiques

- 3 On peut distinguer plusieurs types de réseaux sociaux scientifiques ou dont l'usage peut être scientifique : réseaux sociaux scientifiques, réseaux sociaux généralistes dont un chercheur peut faire un usage utile, applications spécifiques ayant une « couche » réseau social, mais aussi plate-forme de contenus sur lesquelles l'internaute a un profil et des interactions, mais si la plate-forme n'est pas présentée comme un réseau à première vue.
- 4 Researchgate est peu investi par les chercheurs en SHS. Il n'y a pas de transparence sur le nombre de membres et sur leur répartition disciplinaire. Deux questions se posent sur les réseaux sociaux scientifiques portés par des entreprises privées ou des start-up, qui ont levé des fonds et dont l'objectif n'est donc pas philanthropique. Quel est le modèle économique de Researchgate ou des autres réseaux de ce type ? Comment vont-ils monétiser les données et les interactions ? Se pose aussi le problème du lieu de stockage des données. Les entreposer au sein de sociétés privées pose un problème dans le cadre de la recherche publique, en termes de sécurité des données, d'accès et d'usage, de pérennité... Les réseaux de ce type sont des « Facebook de la recherche », ils miment le fonctionnement de Facebook (mise à jour d'un « statut » sur un « mur ») en ajoutant des fonctionnalités propres au monde de la recherche.
- 5 Un enjeu important est celui de la connexion du réseau à des bases de données (publications en ligne, bibliographie, archives ouvertes, ...). On apprend ici que Researchgate a pris contact avec HAL et Isidore pour étudier les modalités d'indexation de leurs données dans le réseau social.
- 6 Academia.edu ne semble pas non plus très utilisé au quotidien, même si davantage de comptes ont été créés. Ce réseau permet de suivre les publications scientifiques de collègues ou de professeurs. C'est sans doute son ergonomie qui fait qu'on y revient peu. Academia.edu avait tellement peu de visites de ses membres qu'ils ont imaginé envoyer un courriel automatique à chaque fois qu'une recherche Google aboutissait sur un profil, en indiquant via quel(s) mot(s). Academia sert plutôt à diffuser sa propre production scientifique et éventuellement à suivre celle d'autres personnes.
- 7 Google Scholar s'interroge sur l'adjonction d'une couche réseau social. Il ne s'agit pas ici d'un réseau d'interaction, mais plutôt d'opérer une identification des auteurs. L'identification des auteurs est un enjeu important pour les réseaux sociaux scientifiques ; en effet, les articles se voient dotés d'un identifiant unique, mais les chercheurs ne devraient-ils pas l'être aussi ? La question de l'identité du chercheur est primordiale : il s'agirait d'un profil unique, agrégeant les données venant de différentes sources (revues en ligne, archives ouvertes, blogs, ...). Centraliser les informations sur un chercheur amène aussi la question de la métrique, de la mesure de la citation scientifique, comme sur une plateforme comme CiteUlike. Les acteurs privés sont en train de travailler sur cette question de l'identification des chercheurs. Même si

certains peuvent être dérangés par le côté « mise en scène intellectuelle », affichage de soi.

- 8 Un réseau social scientifique doit-il être rentable ? Les acteurs privés regardent les profits des éditeurs scientifiques et espèrent capter une partie de cette manne.
- 9 Un réseau social scientifique alternatif, portés par des institutions publiques, devrait-il être, dans le contexte actuel, économiquement viable : il s'agit alors de réfléchir soit à des publicités ciblées (pour des publications, des colloques), soit à des fonctionnalités payantes, sur le principe du freemium. L'usage des fonctionnalités de base serait gratuit, mais certaines fonctions (visualisation de données, interaction) pourraient demander un abonnement, ou certains développements spécifiques pourraient être payants ou vendus aux institutions qui en ont besoin. La question de la valorisation des données stockées et liées dans le réseau social se pose aussi. Sur la question des publicités ciblées, certains pourraient préférer être fiché à leur insu sur Academia que d'avoir sur son espace dans un réseau social public, des publicités. Et se pose également la question du seuil critique : à partir de combien de membres le réseau peut s'autofinancer par la vente de certaines fonctions ?
- 10 Un autre exemple est le projet actuel du CNRS de doter les laboratoires d'un environnement collaboratif susceptible de recevoir une couche réseau social. Cependant, la raison première de ce projet est le partage collaboratif de documents. La sécurité des données, l'usage de Sharepoint, logiciel acheté à Microsoft, le problème de l'identification des membres n'appartenant pas au CNRS ni à l'université (par exemple étrangers ou professionnels), s'ajoutent au fait que les tests semblent prendre plus de temps que prévu. Peut-être ce projet sera-t-il un paquebot de plus échoué sur la plage des bonnes intentions.
- 11 La plupart des réseaux existants sont anglo-saxons et souvent dans les sciences "dures". Les SHS sont réduites à la portion congrue. Se faire une place sur une plateforme où les principaux flux d'information sont occupés par les sciences de la vie et de la nature n'est pas évident, d'où l'idée de créer un réseau spécifique aux SHS.

Les pratiques des chercheurs

- 12 Chaque chercheur a déjà des pratiques multiples sur des outils ayant une dimension de réseau social, que ce soit pour faire de la veille (Twitter, Academia), de la bibliographie partagée (Zotero, Mendeley), du partage de signets seul ou collectivement (Delicious, Diigo), de la veille et du partage (Facebook, Twitter), de la recherche de compétences ou de profils (LinkedIn, Viadeo), du partage et du débat (Wikipedia), de la diffusion de vidéo (Youtube, Dailymotion)... Il peut apparaître un certain désarroi face à la multiplication des comptes et des mots de passe. Faut-il imaginer un réseau qui comporte toutes ces fonctionnalités (un peu comme colwiz¹), ou un environnement qui intègre des comptes ouverts sur d'autres réseaux, qui permette de centraliser les informations et les modalités d'interaction dans un espace unique ?
- 13 Une problématique des réseaux sociaux est également la question du temps passé dessus, le temps du chercheur n'étant pas extensible. Beaucoup de chercheurs ne vont tout simplement pas sur les réseaux sociaux existants car ils ont le sentiment que leur fréquentation est chronophage. Un nouveau réseau social devrait donc, non pas s'ajouter à d'autres activités numériques, mais en remplacer (avantageusement)

certaines pour pouvoir trouver sa place. Le temps de l'investissement du réseau social ne peut être pris que sur un temps consacré à une autre activité. Donc il faut que le réseau social apporte un gain, fasse gagner du temps, simplifie une activité nécessaire. Cela pourrait être celle de la veille, avec par exemple un fil d'actualités qui pourrait être filtré finement selon ses intérêts ?

- 14 Une autre piste consiste à qualifier plus finement les relations entre chercheurs et entre chercheurs et articles. Dans un réseau social, on se met en lien avec des personnes de qui on est proche, à qui on ressemble. Ce serait intéressant d'avoir un réseau social qui permette de qualifier la relation à un chercheur ou entre deux chercheurs (je ne suis pas d'accord avec Untel, Untel critique Unetelle sur tel point, je trouve qu'Untel n'est pas sérieux sur le plan méthodologique...), ou par rapport à un article (je ne suis pas d'accord avec cet article, cet article est contredit par tel autre, cet article comporte un problème de logique, cet article cite tel autre article...). Il faudrait pouvoir se mettre en relation à plusieurs niveaux, chercheurs, articles etc. Dans un réseau social de chercheurs, il faut que les liens dépassent le niveau de la simple citation, mais aussi créer des débats sur ces liens. La simple métrique est trop quantitative, elle ne rend pas compte de la complexité des relations (quand tel article cite tel autre, est-ce pour s'appuyer dessus ou pour démonter ses conclusions ?).

Prédominance des besoins de veille

- 15 Les réseaux sociaux comme Twitter sont en particulier utilisés pour effectuer une veille scientifique, partagée et collaborative ou seulement consommatrice. Le problème de la qualification de l'information, de la curation est un problème central dans la profusion d'information. Chacun se bricole un écosystème informationnel qu'il est obligé de remettre sur l'établi régulièrement.
- 16 Ne pourrait-on pas imaginer un flux intelligent d'actualités, que l'on pourrait affiner et qui apprendrait à s'affiner en fonction de nos actions dessus, un flux d'actualités écrémé, qui sélectionne les résultats à présenter plus finement que les listes de diffusion auxquelles les chercheurs sont abonnés. On pourrait imaginer un moteur de suggestions, un peu comme les suggestions d'Amazon (ceux qui ont consulté cet ouvrage ont aussi consulté ceux-là, ceux qui n'ont pas aimé cet ouvrage n'ont pas aimé non plus ceux-là). Un tel réseau social pourrait alors contribuer à identifier une communauté en émergence, une rencontre interdisciplinaire, un courant ou une thématique émergente ? Pour les communautés, cela peut déjà se voir dans la confection de listes Twitter par certains acteurs. Twitter permet par exemple d'identifier des acteurs américains des DH, qui ne mettent pas les pieds en Europe mais avec qui des échanges sont possibles via ce réseau. Des groupes peuvent alors s'agréger.
- 17 Sur des réseaux sociaux scientifiques se pose aussi la question du dépôt d'articles, comme y incite par exemple academia.edu. Cela revient à dupliquer des contenus. On pourrait imaginer plutôt la suggestion automatique (à valider) d'articles à partir de bases existantes (revues, archives) et la possibilité de créer des notices, en recommandant par exemple plutôt le dépôt sur une archive ouverte (avec les contraintes mais aussi les avantages que cela comporte).
- 18 Ne pas oublier que les réseaux sociaux peuvent aussi être des sources, voire des terrains.

Besoins et activités

- 19 Ceci dit, il est difficile de prévoir les besoins des chercheurs à l'avance, sans oublier que l'outil contribue aussi à créer le besoin.
- 20 Question de la frontière entre réseau social et outils collaboratifs : un réseau social scientifique doit-il proposer aussi des outils permettant de créer des groupes et des sous-groupes, de partager des documents, de rédiger collectivement des documents tels des pads, d'échanger par chat ou vidéo,...
- 21 L'objectif est le gain de temps. Des outils collaboratifs peuvent être appréciés parce qu'ils font gagner du temps. Zotero utilisé en groupe en est un bon exemple.
- 22 On peut aussi citer un réseau comme Ravelry, un réseau autour du tricot et du crochet, qui permet le partage de trucs, de patrons... A partir d'une activité partagée, d'un réseau de pratiques de travail, on peut construire des relations, des échanges, d'autant que les frontières entre loisir et travail ont tendance à devenir poreuses, difficiles à établir².
- 23 Peut-être aussi faut-il que ce réseau ait une dimension ludique, ou une dimension esthétique, autour de la visualisation des données. Autour du réseau social pourraient se greffer des projets de recherche sur la visualisation, etc.

Périmètre

- 24 Enfin se pose la question du périmètre d'un réseau social scientifique : faudra-t-il prouver une affiliation dans une institution de recherche ou d'enseignement supérieur ? Au-delà des chercheurs et enseignants-chercheurs, les doctorants, les docteurs non titulaires, les étudiants, les retraités, les amateurs, les experts, les passionnés exerçant un métier hors de la recherche, seront-ils acceptés, et selon quels critères et procédures de validation ? S'il en faut. Le réseau peut devenir « incontrôlable » mais que veut-on contrôler ? Que doit-on contrôler ?
- 25 La question du périmètre est liée à celle du spam ou du troll, sans parler de la publicité dissimulée. Peut-on éviter ces comportements ? Comment ? Le troll n'est-il pas, dans le domaine scientifique, juste la critique ?
- 26 Elle pose aussi la question du rapport science/société, de la médiation scientifique. Qui sont les experts, les médiateurs. Est-ce que ce réseau doit être replié sur lui-même, les contenus étant réservés à ceux qui se créent un compte (contrôlé ou non), ou ouvert (en tout ou en partie) à un public plus large. Cet enjeu recoupe aussi celui de la nécessité d'attirer des membres. Si le réseau est trop fermé, il ne sera pas attractif. Où placer le curseur entre le degré d'ouverture ou de restriction à l'entrée du réseau, la pertinence de l'information qui y circule et la réduction du brouillage ?

NOTES

1. Ce réseau ne semble pas avoir rencontré le succès escompté par ses promoteurs oxfordiens en mars 2011.
 2. Olivier Le Deuff, « Réseaux de loisirs créatifs et nouveaux modes d'apprentissage », *Distances et savoirs*, 2010/4 (Vol. 8), p. 601-621.
-

RÉSUMÉS

Que ce soit Research gate, academia.edu ou des sites avec une fonction sociale (Mendeley, Zotero), les outils se diversifient. Sont-ils nécessaires ? Utiles ? Qui est vraiment demandeur ? Prennent-ils en compte les besoins des chercheurs en SHS ? Quelles fonctions seraient les plus indispensables à la communauté des SHS ?

INDEX

Mots-clés : réseau social, réseaux sociaux, pratiques de recherche, veille, identité du chercheur

Sciences participatives

Proposé par Jean-Pierre Girard

Collectif

Introduction

- 1 Jean-Pierre, qui a proposé cet atelier, participe au projet « Truelles et pixels » soutenu par le CNRS (Centre national de la recherche scientifique), l'université de Lyon 2 et la Maison de l'Orient et de la Méditerranée. Ce projet tente d'établir des rapports avec le grand public pour construire des liens entre science et société (dont la valorisation des savoirs scolaires vis-à-vis du public des collègues) à travers des contenus et services numériques ludiques et interactifs (par exemple des documentaires Internet ou un « simulateur d'archéologie » multi-supports).
- 2 L'objectif est de déconstruire la mythologie de l'archéologue (alimentée par l'imagerie cinématographique dans la lignée du personnage d'Indiana Jones) pour montrer la réalité de son travail scientifique, qui regroupe une grande diversité de métiers, de disciplines et d'outils.
- 3 Il poursuit avec la présentation de plusieurs textes et références sur les sites collaboratifs dont l'article de Pierre Merckle, « Vers des sciences sociales citoyennes ? » Dans cet article, l'auteur évoque la notion d'« interprétation partagée » et pose une question : « Dans quelle mesure le public lui-même peut-il aider les scientifiques en apportant des compétences que les scientifiques n'ont pas ? »
- 4 Jean-Pierre cite aussi les travaux de Paola Tubaro et d'Antonio Casilli sur les émeutes de Londres de 2010, présentés par Pierre Mounier (Homo numéricus, 28 août-9 septembre 2011). Ces travaux ont montré comment « l'usage de méthodes computationnelles peut aider à réconcilier recherche et demande sociale (à condition de ne pas verser dans le fétichisme de l'outil informatique). »
- 5 Ensuite il présente rapidement plusieurs sites construits sur les logiques participatives ou collaboratives :
 - Site Babelio : les lecteurs sont encouragés à émettre des critiques des livres qu'ils lisent, non pas en concurrence de la critique professionnelle mais en complément de celle-ci. Le site

invite également à qualifier les lectures, créant ainsi une « folksonomie » différente des mots-clés du catalogage utilisés par la Bibliothèque nationale de France (BNF) et la base professionnelle Electre, ce qui ouvre de nouvelles possibilités de choix de lecture ;

- Site ArchéoProvence : le site propose à des archéologues amateurs de contribuer à un inventaire de leurs découvertes, notamment à travers une carte archéologique participative. Mais dans ce cas il existe un enjeu légal : la pratique de l'archéologie en France est très encadrée par des textes de lois (cf. *Code du Patrimoine*). Seules des personnes compétentes et dûment autorisées par l'État peuvent effectuer des fouilles archéologiques. Le risque de ce type de site collaboratif est qu'il puisse alimenter, quelles que soient les intentions des participants, des pratiques illégales et des trafics de vestiges archéologiques, qui se traduisent parfois par de véritables pillages de sites. C'est pourquoi la carte archéologique de la France (officielle) n'est, quant à elle, pas publique¹ ;
 - Le site Artefacts propose de collecter des informations sur les petits objets archéologiques recueillis par tout un chacun. Ce type de prospection est peu, voire pas du tout pris en compte dans les inventaires et les bases de données officielles : l'étude du corpus ainsi généré fait polémique au sein de la recherche ;
 - Le site Archéozoo propose une plate-forme collaborative permettant aux chercheurs en archéozoologie de décrire les types de restes osseux d'animaux sur lesquels ils travaillent ;
 - Le site du National Geographic propose d'utiliser les images de *Google Earth* pour que les internautes identifient sur les images satellitaires des formes susceptibles d'être des sites archéologiques ;
 - Site des Archives départementales de la Vienne met en ligne un outil d'indexation collective pour enrichir l'indexation des recensements de la Vienne.
- 6 Les sites collaboratifs de généalogie se multiplient. Même si la généalogie n'est pas une discipline scientifique, ces sites démontrent les possibilités de collaboration en ligne massive dont certaines disciplines scientifiques pourraient peut-être s'inspirer.
- 7 D'autres sites collaboratifs à visée scientifique sont mentionnés dans le domaine des sciences de la nature :
- L'observatoire des escargots ;
 - Missions Printemps : une expérience de dénombrement massif des espèces naturelles (faune et flore). Chaque mesure est localisée en vue d'analyse comparée avec les évolutions climatologiques ;
 - Le programme européen « Urbanbees » : à Lyon et dans la communauté urbaine des « nichoirs » à abeilles ont été installés pour pouvoir observer à moyen terme les populations ainsi fixées. Les internautes sont par exemple invités à poster leurs photos d'abeilles, de nidification etc. et de participer ainsi à une exposition en ligne.
- 8 Plusieurs notions sont abordées pendant les échanges avec les participants présents dans la salle : *crowdsourcing*, analyse encadrée des données, enrichissement collectif, taxonomies collectives, indexation collaborative, *big data*.
- 9 L'intérêt majeur des sites collaboratifs, à caractère scientifique ou lancés par des communautés scientifiques, est de démultiplier la capacité d'une communauté scientifique à aborder et à alimenter un corpus, grâce à la participation d'un public d'amateurs. La réussite des sites collaboratifs dépend donc de la capacité des réseaux à récupérer durablement des informations utiles à une communauté. Il conviendrait d'explorer les marges du concept de « site collaboratif », d'identifier les initiatives qui fonctionnent et celles qui échouent, et de tenter de comprendre les raisons des succès et celles des échecs de ce genre d'entreprise.

Tour de table et discussions

- 10 Après cette riche introduction, chaque participant de l'atelier se présente en quelques mots et évoque ses questionnements autour du thème des sciences participatives, et éventuellement son expérience dans ce domaine. Des points de vue différents se dégagent de la discussion, issus de corps de métier multiples.
- 11 Les sciences participatives intéressent le doctorant en médecine qui assiste à l'atelier pour découvrir des expériences de sites collaboratifs dans les sciences humaines et sociales. Il s'interroge sur les modalités d'appropriation par les publics de sites collaboratifs (il travaille sur un projet de dossier médical en ligne). Il évoque la notion d'« *empowerment* » (engagement) du citoyen.
- 12 Pour le domaine de l'édition et de la traduction des langues (langues étrangères vers le français), l'application des procédures collaboratives peut s'avérer intéressante. Les pads (prises de notes collaboratives) réalisés pendant les séances du ThatCamp sont un exemple de travail participatif réalisé par la communauté pour elle-même ou utilisable par d'autres communautés. Dans des projets de traduction participative, il peut être utile de mettre en œuvre le système du *wiki*.
- 13 Le journaliste scientifique s'intéresse notamment au *data journalism*, aux données qui peuvent être recueillies de manière collaborative sur des sujets pouvant intéresser les journalistes scientifiques. Il cite plusieurs exemples de ce type d'initiatives :
 - Le site Prix de l'eau ;
 - Le site Mémoire d'Algérie . Des journalistes ont exploité des données provenant d'archives non publiées, habituellement difficiles à récupérer pour des historiens. Ce site comporte une cartographie et une frise chronologie (*timeline*) permettant de visualiser les données recueillies ;
 - Le site Journalism ++, fondés par d'anciens membres d'OWNI.
- 14 Le mouvement de l'*open access* qui permet la circulation des données pour diverses communautés, intéresse particulièrement le journaliste.
- 15 Une interrogation sur la question de la validation des données publiées sur des sites collaboratifs émerge dans la discussion.
- 16 Les personnes impliquées dans des programmes de recherche, par exemple des ANR, s'interrogent sur la façon d'adapter les outils des chercheurs à des populations de non-chercheurs. Des problèmes d'ordre méthodologique et de compétences scientifiques se posent. Il faudrait dépasser les craintes et appréhensions entre les communautés, pour pouvoir aller plus loin dans ces collaborations entre scientifiques et non-scientifiques.
- 17 Ces liens entre population scientifique et non-scientifique intéressent également les personnes travaillant dans le domaine de l'édition. Dans ce contexte on peut évoquer un projet de redocumentalisation comme le projet de partage de photos du débarquement de Normandie lors de la Seconde guerre mondiale qui consiste en une gestion collaborative de métadonnées effectuée conjointement par des spécialistes et des non-spécialistes. PhotosNormandie est un projet collaboratif portant sur 3 044 photos historiques de la bataille de Normandie (6 juin à fin août 1944). Il a pour but d'améliorer (corriger, compléter, voire créer en identifiant les personnes) les légendes

des photos. Depuis le 29 janvier 2007, plus de 7 100 descriptions de photos ont été complétées et corrigées (certaines ont été mises à jour plusieurs fois).

- 18 Un questionnement sur la notion de science citoyenne, qui ne paraît pas toujours adaptée et qui peut, parfois, se traduire par des erreurs dans l'évaluation des contributions des non-scientifiques émerge dans la discussion. À titre d'exemple, en astronomie, de nombreuses « découvertes » sont déclarées par des amateurs, mais qui se révèlent être déjà connues des scientifiques. Certaines expériences de transcription de manuscrits réalisées (dans certaines limites) par le grand public n'apportent pas toujours les résultats escomptés par les initiateurs de projets. Une nouvelle qualification des phénomènes semble nécessaire : on devrait parler de « science citoyenne » et de « science participative » (qui semble préférable), car ces termes ne sont pas équivalents.
- 19 Une autre question concerne la façon dont on peut connaître ou faire connaître ce genre de projets. La création d'un nouvel outil qui inventorie les sites et les projets collaboratifs, notamment en sciences humaines et sociales, semblerait nécessaire à l'heure actuelle.
- 20 Le point de vue de l'ingénieur de recherche en archéologie vient nourrir la discussion. Des données archéologiques descriptives qui sont les données issues des carnets de terrain des archéologues qui deviennent, de plus en plus, des carnets numériques sont évoquées. Aujourd'hui l'utilisation de tablettes est une nouvelle étape dans le développement de ce processus. Ces dispositifs, certes innovants, posent des problèmes de méthodes, car ils introduisent des changements dans le rapport au terrain et à sa transcription. Pour certains archéologues, l'utilisation de tablettes numériques peut se traduire par des risques de perte d'information : pérennité des données numériques, mais aussi perte de « sensibilité » due à la normalisation induite par les interfaces d'enregistrement (ce qui amène à envisager de faire coexister deux modes d'enregistrement des données du terrain : le mode traditionnel sur support papier et le mode numérique).
- 21 Les projets de cartographie collaborative comme OpenStreetMap méritent également l'attention. Il s'agit d'un projet qui se présente comme une alternative à Google Map, opérée par les internautes à travers une interface normalisée (cf. *infra*).
- 22 L'un des intervenants qui collabore au Centre virtuel de la connaissance sur l'Europe, est engagé dans divers projets participatifs. Il a commencé par s'intéresser à SecondLife qui, pour lui, est une sorte de plate-forme collaborative. Il a suivi le projet *International Spaceflight Museum* qui produit un inventaire d'objets en relation avec le domaine spatial. Des projets participatifs, comme le projet de *Virtual Museum*, soulèvent la question de l'identité des gens qui contribuent à de tels projets, mais aussi de leur motivation. Une question éthique émergente dans ces situations collaboratives est mise en avant. Il s'agit souvent de professionnels qui, dans leurs domaines respectifs, font le choix de travailler volontairement pour la collectivité.
- 23 Une conceptrice de projets multimédia, s'interroge sur la façon d'intégrer la dimension participative à partir de l'exemple de deux projets :
- 24 1. Le site « Musette-nature » conduit avec le parc naturel régional des volcans d'Auvergne, qui a permis la rédaction de fiches découvertes utilisées sur le terrain par des publics scolaires (enseignants et élèves, en lien avec le programme en sciences de la vie et de la terre), mais aussi par le grand public, dans le cadre de découvertes

touristiques. Une plate-forme d'échange, pour évaluer les connaissances acquises sur le site, a été mise en place. Son objectif est d'évaluer la capacité de passer du contenu d'un site à une observation sur le terrain. L'évaluation s'appuie sur des questionnaires destinés à des enseignants. Le problème est que la plate-forme s'adresse en réalité à des publics très différents. Le retour d'expérience a donc imposé au site d'évoluer. Un participant exprime l'importance du rôle des modérateurs pour ce type de site, et la difficulté pour un projet hybride (tourisme et scolaire) à satisfaire deux publics (enseignants et touristes).

- 25 2. Le centre d'interprétation du temple de Mercure (deuxième siècle après J.-C.) situé au sommet du puy de Dôme. Il s'agit d'un site archéologique qui fut longtemps peu valorisé. Un centre d'interprétation a été créé avec des archéologues, pour montrer comment le temple a été construit. Des problèmes de représentation du temple (pas de consensus scientifique) se sont posés. De là est née la question « Peut-on proposer aux visiteurs de suggérer leur propre interprétation du temple ? » Autrement dit : « La contribution des visiteurs peut-elle venir en complément (à la place ?) des interprétations des scientifiques ? », et, si oui, comment et à quelles conditions ?
- 26 Le projet *OpenStreetMap* (OSM) apparaît dans la discussion car il pourrait être une alternative à d'autres projets, comme celui de Google qui abuserait selon certains de son rôle.
- 27 Devant cet exemple et d'autres évoqués précédemment, des questions portant sur les compétences nécessaires et les contraintes de temps auxquelles les acteurs impliqués doivent faire face pour produire et valider des données surgissent. Les projets existants sont variés et les perspectives et les méthodes mises en œuvre sont variées. Du point de vue de la méthode, si l'on parle de cartographie, on constate une prépondérance des thèmes liés à la circulation automobile. À l'inverse, les *items* permettant la description de l'environnement se sont révélés d'une grande pauvreté.
- 28 La variété des projets peut être illustrée par des sites comme : <http://www.spipoll.org/> (dénombrement d'insectes) dont les consignes méthodologiques sont d'une grande complexité (contraintes de formats et de taille des photos à déposer sur le site).
- 29 Face à tous ces projets, cependant, il est regrettable de constater le manque d'échanges humains dans l'apprentissage de l'utilisation des outils collaboratifs.
- 30 D'autres intervenants s'intéressent aux possibilités de faire découvrir au grand public les travaux des sciences humaines et sociales par le biais de ce genre de sites.
- 31 La question de la négociation entre ceux qui recueillent les données et ceux qui sont en mesure de les exploiter, notamment pour des besoins d'information environnementale et de santé est évoquée. Une négociation semble nécessaire entre le « donneur d'ordres » et le participant sur le protocole de recueil des données, mais aussi sur la propriété et la diffusion ultérieure des données rassemblées/apportées à la collectivité.
- 32 La dernière intervenante du tour de table travaille en ethnomusicologie au CNRS (Centre de Recherche en Ethnomusicologie). Elle a participé à une plate-forme de diffusion d'archives sonores des chercheurs, *Telemeta*. Il en ressort un besoin de contribution des populations pour décrire et enrichir les archives, car les porteurs de ce projet n'ont pas la capacité pour traiter les volumes de données recueillies par ce projet (manque de temps, de personnel et de compétences). Un wiki a donc été mis en place pour ce projet : <http://www.telemeta.org/>. Sur ce site, il est possible de signer les contributions, de les annoter et de les identifier. Les membres de ce projet réfléchissent

à la réalisation de blogs et d'autres outils d'échange pour les utilisateurs (forum, messageries).

- 33 En poursuivant la discussion, on cite les études faites sur les tendances de recherche des internautes avec GoogleTrends et de la capacité prédictive (en suivi d'épidémie, par exemple) d'une analyse de ces *big data*. Il propose de s'inspirer de ce qui est fait par les sociétés privées présentes sur l'Internet dans un but commercial, pour l'appliquer dans le cadre des recherches scientifiques sans but commercial.
- 34 L'utilisation du travail collaboratif de manière implicite est flagrante dans l'utilisation faire des codes CAPTCHA. Il s'agit de lettres en format image que les internautes retranscrivent en caractères alphanumériques pour distinguer des usages volontaires de pages de sites Internet entre des humains et des robots. Le système de reCAPTCHA, mis au point par des chercheurs de l'université Carnegie-Mellon, appartient à Google depuis fin 2009 et contribue à déboguer la transcriptions OCR des livres numérisés par le programme Google Books.
- 35 La question qui se pose est celle de l'information des utilisateurs concernant l'utilisation de ces données qui proviennent de leur usage d'internet.
- 36 D'où la question : dans quel contexte juridique, éthique, etc., les données des sites participatifs, fussent-ils scientifiques et publics, sont-elles exploitées ?
- 37 La question éthique qui émerge des méthodes et des stratégies différentes susceptibles d'être utilisées pour arriver à des résultats scientifiques est inévitablement posée.
- 38 À ce propos, on évoque un projet NutriNet-Santé, projet de recherche épidémiologique (« santé participative ») pour lequel les internautes sont invités à déclarer les médicaments qu'ils prennent.
- 39 On revient sur la question des intentions et des responsabilités de ceux qui participent à un ensemble collaboratif. Le travail effectué par Thierry Joliveau sur son blog à propos de la carte publiée par Facebook en décembre 2010 sur les membres connectés est cité : <http://mondegeonumerique.wordpress.com/2011/01/10/500-millions-damis-la-carte-de-facebook-1-deconstruction/>
- 40 Le travail de déconstruction-analyse-étude de l'évolution effectué par T. Joliveau sur les représentations cartographiques publiées par Facebook montre comment des images produites à partir des données de connexion sont instrumentalisées pour créer ce qui est censé être une donnée objective, dans le but de démontrer une omniprésence de Facebook dans le monde.
- 41 En conclusion de ce tour d'horizon, au cours duquel les sujets abordés ont été très variés, les projets collaboratifs apparaissent comme des sortes de « capteurs sociaux », ce qui soulève des questions parmi lesquelles celle-ci : quelle est la capacité explicative du chercheur face aux données déposées par les internautes ? De même, la question de la propriété des données et de leur partage ainsi que la responsabilité associée, reste entièrement posée. Enfin, il faut peut-être accepter un niveau d'incertitude sur le devenir des données, lorsqu'on s'engage dans des projets participatifs.

Pour conclure : quelques pistes de réflexion

- 42 Ces questions peuvent être synthétisées en quelques grands thèmes, dont certains recourent les idées émises lors de la non-conférence d'ouverture de ce THATCamp :

- 43 **Procédures** : On doit souligner l'importance et la complexité des procédures, et montrer le risque (généralement implicite mais masqué) de simplification des données. Ce qui rejoint l'une des principales interrogations de la non-conférence : qu'est-ce donc cette donnée que les *digital humanities* se proposent de traiter ?
- 44 **Données** : Elles sont à interroger sous tous les angles : qualité des contributeurs, qualité de l'interface, degré de compétence scientifique implicite requise par le projet, homogénéisation ou, pire, surdétermination par l'outil des contributions recueillies... Toutes ces questions doivent être posées en amont, par le concepteur de l'expérience, et les réponses réinterrogées en aval, par les analystes du corpus recueilli.
- 45 **Appropriation par le public** : Elle est indispensable parce qu'elle conditionne le sentiment de responsabilité des utilisateurs/internautes participants à ces projets collaboratifs. Il importe donc de fonder leur motivation sur un socle solide. On doit s'interroger sur cette motivation : pourquoi participer ? Qu'attend l'internaute de sa participation à « ce » projet collaboratif, à court terme, à moyen terme ? Quel retour lui offrir ?
- 46 **Science ou médiation ?** Peut-être est-ce une partie de la réponse à la question précédente. Si on considère l'internaute comme une ressource corvéable et gratuite, éventuellement récompensée par le sentiment de servir la science, la transmission de savoir/savoir-faire, voire la diffusion en retour ou le partage de l'accès à des informations l'enrichissant personnellement apparaît important.
- 47 **Déontologie** L'internaute devenant un participant à part entière du projet scientifique, il doit être traité comme tel : transparence des intentions et des demandes du « donneur d'ordres », transparence aussi du traitement qui sera fait, ou peut être fait ensuite, des données recueillies, enfin clarté du cadre juridique du projet, ce qui inclut les responsabilités respectives des uns et des autres, mais aussi les questions de propriété, de droit d'usage, des données (les données individuellement transmises par l'internaute, le corpus global).
- 48 *In fine*, le consensus se fait autour l'idée de la mise en place d'un outil de recensement des expériences scientifiques participatives... qui devrait lui-même être un outil participatif. Dans ce domaine, on évoque le site Europeana et les projets d'inventaire de projets collaboratifs. Il s'agira donc de vérifier si un tel site n'existe pas déjà.
- 49 Dans l'immédiat, une suite sera donnée à cet atelier par l'ouverture d'un carnet de recherches sur *Hypothèses.org* consacré aux expériences participatives.
-

NOTES

1. Pour plus de détails cf. http://www.culture.gouv.fr/picardie/les_services/definitions/archeo/carte_archeo.html

RÉSUMÉS

Une proposition de transfert de non-savoirs faire ! Un programme non pas de recherche mais de valorisation de l'archéologie. Nombreuses pratiques interdites au public. Plusieurs expérimentations lancées ou identifiées. L'objectif est de partir de cet état des lieux pour travailler ensemble au problème que pose une discipline très très technique, très très terrain et à ses limites et comment peut étendre la réflexion à d'autres disciplines.

INDEX

Mots-clés : science participative, sites collaboratifs, archéologie, sciences humaines

Plate-formes participatives

Proposé par Marie-Madeleine Mervant-Roux, ARIAS

Collectif

Présentation

- 1 Comment favoriser l'exploitation optimale d'un outil numérique par des chercheurs, au-delà de l'aspect pratique que ceux-ci peuvent spontanément lui trouver dans la perspective qui est déjà la leur, fondée sur l'usage d'outils traditionnels ?
- 2 Autrement dit, comment conduire les chercheurs à utiliser l'outil en tant que nouveau média permettant au groupe de recherche de changer de perspective et finalement de modifier le projet lui-même, de se laisser faire par l'outil et retravailler à partir de ce que l'on découvre de lui ?
- 3 Pour cela une interaction d'un certain type est nécessaire entre ceux qui connaissent l'outil et ceux qui ont conçu le projet de recherche, très différente d'une simple prise en main technique. C'est la question de la relation ingénieur-chercheur évoquée dans d'autres ateliers, et celle de la formation de l'un et l'autre qui sont en jeu.

Un cas

- 4 L'exemple présenté est celui d'une recherche inscrite dans un Labex (réunissant 13 laboratoires) dont la problématique centrale est celle du transfert culturel, avec un axe « humanités numériques ». Ces laboratoires sont invités à s'intéresser à ces nouvelles pratiques (recherche et transmission).
- 5 Dans ce cadre, une plate-forme participative (hébergée par Adonis) a été élaborée par des chercheurs du laboratoire ARIAS. Elle est destinée à accueillir les chantiers de plusieurs glossaires internationaux conçus pour présenter la circulation des mots et des concepts (traductions, translations) dans un champ et une période donnés. Un des glossaires (« Le son du théâtre : mots et concepts ») accueille les termes les plus importants pour décrire les dimensions sonores du théâtre (XIX^e-XXI^e siècle). Il

comporte une vingtaine d'entrées et circule entre cinq langues. Chacune des entrées est conçue comme une sorte de rubrique accueillant plusieurs contributions d'auteurs différents qui se répondent éventuellement les uns les autres.

- 6 La plate-forme participative constitue un chantier unique : il n'y a pas d'écriture collective de textes, mais chaque rédacteur peut accéder à tout ce qui s'écrit, à ce qui est fait dans l'entrée d'à côté. Chacun peut lire, commenter, écrire en regard des textes des autres (entre autres grâce à deux forums internes).
- 7 Pour l'instant le chantier n'est pas public. Il est prévu de le publier en ligne et en libre accès dans trois ans, lorsque suffisamment de travail aura été accumulé.
- 8 Étant donné la méthode de travail choisie, les avantages de la plate-forme semblent évidents : possibilité d'interagir entre contributeurs pendant toute l'élaboration du glossaire, d'insérer des images et surtout des sons (archives audio du groupe, ou liens vers des sites fiables possédant des archives audio comme Gallica), de créer des liens internes au glossaire. C'est le logiciel Spip 3 qui a été choisi, car il permet d'adapter l'interface de travail aux besoins des contributeurs.

SPIP est un système de publication pour l'Internet qui s'attache particulièrement au fonctionnement collectif, au multilinguisme et à la facilité d'emploi. C'est un logiciel libre, distribué sous la licence GNU/GPL. Il peut ainsi être utilisé pour tout site Internet, qu'il soit associatif ou institutionnel, personnel ou marchand. SPIP est développé (programmé, documenté, traduit, etc.) et utilisé par une communauté de personnes que chacun est invité à rejoindre (ou simplement à contacter) sur différents sites Web, listes de discussion par email et rencontres (les fameux « Apéros-SPIP »). Le programme est né en 2001 d'une initiative du minirézo, un collectif défendant le Web indépendant et la liberté d'expression sur Internet. Il est actuellement utilisé sur des dizaines de milliers de sites très divers. <http://www.spip.net>



La présentation de l'outil

- 9 Pendant la conception scientifique du projet, les présentations techniques de la plate-forme aux chercheurs ont eu des effets sur la conception de l'écriture du glossaire, de son espace et finalement de sa structure, au-delà des bénéfices attendus de ce mode de travail : d'où le constat que ce moment d'information technique est décisif : ce que le chercheur doit savoir au niveau technique est déterminant dans son usage plus ou moins innovant de l'outil.
- 10 Ce moment où se fait le dialogue entre le savoir technique et le projet devrait être prolongé. Il correspond à l'idée de « non-laboratoire » proposé en introduction du

THATCamp Paris 2012 par Paul Bertrand, c'est-à-dire un espace de réflexion, pour le plaisir.

- 11 L'outil se met en effet à jouer un rôle actif : comment exploiter ce mode d'écriture multimédia pour présenter par exemple la notion de « résonance » ? pour rendre une part du glossaire audible, plutôt que lisible, ou les deux (il s'agit de théâtre, de vocalisation) ? Comment exploiter les hyperliens pour inscrire concrètement les questions de « transferts » dans la lecture du glossaire ? Le fait d'avoir ce projet de plate-forme collaborative a attiré de nouveaux chercheurs et suscité un nouvel intérêt chez d'autres. Expérimenter cette relation à l'outil s'inscrit dans l'exploration d'une question posée par le Labex : comment renouveler les pratiques de recherche, d'édition, de publication, de transmission ?

Comment le projet peut-il se transformer avec l'outil ?

- 12 Il semble nécessaire d'organiser un temps de redéfinition du projet, entre ingénieurs ou bons connaisseurs du numérique d'un côté, et chercheurs de l'autre. Mais comment faire ? Où trouver, comment rétribuer et comment former les personnes capables d'établir ce dialogue ? De la même manière, il peut être intéressant de chercher à documenter le processus de recherche lui-même, cette transformation des pratiques à l'œuvre. Le format carnet de recherche semble bien adapté à cet objectif et permettrait à tous de bénéficier du retour d'expérience.
- 13 Un aspect important du projet est le design de l'interface. L'objectif du glossaire est de travailler sur les concepts et sur le vocabulaire relatifs aux sons. L'interface doit permettre de penser différemment ce lien. L'équipe a repéré des sites Web innovants pour ce qui concerne la visualisation de l'information dans le champ sonore :
- 14 Les laboratoires d'Aubervilliers – <http://www.leslaboratoires.org/>
- 15 Le site Archipel, site belge de musique contemporaine – <http://www.archipels.be/>

Archipel

- 16 Dans un autre domaine, un projet porté par l'école d'Art de Valence partage les mêmes problématiques : trois journées ont été organisées cette année sur la contribution du design aux pratiques de lecture dans l'environnement numérique. De ces journées sont nés un projet de publication imprimée et un projet d'édition numérique. Les chercheurs ont tenté de trouver des outils existants sans trouver celui qui correspondait à leurs besoins. Ils ont donc eu recours à un développement *ad hoc*.
- 17 Ce choix, bien compréhensible, n'est pas sans poser quelques questions malgré tout ; en particulier en ce qui concerne la pérennité de l'accès à l'information une fois publiée. L'interface de visualisation est conçue en fonction d'écrans et de systèmes de navigation tels que nous les connaissons aujourd'hui. Mais qu'en sera-t-il dans les années à venir ? Il n'est même pas certain que nous continuerons à utiliser des ordinateurs personnels pour y accéder.

L'ouverture de la plate-forme

- 18 Dans l'exemple de Valence, la question s'est posée du degré d'ouverture à la contribution : qui peut contribuer au projet ? On peut craindre qu'une ouverture totale conduise à un afflux de contributions non pertinentes difficiles à modérer. Pourtant une étudiante dans le cadre de son diplôme a décidé d'ouvrir un espace infiniment extensible où les gens qui voudront contribuer peuvent ajouter des objets. Cet espace est ouvert à tous, à la condition qu'ils acceptent de s'inscrire. Il s'agit de créer une communauté.
- 19 C'est un peu la même politique qui a été adoptée pour le projet d'archives sonores Telemeta : l'accès à certaines informations est libre, mais il faut appartenir au groupe de chercheurs autour du projet pour aller plus loin.
- 20 En réalité, la question de la modération des contributions est en partie déterminée par le cadrage qu'impliquent la mise en page ou la qualité des textes déjà présents sur la plate-forme. Ces éléments suscitent ou non des contributions pertinentes.

La question du temps

- 21 Finalement, opérer un retour réflexif sur ces outils collaboratifs et leur mise en œuvre pour la réalisation de projets de recherche et de publication, c'est être confronté à la difficile question de la gestion des temporalités.
- 22 Pour les plate-formes, se pose la question de la pérennité : si elles sont liées à un projet de type Labex, au bout d'un moment il n'existe plus de financement pour permettre l'entretien du site, or, si le site permet des commentaires, il faut des modérateurs. La question de la technologie choisie est elle aussi importante, car il faut faire évoluer les outils (exemple : pour que les travaux soient lisibles sur tablette), mais cela nécessite aussi des moyens permanents. Sans financement prenant le relais, comment continuer à assurer une actualité éditoriale et un développement au projet ?
- 23 Inversement : à partir de quand faut-il publier les premiers résultats du projet ? L'équipe du glossaire avait d'abord pensé à un délai de trois ans. Pourquoi cette durée ?

Essentiellement parce que cela correspond à la durée de préparation de ce qui serait son équivalent imprimé.

- 24 Le problème est que les temporalités ont changé. Le nouvel environnement numérique conduit à rendre public le travail en cours dès son origine, sans attendre qu'il atteigne un point de maturité difficile à définir. Le numérique implique une continuité documentaire différente de l'imprimé. Il s'agit d'une évolution très particulière pour les sciences humaines où l'on a l'habitude de ne publier que des travaux extrêmement achevés.
- 25 Cette question a également un impact important sur la motivation des contributeurs. De nombreux projets n'ont jamais été nourris par d'autres que par la personne qui les a créés. Il est donc toujours nécessaire de penser à la motivation, dans la durée, des personnes que l'on souhaite voir contribuer. La facilité d'utilisation de l'outil et sa nouveauté suscitent souvent beaucoup d'enthousiasme au début d'un projet, mais d'autres motivations doivent ensuite prendre le relais sous peine de voir celui-ci retomber assez rapidement. Il faut toujours penser à « rémunérer » les contributeurs et cette « rémunération » peut prendre plusieurs formes. Dans le cas d'une publication numérique, la visibilité qu'offre la diffusion en ligne est importante, et c'est elle qui est susceptible de motiver les contributeurs. Dans ce cas, l'imposition d'un délai de trois années avant publication semble assez artificielle et peut-être contre-productive. Il paraît en ce cas nécessaire pour les contributeurs du glossaire de décider ensemble de maintenir le chantier « fermé », et de proposer en parallèle un carnet de recherche ouvert, ou de graduer l'ouverture du chantier, ce que permet SPIP.

RÉSUMÉS

L'atelier auquel j'espérais pouvoir m'inscrire, ayant en ce domaine plus de questions que de réponses, serait un atelier consacré au dispositif appelé « plate-forme participative », à ses usages effectifs par les chercheurs en SHS et plus précisément à ses usages spécifiques, c'est-à-dire ceux qui exploiteraient systématiquement les caractères nouveaux de ce nouveau média (on en fait beaucoup d'usages « à l'ancienne »), usages qui comporteraient éventuellement cette expérimentation dans leurs projets eux-mêmes. Dans la continuité de l'atelier « Outils collaboratifs » du THATCamp 2010, dont on peut lire le CR, il s'agirait de confronter les discours, souvent à la fois utopiques et limités, aux pratiques des membres de l'atelier, qui pourraient s'avérer plus modestes en rhétorique et plus inventifs en théorie. Élaborant dans notre laboratoire un glossaire (historique et conceptuel) international, multilingue, en ligne, à l'aide d'une plate-forme, mettant l'accent sur les questions de traduction et translation, nous sommes particulièrement intéressés par certaines questions. Quelles nouvelles formes de chantier pour ce type de recherche ? Quel gain scientifique et intellectuel attendre de ce que l'outil peut produire de lui-même, et comment ne pas passer à côté de ces pistes ? Les temps de présentation technique en particulier semblent décisifs. Sans doute des participants plus expérimentés pourraient-ils animer un tel atelier ?

INDEX

Mots-clés : plate-forme, outils collaboratifs, communauté, laboratoire

Quelles coopérations d'acteurs dans une politique d'édition numérique pour un laboratoire en SHS ?

Dorian Ryser et Jean-Pierre Masse

Collectif

L'exemple d'un laboratoire en SHS

- 1 Le point de vue de départ est celui des chercheurs du Centre d'études et de recherches internationales (CERI) un laboratoire Sciences-Po/CNRS, qui présentent leur laboratoire et les questions particulières qui émergent de leur travail. Il s'agit ici de faire participer tout le monde afin d'introduire une problématique, de lancer des pistes de réflexion et de faire surgir de nouvelles idées ou des solutions, dans le cadre d'un échange autour d'expériences différentes.
- 2 Le CERI publie aussi bien sous forme « papier » qu'« en ligne ». Depuis quelque temps, l'injonction lui a été donnée de passer au numérique, sans pour autant lui allouer plus de moyens et lui donner des indications sur « comment faire », même si, depuis trois à quatre ans, il « décolle » littéralement en termes de formations et compétences. D'autres se trouvent dans des situations similaires. Le CERI bénéficie de ressources humaines importantes (documentalistes, informaticiens développeurs) même si les moyens financiers mis à sa disposition sont insuffisants. Les membres du laboratoire ont la chance de travailler avec des équipes extérieures mais intégrées à Sciences-Po, comme avec l'atelier de cartographie.
- 3 Dans ce contexte de transformation, les chercheurs ont dû apprendre à travailler ensemble et à mettre en commun leurs compétences. Au départ, il n'y a pas eu d'encadrement. La direction du CERI a manifesté sa volonté de développer des ressources numériques mais n'a cependant pas mené de vraie réflexion autour de la question. Au début de tous ces changements, il y avait deux collections « papier » mises

en ligne et, depuis quatre ans, de nouveaux projets sont apparus avec des lignes éditoriales plus claires.

- 4 Le rôle joué par les auteurs et les chercheurs dans cette dynamique est un point important. Ils constatent la persistance d'une situation assez paradoxale : « nous avons tous une injonction forte de passer au numérique, mais au niveau institutionnel, dans la gestion de carrière des enseignants-chercheurs par exemple, le numérique ne bénéficie d'aucune reconnaissance ». D'un point de vue institutionnel, une publication numérique a « moins de valeur » qu'une publication papier.
- 5 Devant ce constat, il est important de se demander comment, avec cette demande forte de passer au numérique, on peut convaincre les auteurs, généralement des chercheurs, de participer à nos activités de publication numérique, alors que ça ne leur « rapporte rien » ; comment faire évoluer la situation ? S'agit-il d'une problématique qui doit être discutée directement avec les chercheurs ou à résoudre en interne ?
- 6 Depuis quelques années, nous constatons des évolutions dans les positions des chercheurs. La collection « Working papers » dirigée par Nicolas de Lavergne en est un exemple. Il s'agit du premier « jet » des travaux de recherche, publié en format numérique sur le site de la Fondation de la Maison des sciences de l'homme (FMSH) qui peut permettre à un chercheur de retravailler son texte selon les commentaires de la communauté. La question de la légitimité apportée par l'institution qui héberge ces travaux émerge, car il apporte aux chercheurs un « plus » scientifique mais aussi une reconnaissance de ce type de publication.
- 7 En la matière, les membres du laboratoire soulignent que le CERI a la chance d'avoir un « vivier » important de chercheurs et de chercheurs-associés. Il y a une réelle reconnaissance du « label » CERI, qui leur donne envie d'écrire pour le compte du laboratoire. Cependant, il n'y a pas de réflexion collective en interne pour analyser véritablement cette problématique.
- 8 De manière générale, il est assez difficile de développer des projets intéressants sans de vraies possibilités offertes par la suite. Un effort devrait être porté au niveau de la vulgarisation : quelque chose qui ressemble aux articles des chercheurs publiés dans la presse nationale, les émissions de télévision, de *France Culture*. Un ciblage est peut-être à faire sur « pourquoi veut-on communiquer ? » Dans quel but publier sur le Web ? Un début de réponse pourrait être apporté par le fait que certains chercheurs considèrent ces publications Web comme un tremplin, créant une dynamique positive dans le monde de la recherche.
- 9 D'autres personnes impliquées dans ces projets retiennent avec plaisir la coopération et l'évolution du dialogue entre les différents corps de métiers, vécus par la plupart des membres du laboratoire. L'équipe s'est auto-formée. Cette situation idéale est rendue possible par la taille de l'équipe ce qui a permis de nombreux apports de compétences.
- 10 Le projet CERISCOPE est évoqué. Il s'agit d'un site internet qui publie chaque année trente à quarante articles autour d'un thème (la pauvreté, les frontières, etc.). Au départ, ce projet est parti d'une publication papier dont le contenu devenait obsolète après deux à trois ans. L'intérêt du site internet est qu'on peut faire vivre le contenu et continuer à l'actualiser. Mais, quand ce projet a été lancé, l'équipe ne savait pas du tout où cela allait la mener. Actuellement, le lancement de la troisième édition est en cours et des problèmes demeurent au niveau du *planning* et du cahier des charges.

La communauté s'interroge

- 11 Ces questions ont nourri les débats et les interrogations des éditeurs depuis une dizaine d'années, les premières discussions ayant débuté dans les années 2000 autour des projets menés par l'équipe de Revues.org. Depuis lors, une évidence semble se confirmer : nos métiers d'éditeur n'ont en fin de compte pas changé. En revanche, les frontières entre nos métiers sont beaucoup plus floues (on a vu beaucoup de chercheurs qui étaient aussi des blogueurs), d'autres types d'expertise émergent. L'éditeur a aussi gardé sa compétence qui est de sélectionner, de mettre en forme, au bout de la chaîne. Certains éditeurs ont fait le choix de l'édition multi-supports.

Le métier d'éditeur en question

- 12 Que met-on derrière le mot « édition » ? Les *position papers* et les *working papers* par exemple ont déjà deux statuts différents (plus ou moins avancés), et ils ne sont pas éditorialisés. Le terme « multi-support » signifie que, pour un même travail éditorial fourni (ressources, compétences, argent dépensé), on investit seule fois. Il y a eu un long travail mené par la collectivité d'éditeurs avec des formats TEI-XML et aussi dans le cadre de l'Association des éditeurs de la recherche et de l'enseignement supérieur (AEDRES). Le pôle d'édition aux Presses universitaires de Caen, a travaillé pour la collectivité des éditeurs à la description des unités documentaires et au partage des connaissances et des méthodes pour que les avancées dans le domaine du travail sur la chaîne éditoriale « xml » bénéficient à tous.
- 13 La longue tradition éditoriale des maisons d'édition a profité au programme. La mutualisation et l'adoption de standards communs ont donné de la force à la communauté.
- 14 Dans le cadre de cette coopération d'acteurs et de corps de métiers, quelles articulations trouver entre des cultures professionnelles différentes (partager, expliquer, revenir sur des choses qui peuvent paraître évidentes) ? La nécessité de comprendre les contraintes de chacun pour pouvoir y répondre est à prendre en compte : il faut beaucoup expliquer, à la fois ce qu'on fait et les enjeux et les contraintes auxquelles on est confrontés. Des compromis sont nécessaires pour parvenir à la réalisation de projets communs, et cela prend du temps. La connaissance et la compréhension du métier de l'autre s'inscrivent dans la durée.
- 15 Face à ces résultats, il est utile de regarder quelle est l'implication des chercheurs dans les projets éditoriaux. Ont-ils envie de s'occuper d'édition numérique ? Y a-t-il une vraie volonté de leur part, de le faire en dehors de leur travail de chercheur ?
- 16 La réponse paraît évidente : les chercheurs doivent publier, l'éditeur arrive en bout de chaîne. C'est au niveau de l'expertise qu'on trouve des croisements entre les métiers et une vraie coopération. La communication scientifique est un projet en soi mais ce n'est pas là qu'éditeurs et chercheurs vont se rencontrer.
- 17 Cependant, il ne faut pas faire porter au numérique des responsabilités qu'il n'a pas. Est-ce qu'on poserait ces mêmes questions de l'implication du chercheur par rapport à l'édition papier ? Il faut garder à l'esprit l'idée que chaque acteur de cette relation a son métier : au chercheur de faire des recherches, et aux autres corps de métier de lui permettre de publier. Pourquoi demanderait-on au chercheur, dans un contexte

électronique, de changer de métier ? Des possibilités leur sont données de publier eux-mêmes¹, mais on ne peut appeler cela de l'édition. C'est une situation similaire aux *working papers*, où le chercheur s'adresse en son nom à la communauté.

Quel impact pour le chercheur

- 18 Dans ce paysage de l'édition, il est utile de regarder quel est l'impact des changements apportés par le numérique à la diffusion.
- 19 Il s'agit de rendre encore plus visible une production intellectuelle. Mais, malgré l'évidence de cette situation, il est des cas où les choses ne sont pas si simples. Il suffit pour cela de citer l'exemple du projet européen CHALLENGE (Liberté et sécurité en Europe), où les chercheurs devaient publier sur le site. Pour eux, deux options sont possibles : publier un livre à 200 exemplaires et diffuser de façon confidentielle ou mettre un texte en ligne et avoir environ 30 000 visites par jour. Le choix paraît évident, mais malgré cela, beaucoup choisissent la publication papier.
- 20 Cette situation somme toute étonnante peut être en partie expliquée par la contrainte exercée par le système lui-même : si le texte a déjà été diffusé en ligne, il est souvent refusé pour l'édition papier. Si cette dernière existe, la question des droits interdit de mettre le texte en ligne.
- 21 La communauté doit s'interroger sur la façon de gérer ce genre de situation.
- 22 À l'heure actuelle, ce sont des choix de personnes : certains ont décidé que plus on donnerait à voir (en numérique : métadonnées, premières pages, textes intégraux, etc.), plus on vendrait. Dès lors, il est naturel de se lancer dans le projet multi-support. Cela aboutit à la vente de supports (avec l'offre *freemium* par exemple).
- 23 Pour revenir à la situation du CERI et réfléchir à des solutions, on évoque le statut du CERISCOPE au sein du laboratoire.
- 24 Le CERISCOPE n'est pas une revue. C'est un ouvrage collectif, évolutif, périodique sur un thème précis. La façon dont il a été pensé, ce qui rend ce projet même un peu surréaliste, suppose une actualisation simultanée de tous les thèmes déjà abordés. Au départ, les deux directeurs de publication étaient aussi relecteurs. Autour d'eux, l'équipe des documentalistes, des éditeurs et des informaticiens travaillait de manière collégiale. Le sommaire était élaboré de façon collective. La proposition a été faite à la direction d'essayer d'impliquer des chercheurs dans le projet, en désignant un « responsable » pour chaque grande partie. Ceci s'est avéré très compliqué, car ce travail demande du temps et cette publication électronique pose des problèmes au niveau de l'évaluation notamment. À Sciences-Po, le CERI est le seul laboratoire sur ce créneau et il a très peu de contacts avec le CNRS. En dehors de la question de l'évaluation, les chercheurs sont plutôt satisfaits du CERISCOPE mais la question du savoir-faire et des compétences nécessaires pour effectuer le travail demeure.

La coopération interprofessionnelle : à chacun son métier

- 25 Pour améliorer ses compétences, la communauté autour de la table rappelle qu'il existe un réseau interdisciplinaire et inter-organisme des métiers de l'édition scientifique publique au CNRS, le réseau Médiçi, et qu'une aide sur des projets spécifiques peut être accordée aux institutions qui la sollicitent.

- 26 Par ailleurs, beaucoup de formations existent (dans le réseau des MSH, dans le réseau Médecin, etc.), il y a même une incitation à la formation. Il semble manquer un réseau d'échanges d'informations, de savoirs et de pratiques pour que les personnes concernées n'aient pas l'impression d'être seules dans leur laboratoire. Le réseau Médecin souhaiterait le créer mais le dialogue semble difficile. L'environnement permet de se rencontrer, de produire des choses ensemble, mais le plus compliqué est de faire comprendre l'existence du métier, mais ce n'est pas propre aux métiers de l'édition.
- 27 L'évidence de la (non-)reconnaissance institutionnelle des disciplines émerge : il n'y a pas de retour de la part des institutions ou des tutelles.
- 28 Pour avoir une réelle reconnaissance, il faut aller faire du *lobbying* dans les sections du comité national CNRS et dans les Conseil national des universités (CNU) afin que ces compétences dans l'édition numérique soient reconnues en tant que telles. Mais le temps de cette reconnaissance est très long, beaucoup plus long que celui de l'édition.
- 29 Les personnels travaillant dans les sections ont mis 20 ans à apprendre leur métier et ne sont pas prompts à changer d'optique. Dans le domaine de la physique, les choses sont très différentes.
- 30 Il y a également une nécessité de reconnaissance de certains métiers. Concernant les développeurs, le CNRS ne mise pas beaucoup dessus. Il en propose peu aux laboratoires qui sont alors obligés de faire appel à des prestataires externes.
- 31 Ces constats et ces expériences montrent la nécessité de faire coopérer les corps de métier, si possible en amont de chaque projet, ce qui est une manière en soi de valoriser le projet lui-même. Identifier en amont les différents acteurs et corps de métiers nécessaires au travail peut être une manière de valoriser le projet lui-même.
- 32 Dans le cas d'un laboratoire comme le CERI, comme dans celui des maisons d'édition, les métiers n'ont pas véritablement changé et les contraintes restent les mêmes. Les auteurs continuent à produire leurs textes et les éditeurs à faire leur métier de sélection. La seule différence dans des publications comme le CERISCOPE est que les auteurs peuvent toujours continuer à revenir sur leurs textes, les actualiser, les réviser, modifier même leur point de vue. C'est le caractère spécifique de ce genre de publication par rapport à l'édition multi-supports. Dans l'édition plus « traditionnelle », le seul changement est le caractère incontournable de la maîtrise de la chaîne TEI-xml dans le métier de secrétaire d'édition/rédaction.
- 33 Dans ce dernier cas de figure, il arrive le plus souvent que le support doive être enrichi pour l'édition multimédia. La question de la contrainte est la même au départ : décide-t-on de publier telle ou telle chose, donc de le rendre visible ? Ce qui est notre métier au sens propre. Les mêmes questions se posent dans l'édition électronique et dans l'édition papier. Si on décide de ne pas mettre de photos dans un livre papier, c'est parce qu'on juge que leur présence n'est pas pertinente ou nécessaire.
- 34 Pour une édition augmentée multimédia il faut très tôt faire intervenir beaucoup plus de corps de métier. Les projets multimédias sont pensés bien en amont. Si on peut développer un réseau de compétences et envisager un modèle éditorial où des personnes externes peuvent venir verser d'autres types de contenus, ce serait intéressant. Pour le moment, ce genre de projet ne peut voir le jour car les projets sont très complexes et très différents.

- 35 En poussant la réflexion, on se demande qui, du chercheur ou de l'éditeur parmi les acteurs de l'édition numérique est censé conduire le projet et en exprimer les besoins ?
- 36 La réponse reste la même, chaque acteur garde son métier. Les éditeurs ne demandent pas aux chercheurs d'entrer dans la feuille de style ni d'encoder des unités documentaires. La spécificité de chaque corps de métier est préservée.

Pour conclure : une question de gouvernance

- 37 De manière générale, il est nécessaire de collaborer entre les différents corps de métier et de faire converger les projets en humanité numérique et éviter que chacun « bricole dans son coin ». Mais qui est à même de coordonner tous les projets ?

NOTES

1. cf. le site du Cléo sur <http://hypotheses.org>

RÉSUMÉS

À partir de l'expérience d'un laboratoire en SHS, l'atelier vise à s'interroger sur les modalités de coopération entre différents corps de métiers (métiers de l'édition, de la documentation, de l'informatique, etc.) dans l'élaboration et la mise en place d'une politique d'édition numérique. Quels acteurs sont concernés et comment les identifie-t-on ? Ce type de problématique existe-t-il dans d'autres laboratoires et quelles seraient les coopérations envisageables afin de bénéficier des retours d'expériences des uns et des autres...

INDEX

Mots-clés : coopération, édition, laboratoire, métier

Comment écrire pour un carnet de recherche ?

Proposé par Mélodie Faury et Pierre Mounier

Collectif

NOTE DE L'ÉDITEUR

Ce texte a été rédigé par Mélodie Faury

- 1 Les pistes de discussions proposées par Pierre Mounier dans le cadre du THATCamp Paris 2012 ont suscité des échanges passionnants au sujet du rapport à l'écriture et à la lecture des blogs chez les chercheurs en sciences humaines et sociales, sur le rapport à la subjectivité dans l'écriture, particulièrement présente dans les textes en ligne, sur la forme intermédiaire entre oral et écrit que représente souvent le billet de blog ou de carnet, mais aussi sur les critères de reconnaissance, par l'institution et par les collègues, de ce qui constitue une « vraie » écriture scientifique¹, individuelle ou collective, sur les possibilités et les obstacles liés à l'évaluation des carnets de recherche et des billets de blogs écrits par des chercheurs. En moins de deux heures nous avons ainsi parcouru un nombre important des problématiques face auxquelles nous met l'écriture en ligne, et en particulier celle du blog et du format billet. L'entrée, proposée par Pierre Mounier, de la spécificité de l'écriture permise par le format billet est tout à fait intéressante pour explorer à la fois les potentialités et les difficultés posées par une pratique qui, de fait, s'installe de plus en plus en sciences humaines et sociales², et qui vient s'articuler avec une longue tradition de publications scientifiques, sous la forme de la publication d'ouvrages après comités de lecture. Je reviens ici sur ce qui m'est apparu comme des éléments de réponses, ouvrant sur de nouveaux questionnements, à la question initiant l'atelier : « comment écrire pour un carnet de recherche ? »³.

Illustrations

- 2 Quels sont les liens possibles entre l'écriture sur un carnet de recherche et les formes de publication habituelles de l'activité académique ? – Quelques exemples :

Pourquoi la question se pose-t-elle ? Démêler quelques enjeux

- 3 Les carnetières et carnetiers qui écrivent sur Hypotheses.org trouvent souvent *a priori*, ou au fil de l'expérience d'écriture, une réponse empirique à cette question : ils écrivent soient comme ils en ont déjà l'habitude dans d'autres contextes, soit de manière autre, selon le rôle que vient jouer le carnet pour eux dans le cadre d'une pratique de recherche individuelle ou collective. Lorsque l'écriture est différente, d'un article, d'un chapitre d'ouvrage, d'un chapitre de thèse, elle prend la plupart du temps en compte les potentialités de l'écriture numérique, à des degrés divers. Les fonctionnalités jouent ainsi des fonctions différentes dans l'écriture, pensée en contexte, celui de son inscription dans une pratique de recherche qui n'est pas seulement numérique.
- 4 Les blogs et les carnets existent et les formes d'écriture, les choix d'éditorialisation et de rythme de publication se diversifient à mesure que le nombre de carnets ouverts augmente. Pourquoi ressent-on le besoin de préciser en quoi consiste l'écriture dans les carnets et d'identifier les formes existantes ou possibles ?
- 5 Une réponse me paraît incluse dans les questions de Pierre Mounier initiant l'atelier du THATCamp : ces formes d'écritures sont nouvelles et elles déstabilisent, elles brouillent les pistes et nous renvoie à des questions cruciales pour nos pratiques de recherche. Si elles existent, c'est vraisemblablement qu'elles remplissent des fonctions, qu'elles répondent à des besoins, variables selon les projets de blogs et de carnets. Et on a encore du mal à identifier leur place, leur position vis-à-vis des formes d'écritures académiques, classiques, c'est-à-dire celles qui sont bien reconnues, notamment par les instances d'évaluation et les comités de lectures de revues scientifiques.
- 6 Derrière ce premier élément de réponse, je vois un deuxième aspect, peut-être encore plus problématique et que j'énonce sous la forme d'une nouvelle question : peut-on dire que l'écriture sur un carnet de recherche est scientifique ? Fait-on de la science lorsque l'on écrit sur un carnet de recherche ? Je réponds d'emblée à cette question par une nouvelle interrogation : qu'entend-on par faire de la science ? Si l'on pense à la production d'écrits qui participent à la construction de connaissances. Alors la réponse est oui, l'écriture sur un carnet peut être scientifique. Si l'on pense à l'écriture permettant de créer du lien entre des travaux de recherche, au sein d'une institution, d'un groupe de recherche ou entre chercheurs et équipes travaillant sur des thématiques proches, ou encore, de manière décloisonnante et indépendamment des inscriptions institutionnelles, entre disciplines, la réponse est encore positive : la recherche peut se construire, se développer, se tisser, échanger numériquement. Si l'on entend maintenant faire de la recherche sur un carnet de recherche, et y développer les grandes missions de l'université, des institutions de recherche et de ses acteurs (recherche, enseignement et diffusion des savoirs) alors là encore, les billets peuvent permettre une diffusion des connaissances, constituer des ressources pour les enseignements et endosser les enjeux divers des pratiques de communication dans nos

pratiques de recherche. Le billet peut ainsi être abordé comme une forme de communication, en contexte, celui des pratiques de recherche.

- 7 *On a l'habitude d'opposer communication interne (publication reviewed avec notes, neutralité de point de vue, formalisation du raisonnement) et communication externe ou de vulgarisation (peu de note ou de référence, éventuellement énonciation subjective, pas de démonstration). Mais il y a aussi des modes de communication entre chercheurs qui peuvent prendre d'autres formes que l'écriture académique classique. Quelle place pour ces modes de communication ?*
- 8 (Extrait de l'introduction de P. Mounier)

Figure 1

Analyse du discours numérique

Blogging scientifique et carnet de recherche en SHS



Accueil À propos Etudes Empiriques Outils de travail Crédits

À propos

Metablog, un espace numérique où les discours numériques sont étudiés sous leur formes linguistique, discursive, ergonomique et sociale. Le nom de ce carnet composé du préfixe "Meta" qui fait allusion à la réflexion, le changement, la succession et l'au-delà des blogs scientifiques, ou autrement appelés, les carnet de recherche.

Les objectifs de ce carnet de recherche sont les suivants:

- 1- Diffusion et mise en valeur d'informations scientifiques relatives au blogging scientifique,
- 2- mise en réseau avec les carnetiers de la plateforme Hypothèses et avec des membres extérieurs participant au blogging.
- 3- Mise en relation avec d'autres travaux se rapprochant au même sujet du présent blog.
- 4- partage d'hypothèses de travail et confrontations d'idées grâce à l'échange et à la discussion via les commentaires.

Présentation du carnet

Ce carnet de recherche se veut d'être un lieu d'échange de réflexions autour des discours nativement numériques, de l'écriture scientifique en sciences humaines et sociale à l'ère du web 2.0 et des Digital Studies.

Mots-clés

analyse de discours numérique
blogging scientifique Carnet de recherche
CIFRE Commentaire identité numérique
Mutations intellectuelles recherche
empirique réseaux sociaux
Stratégies discursives texte écriture

Articles récents

- Etudier les réseaux sociaux
- Du papyrus au numérique: visite aux Archives nationales de Paris.
- A la recherche d'un CIFRE (2): compétences et projet de recherche

Enjeux de communication exprimés par Aboubekour Zineddine, doctorant en linguiste écrivant dans un carnet de recherche personnel (metablog.hypotheses.org)

- 9 Chaque billet, chaque écriture ne remplit bien sûr pas toutes ces fonctions simultanément. L'enjeu communicationnel ou conversationnel des billets est notamment très variable, et c'est parfois l'enjeu d'élaboration d'une idée, pour celui même qui écrit, qui prime, se rapprochant en cela du carnet de note papier. À l'inverse, ou en regard, le lecteur ne cherche pas nécessairement à entrer en conversation avec celui qui écrit. De plus, la diversité des fonctions recouvertes par les billets des carnets de recherche et des blogs de chercheurs diffère d'un projet à l'autre.

Figure 2

Pourquoi lisez-vous les carnets de recherche de la plate-forme Hypothèses ?

– Question posée sur Twitter en juillet 2012, voici quelques réponses

Profil des répondants : enseignants du secondaire, chercheurs, bibliothécaires, lecteurs du numérique

- 10 Doit-on dès lors différencier explicitement les types d'écriture, définir leur statut, voire classer et hiérarchiser la qualité, en particulier scientifique des billets, et si oui, quels seraient les enjeux d'une telle échelle de valeur ?
- 11 De manière sous-jacente aux questions posées par Pierre Mounier et les discussions qu'elles ont suscitées pendant l'atelier, je vois un troisième aspect, peut-être plus pragmatique : celui du temps et de la place que peut prendre un carnet de recherche dans le travail d'un étudiant, d'un chercheur, d'une équipe de recherche, etc. Comment peut-on convaincre les individus, les collectifs, les institutions de l'intérêt d'un carnet de recherche ? Et pour cela, comment donner de la légitimité à ce format d'écriture, au sein des critères actuels d'évaluation, de publication, de légitimité de ce qui constitue les pratiques de recherche ?
- 12 Pourquoi convaincre d'abord ? Par conviction peut-être, celle de l'importance de l'accès et du partage de la connaissance que l'on retrouve affirmée par Pierre Mounier dès les remerciements du Read/Write Book 2, et que je crois retrouver chez la grande majorité des participants au THATCamp Paris 2012.
- 13 Se poser la question de l'écriture dans les carnets des recherches n'est donc pas neutre et nous confronte d'emblée à la question de la reconnaissance institutionnelle, de la validation et de l'évaluation de pratiques qui sont déjà reconnues de fait, par l'usage, par les chercheurs qui écrivent, lisent, relayent, et utilisent les blogs de science, en tant que ressources ou pour ses fonctions conversationnelles. La tension entre ces deux formes de reconnaissance, crée notamment une tension bien visible entre :
 - 14 engagement personnel et/ou collectif dans la pratique d'écriture sur un carnet de recherche, associé souvent au plaisir d'écrire sous des formes différentes de celles des

articles ou des ouvrages, mais souvent exigeante en temps et ressources humaines, d'une part ;

- 15 et nécessité de témoigner de l'activité d'écriture effectuée, quand celle-ci prend une place indéniable dans la pratique de recherche, de la rendre « viable », de trouver les espaces institutionnelles où elle peut se développer, selon les critères de reconnaissance existants, et dans la mesure où l'écriture sur carnets/blogs est considérée souvent par ceux qui la pratiquent comme faisant partie de leurs missions.
- 16 Faut-il d'ailleurs valider et évaluer les pratiques d'écriture en ligne, comment trouver les « indicateurs » pertinents (qui feraient sens) et quels en seraient les effets sur les pratiques ? C'est une question qui mérite sans doute d'être posée, car la réponse n'est pas évidente.

Des fonctionnalités aux fonctions des carnets de recherche : de nouvelles formes d'écriture

- 17 On voit ainsi émerger de nouveaux modes d'énonciation, des styles propres aux blogs scientifiques, qui n'ont pas encore été étudiés de manière approfondie, à ma connaissance⁴. Ces nouvelles formes d'écriture mobilisent des fonctionnalités propres au format blog (Deseilligny, 2006) qui s'articulent avec des fonctions de l'écriture scientifique, déjà endossées par les formes classiques de publication, mais qui ouvrent aussi potentiellement d'autres fonctions⁵, qui posent des problèmes spécifiques.

Une écriture brève

- 18 La brièveté de l'écriture ne s'associe pas nécessairement avec une absence de rigueur dans l'écriture, même si elle induit un allègement, ne serait-ce qu'en taille, de celle-ci, par rapport à l'article scientifique. Le billet de carnet ou de blog est pris dans la tension entre rigueur, exactitude d'une part, et justesse, format court, soit l'adaptation à un format particulier d'autre part. On retrouve une problématique commune à l'étude des discours de vulgarisation scientifique et l'idée fréquemment partagée que rendre accessible (longueur, teneur du propos, langage, etc.) serait assorti d'une moindre pertinence scientifique. Si l'on pense le billet dans le contexte spécifique de sa publication, dans celui, défini par l'auteur (ou les auteurs) de ses objectifs communicationnels et enfin de sa lecture potentielle, alors l'écriture peut être avantageusement pensée en terme de justesse plutôt que d'exactitude (Labasse, 2001). Autrement dit, être lisible, accessible, compris et bref n'amène pas à réviser ses exigences en termes de qualité et de rigueur scientifique.
- 19 Cette forme d'écriture pose des questions spécifiques en termes de choix éditoriaux : rythme, programmation, organisation (catégories, etc.), style, fragmentation des textes éventuellement republiés, etc.

« Quelles tendances se dégagent : plus de brièveté, moins fouillé ? On tendrait a priori à penser que l'on s'éloigne des prérequis de l'écriture académique. D'autres prérequis n'émergent-ils pas de ces nouvelles formes ? De nouvelles règles ? D'autres valeurs ajoutées ? Plus ou moins d'intertextualités, de renvois ? »

Questions lancées par Pierre Mounier

Une écriture hyper-reliée

« Une écriture « hypertextuelle » qui fonctionne plus suivant la logique de l'association d'idées (par l'auteur et par la lecture), donc plus proche de la pensée en cours de formation. »

Intervention d'une participante à l'atelier

- 20 L'écriture sur un carnet ou un blog peut s'enrichir de liens hypertextes qui créent des liens, pointent sur des ressources, changent le mode de discours, donnent de l'épaisseur à l'article, se mettant en relation avec des travaux académiques (articles en lignes, billets, ouvrages numérisés, événements scientifiques, etc.) ou avec des textes précédemment écrits par l'auteur du billet lui-même (sorte d'auto-éditorialisation, de construction d'une articulation entre des écrits successifs, qui n'est pas nécessairement anticipée).

Figure 3

| |
|--|
| <p>Les liens hypertextes comme des portes ouvertes sur les réflexions et les rebonds de l'auteur du billet</p> <p>http://reflexivites.hypotheses.org/2311</p> |
| <p>Les liens hypertextes comme des ressources vers des références et des corpus en ligne</p> <p>http://penseedudiscours.hypotheses.org/10578</p> |

Les épaisseurs de billets hyper-liés – Deux exemples.

- 21 Cette écriture reliée, référencée, appuyée sur des ressources, des références peut permettre d'associer brièveté et densité. D'une manière spécifique à l'écriture numérique, et courte, elle rend perceptible le maillage et les interactions de l'écriture développée dans le cadre du billet avec une multitude de sources, d'autres réflexions et de la positionner dans un contexte de recherche. Les liens hypertextes, qu'ils soient explicitement présentés comme des ressources bibliographiques, ou non, peuvent construire en quelque sorte le cadre de référence du billet et donner accès aux ressources permettant au discours de se construire.
- 22 Par ailleurs, et de manière spécifique, la notification des liens entrants sur un carnet de recherche permet des échanges inédits entre auteurs : ces liens matérialisent, de manière quasi instantanée, la résonance d'une écriture avec une autre, l'articulation d'une pensée ou d'une information à une autre – et peuvent être à l'origine de rencontres scientifiques inattendues et riches⁶, de coopération, d'échos interdisciplinaires.

Une écriture multimédia

« C'est un espace plastique, qui favorise des formes d'écritures nouvelles (multimédia), qui ne passent pas forcément par les mots (images, vidéos, sons...). »

Intervention d'une participante à l'atelier

- 23 Même si cette fonction n'a pas été au centre de nos discussions, il est clair qu'elle constitue l'une des grandes spécificités de l'écriture possible sur un carnet de recherche. L'habitude de l'utilisation du texte dans les pratiques de communication de

recherche en sciences humaines et sociales (SHS) expliquerait *a priori* que ce ne soit pas la fonctionnalité la plus investie par les enjeux académiques. Elle ouvre des possibilités qui seraient pourtant particulièrement intéressantes à explorer plus avant, ce que font déjà certains carnets de recherche, notamment quant à la mise à disposition de ressources audio ou vidéo issues de terrains de recherches (voir les carnets de la phonothèque).

Une écriture sans évaluation préalable à la publication

- 24 Les avantages et intérêts d'une telle publication, qui sont souvent mis en avant par les acteurs des blogs scientifiques sont notamment la réduction des délais de publication, la possibilité d'un accès gratuit aux productions académiques (ou *open access*), qui s'affranchissent des coûts liés à l'édition et à l'accès aux revues payantes, un accès à la recherche « en train de se faire », selon une expression de Bruno Latour, aujourd'hui largement reprise au sujet des blogs scientifiques, etc.

Figure 4



Bandeau de présentation du carnet de recherche de l'Institut français du Proche-Orient (<http://ifpo.hypotheses.org/>)

- 25 Les textes publiés sur les blogs scientifiques sont publiés directement par l'auteur ou par celui/celle qui se charge de l'édition et de la mise en ligne. On peut donc les considérer comme des communications scientifiques directes, sans soumission préalable aux pairs. Ce n'est pas le cas exclusif des publications de type billets et on rencontre dans le cas de ces publications des problématiques proches des cas de publications d'articles, de thèses sur des plateformes d'archives ouvertes ou encore sur Wikipédia qui ne fonctionne pas sur la base d'une validation *a priori*. Cette question revient souvent au cœur des discussions concernant le statut et la valeur des billets publiés. Elle interroge directement notre rapport au savoir académique et repose de manière vive la question de la garantie de la qualité scientifique des écritures en ligne, non préalablement validées.
- 26 Comment est-il possible d'articuler la reconnaissance de fait d'un article, c'est-à-dire en un sens sa *valeur* d'usage et les recommandations entre lecteurs, à sa reconnaissance institutionnelle? Ou pour pousser plus loin la question: la reconnaissance institutionnelle serait-elle la seule garante de la légitimité scientifique d'un billet, d'une écriture, d'un carnet? On tombe à ce stade dans des discussions brûlantes et beaucoup plus générales en termes d'évaluation scientifique de la qualité des travaux de recherche par le système actuel de publications avec comité de lecture.

Une écriture ouverte à tous

- 27 Enfin, la question du lectorat potentiellement large des billets se pose. Pour qui écrit l'auteur ? Pour ses pairs, pour un lectorat préalablement défini et/ou qui se construit au fil de la publication des billets ? L'ouverture du billet, si elle remplit une fonction certaine en termes de diffusion large du propos, induit aussi de nouvelles formes d'écritures, différentes des formats académiques « traditionnels ».

« le carnet est ouvert, plus coopératif, on écrit des choses plus générales, à destination d'un public plus vaste, profane. On s'oriente vers du moins spécialisé mais aussi plus participatif et interdisciplinaire. »

Un participant de l'atelier

- 28 L'ensemble des « fonctions » du carnet de recherche parcourues jusque-là sont bien connues (espace de liberté, interdisciplinarité, ouverture vers d'autres publics, réflexivité pour le chercheur...). Et on nous renvoie aux types d'écriture que ces fonctions induisent :

« Comment connecter la question des fonctions à la question plus technique de l'écriture elle-même ? Que pourrait être l'écriture académique de carnet ? Quels pourraient être des prérequis de l'écriture du carnet ? Quelques pistes : l'abondance des liens (les hypertextes) ; la subjectivité (la première personne) : quelles sont les fonctions de ce « je » ; la brièveté : quels sont ses avantages, à quoi sert-elle ? »

Un animateur de l'atelier

L'écriture sur un carnet : subjectivité, oralité et réflexivité

« La comparaison avec un article achevé est impossible : dans un carnet de recherche, le degré d'achèvement n'est pas le même. Le carnet est l'occasion de reprendre et de formaliser des notes éparses. Ce stade de la réflexion est inédit. »

Une participante de l'atelier

- 29 Pierre Mounier présente l'écriture sur les carnets de recherche comme une forme de développement d'une tradition intermédiaire entre blog et publication scientifique classique. Je trouve l'idée de continuum discursif, développé par D. Jacobi (1988), à propos des discours de vulgarisation, particulièrement intéressante pour penser les formes d'écriture que l'on peut lire sur les carnets de recherche. La diversité, l'hybridité des formes que l'on rencontre sur les carnets de recherche matérialise en quelque sorte ce *continuum*.

Le cas de l'article d'Aurélien Berra dans le Read/Write Book 2 est un exemple intéressant à au double titre au moins : il constitue l'exemple d'une présentation orale, transcrite, qui donne lieu ensuite à un article d'ouvrage, d'une part, et il montre bien l'effacement des frontières délimitées entre oral et écrit auquel participent les écritures initiées en ligne, et pensé pour leur mise en ligne au moment de leur élaboration.

Des formes d'écriture assumant la subjectivité

« Dans les SHS, il existe des « traditions », des modes, des contraintes imposées par des revues, des travaux collectifs, et qui s'imposent au « je » de l'écriture, aussi des « nous » qui disent « je » et de « je » qui se masquent derrière le « nous »... »

Une participante de l'atelier

- 30 L'écriture du carnet de recherche laisse à la subjectivité la possibilité de se développer, elle donne une place légitime à la perspective individuelle de celui qui écrit : l'écriture d'un billet est souvent issue de la lecture d'un ouvrage, d'une réflexion suscitée par l'actualité, de la participation à une conférence, ou de travaux menés sur le terrain.
- 31 Dans le processus de construction d'un propos dans le billet, l'effacement de la subjectivité, la recherche d'une certaine objectivité, n'est pas une valeur en soi. Au contraire même pourrait-on dire. Le blog, dans ses formes initiales, laisse la part belle à l'individu, qui se raconte. Le blogging scientifique garde cette dimension fortement située, à l'échelle du billet ou du blog, dans la mesure où l'on identifie, en venant lire un blog de chercheur, de structure, de groupe de recherche, de bibliothèque, où l'on se trouve et qui parle (l'individu, l'institution, le collectif). L'écriture sur un carnet de recherche donne une place privilégiée à la subjectivité, notamment via des marques d'énonciation plus apparentes que dans la plupart des articles scientifiques. En cela, elle s'inscrit dans la tendance de ré-interrogation multidisciplinaire de la place du « je » dans l'écriture scientifique, de la posture du chercheur vis-à-vis de son terrain et de ses recherches, ainsi que dans la démarche réflexive permettant de mieux expliciter « d'où parle le chercheur ? ».
- 32 Une écriture qui assume la subjectivité, sans pour autant renoncer à l'enjeu de scientificité, peut composer avec l'idée d'engagement. En explicitant le positionnement de l'auteur, le statut et la fonction du discours développé, le carnet engage plus que la publication « traditionnelle ». Le carnet de recherche réactiverait-il ponctuellement une certaine forme de parole de l'intellectuel dans les débats de société ?

Une écriture oralisée

- 33 Il est alors proposé de considérer le billet comme une forme d'oralisation de l'écriture scientifique, comme une forme intermédiaire entre oral et écrit. De cet entre-deux surgit une tension : celle de l'exigence de références scientifiques explicites dans la forme orale d'une communication scientifique. Les liens hypertextes pourraient-ils réintroduire dans une écriture oralisée un second niveau de lecture qui donnerait accès à cette exigence de référencement, de situation et de contextualisation du discours ?
- 34 L'écriture conversationnelle, intégrant elle aussi une dimension orale forte, pourrait-elle contribuer à la construction d'une intersubjectivité qui participerait à la légitimation scientifique du propos écrit dans le carnet ?
- 35 En parlant la science, et en introduisant une forme d'oralité dans l'écriture scientifique, l'écriture des billets serait-elle enfin l'occasion d'une forme de réflexivité différente⁷ de celle à l'œuvre dans les formes de publication « classiques » ?
- 36 Quoi qu'il en soit, il serait intéressant d'étudier particulièrement les spécificités d'une écriture oralisée telle que celle que l'on trouve

L'écriture sur un carnet de recherche favorise-t-elle la réflexivité ?

- 37 L'écriture sur un carnet de recherche met face à des choix : quant aux objectifs de l'écriture, quant à l'éditorialisation, quant à l'utilisation des fonctionnalités disponibles, etc. Ces choix s'explicitent dans des rythmes de publications, des catégories et des mots-clés, des liens affichés, des adresses au lectorat éventuellement

ciblé, etc. Par ailleurs, l'écriture d'un carnet de recherche assume une forme de subjectivité, à l'échelle du billet ou du blog, ne serait-ce que par la cohabitation de différents registres de discours dans un même lieu numérique, identifiables et dont les propos sont situés (équipe de recherche, équipe de rédaction, institution, etc.). Enfin, par son caractère oralisé, publique et souvent adressé l'écriture de billet renvoie potentiellement l'auteur aux questions suivantes : pourquoi j'écris/je publie ? Pour qui (y compris pour soi) et comment ? Pourquoi dans un carnet plutôt que sur le papier, et pour soi seul ? Pourquoi un billet plutôt qu'un article ? Pourquoi en accès ouvert ou en accès restreint ? Pourquoi autoriser les commentaires ou les refuser ?

- 38 Face à l'absence de normes définies *a priori*, et aux potentialités offertes par des fonctionnalités de l'outil, l'écriture sur un carnet de recherche peut être l'occasion de réinterroger ce que l'on a tendance à ne plus réinterroger lorsque l'on publie de manière plus « classique ». Et d'une réflexivité individuelle on passe parfois, comme dans le cas de l'atelier du THATCamp Paris 2012 dont est issu cet article, à la mise en œuvre d'une réflexivité collective sur nos pratiques de publication scientifique, d'évaluation et sur notre rapport à la scientificité (la recherche du billet « vraiment » scientifique) et à la qualité ou la valeur d'une écriture.

Conclusion

- 39 Au fil de l'atelier, et des problématiques évoquées, plus ou moins rapidement, de l'exploration des fonctions du carnet de recherche, c'est-à-dire de son utilité, de sa spécificité et de son positionnement vis-à-vis d'autres publications, plusieurs caractéristiques ont été mises en évidence : l'écriture sur un carnet offre un espace de liberté, de la visibilité, une lecture-écriture-conversation interdisciplinaire et pouvant parfois sortir de l'entre-soi académique et la possibilité à un discours réflexif, individuel ou collectif, de se déployer, notamment du fait de la place assumée de la subjectivité et de sa position intermédiaire entre communication orale et écrite.
- 40 Pourrait-on se mettre d'accord sur un format d'écriture qualifié d'académique qui rende compte des fonctions particulières que sert un carnet de recherche, qui serait dès lors reconnaissable facilement et objectivement (ou inter-subjectivement) et que l'on pourrait expliciter ? La discussion de l'atelier du THATCamp Paris 2012 nous amène à mettre en avant quelques critères fondamentaux sans pour autant répondre de manière définitive à cette question.

« Quels critères : qualité de la réflexion, de la mise en mots, qualité des liens (bibliographie) ou capacité à produire du réseau, du rebond. »

Intervention d'une participante à l'atelier

- 41 Situer le discours, construire une problématique, rendre compte d'une démarche, présenter une argumentation claire et cohérente, donner accès aux ressources et cadres de référence participant à la construction de la réflexion partagée dans un billet, donner à penser : autant d'éléments qui participent à ce que l'intersubjectivité soit possible et qui pourraient contribuer à expliciter ce qui pourrait être attendu du point de vue académique d'un carnet de recherche. Au cours de l'atelier, on retrouve aussi l'idée, commune à différents modes de communications écrites ou orales de nos pratiques de recherche, que la qualité scientifique d'un billet est opposée directement au registre de l'opinion, et celle de la nécessité que celui-ci rompe avec le sens commun pour acquérir une valeur scientifique.

- 42 Faut-il chercher absolument une dimension « scientifique » dans chaque billet, au sein d'un carnet de recherche dont le projet scientifique est clair ? Le billet est-il la bonne unité à considérer pour juger de la pertinence académique d'un carnet de recherche ? Peut-on/doit-on enfin établir des critères objectifs permettant de juger de la qualité d'un carnet académique ou distinguer des genres ?

« Mettre en place un système d'évaluation plus lourd ne conduirait-il pas à ralentir et freiner la production de carnets, à « alourdir » le processus d'écriture ? »

Un.e participant.e à l'atelier

- 43 Le risque de normaliser et de différencier les carnets « valables » (d'un point de vue scientifique, ou institutionnel) des autres et de perdre la créativité dans l'écriture qui s'observe actuellement existe. Même si les enjeux de reconnaissance sont incontournables et qu'il importe de pouvoir se repérer dans les types de publications en ligne, je ne peux m'empêcher de poser une dernière question : et si un carnet « valable » (ce qui ne signifie pas pour autant, et la nuance est importante, « dont les billets sont scientifiques ») était celui qui fait sens pour ceux qui y contribuent et ceux qui le lisent ? Étudier de manière détaillée les formes d'écriture existantes dans les carnets de recherche permettrait d'interpréter le sens, dans le contexte des pratiques académiques, de ces pratiques individuelles et collectives, tout à la fois d'écriture et de communication.

BIBLIOGRAPHIE

CASILLI, Antonio, 2010, *Les Liaisons numériques*, Paris, Seuil.

DESEILLIGNY, Oriane. L'écriture de soi, continuités et mutations. Du cahier aux journaux personnels sur le web (1998-2003). Thèse.

JACOBI, Daniel (1988). « Le discours de vulgarisation scientifique », in Jacobi (D.) et Schiele (B.), éd. *Vulgariser la science. Le procès de l'ignorance*. Éditions Seyssel : Champ Vallon. p. 87-117.

JURDANT, Baudouin (22 février 2012) « Communication scientifique et réflexivité », *Espaces réflexifs*, <http://reflexivites.hypotheses.org/695>, consulté le 29 septembre 2012,

JURDANT, Baudouin (2006). Parler la science ? *Alliage n° 59*, <http://www.tribunes.com/tribune/alliage/59/page6/page6.html>, consulté le 20 avril 2012.

URL : <http://www.tribunes.com/tribune/alliage/59/page6/page6.html>

LABASSE, B. (2001). *La communication scientifique. Logiques et méthodes*. Chassieu, Pôle universitaire de Lyon.

LE MAREC, J. (2002). Situations de communications dans la pratique de recherche : du terrain aux composites, *Études de communication n° 25 - Questions de Terrains*.

NOTES

1. La question du carnet « vraiment » scientifique a émergé très rapidement au cours des discussions de l'atelier.
 2. Même si l'on remarque sur Hypotheses.org (voir la liste des carnets Hypothèses) que certaines disciplines développent le format blog en ligne plus fréquemment que d'autres (voir le carnet La pensée du discours).
 3. Je m'appuie sur mon expérience d'écriture individuelle et collective dans plusieurs carnets de recherche, sur ma lecture régulière de carnets et de blogs scientifiques ainsi que sur mon observation des billets publiés sur la plate-forme Hypothèses dans le cadre de son conseil scientifique (voir le carnet fr.hypotheses.org).
 4. Une thèse portant sur les discours de blogs scientifiques en SHS est actuellement en préparation (L'analyse des discours numériques: le cas du blogging scientifique en sciences humaines et sociales).
 5. Ces fonctions sont parfois matérialisées dans les carnets sous la forme de différentes catégories, ou parfois de mot-clefs, associées aux billets.
 6. Dans le cas des *Espaces réflexifs* par exemple (<http://reflexivites.hypotheses.org/>), les liens hypertextes matérialisent la lecture et la discussion ou l'articulation des écrits des mois précédents avec la conception de la réflexivité développées par l'un des « locataires » mensuels, dans l'écriture de ses billets au cours du mois où il « habite » la « Villa réflexive » : les liens hypertextes, avec les commentaires, tissent les échanges interdisciplinaires.
 7. J'utiliserais volontiers ici le terme de réflexivité par oralisation, suite aux travaux de Baudouin Jurdant sur la réflexivité liée à la communication scientifique orale (<http://reflexivites.hypotheses.org/695>), dans le cadre des sciences exactes et expérimentales, et en particulier dans le contexte de situation de vulgarisation.
-

RÉSUMÉS

Un atelier proposé à cause du constat concernant le développement de nouvelles pratiques d'écriture sur la plateforme Hypothèses. Des carnetiers qui réfléchissent actuellement à la manière de produire des livres à partir de notes de blog. Cet atelier est conçu pour permettre une discussion sur ces questions. Écrire pour un carnet de recherche est très différent de ce qui est attendu pour une revue à comité de lecture. Quelles sont les particularités de ce mode d'écriture ? Comment concilier précision et brièveté de la forme ? Le style oral et conversationnel est-il compatible avec la rigueur propre au raisonnement scientifique ? L'idée est que l'atelier soit basé sur un échange d'expérience et une analyse en commun de quelques billets significatifs.

INDEX

Mots-clés : carnet hypothèses, nouvelles pratiques d'écriture, billet, style bref

La diffusion de la recherche

Comment commencer, poursuivre et pérenniser une revue en ligne ?

Proposé par Paul-Henri GIRAUD

Collectif

NOTE DE L'ÉDITEUR

Récemment l'université de Lille a encouragé Paul-Henri Giraud à créer une revue en ligne, et ses collaborateurs et lui-même ont opté pour la création d'une revue sur les langues romanes. Le parti-pris est de garder les articles dans leur langue d'origine (italien, espagnol et portugais). Les revues multilingues étant confrontées à des difficultés particulières, l'édition de chaque numéro thématique se fera sous la responsabilité d'un chercheur multilingue.

Pourquoi une revue en ligne ?

- 1 Pourquoi en effet, alors que nombreux sont ceux pour qui le rôle symbolique du papier est encore très prégnant et que le passage en ligne d'une revue est souvent considéré comme une dévaluation de cette dernière ?
- 2 La lenteur des publications papier, leur diffusion médiocre, les coûts d'impression et le moindre coût ont guidé ce choix. En publiant sur une plateforme, grâce au référencement, la revue bénéficiera en outre d'une visibilité que la version papier ne peut pas offrir. De plus, ce format permettra de fédérer un maximum de personnes autour d'un projet, dépassant ainsi les cloisonnements disciplinaires.

Complémentarité entre édition papier et édition numérique

- 3 Il ne faut pas penser le numérique et le papier comme alternatifs en bloc. Loin de s'exclure mutuellement, il existe une complémentarité entre les deux, et les universitaires sont souvent amenés à choisir ces modes complémentaires pour valoriser leurs recherches : le Web et le livre. La collection de livres Lire les Sciences Sociales en est un exemple. Actuellement imprimée, l'idée est de la faire passer en ligne, avec une dimension participative possible. Une anthologie papier est en outre envisagée.

Il est très difficile, à l'heure actuelle, de publier une revue en ligne en dehors d'une plate-forme.

La question de la pérennisation

- 4 On pense souvent qu'une revue en ligne est pérenne, mais qu'en est-il si par exemple elle disparaît du serveur de l'université ? Un des gros problèmes des publications électroniques est de trouver un lieu sécurisé pour publier. Cairn : <http://www.cairn.info/>, Persée : <http://www.persee.fr> ou encore Revues.org <http://www.revues.org/> peuvent être considérés comme des réponses à ces problèmes de pérennisation. Sur les plate-formes, tous les textes sont convertis en XML, ce qui permet une conservation à long terme quels que soient les changements de paradigmes numériques à venir.

La diffusion

- 5 La diffusion des revues visant un lectorat très disséminé et spécialisé peut être optimisée grâce à leur mise en ligne. Revues.org et Persée permettent de diffuser la revue papier en ligne. Le Bulletin de l'École française d'extrême orient, par exemple, est diffusé par Persée.
- 6 Certains grands groupes éditoriaux internationaux ont bien identifié les difficultés des structures universitaires et font aux revues des propositions de prise en charge mais cela implique par ailleurs pour les revues une perte de leur autonomie.

L'édition

- 7 Il faut bien distinguer édition et diffusion électronique. Si le numérique permet de surmonter les problèmes de diffusion, la problématique est différente en ce qui concerne l'édition. Ici l'objectif est de réaliser une seule fois un bon travail éditorial avec des systèmes de balises *ad hoc* (comme le XML TEI) qui seront utilisables pour la publication sur plusieurs supports (papier, mais aussi liseuses, tablettes, téléphones, grâce au format EPUB). Si, comme on l'a dit plus haut, des économies peuvent être faites sur l'impression, il ne faut pas croire pour autant que la publication d'une revue

en ligne ne coûte rien. Dans certaines revues SHS, le poste de secrétaire de rédaction est assuré par un enseignant chercheur, qui ne peut s'investir de façon régulière. Or, ce poste est stratégique. Si, après trois ou quatre ans, il y a moins de moyens et d'investissement que lors du lancement, cela donne lieu à un essoufflement ; s'ensuit alors une perte de régularité et de qualité. L'idéal est qu'un professionnel de l'édition prenne en charge l'ensemble du travail éditorial. Mais si ce poste vient à disparaître, cela peut mettre en péril la revue elle-même.

Pratiques de lecture opportunistes

- 8 Un des écueils de la diffusion électronique est qu'elle engendre des pratiques de lecture opportunistes, qui consistent à piocher uniquement l'article qui intéresse. Comment dans ce cas rendre compte de l'unité éditoriale d'un numéro ? Il n'y a pas de réponses à ce problème qui d'ailleurs ressemble à s'y méprendre au « photocopillage », toujours d'actualité. Aux Éditions de la Maison des sciences de l'homme on a contourné ce problème pour les revues en vendant les articles à l'unité mais on n'a pas encore « accepté l'idée » de vendre les contributions à l'unité pour les livres collectifs.

Multilinguisme

- 9 La publication multilingue implique des contraintes particulières, comme celle de disposer d'un secrétariat de rédaction compétent dans les différentes langues de travail. Certaines revues acceptent des articles écrits dans d'autres langues que la langue principale de travail mais les auteurs peuvent avoir des surprises. Leur texte n'est parfois pas corrigé voire publié avec des normes typographiques fautives. D'autres jeunes revues toutefois, comme la revue *Diacronie*, savent gérer la publication d'articles en plusieurs langues, en s'appuyant sur un réseau scientifique étendu dans plusieurs pays. Il s'agit d'une revue électronique. Et ceci n'est sans doute pas sans rapport avec cela.

RÉSUMÉS

Récemment l'université de Lille a encouragé Paul-Henri Giraud à créer une revue en ligne, et ses collaborateurs et lui-même ont opté pour la création d'une revue sur les langues romanes. Le parti-pris est de garder les articles dans leur langue d'origine (italien, espagnol et portugais). Les revues multilingues étant confrontées à des difficultés particulières, l'édition de chaque numéro thématique se fera sous la responsabilité d'un chercheur multilingue.

Éditeurs SHS et humanités numériques

Proposé par Marion Colas et Maria Vlachou

Collectif

Introduction

- 1 Cet atelier a été proposé par Marion Colas, responsable du département de cessions de droits étrangers et Maria Vlachou, chargée de cessions de droits pour l'Espagne et l'Amérique latine (sauf le Brésil), l'Allemagne, les Pays-Bas, l'Europe du Nord et l'Extrême-Orient au sein des Presses universitaires de France (PUF). Marion et Maria ont eu, au sein de leur pratique quotidienne, la possibilité de voir l'évolution du secteur de l'édition électronique à l'étranger.
- 2 Le THATCamp leur a semblé une bonne occasion pour lancer une discussion sur les pratiques des éditeurs et notamment celles des éditeurs en sciences humaines et sociales face aux changements impliqués par le numérique, d'explorer aussi les liens de plus en plus tenus entre édition et humanités numériques.

L'édition, différentes manières de faire

- 3 Au sein du monde de l'édition coexistent différents acteurs, les éditeurs privés ou commerciaux et les éditeurs publics et subventionnés. Ces différents acteurs font-ils de l'édition différemment ? Quelles sont leurs pratiques et leurs missions respectives ?
- 4 Pour Marion et Maria, il est nécessaire de distinguer l'éditeur privé de l'éditeur public dont les logiques diffèrent. L'éditeur privé se caractérise par une logique commerciale nécessaire au financement de l'entreprise, ce qui s'oppose à la logique des éditeurs publics, subventionnés.
- 5 Avec l'évolution du support papier vers le numérique, on assiste à un renouvellement des compétences au sein du métier d'éditeur - et ce, qu'il s'agisse d'éditeur public ou d'éditeur privé. Les éditeurs sont tous dans une « niche économique ». La bascule

numérique s'impose aussi de cette façon. Les usages s'adaptent, les éditeurs naviguent à vue.

L'édition et les humanités numériques

- 6 Les humanités numériques constituent une communauté de pratiques « à l'intersection des sciences humaines et sociales et des technologies numériques ». L'expression peut désigner tout à la fois le travail, par définition réflexif, de collecte, d'exploitation et/ou de diffusion de données, qu'elles soient nativement numériques ou non. À la fois nouveau champ disciplinaire – ce que le numérique fait aux sciences humaines et sociales, et comment, en retour, celles-ci s'en emparent – et utilisation de nouvelles techniques numériques de diffusion des savoirs, ce champ émergent de création et de diffusion des savoirs offre donc un lieu privilégié pour essayer de réfléchir aux mutations numériques du champ éditorial aujourd'hui.

De l'édition en sciences humaines et sociales sans éditeur

Quand le chercheur devient éditeur : récits d'expériences

- 7 Les nouveaux outils de publication en ligne (blogs, archives ouvertes, sites internet, etc.) peuvent poser la question de savoir si le chercheur-auteur a encore besoin de l'éditeur pour diffuser ses écrits, pour les promouvoir - bref, pour être lu. Finalement, aujourd'hui, avec les nouveaux outils de publication en ligne, le chercheur a-t-il encore besoin d'un éditeur ? Est-ce que l'éditeur aura le même rôle à l'avenir ? Et de manière générale, est-ce que l'édition scientifique en sciences humaines et sociales peut se passer des éditeurs¹ ?
- 8 Certes, chacun conviendra qu'il existe un besoin d'interaction entre les chercheurs et les éditeurs, et, plus largement, le monde de l'édition. Si le chercheur peut mettre lui-même sa publication en ligne après avoir acquis les compétences techniques nécessaires pour le faire – et oui, on demande de plus en plus souvent au chercheur-auteur d'être également technicien ! – il n'en reste pas moins qu'il a, la plupart du temps, besoin du savoir-faire de l'éditeur pour éditer son texte, tant pour le « mettre au jour » que pour le mettre en valeur. En édition de sciences humaines et sociales, il est donc essentiel que les auteurs-chercheurs, les éditeurs ainsi que les autres acteurs de la chaîne du livre travaillent ensemble. Les métiers doivent s'articuler entre eux afin de permettre la diffusion la plus large possible de la production du savoir.
- 9 Mais, parfois, cette interprofessionnalisation n'est pas possible. Alors, certains chercheurs, dans les laboratoires de recherche, s'improvisent éditeurs.

Squintar

#Jourdain RT @solen_r: Peut-on faire de l'édition sans le savoir ? #tcp2012

26/09/12 17:12

- 10 Une participante à l'atelier, chercheuse, fait part de son expérience d'éditrice... acquise sur le tas et par la force des choses ! Son équipe et elle souhaitaient éditer une

publication, mais n'ont trouvé aucun éditeur pour le faire. C'est donc les membres du laboratoire qui ont pris en charge le processus d'édition de la publication, du travail éditorial sur le texte, à la génération du PDF pour impression. Reste le problème de la diffusion de la publication en ligne et dans les circuits traditionnels.

- 11 Le débat s'oriente ensuite sur l'offre de formation existante proposée aux chercheurs et leur permettant de mettre en ligne leurs publications.
- 12 Les formations Lodel utilisateurs proposées par le Centre pour l'édition électronique ouverte ont été suivies par certains éditeurs ou chercheurs participants à l'atelier.
- 13 Lors de ces formations, qui durent deux jours, les chercheurs, chargés d'édition ou éditeurs apprennent à se servir de Lodel, un logiciel de gestion de contenus ou CMS (*Content Management System*). Ce logiciel a été pensé pour les chercheurs ou chargés d'édition ne disposant pas de compétences techniques particulières. Lodel est donc un outil dont l'utilisation est simple, facile à prendre en main et qui repose sur le mécanisme suivant : il s'agit de préparer des documents, de les sémantiser et de les structurer. La structuration réside notamment en la distinction entre les métadonnées du document et le texte courant. Cela se fait dans un logiciel de traitement de textes – Word, OpenOffice ou autre. Il s'agit d'utiliser un modèle de document servant à la fois pour sémantiser l'information et établir une correspondance entre le fichier Word et le template (la maquette) auquel le fichier sera associé et nettoyer le document à l'aide de macros avant sa publication sur la plate-forme de Revues.org.
- 14 Le document ainsi préparé est prêt à être lodélisé, autrement dit à être chargé via l'interface de Lodel sur la plate-forme de Revues.org pour être mis en ligne.
- 15 En plus de ces formations, un site d'accompagnement éditorial – La maison des revues – a été créé à destination des utilisateurs de Lodel et est régulièrement mis à jour par l'équipe du pôle édition d'OpenEdition.
- 16 Cette volonté de mettre les outils de publication à disposition des chargés d'édition se fonde sur le principe d'appropriation. Il s'agit de donner les moyens aux chargés d'édition de mettre en ligne eux-mêmes leurs publications, sans pour autant qu'ils se transforment en informaticiens.
- 17 Lodel est un CMS permettant la mise en ligne d'une publication déjà éditorialisée ; c'est donc un outil de mise en ligne et non d'édition. L'édition de la publication doit être faite en amont.
- 18 Bien que la distinction soit importante entre éditer une publication et la mettre en ligne, on tend de plus en plus vers une action commune de valorisation des contenus. La dernière version de Lodel (Lodel 1.0) permet d'atteindre un niveau de granularité extrêmement poussé et donc, de bien valoriser les contenus mis en ligne sur la plate-forme.
- 19 Parmi les participants à l'atelier, certains, exerçant au sein de laboratoires, racontent leurs expériences d'« auto-édition » : fabrication de la revue, auto-diffusion, auto-distribution. Pour réaliser leur projet éditorial, ces personnes ont acquis les compétences nécessaires sur le tas. D'autres participants, encore, font part de leurs différents projets de revues électroniques².
- 20 D'autres projets d'édition électronique, comme ceux de textes en TEI, réunissent tous les corps de métier qui travaillent ensemble : développeur pour créer l'interface, chercheur pour rassembler les sources, secrétaire de rédaction pour réaliser la

transcription et la saisie des données. Le laboratoire est son propre éditeur et diffuseur avec la mise en ligne d'un site web dédié.

- 21 Cette pratique d'auto-édition a toutefois des limites car les compétences requises pour mener à bien les projets d'édition sont celles d'un éditeur, ce qui implique un travail de mise en valeur de ces textes. Le numérique ne change pas le métier d'éditeur profondément. Le premier rôle de l'éditeur, sélectionner des textes en vue de leur publication, reste le même.
- 22 Au-delà du métier d'éditeur, les statuts occupés dans la chaîne du livre sont divers : éditer, bien sûr, mais aussi diffuser et distribuer. L'inter-professionnalisation entre les différents acteurs de la chaîne du livre est essentielle afin qu'ils puissent réfléchir ensemble aux problématiques inhérentes à la chaîne du livre en prise avec le numérique.

Quel(s) modèle(s) économique(s) pour l'édition électronique ?

- 23 À l'heure où les reconfigurations actuelles impliquent de repenser le statut des différents acteurs de la chaîne du livre, et la collaboration des uns avec les autres, il semble primordial de s'interroger sur les différents modèles économiques pour l'édition électronique.
- 24 L'un des modèles économiques existant et présenté lors de l'atelier est celui du *freemium*. Ce modèle hybride repose sur l'association d'une offre gratuite, *free* et d'une offre payante, *premium*.
- 25 Le *freemium* a été adopté par le Centre pour l'édition électronique (Cléo) pour la diffusion des publications en libre accès. Concrètement, cela repose sur la vente d'un bouquet de revues, de livres et de services associés aux bibliothèques qui se sont abonnées au programme *OpenEdition freemium*. Une partie des revenus générés par la vente des bouquets est reversée par la suite aux revues.
- 26 Avec ce modèle économique se pose la question de ce que les éditeurs proposent à l'achat aux bibliothèques. Celles-ci achètent – ou louent – des contenus et des services dont le principal est l'accès aux contenus. Dans le cas d'un achat, par une bibliothèque par exemple, d'un accès aux formats détachables des publications en texte intégral (PDF, Epub) se pose la question de la pérennité de cet accès : une fois l'abonnement arrêté, la bibliothèque reste-t-elle propriétaire des fichiers qu'elle a achetés ? Il est essentiel que les éditeurs ou diffuseurs réfléchissent à ce qu'ils vendent aux bibliothèques.
- 27 L'adoption du *freemium* par le Cléo répond à la nécessité de trouver des solutions de financement pour le libre accès par la vente de services associés, qui eux sont payants : formats détachables (PDF et Epub), mise à disposition de statistiques de consultation, services de données, actions de formation et d'assistance...
- 28 Pour les éditeurs qui ont fait le choix d'adhérer au programme *OpenEdition freemium*, de nouveaux services leur sont proposés : une diffusion de leurs publications en librairies électroniques et de l'impression à la demande notamment.
- 29 Les collections des éditions de la Maison des sciences de l'homme (MSH) vont rejoindre le catalogue *OpenEdition Books*. Une moitié de la collection sera en libre accès, l'autre

moitié en accès restreint. Le choix est ainsi laissé à l'éditeur d'équilibrer son catalogue entre diffusion en accès libre ou en accès restreint.

- 30 Les publications du Collège de France sont également disponibles en *freemium* via le portail Openedition.org.

L'impression à la demande, un service demandé et apprécié

Qu'est-ce que l'impression à la demande ?

- 31 L'impression à la demande est l'impression à courts ou moyens tirages et permet d'avoir un petit stock de quelques cartons. Cette solution d'impression est choisie lorsqu'un éditeur souhaite minimiser l'investissement dans un titre mais être vendu en librairie. Il est préférable de lancer une impression à courts tirages si l'on sait que 100 à 1 000 exemplaires de livres sont attendus. Cette solution entraîne en effet des coûts d'entreposage et d'inventaire qu'il faut être sûr de pouvoir amortir par la vente future des ouvrages. Chaque année, les éditeurs publient de plus en plus de titres. Face à cette offre abondante, on assiste à une augmentation importante du nombre de titres invendus. Alors même que le tirage moyen de ces titres est en baisse constante. L'écrasante production de livres entraîne un roulement important de ceux-ci. Le cycle de vie des livres se réduit considérablement (compter 4 à 5 mois en librairie).
- 32 Ainsi, c'est tout naturellement que certains éditeurs ont fait le choix d'expérimenter une solution d'impression alternative, plus souple, et permettant, dans certains cas, de faire des économies. En effet, en choisissant d'imprimer de courts tirages, l'éditeur évite de se retrouver avec un stock d'invendus trop important, générateur de frais de stockage. Par ailleurs, l'éditeur évite également des frais supplémentaires dûs aux invendus.
- 33 Avec le POD (*print on demand*), on assiste donc à un changement de paradigme : l'éditeur ne s'engage plus à imprimer des exemplaires qu'il n'est plus sûr de vendre mais imprime des exemplaires à l'unité ou en petit nombre qu'il a déjà vendus.
- 34 Cette solution convient particulièrement aux éditeurs qui ont un catalogue d'anciens titres dont la future demande est difficile à prédire, comme les éditeurs universitaires par exemple.
- 35 Le cas d'OpenEdition Press est intéressant à ce titre. L'éditeur a fait le choix de proposer tous les livres qu'il publie à la fois au format HTML, sur le site des éditions OpenEdition Press, aux formats Epub et PDF en librairies électroniques et au format papier en impression à la demande.
- 36 L'expérience a notamment été menée avec le Médialab de Sciences-Po à l'occasion de la publication de l'ouvrage *Opinion Mining et Sentiment Analysis*³. L'ouvrage est disponible en libre accès sur le site de la collection Sciences-Po Médialab, au format Epub et PDF en librairie électronique et en version papier (impression à la demande) auprès de I6doc.com. Six mois après la mise en ligne de l'ouvrage en libre accès sur la plate-forme, une vingtaine d'ouvrages a été vendue en librairies électroniques, une dizaine d'exemplaires papier ont été vendus par le biais du prestataire d'impression à la demande.

- 37 Lulu.com et Lightning Source font partie des prestataires d'impression à la demande les plus importants du secteur

Lulu.com

- 38 Fondée en 2002, aux États-Unis, par Bob Young, l'entreprise s'est installée en France en 2006 et apparaît depuis comme un acteur solide du secteur. Depuis cette date, l'entreprise a su asseoir sa position et diversifier son offre à destination des auteurs, petits et moyens éditeurs en proposant un service d'impression à la demande et un service de création d'autres produits.

Lightning Source

- 39 Lightning source est une filiale du groupe Ingram, fondée en 1997. L'entreprise est implantée aux États-Unis, en Grande-Bretagne et en France. Elle s'appuie sur les infrastructures du groupe Ingram. Lightning Source oriente prioritairement ses services à destination des éditeurs.

- 40 Un autre acteur important du secteur de l'édition électronique en France, Publie.net, s'est lancé dans l'impression à la demande récemment. C'est grâce au partenariat avec Hachette Livre que François Bon a choisi d'expérimenter ce nouveau service notamment « dans une logique de distribution publique tous libraires⁴ ».

- 41 Publie.papier - c'est ainsi que se nomme cette récente initiative - repose sur la vente couplée de la version électronique de l'ouvrage avec la version papier (à des prix différents).

« D'autre part, nous avons décidé une *vente couplée* : mort à ce discours des éditeurs frigoristes qui parlent de *décote* du prix du livre numérique par rapport au livre papier – nous maintiendrons nos prix de diffusion numérique tous formats à 4 euros (ou 0,99, 1,99, 2,99 selon l'ouvrage), le livre papier sera vendu aux prix de 12 à 18 euros selon le format et complexité (et donc prix de revient de la préparation éditoriale), mais un code inséré dans le livre permettra l'accès libre et gratuit à la version numérique⁵. »

- 42 Le prix du livre au format papier est supérieur au prix du livre au format électronique. Cela permet de soutenir les libraires qui subissent de plein fouet ce changement de paradigme.

- 43 Par ailleurs, on constate que ce système de vente couplée se généralise : de plus en plus d'éditeurs proposent à la vente le livre papier enrichi de l'adresse de la version électronique du livre.

- 44 Toutefois, cette politique tarifaire en faveur du livre papier n'est pas toujours adoptée. Par exemple, Oxford University Press propose un prix de l'exemplaire numérique supérieur au prix papier, considérant que l'édition la plus riche – et elle l'est de fait ! – est la version numérique.

Et le libraire dans tout cela ?

- 45 Le libraire est beaucoup plus qu'un vendeur de livres ; il doit proposer des événements, de services, des lectures, faire vivre le quartier... Actuellement, 70 % des achats de livres des éditions de la Maison des sciences de l'homme seraient réalisés chez les

libraires. Ce gros système de diffusion existe et pèse lourd. Les éditeurs doivent donc soutenir les libraires et au besoin, proposer un prix différent en cas de vente couplée papier et format électronique (Epub).

L'édition multi-supports et de la diffusion multi-canaux

- 46 Les Presses universitaires de France n'ont pas encore franchi le pas de l'édition simultanée sur tous les supports. Actuellement, les ouvrages de la collection des « Que sais-je ? » sont publiés en numérique homothétique, autrement dit, ils présentent une forme similaire aux livres papier. *A contrario*, les livres publiés par OpenEdition Press ont d'emblée été destinés à être imprimés à la demande mais aussi publiés en ligne en texte intégral et également disponibles aux formats Epub et PDF en librairies électroniques.
- 47 Dans cette logique de développement de l'activité de l'éditeur, un critère est à prendre en compte : les aides publiques sont destinées en priorité aux éditeurs papier, qui publient des ouvrages à un tirage minimal de 500 exemplaires. Ce critère est d'autant plus difficile à remplir que l'on assiste à une érosion progressive du nombre de tirages. Alors qu'il y a encore quelques années, le tirage moyen d'un livre en sciences humaines et sociales atteignait 1 000 exemplaires, aujourd'hui il peut parfois n'être que de 200 exemplaires.

Le métier d'éditeur perdurera

- 48 Les nouvelles formes d'édition qui émergent remettent en cause non seulement le métier d'éditeur qui doit se réinventer, mais rendent également floues les frontières entre les différents acteurs de la chaîne du livre. Une certitude toutefois : le métier d'éditeur perdurera !

NOTES

1. Voir http://fr.wikipedia.org/wiki/Andr%C3%A9_Schiffrin
2. À ce sujet, voir le chapitre « Comment commencer, poursuivre et pérenniser une revue en ligne ? ».
3. Dominique Boullier, Audrey Lohard, *Opinion mining et Sentiment analysis*, Collection « Sciences Po | médialab », 2012.
4. Voir <http://www.tierslivre.net/spip/spip.php?article2856>
5. Ibid.

RÉSUMÉS

Dans *La Grande conversion numérique*, Milad Doueïhi écrivait que les éditeurs traditionnels sont des touristes découvrant un pays étranger qui les dérange et qu'au lieu de résister des quatre fers au changement et de pérenniser arbitrairement les anciennes normes (parfois avec aveuglement) dans un contexte où elles ne peuvent plus fonctionner telles quelles, ils doivent imaginer de nouvelles structures (Paris, Le Seuil, 2011, p. 303). Que les éditeurs SHS doivent suivre les manifestations des *humanités numériques* est une évidence. Comment doivent-ils se former pour créer de « nouveaux » livres ? Trouveront-ils goût aux changements rapides de la réalité numérique ? Continueront-ils d'être les partenaires privilégiés des auteurs, garants de leurs droits vis-à-vis des tiers, ainsi que de la qualité éditoriale et technique de leurs travaux ?

INDEX

Mots-clés : édition traditionnelle, humanités numériques, édition numérique

Évolution de l'archive ouverte HAL-SHS

Proposé par Laurent Capelli

Collectif

- 1 HAL-SHS est le portail en sciences humaines et sociales de la plate-forme d'archives ouvertes HAL (hyper articles en ligne), créé en 2001. La plate-forme SHS a, pour sa part, été lancée en 2004.
- 2 Le directeur adjoint du Centre pour la Communication Scientifique Directe (CCSD) qui gère et développe HAL depuis le début, Laurent Capelli, présente les évolutions prévues de la plate-forme HAL au printemps 2013, lors de la sortie d'une version 3 complètement réécrites. Comme toute application, HAL a vieilli, même si le code PHP a été retouché au fur et à mesure. Il était donc temps de faire une nouvelle écriture de l'application. Cette réécriture ne sera pas fondée sur un dépôt d'archives existant (type DSpace), mais sur le même cœur (*core*) que Sciencesconf.org, développé ces dernières années par le CCSD en parallèle avec HAL. Les évolutions les plus importantes concernent l'élargissement des types de documents déposables aux vidéos, la création d'un moteur de recherche, la possibilité de créer une page chercheur. Ces évolutions mènent à des interrogations plus globales sur le statut des documents déposés dans une archive ouverte et sur l'usage (détourné ?) de la plate-forme à des fins d'évaluation.

Intégration d'un moteur de recherche

- 3 Actuellement, il n'y a pas de moteur de recherche associé à HAL, au sens d'une indexation des métadonnées et des données déposées. En fait, le moteur actuel de HAL effectue de simples requêtes SQL dans la base de données MySQL, sans entrer dans le contenu des documents déposés. Il est donc prévu de développer un véritable moteur de recherche à facettes. Le software n'est pas encore choisi, entre AFS (une des briques d'Isidore) et Solr (moteur de recherche Open Source). Le futur moteur de recherche interrogera également les données Exif des images et des vidéos. Il y aura des limites

dans la recherche en plein texte dans les documents textuels, dans la mesure où les pdf ne sont pas structurés.

Élargissement de la typologie de documents déposables aux vidéos

- 4 La nouvelle version de HAL sortira au printemps 2013, sous la forme d'un nouveau portail dans lequel on pourra directement déposer tous les types de documents. La grande nouveauté est l'extension du type de documents déposables aux données de la recherche et en particulier à la vidéo, après l'expérience réussie de l'élargissement aux images (à partir de 2008, portail MediHAL créé en 2010). HALvidéo est développé dans le cadre de la v3 et sera intégré dans la nouvelle plateforme. La vidéo pose un problème spécifique de stockage, les fichiers vidéos étant beaucoup plus volumineux que des fichiers texte, PDF ou même image. Actuellement, le CCSD a la capacité d'héberger 5 000 heures de vidéos sur ses serveurs, mais cette capacité de stockage n'est pas limitée et peut être augmentée - rappelons que le CCSD se situe dans les locaux de l'IN2P3. 5 000 heures de vidéos correspondent à une journée d'enregistrement du LHC. Même si au départ le CCSD ne disposait pas de penta octets pour accueillir des dizaines de milliers d'heures de vidéos, le ministère a assuré mettre les moyens pour un portail vidéo conséquent.
- 5 Différents types de vidéos pourront être déposés : enregistrement d'une conférence ou d'un colloque, vidéo expliquant un article, etc.
- 6 Beaucoup de centres de recherche ont accumulé des heures de vidéos et les ont hébergées de manière disparate. Par exemple, le Centre Edgar Morin, qui a une cellule audiovisuelle depuis 20 ans, a déposé une partie de ses vidéos sur canal-u, une partie sur le réseau public parisien. Ce corpus a-t-il vocation à être hébergé sur HALvidéo ? Quels types de vidéos peuvent être considérés comme des archives ? On peut en effet distinguer divers types de documents et questionner leur caractère scientifique.
- 7 Pour saisir les enjeux liés au dépôt de vidéos, on peut se rapporter aux pratiques déjà en place avec la photographie, les images fixes. Par exemple, quel est le statut d'une reproduction numérique d'un manuscrit médiéval réalisée à l'IRHT ? S'agit-il d'archive de la recherche ou d'archive tout court ? Autre exemple, l'archéologue revient du terrain de fouille avec des photos : il peut les confier à son laboratoire ou les déposer dans MédiHAL.
- 8 La plateforme HAL n'est pas un dépôt légal, il y a une autre institution pour cela. Le dépôt sur HAL peut tout à fait coexister avec le dépôt légal.
- 9 On peut prendre l'exemple de MédiHAL, qui est utilisé par des laboratoires comme un complément à une plate-forme existante comme une photothèque internet mais ne gérant pas l'archivage ou la diffusion. MédiHAL sert alors à la fois à assurer l'archivage à long terme et à mieux diffuser les données. Dans d'autres cas, c'est l'ensemble de la photothèque qui a été basculée dans MédiHAL parce que le laboratoire n'arrivait plus à maintenir son application.
- 10 HALvidéo permettra ces deux usages : le dépôt individuel de vidéos liées à la recherche, comme source ou comme explication mise en scène, de conférences et colloques, mais

aussi le dépôt de collections de vidéos par des laboratoires ou des institutions de recherche

- 11 Dès l'origine, HAL a été pensé pour accueillir des articles et les données liées à cet article, en permettant le dépôt d'annexes. Dans la nouvelle plate-forme, il sera possible de faire des liens entre des productions scientifiques et des documents, des images, des séquences audiovisuelles.
- 12 HAL, pour aucun type de document, n'est un outil de traitement ou d'analyse des données. Il est seulement un outil de dépôt et de diffusion des données. Il stocke des documents avec des métadonnées minimales. Toute analyse ne peut se faire que dans une application tierce qui se branche sur HAL via une API. Cette API, qui existe déjà, sera également améliorée dans la version 3 de la plate-forme. Les métadonnées n'ont pas à être polluées par telle analyse d'untel à tel moment. De même, en ce qui concerne la vidéo, HAL fournira un outil de diffusion des vidéos : une fonctionnalité simple de chapitrage et d'annotation minimale des vidéos et la possibilité d'incorporer (*embedder*) la vidéo sur une page web (fonction qui sera automatisée pour les carnets de la plate-forme Hypotheses.org, en insérant juste l'identifiant de la vidéo, un lecteur sera affiché).
- 13 HAL fonctionne ainsi comme un gros disque dur virtualisé quelque part, que l'on peut interroger *via* une API, un protocole comme l'OAI-PMH, et diffuser via un flux RSS ouvert et personnalisable.
- 14 L'élargissement de la typologie des documents aux vidéos pose aussi la question d'autres types de documents : rapports ANR, mémoires de recherche, rapports de fouilles, articles de vulgarisation, billets de blog.

Le référentiel chercheur et la question de l'évaluation

- 15 Une autre fonctionnalité, demandée depuis longtemps par certains utilisateurs de la plate-forme, sera implémentée : l'ajout d'un référentiel chercheur permettra de rassembler les publications de l'auteur sur une page sur laquelle le chercheur pourra également ajouter son CV, sa bibliographie et où il pourra afficher quelques éléments « statistiques » sous la forme d'un graphe : avec qui il publie, de quels pays sont les co-auteurs, dans quelles revues, dans quelles catégories ou disciplines, à quel rythme. Ces éléments ne seront affichés sur la page du chercheur que s'il le souhaite. Ces statistiques ne valent évidemment que pour les documents effectivement déposés dans l'archive HAL.
- 16 Mais la question se pose de l'usage possible qui peut être fait de ces statistiques du point de vue de l'évaluation de la recherche et des chercheurs. On peut pointer l'incomplétude des données en question, critiquer le fait que HAL puisse être utilisé pour l'évaluation de la production scientifique des chercheurs alors qu'il n'est pas fait pour cela. Le simple fait de donner la possibilité d'afficher ces données sous la forme d'un graphe sur la page personnelle du chercheur met en lumière un usage que peuvent faire les instances d'évaluation d'une archive ouverte. On peut cependant remarquer que lesdites données existent de fait dans un dépôt d'archives ouvertes et qu'il suffit de les extraire et de les exploiter, ce que certaines instances font déjà. Même si HAL est un outil d'auto-archivage et non d'évaluation, il est de fait utilisé dans ce sens. Dans un climat général où les instances sont à la recherche de tous les indicateurs et de tous les

chiffres possibles, HAL est bien une source de données, parmi d'autres, servant à l'évaluation.

- 17 D'ailleurs, dès la création de HAL, les laboratoires ont demandé à pouvoir afficher les productions de leurs chercheurs. Dès 2000, la demande du CNRS portait sur l'estimation du nombre de publications du CNRS dans arXiv.org, l'archive des physiciens. Il fallait donc pouvoir afficher les publications de chaque institution, générer des listes bibliographiques. Par la suite, certains établissements ont mis en œuvre une politique volontariste d'incitation voire d'obligation destinée aux chercheurs afin qu'ils déposent l'intégralité de leur production dans HAL. C'est le cas, par exemple, de l'université de Liège ou de l'Institut national de recherche en informatique et en automatique (INRIA, France), dont les chercheurs seront poussés à déposer tous leurs articles à partir de 2013 sur son site propre sur HAL. Concrètement, un chercheur de l'INRIA ne sera pas autorisé à publier un article si son *preprint* n'est pas déposé au préalable dans HAL. Pour conclure sur ce point, il semble important d'indiquer que HAL n'est pas un outil d'évaluation, mais un réservoir d'articles scientifiques décrits de la façon la plus correcte. Concernant les statistiques affichées pour les dépôts, on peut suggérer certains affinements et un ajout : lors de la consultation des fiches, il faudrait que le passage des robots soit différencié des visites ; et qu'une statistique du moissonnage de l'article via le protocole OAI-PMH soit affichée, permettant de savoir dans combien de bases, voire dans lesquelles, l'article a été moissonné et est donc présent.

Dépôt du texte intégral et de notices simples

- 18 Le dépôt d'un article sur une archive ouverte comme HAL n'est pas sans poser de difficultés dans certaines disciplines, en particulier vis-à-vis de certains éditeurs et revues de SHS qui n'acceptent pas d'articles ayant déjà été déposés dans HAL. De ce point de vue, le rapport de force actuel, en particulier avec les éditeurs privés de revues, n'est pas forcément à l'avantage des chercheurs. On peut cependant inciter les chercheurs à négocier avec leurs éditeurs la possibilité de déposer la version d'auteur dans une archive ouverte comme HAL. La question demeure de la manière de valoriser pour le chercheur également les notices correspondant à des articles qui n'ont pas pu être déposées par l'éditeur.

Épisciences.org : création d'épi-revues

- 19 Il est également prévu, dans la nouvelle version de HAL, de pouvoir créer des épi-revues qui se distinguent des simples collections existantes. Un groupe scientifique pourra constituer un comité de lecture qui sélectionnera des papiers déposés dans le réservoir, les évaluera et proposera à leur(s) auteur(s) de les publier dans ladite épi-revue.
- 20 Une épi-revue se distingue d'une revue à qui les articles sont soumis, via un appel à contribution ou de manière volontaire. Dans le cas d'une épi-revue c'est en effet le comité de rédaction qui effectue une veille ou une recherche sur ce qui est effectivement déposé dans l'archive ouverte et qui va proposer à l'auteur d'intégrer son article dans l'épi-revue en question. Dans d'autres cas, un appel à contribution classique peut être lancé et la soumission des articles proposés s'effectue alors

simplement par un dépôt en archive ouverte. Un travail éditorial (sur le fond et la forme) va alors être réalisé sur l'article au format papier, et l'auteur pourra déposer une version 2 – la version 1 étant conservée dans HAL suivant le principe qu'aucun document déposé ne peut être retiré. On rappelle que le CCSD ne valide pas les articles déposés après évaluation scientifique, mais seulement après évaluation sur le plan technique (format, fichier, métadonnées, type de document). L'idée d'une épi-revue est donc de faire quelque chose, à un niveau scientifique, des contenus déposés dans le réservoir. Cela ne revient donc pas au même que d'effectuer un tamponnage des articles, puisqu'il y a un travail de sélection scientifique.

- 21 Cette fonctionnalité est née d'une demande, d'un besoin de créer des journaux scientifiques en validant des articles déposés dans HAL – ou dans un autre réservoir. Il faut noter qu'ArXiv.org, qui a inspiré la création de HAL, n'a pas évolué vers ces nouvelles fonctionnalités.
- 22 Ce nouvel objet éditorial a répondu à une demande de chercheurs de certaines disciplines, en particulier en mathématiques et en physique alors que HAL n'en était qu'à ses débuts. L'INRIA y a joint sa propre demande. Les sciences humaines et sociales sont moins concernées par cette fonctionnalité, hormis la revue TAL (Traitement automatique des langues), qui utilise d'ores et déjà de manière détournée Sciencesconf.org pour gérer l'évaluation des articles. Cette demande vient de revues hors SHS qui ne disposent pas d'une plateforme de diffusion électronique telle que Revues.org destinée aux revues en SHS.
- 23 Il s'agit donc d'un nouvel outil développé par le CCSD, Episciences.org, qui sera connecté à HAL et permettra de développer des épi-revues. Son *workflow* se calera sur les pratiques courantes d'évaluation des publications scientifiques. Il sera donc proche de celui sur lequel repose OJS (Open Journal Systems) utilisé par le service de Manuscrits proposé par Revues.org. Il permettra de nommer des lecteurs externes, de créer une grille de lecture, d'échanger avec l'auteur qui pourra faire une version 2 ou 3 modifiée et la déposer sur HAL, de créer un numéro, etc.
- 24 Pour aller plus loin, il est possible de consulter un article sur la question des épirevues à partir des archives ouvertes comme HAL, « De la numérisation des revues à leur déconstruction numérique »¹. Cette initiative peut être rattachée au Budapest Open Access Initiative (BOAI) de 2001 : mercredi 13 février 2002, un collectif de chercheurs a déclaré la guerre aux revues scientifiques éditées par des « grands éditeurs » comme Springer-Verlag ou Elsevier. La création de revues alternatives, en ligne et gratuites, est pensée comme un moyen pour empêcher les grands éditeurs de prendre à terme le contrôle de la production de la littérature grise. Ce débat reste d'actualité en 2012, avec d'une part, le lancement d'une pétition de chercheurs appelant à ne plus soumettre d'articles aux revues de Springer ou d'Elsevier et d'autre part, la multiplication d'appels à la publication des résultats de la recherche en open access.

Conclusion

- 25 Pour conclure, cette nouvelle version de HAL comprend un élargissement de la typologie des documents déposables, avec l'introduction de la vidéo (et des services afférents : chapitrage et lecteur exportable), d'un véritable moteur de recherche, et un développement des fonctionnalités (référentiel chercheur et pages personnelles, Episciences.org). Ces élargissements suscitent des interrogations légitimes sur la

frontière entre les documents recevables et ceux qui ne le sont pas - ce qui est particulièrement aigu pour la vidéo où les fichiers sont lourds et pour les questions de droits à l'image - et sur les usages de la plate-forme pour l'évaluation, une demande émanant des institutions, mais soumise à la critique par les chercheurs.

NOTES

1. Jean-Max Noyer, Gabriel Gallezot, Olivier Ertzscheid, Ghislaine Chartron, "De la numérisation des revues à leur déconstruction numérique", *Les Cahiers de la SFSIC*, n°2, avril 2008, p.15-17.

RÉSUMÉS

Le CCSD est en train de repenser son archive ouverte HAL. Il propose d'en dévoiler les futurs contours : le chercheur au centre du dispositif, moteur de recherche, dépôt de vidéo, épirevues, etc.

INDEX

Mots-clés : archives ouvertes, HAL, HAL-SHS, Open Archives, vidéo, identité du chercheur, évaluation

Open data en SHS

Proposé par Cynthia Pedroja, Elifsu Sabuncu, Anne-Laure Stérin

Collectif

NOTE DE L'ÉDITEUR

Cet atelier résulte d'une fusion entre les propositions « Open data en SHS. À l'heure où l'on ne parle que de l'ouverture des données de la recherche, comment les chercheurs en SHS envisagent-ils cette question ? » et « Comment diffuser en ligne les résultats de la recherche. Questions de droit et d'éthique ».

- 1 Open data d'après Wikipedia : Une **donnée ouverte** (en anglais *open data*) est une information **publique brute**, qui a vocation à être **librement accessible** et **réutilisable**. La philosophie pratique de l'open data préconise une libre disponibilité pour tous et chacun, sans restriction de copyright, brevets ou d'autres mécanismes de contrôle.¹

Introduction

« Open data en SHS : À l'heure où l'on ne parle que de l'ouverture des données de la recherche, comment les chercheurs en SHS envisagent cette question ? »

- 2 La première proposition d'atelier concernait la diffusion en open data des données numériques issues de la recherche en sciences humaines et sociales (SHS). Quel lien peut-il y avoir avec le mouvement de l'open data, dans un contexte de volonté politique et sociale de mise à disposition des données publiques² ? Les données des SHS relèvent-elles du même statut ? Quel type de données diffuser ? Est-il pertinent de les rendre accessible à plus ou moins grande échelle ? En prenant l'exemple d'un travail sur un projet d'histoire des sciences, où les données sont principalement des enregistrements d'entretiens, quelle forme devrait prendre l'ouverture de ces données (publication en ligne, mise à disposition des données au laboratoire, etc.) ? Le chercheur peut-il conserver ses données lorsqu'il change de laboratoire, ou bien doit-il les laisser au laboratoire ? À qui appartiennent réellement les enregistrements d'entretiens ? Dans

quelle mesure un laboratoire peut-il revendiquer la propriété des données alors que le chercheur est tenu par un contrat moral avec l'interviewé ?

« Comment diffuser en ligne les résultats de la recherche. Questions de droit et d'éthique ».

- 3 La deuxième proposition d'atelier recoupait la première, au sens où des questions de droit et d'éthique sont soulevées par la mise à disposition des données de la recherche en SHS. Depuis 2011, un groupe de travail transversal et ouvert a été créé à la MESHS de Lille. Ce groupe, initié par Véronique Ginouvès³ et Pascal Garret de Créville, rassemble des chercheurs, des archivistes, des documentalistes, des juristes, etc., autour de ces questions. Des journées d'études valorisant les retours d'expériences ont déjà été organisées, et continueront à l'être régulièrement. Toutes les disciplines de SHS ainsi que tous les types de supports (sonores, visuels) sont étudiés. Une des principales difficultés soulevées par ce travail réside dans la définition même de ce qui constitue les données en sciences humaines et sociales.
- 4 Comme les enjeux de l'open data et de la publication des données en SHS sont à la source de nombreux questionnements, il est important de définir clairement l'objet de cet atelier, né d'une fusion entre ces deux propositions initiales : La première proposition consiste à déterminer « à l'heure où l'on ne cesse de parler d'ouverture des données et d'open data, comment les chercheurs en SHS envisagent cette question ? ». On se situe ici dans une logique où les chercheurs vont produire des données (c'est-à-dire mettre en place des protocoles qui vont permettre de produire les données). D'autres chercheurs peuvent vouloir avoir accès à ces données, alors de nouveaux protocoles sont à inventer et à mettre en place. La deuxième proposition concerne de ce fait « la manière de diffuser en ligne les résultats de la recherche. C'est une question juridique et éthique ». Dans le cadre du groupe de travail constitué à la MESHS de Lille, cette même problématique est abordée, mais à l'échelle de centres de recherche ou d'archives qui ont des fonds et se posent la question de leur mise en ligne.
- 5 L'atelier porte-t-il sur des problématiques juridiques concernant la mise à disposition de données en SHS ou pose-t-il plus généralement la question de la pertinence de la diffusion des données en open data ? À certains égards, les chercheurs étant des fonctionnaires publics, on pourrait considérer que la diffusion des données en SHS devrait être automatique. Cependant, tout le monde n'est pas partisan de l'open data. L'idée de cet atelier est aussi de se faire l'avocat du diable pour poser les bases d'un argumentaire : « Prouvez-moi que vous voulez avoir mes données et que vous allez en faire bon usage ! ». L'objectif final de cet atelier serait de construire un argumentaire qui permettrait de dire en quoi il est pertinent de diffuser les données des SHS et comment on peut le faire (en pratique et selon quels protocoles).

Quelle est la pertinence d'une diffusion des données SHS en open data ?

- 6 La discussion s'engage sur les intérêts et les inconvénients pour un chercheur en SHS à partager ses données et la possibilité de mise en place de barrières mobiles pour protéger le travail en cours. Dans certains cas, le choix de diffuser peut également représenter une prise de risque assumée en termes légaux (par exemple pour la publication d'un cliché dont on n'a pu retrouver toutes les personnes représentées).

- 7 Déterminer l'intérêt pour le chercheur de mettre à disposition ses données en open data est-il un point crucial ? Tous les freins juridiques mis en avant semblent souvent être des paravents. La question que se posent en fait les chercheurs est plutôt de savoir *qui va exploiter les données diffusées ?* La deuxième question est de savoir dans quel état d'esprit ouvrir les données ? Deux positions sont envisageables : La première est de le faire seulement dans le cas où toutes les garanties (juridiques) sont possibles. L'autre position consiste à mettre en ligne d'abord, et voir s'il y a un problème ensuite.
- 8 Lorsque la donnée est accompagnée de son contexte de production, un grand nombre des questions éthiques soulevées par la publication ne se posent plus. En effet, contextualiser les données permet de donner les clefs pour leurs utilisations futures. On évoque au cours de la discussion la possibilité de passer par la création de licences spécifiques, du type creativecommons⁴, qui permettrait d'offrir un cadre pour la réutilisation des données et ainsi rassurer certains chercheurs. Cependant, même si les licences sont des dispositifs ingénieux, leur multiplication à l'infini sur chaque problématique juridique ne constitue peut-être pas une solution. L'enjeu de l'open data est précisément de permettre la réutilisation des données, et certaines licences peuvent par trop restreindre ce réemploi.

À qui appartiennent les données ?

- 9 Dans un certain nombre de domaines, comme le médical ou la biologie, les carnets de recherche appartiennent aux laboratoires. Il est frappant de constater qu'en SHS, on considère souvent que les données appartiennent aux chercheurs qui, bien souvent, quittent leur laboratoire avec. Dans bien des cas il n'y a pas de capitalisation au niveau du laboratoire sur le travail des chercheurs. Les générations successives d'un laboratoire d'histoire de l'art réaliseront par exemple successivement les mêmes clichés sans jamais les partager.
- 10 Plus généralement, les archives des chercheurs sont souvent considérées comme des données privées par les institutions de rattachement elles-même. Pourtant, d'un point de vue strictement légal, ces archives sont publiques ! L'institution qui paye le chercheur peut, théoriquement, se retourner contre le chercheur en cas de destruction.
- 11 Cependant, de réelles questions se posent concernant la diffusion de ces archives. Ainsi, les archivistes des archives publiques (départementales) sont en pleine discussion sur le fait de savoir ce qu'ils veulent/peuvent ou pas diffuser. Le risque est de se retrouver dans la même situation avec des chercheurs qui pourraient se trouver privés de statut. Lorsqu'on parle de partage et d'open data, une manière de sortir du problème consiste peut-être à considérer que la publication de sources est partie prenante du travail scientifique. Il faudrait donc distinguer le travail de recherche en train de se faire du travail de recherche abouti. Dans le cas des SHS, cela pourrait notamment permettre la mise en place d'une véritable critique des sources.

Comprendre l'open data au-delà de la seule recherche académique

L'importance de la restitution aux populations

- 12 Dans certaines disciplines, telle que l'ethnologie, il est important que les données soient restituées aux populations étudiées et qu'elles puissent éventuellement se les approprier. Cette demande de mise à disposition des données est d'ailleurs très récurrente lorsque les études suscitent une participation des citoyens. Une politique d'open data en SHS aurait l'avantage de répondre à cet enjeu.

Responsabilité sociale du chercheur et participation du public

- 13 Les chercheurs ont aussi une responsabilité sociale qui implique une diffusion de l'information. Quand on parle d'Open data, on parle également du grand public et il est important de ne pas rester trop « chercheur centré ». Comment transférer de l'information scientifique validée à d'autres communautés ? Comment permettre, et même favoriser des réappropriations ? Bien sûr, une telle ouverture des données peut soulever de nouvelles difficultés. Par exemple, sollicité au sujet du projet wikidata, le CNRS a répondu que les chercheurs avaient des droits d'auteurs sur leur contribution, or une telle réponse impliquerait de contacter un par un les 8 800 chercheurs du CNRS afin de pouvoir réutiliser leurs contributions. Lorsque ces mêmes chercheurs publient leurs données sous licence libre, dans PLOS⁵ par exemple, on est obligé de passer par un intermédiaire privé (une fondation) pour récupérer des données financées sur des fonds public...

Un changement des mentalités nécessaire

- 14 Un travail de fond doit être mené pour faire évoluer les mentalités. Un tel changement passe sans doute par des évolutions concernant l'évaluation des thèses. Tant qu'un doctorant n'a pas publié ou remis son travail, il est normal qu'il ne souhaite pas mettre ses sources à disposition. Cependant, on pourrait tout à fait imaginer une sorte de contrat doctoral stipulant que les données soient rendues accessibles au terme de la thèse. Il serait intéressant que ces nouveaux modèles de diffusion commencent à prendre une réelle dimension dans plusieurs champs disciplinaires. Pour l'instant, la seule discipline mettant en pratique un tel principe est l'archéologie préventive. L'inventeur ou le découvreur d'un artefact archéologique original possède l'exclusivité pendant deux ans, cette exclusivité est éventuellement prolongeable sur cinq ans. Au terme de cette période de cinq années, et dans tous les cas, les données sur l'artefact doivent être rendues accessibles aux autres chercheurs.

La nécessaire définition de protocoles et de formats pour la diffusion des données

La diffusion des sources peut être sensible

- 15 Certaines données ne sont pas diffusables. Par exemple, on ne peut pas rendre compte de la même manière des choses lorsqu'il s'agit de périodes contemporaines avec des témoins de violences, génocides, etc. Pour les périodes anciennes, la diffusion des sources est beaucoup moins problématique.

La définition des protocoles

- 16 Dans le groupe de travail de la MESHS de Lille, des propositions de protocoles ont commencé à être envisagées avec l'idée qu'ils puissent être utilisés par d'autres. Par exemple, la prochaine journée d'étude sera consacrée à la mise à disposition de données de la recherche concernant les fichiers d'enquêtes produits par des organismes publics ou parapublics (INSEE, INED, Ministères, CNRS)⁶.
- 17 L'idée est de permettre aux gens de savoir comment diffuser leurs données soit selon des modalités restreintes, soit en ayant recours à des modèles plus ouverts. L'idée est avant tout de décrire des situations concrètes, de partir des pratiques, et d'envisager les diverses solutions possibles, en présentant leurs avantages et leurs inconvénients.

Trouver des formes de collecte et des formats pour permettre à terme le partage des données

- 18 Définir des protocoles et des formats pour le partage de données suppose également de trouver des formes de collectes qui permettent à terme ce partage. Il est de ce fait important de partager les expériences de base de données collaboratives⁷ et de valoriser le capital énorme de collecte individuelle de toutes les données primaires produites par le chercheur, stockées sur son ordinateur personnel et qui seront perdues après publication de travaux. Au bout de quelques années, certaines données ne sont même plus lisibles car les technologies utilisées pour les produire n'existent plus !
- 19 Pour diffuser les données, il faudrait d'abord s'accorder sur la manière dont on pourrait structurer ces informations pour pouvoir les partager, ce qui n'est pas sans difficultés, car l'apprentissage de nouvelles méthodes ou de nouveaux outils est souvent laborieux. Par exemple, dans un laboratoire CNRS, l'intervention d'ingénieurs spécialisés en traitement des données est souvent perçue comme une intrusion dans le travail des chercheurs. Lorsqu'on propose de passer à un nouvel outil, cela pose problème par rapport à un processus mental qui s'est mis en place depuis des années. Proposer un nouvel outil, c'est souvent remettre en cause la méthodologie du chercheur ; ce qui n'est d'ailleurs pas complètement faux.
- 20 L'autre difficulté provoquée par le changement des méthodes de travail relève d'une certaine peur quant à la mutualisation – même si en échange le chercheur pourrait profiter de ce qu'ont récolté ses collègues. La seconde étape, après la mutualisation, réside dans la publication sur Internet. Lorsque l'on rend les données accessibles, il se

peut qu'un autre groupe de recherche travaillant sur la même population puisse accéder à ces données et les rende visibles sur un site tiers – mais la réciproque est également possible. Comment gérer ce défi ?

- 21 Le groupe de la MESHs de Lille n'a pas encore vraiment abordé la question des formats. La question de la mise à disposition des données est cependant structurée en cinq temps : préparer, collecter, traiter, diffuser et réutiliser les données/documents/etc.

Création d'une plate-forme pour l'échange de bonnes pratiques

- 22 Il serait intéressant de disposer d'un lieu au niveau national, ou de la communauté des humanités numériques francophones, pour l'échange de bonnes pratiques en matière de production des données. L'objectif serait de se mettre d'accord sur la manière de les publier et de les rendre interopérables. On pourrait les mettre à disposition dans une sorte d'« entrepôt général ». Une des difficultés réside dans le fait que la diffusion des données n'est pas valorisée dans la carrière des chercheurs. C'est un véritable travail que de diffuser des données et cela prend énormément de temps. En terme de valorisation de ce travail de diffusion, il existe l'exemple d'une plate-forme qui s'appelle NUMES (abes/adonis)⁸. Cette plate-forme est mal connue mais est interopérable avec le projet patrimoine numérique⁹. L'objectif de NUMES est d'inventorier les différents corpus de données mis en ligne.

Former les étudiants au traitement des données

- 23 Afin de changer les mentalités, la proposition d'un « contrat » entre l'étudiant et son laboratoire semble envisageable. Cela pourrait prendre la forme d'un engagement réciproque : l'étudiant devrait libérer ses données dans un délai raisonnable après la soutenance de sa thèse, mais en retour l'école doctorale aurait l'obligation de fournir à l'étudiant un accompagnement et une formation pour préparer ses données (équipe informatique, formation sur encodage, logiciel, etc.) Plus généralement, cette idée de contrat doctorant/laboratoire soulève la question de l'apprentissage et de la formation des étudiants. Il faudrait même envisager des formations avant le niveau du doctorat. En effet, comme le soulignait Paul Bertrand lors de sa non-conférence, *Les digital humanities sont-elles solubles dans le steampunk*, les étudiants d'aujourd'hui sont plutôt « facebook native » que « digital native ».
- 24 Changer les mentalités et les pratiques passe par une transformation radicale de tout le système universitaire. On pourrait imaginer n'accepter en thèse que des étudiants ayant deux diplômes ou, plutôt que de rallonger les parcours, penser à intégrer des formations de traitement et valorisation des données dans les cursus. Dans tous les cas, que cela soit pour un double cursus ou une formation spécifique proposée au cours des premiers cycles, il est important que des intervenants extérieurs au monde académique puissent assurer des cours afin de favoriser les échanges de compétences. Par exemple, dans les IUT, un minimum d'intervenants doivent venir de l'extérieur du monde académique, ce qui est important pour qu'un échange se crée avec l'extérieur. L'université ne doit pas rester un monde clos, et les industriels pourraient venir présenter les dernières nouveautés dans le domaine du numérique. Il est cependant important que les étudiants ne soient pas formés exclusivement sur des aspects

techniques, ils doivent être sensibilisés à la qualité des données, aux aspects juridiques, etc.

Engager une réflexion d'ensemble dans les SHS sur les pratiques de collecte, la formation

- 25 Les discussions de cet atelier sur l'Open Data et les SHS ramènent sans cesse à la toute première discussion du THATCamp Paris 2010. On est en fait en train de changer plusieurs paradigmes à la fois dans le monde des sciences humaines et sociales. D'une part les données accessibles pour le chercheur sont aujourd'hui devenues des flux d'information. D'autre part, le travail classique du chercheur consistait à produire une somme de connaissances puis de la publier. Ce qui est en train de changer, c'est que cette publication devient en elle-même un flux aujourd'hui. D'où les difficultés exprimées par les chercheurs tout à l'heure, puisque cela ne correspond plus à la représentation que l'on se fait d'une publication scientifique.
- 26 La publication scientifique aujourd'hui n'est plus la somme des connaissances à un instant t . C'est faute d'une réelle prise en compte institutionnelle de cette dimension qu'on a du mal à appréhender les difficultés que pose la mise à disposition des données. Cette notion de flux de connaissances est très visible en histoire de l'art avec l'exemple de la rédaction des catalogues raisonnés. Les chercheurs souhaitent souvent absolument pouvoir figer la recherche sur le papier afin de pouvoir y mettre leur nom, quand bien même cette publication deviendrait rapidement obsolète. La production d'un objet numérique oblige une réelle prise de conscience par le chercheur du fait qu'il n'est pas le seul à avoir produit la donnée scientifique, et que beaucoup d'autres personnes ont travaillé les décennies précédentes sur ce sujet avant lui.

Conclusion

- 27 Le monde de la recherche en SHS est confronté au changement des règles de la publication scientifique et au changement même du sens de cette publication des données. S'il est pertinent de vouloir rendre accessibles ces données, on doit distinguer celles concernant les recherches en cours et celles concernant des recherches abouties. Afin de réduire la méfiance que peuvent éprouver les chercheurs en SHS par rapport à la diffusion de leurs données, les efforts des prochaines années devront sans doute porter sur la responsabilisation des chercheurs ou la contractualisation de ces questions dans le cadre des formations avec la mise à disposition de moyens pour les aider. Il sera également important de rassurer les producteurs de données à propos de la mise à disposition de ces dernières. La création de licence ou de barrière juridique pour limiter les usages possibles est un exemple de proposition. Une autre difficulté soulevée au cours de cet atelier concerne la difficulté à valoriser le temps passé à mettre à disposition les sources ou les données. Il faut donc aller vers un contrat, une responsabilisation de toutes les parties.

NOTES

1. http://fr.wikipedia.org/wiki/Open_data.
 2. Par exemple, la plate-forme d'ouverture des données publiques du gouvernement français (Open Data) : <http://www.data.gouv.fr/>
 3. Véronique Ginouvès, phonothèque de la MMSH : <http://phonothèque.mmsh.univ-aix.fr>
 4. <http://creativecommons.org>
 5. <http://www.plos.org>
 6. <http://www.meshs.fr/PUDL/>
 7. Par exemple, http://larhra.ish-lyon.cnrs.fr/Pole_Methodes/symogih_accueil_fr.php
 8. <http://www.numes.fr>
 9. <http://www.patrimoineecrit.cultur.gouv.fr/Num.html>
-

RÉSUMÉS

Question de la collecte des sources et de leur publication en ligne. Question des autorisations. Discuter dans l'atelier avec d'autres chercheurs de ces questions et sortir avec une check list pour discuter de ce qu'on doit faire, des risques, et si l'on décide de le faire savoir comment le faire. À l'heure où l'on ne parle que de l'ouverture des données de la recherche, comment les chercheurs en SHS envisagent-ils cette question ? Si la question des sources primaires (documents d'archive, bibliothèques numériques...) est désormais bien balisée, la question des données de terrain (entretiens, relevés d'observations...) est plus complexe : comment anonymiser les témoignages recueillis ? Des informations recueillies dans le cadre d'un protocole basé sur la confiance peuvent-elles être libérées sans contrainte ? Des entrepôts d'open data en SHS peuvent-ils être mutualisés, et si oui à quelle échelle ? Nous proposons de confronter les expériences et réflexions pour faire le point sur cette question épineuse.

INDEX

Mots-clés : open data, données, ouverture des données, données de la recherche

La diffusion des sciences humaines et sociales vers un large public : panorama des outils de médiation numérique

Proposé par Isabelle Jouve et Lars Wieneke

Collectif

Introduction

- 1 De quels outils avons-nous besoin pour médiatiser nos connaissances vers un plus large public ? Qu'est-ce qui rend un outil plus efficace qu'un autre ? Quelles sont ses contraintes financières ou techniques ?
- 2 L'idée est de confronter les différents moyens de médiation, de les formaliser et de tenter de les évaluer sur des critères encore à établir.
- 3 Le tour de table des 23 participants montre qu'une large palette de médiations est représentées : théâtre, captations (archives), bandes dessinées, expositions, etc. Les acteurs de la médiation dans l'édition sont présents mais pas dominants et ceux qui utilisent les outils numériques sont encore minoritaires autour de la table.
- 4 Il existe une dichotomie entre les expériences de médiation faites dans un espace physique, où l'on s'adresse à un ou plusieurs visiteurs aisément identifiables et les expériences de médiation dématérialisée entre internautes dont la cible n'est pas identifiée. Selon les types de médiation, un musée, une exposition, un réseau social ou un blog, la médiation suit donc des logiques différentes. Quels sont alors les outils à préconiser ?

Un exemple de médiation numérique : Le projet Goulag

Description du projet

- 5 Une équipe internationale de chercheurs a rassemblé de nombreux témoignages d'anciens déportés, des photos prises au cours de leur vie, des documents d'archives privées et publiques, des films. Ce corpus devait être rendu accessible non seulement à la population concernée mais à un public de l'ex-orbite soviétique, l'objectif étant de faire naître une histoire commune européenne, d'aller au-delà des archives sur le Goulag.
- 6 Pour rendre le discours plus « immédiat », la solution choisie a été la création d'un musée en ligne qui privilégie le son et l'aspect visuel mais, malgré la pluralité des médias utilisés, la structure de ce musée virtuel rappelle le musée classique avec des salles thématiques et des salles biographiques. Parallèlement une base de données Drupal a été créée permettant aux chercheurs de compléter les salles du musée par des biographies, des cartes, des témoignages, etc.
- 7 Une des difficultés rencontrées a été de convaincre les chercheurs de la nécessité d'impliquer outre des développeurs, d'autres corps de métiers, comme par exemple des graphistes. Mais, ce sont les chercheurs eux-mêmes qui ensuite, étaient chargés de remplir le contenu. Avec le temps, le problème de la pérennisation de la base de données et du musée virtuel s'est posée. Le site est consultable en quatre langues : français, russe, polonais et anglais.

Questions des participants :

- 8 Quels ont été les retours ? Quelles actions de médiation ont été entreprises ?
- 9 Il y a eu des émissions à la radio (Radio France internationale - RFI), une publication dans la collection *Autrement*, une indexation sur Isidore, des journées d'études avec les témoins. Par contre, il n'y a pas eu de diffusion sur les réseaux sociaux. Chacune des actions a renforcé les visites sur le site. Par ailleurs le projet a été présenté dans les alliances françaises des pays concernés, mais cela a suscité des résistances dans les populations.
- 10 Le suivi du projet a été fait grâce l'exploitation des statistiques sur la consultation du site, de localisation géographique des connexions, des pages les plus consultées chaque jour. Une salle a été « dédiée » aux commentaires des visiteurs (type boîte de dialogue).
- 11 Mais, depuis la fin de l'ANR, l'équipe a été réduite et la question de comment faire vivre le site, comment joindre toujours plus de public et comment interagir avec, se pose de manière plus cruciale.

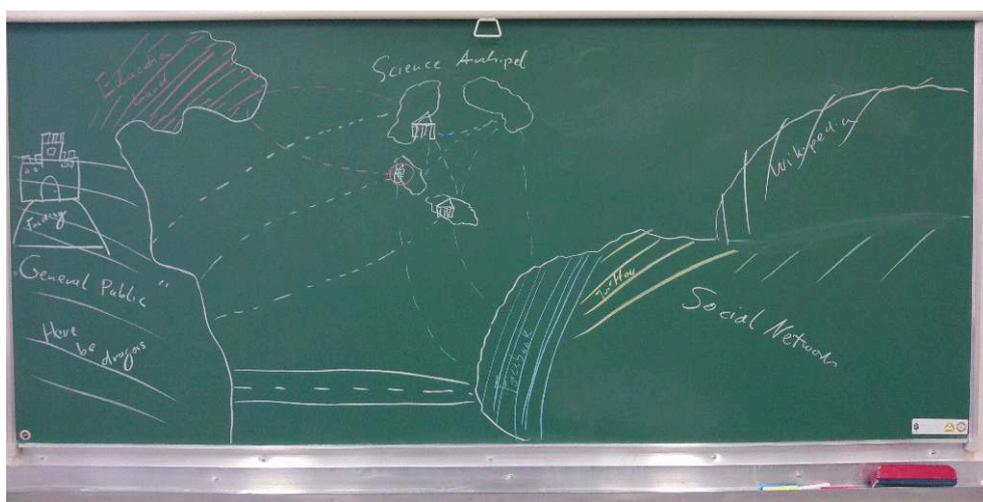
Communication et médiation, quelle différence ?

- 12 À quel public s'adresse-t-on et quel contenu lui offre-t-on ? La médiatisation est d'autant plus efficace que public est bien identifié. Dans un musée physique, le contenu peut être apporté aux visiteurs et le médiateur a la possibilité d'adapter son discours à son public. Dans le cas de la médiation sur Internet, le public doit venir chercher le contenu, sa réception sera donc amoindrie. C'est là qu'il faut bien différencier la

communication de la médiation. Avec le numérique, il est donc nécessaire de créer un lien avec le public et pour cela de bien l'identifier. Le simple fait de numériser ne suffit pas à garantir diffusion et accès, l'aspect humain reste très important.

- 13 Pour étayer ce propos, un chercheur médiateur raconte ses expériences de médiation faites en des classes difficiles ou en prison. Dans les deux cas, le fait d'être allé vers le public et d'avoir adapté son discours à ses interlocuteurs, ont amélioré grandement le dialogue et l'information est beaucoup mieux passée. Au Museolab du centre Erasme, le contenu des visites n'est pas fixé d'avance. Le médiateur peut valoriser tel ou tel contenu sur un I-Pad en fonction de la discussion en cours avec les visiteurs. Ce procédé est évidemment très coûteux et malheureusement, la plupart du temps, ce type d'action n'est pas suffisamment reconnu pour obtenir des budgets.
- 14 La médiation existe aussi sur Internet. Par exemple, la bibliothèque numérique d'histoire des sciences propose un texte classique commenté par un scientifique moderne. Le public peut réagir sur un blog, avec des réactions parfois d'autant plus virulentes que le public est virtuel, ou encore retro-acter sur le contenu.
- 15 À la Bibliothèque nationale de France (BNF), l'accessibilité aux fonds numérisée s'est posée. Comment éveiller l'intérêt du public pour l'énorme quantité de documents numérisés ?
- 16 Le choix de l'équipe de la BNF a été de ne pas se prendre au sérieux, de s'adresser directement au public en jouant avec l'actualité, en particulier sur facebook et Twitter (énigme du vendredi, reconnaissance d'un portrait). Le médiateur peut s'effacer au profit des internautes qui prennent le relais avec leurs échanges. Ce choix d'interaction avec le public associé à l'humour décalé s'est avéré un franc succès. En termes de moyens et de ressources, cinq personnes représentant un plein temps assurent la veille. Pour mettre en œuvre cette politique de médiation commune, elles sont toutes spécialisées sur les contenus et familiarisées avec l'usage des réseaux sociaux.

Figure 1



Une carte de la répartition des différents acteurs de la médiation¹ :

L'interaction dans le numérique

- 17 Le but n'est pas seulement de confronter des points de vue différents, mais de faire en sorte qu'il y ait une incidence sur le contenu lui-même. La médiation est plus que de la diffusion de contenu, elle en favorise son appropriation. Quelle est la place de l'internaute dans ces dispositifs ? À quel endroit peut-il intervenir ? Pour les œuvres d'art, par exemple, on devrait donner aux internautes la possibilité de « toucher », d'interagir avec l'œuvre, voire de la modifier.
- 18 À l'inverse, lors des émeutes anglaises de l'hiver 2011, les acteurs des événements ont réagi par retour aux discours mis en ligne concernant leur action².
- 19 L'exemple l'échec d'une plate-forme participative a été rapporté. Un projet d'interactivité numérique avec des contenus sur la ville de Tunis n'ayant pas réussi, un livre a été publié³ et une exposition de photos organisée (« Entrez ! halls et escaliers tunisois photographiés par Arnaud du Boistesselin »). Ces deux moyens « classiques » de médiation ont suscité de nombreux retours spontanés du public qui voulait apporter des informations supplémentaires. (« *crowdsourcing* sans le vouloir »). L'idée de faire une édition numérique augmentée pour mieux rencontrer les lecteurs s'est alors imposée. Mais, comme l'éditeur du livre n'était pas intéressé par une édition numérique, il aurait fallu repenser tout le projet.
- 20 Aux Etats-Unis, il existe un exemple de livre contributif sur le Musée 2.0. pour accompagner les évolutions culturelles. Ce modèle d'écriture collective soulève le problème de la modération du site en question mais, aussi, de la modération des blogs en général. Quelles sont les règles que l'on devrait imposer aux blogueurs, devrait-il y avoir des règles différentes entre un blog personnel ou officiel ? Est-ce que cette régulation n'intimiderait l'internaute et n'entraverait pas sa spontanéité ?
- 21 Le lancement d'un carnet de recherche autour de bibnum, a été expérimenté pour lancer des discussions, mais ce projet s'est avéré assez difficile (cf. <http://bibnum.hypotheses.org>).

Objectifs et motivations de la médiation numérique ?

- 22 De nombreuses questions ont été évoquées : Qui l'utilise ? Quel type de public est visé ? Cette cible est-elle atteinte ? Quelle est la part de l'obligation (institutionnelle) de « parler » au public ? Cette obligation existe-t-elle ? Se voit-elle dotée d'objectifs précis et formalisés ?
- 23 Pour la BNF, l'analyse de l'échec du fil tweet automatisé a montré que le besoin d'animation humaine du tweet reste fondamental. Mais, il faut laisser un certain temps pour expérimenter et évaluer ces nouveaux médias. La nécessité d'identifier son public au préalable semble évident.
- 24 En prenant l'exemple de l'ouvrage participatif des visiteurs dans les musées⁴ qui permet aux visiteurs de laisser des commentaires sur les enregistrements, on voit que la réponse est en général très timide. Un autre exemple (*A Public tagging experiment* : <http://tagger.steve.museum>) a été évoqué.

La visibilité de la recherche

- 25 L'édition scientifique a pour objectif de rendre accessible des travaux de recherche et de les valoriser. Pour les chercheurs l'enjeu de la visibilité de leurs travaux répond à une demande institutionnelle (par exemple, l'ANR demande maintenant un volet de diffusion de la recherche vers la société).
- 26 La question de savoir si l'édition numérique peut toucher un public plus large que le seul public scientifique se pose. *A priori*, la réponse est négative même s'il est vrai qu'une édition numérique peut « voyager » plus facilement qu'une édition papier. Pour ce faire, il faudrait les chercheurs eux-mêmes mobilisent plus leurs réseaux pour diffuser la recherche.

Les compétences nécessaires

- 27 Au départ, il faut identifier les compétences nécessaires pour une médiation numérique, savoir où les trouver et, ensuite, les articuler avec celles des chercheurs.

Le projet Gallica à la BnF

- 28 Il a fallu des personnes possédant une double compétence : une bonne connaissance des fonds de la BnF et de solides connaissances en informatique. Ainsi, l'équipe Gallica BnF s'est agrégée progressivement les compétences (extérieures à l'équipe) et la bonne volonté de spécialistes des fonds, relativement jeunes, disposés à adopter l'outil numérique pour y contribuer. Des compétences en recherche iconographique et en gestion des droits de l'image étaient également requises car toute bonne médiation scientifique numérique utilise des images.

La pérennité

- 29 Que faire quand facebook et twitter disparaîtront ? Comment récupérer les contributions des utilisateurs, comment les réintégrer sur une autre interface ? Comment les recycler ? Est-t-il possible d'en faire une collection, de recomposer une unité documentaire ?
- 30 Il est nécessaire de saisir la répartition des publics à toucher, de les différencier dans la masse, et d'en affiner les profils. Pour créer des passerelles entre les le monde de la recherche et les différents publics, quel ton adopter ? Plutôt institutionnel ou plutôt amical ? Comment adapter son langage sans appauvrir le contenu ? Pour soutenir cette ouverture la politique institutionnelle devrait encourager les chercheurs à communiquer sur des blogs, des comptes facebook, etc.

NOTES

1. La référence est la stratégie de la Tate Gallery vis-à-vis des médias sociaux qui est à comparer avec la stratégie numérique du Centre Pompidou Virtuel en cours de lancement, qui vise à mettre la collection (non diffusable en ligne !) au cœur d'un échange « social » avec le public d'Internet. *Tate Social Media Communication Strategy 2011-2012* : <http://www.tate.org.uk/research/publications/tate-papers/tate-social-media-communication-strategy-2011-12>

Conclusion : Outil de partage : <http://www.spiderscribe.net/app/?9f74b42b4ebc3598018012d11386ed2a> vision générale des problèmes.

2. (<http://videos.liftconference.com/video/6401015/reading-the-riots-on-twitter>)

3. *Tunis architectures XIX^e-XX^e s.*, Arles : Éditions Honoré Clair ; Tunis : Elyzad, 2011.

4. <http://www.participatorymuseum.org/read/>

RÉSUMÉS

Dans cet atelier, nous proposons d'échanger autour des outils numériques qui permettent de valoriser la recherche auprès des publics non spécialistes. Regroupés sous le terme générique de « médiation scientifique » (qui reste ouvert à discussion), nous souhaiterions croiser les expériences et tenter de définir ensemble quelques critères d'évaluation, avec par exemple : quels outils ont été utilisés ? Dans quels contextes (espaces muséographiés, en ligne) ? Est-il possible de les classer par genre ? La méthodologie de conception : comment écrire un support de médiation ? Les aspects techniques : quelles compétences ont été nécessaires ? Les retours des utilisateurs : quels outils pour les évaluer ? L'atelier étant basé sur le croisement des expériences, nous proposons de formaliser les échanges sous forme de cartographie expérimentale des outils investigués (type *mind mapping*). Cette cartographie pourrait être une piste évolutive pour la création d'un panorama des bonnes pratiques, voire un support d'aide à la décision des outils de valorisation des SHS.

INDEX

Mots-clés : médiatisation numérique, interaction, public, valorisation des SHS

Instrumenter la recherche

Utiliser l'ontologie CIDOC CRM pour l'information relative au patrimoine culturel

Proposé par Anne-Violaine SZABADOS (équipe LIMC-ArScAn / Archéologies et Sciences de l'Antiquité - UMR7041, CNRS), Katell BRIATTE (Ministère de la Culture et de la Communication) et Rosemonde LETRICOT (Irht)

Collectif

Présentation

- 1 Il s'agit de présenter et de discuter d'un modèle, émanant du monde des musées (Icom : the International Council of museums), pour structurer et encoder les données sur les œuvres matérielles et immatérielles du patrimoine culturel (*Cultural Heritage*). Cette norme, surtout utilisée dans le monde des musées et peu dans la communauté scientifique, s'applique à des documents hétérogènes et pourrait servir de modèle pivot. Cependant, elle semble peu connue en France. Il serait souhaitable de voir émerger une communauté d'intérêt et de partage d'expériences.
- 2 Après une introduction au Modèle Conceptuel de Référence CIDOC (ISO 21127:2006), les domaines d'application (musées, sites archéologiques, gestion de collections et de fouilles, corpus d'objets), les notions clés (objet/concept, événement/temps, lieu, intervenant/agent de l'événement) et leurs interactions seront discutés.

Introduction

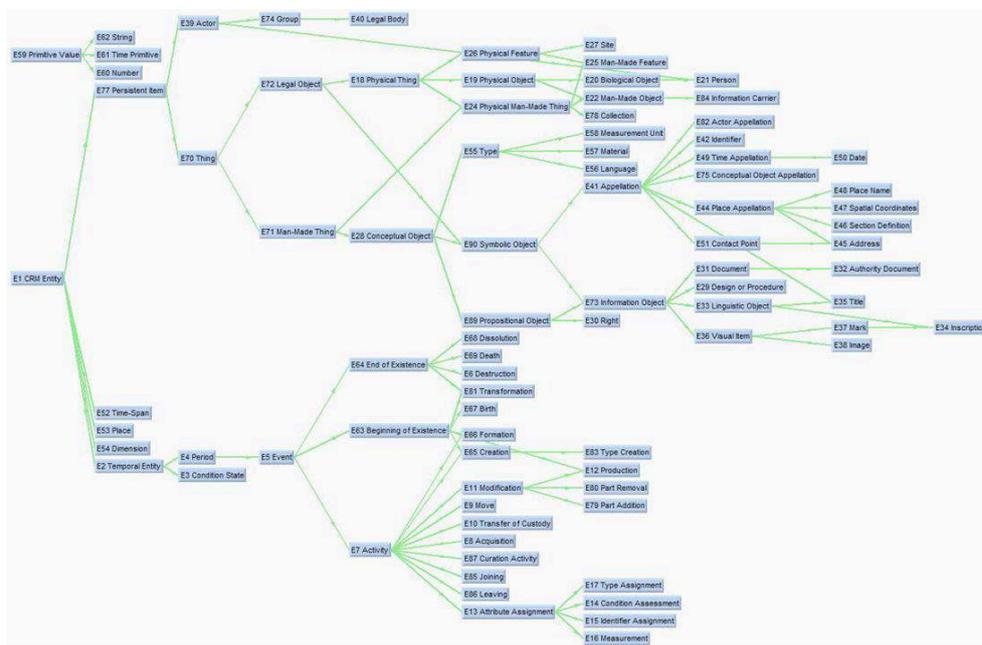
- 3 Le CIDOC CRM est un modèle conceptuel de référence pour l'information relative au patrimoine culturel qui émane du monde des musées¹. Il a acquis le statut de norme ISO en 2006, mais reste en constante évolution et continue d'être mis à jour. La dernière version de l'ontologie (v.5.0.4, décembre 2011) est en ligne sur le site du CIDOC CRM

SIG. La nouvelle version de la norme ISO (et de sa traduction en français) qui prend en compte les récentes évolutions du modèle est en cours de publication.

- 4 La version SIG de l'ontologie, en anglais, est disponible gratuitement sur le site du CIDOC. En revanche, le document de la norme ISO accessible sur le site de l'ISO dans les versions anglaise et française, est payant. Une définition courte de la norme est disponible sur la page de l'ISO :
- 5 « ... Le domaine d'application de principe de l'ISO 21127 : 2006 peut être défini comme toutes les informations nécessaires pour l'échange et l'intégration de la documentation scientifique des collections de musée. [...] »
- 6 Le terme « collections de musées » englobe tout type de matériel rassemblé et exposé par des musées et des institutions apparentées, selon la définition de l'Icom. Ceci inclut des collections, des sites et des monuments en rapport avec des domaines tels que l'histoire sociale, l'ethnographie, l'archéologie, les beaux-arts, les arts appliqués, l'histoire naturelle, l'histoire des sciences et de la technologie.
- 7 La documentation des collections inclut la description détaillée d'objets individuels qui font partie des collections ainsi que des groupes d'objets et des collections dans leur ensemble. L'ISO 21127 : 2006 est spécifiquement censée couvrir des informations contextuelles (c'est-à-dire historiques, géographiques et théoriques qui donnent aux collections de musée leur signification culturelle et leur valeur). [...] »
- 8 Si le modèle émane du monde des musées, il n'est pas limité aux objets de musées ou de collections. Il traite de notions, de types de données, de domaines ou de champs disciplinaires très variés, dont le point commun est l'appartenance au « patrimoine culturel, aussi bien immatériel que matériel ».
- 9 Pouvant s'adapter à des documents et des notions très variés, le CIDOC CRM peut être un pivot pour traiter des données hétérogènes et pour permettre des approches interdisciplinaires. On peut, par exemple, l'employer à deux niveaux pour rassembler un ensemble hétérogène d'informations archéologiques issues d'un chantier de fouilles : d'une part pour structurer et encoder une grande partie de ces données (zones fouillées, géolocalisation, date des fouilles, date et identification des documents trouvés, liste et description des découvertes, état de conservation, référence à un taxon, lieu de conservation, intervenants, droits, etc.), d'autre part pour insérer des liens vers d'autres informations (reproductions iconiques, documents de fouilles, publication, thésaurus...).
- 10 C'est un modèle souple, qui peut être utilisé partiellement ou qui peut être augmenté d'ajouts pour des besoins spécifiques. Le modèle est en effet extensible, pourvu que les extensions respectent les règles de l'ontologie. Patrick Le Boeuf (BnF, membre actif du CIDOC) a en particulier travaillé sur l'extension FRBR *object-oriented* (FRBRoo)[8], une extension du CIDOC pour le monde des bibliothèques. L'articulation entre FRBR et le CIDOC permet de différencier l'œuvre de son contenu, de faire dialoguer le livre en tant qu'objet patrimonial et le livre en tant que contenu. Ainsi s'articulent les différents domaines métier. Une tentative d'extension dans le monde des archives semble pour le moment abandonnée à cause de problèmes de réception au sein de la communauté. Les archives sont en train de créer leur propre modèle conceptuel, mais le risque est de voir proliférer les ontologies et de perdre en interopérabilité.

- 11 Quelques principes permettent de définir le modèle :
- Il est fondé sur cinq notions-clés qui sont l'objet (*Thing*) matériel et immatériel, l'agent (*Actor*) ou l'acteur de l'événement, le lieu (*Place*), le temps/événement (*Time/Event*) et le concept.
 - Il inclut des éléments permettant de traiter des ensembles d'informations comme, par exemple, le droit, la documentation, la gestion de l'objet, etc.
 - Il combine deux types d'éléments : les « classes », qui sont les notions et les concepts (au nombre de 90) et les « propriétés » (prédicats exprimés par un verbe) qui correspondent à une action, un verbe ou une interaction possible entre les classes (149 ont été définies)
 - Il existe une hiérarchie entre les classes ainsi que des interactions classes/propriétés. Par exemple, la combinaison entre classes et propriétés permet de reformuler une information du type : « la statue est actuellement localisée dans tel lieu, tel musée ». La notion « est actuellement localisée » correspond à la propriété « P55_has current location (currently holds) » ; ou encore, « l'objet est en marbre » devient *l'objet* - P45 consist_of - E57 Material *marbre* (P45 étant une propriété et E57 une classe).
 - La notion d'événement (Event) est essentielle et, combinée avec les autres éléments proposés dans le modèle, permet la description du contexte historique et des étapes de la vie de l'« objet ».
- 12 Un exemple de schéma de la hiérarchie des classes montrant l'articulation entre les différents niveaux de l'ontologie est disponible sur le site du CIDOC.

Le site du CIDOC



Le schéma donne une idée des différents types de concepts traités. Sont également disponibles sur cette page internet des schémas thématiques présentant des interactions classes/propriétés, et des modèles d'organisation de l'information (acquisition d'un objet de collection ; phases de création / modélisation d'un objet ; documentation et références ; relation temps / espace ; etc.).

Interopérabilité

- 13 L'Aphrodite Braschi, statue conservée à Munich au Staatliche Antikensammlungen und Glyptothek [inv. Gl 258] est une réplique romaine de l'Aphrodite de Cnide de Praxitèle. Une notice de description en DC (15 métadonnées du *Dublin Core* non qualifié) a été établie. Cependant, le *Dublin Core* « aplanit l'information » et met tout au même niveau comme par exemple, les dates de création de la statue ou de création du document numérique. Se limiter à l'emploi du DC oblige à choisir entre les informations sur l'enregistrement et celles sur le document traité (la statue) au lieu de pouvoir les hiérarchiser dans des types de balises différentes. Cette statue, comme beaucoup de documents du patrimoine culturel, génère beaucoup d'informations pertinentes pour les requêtes des internautes, mais le DC ne permet pas de toutes les traiter.

Exemple de l'Aphrodite Braschi et des métadonnées DC (volontairement trop renseignées)



```
<dc:title>Aphrodite Braschi</dc:title>
<dc:title>Munich, Staatliche Antikensammlungen und Glyptothek Gl
258</dc:title>
<dc:creator>Praxitèle</dc:creator>
<dc:creator>inconnu</dc:creator>
<dc:subject>Aphrodite, Vénus, Antiquité, sculpture, archéologie...</
dc:subject>
<dc:description>statue en marbre d'Aphrodite (type Aphrodite de
Cnide)</dc:description>
<dc:description>Aphrodite nue et debout près d'un vase</
dc:description>
<dc:publisher>LIMC</dc:publisher>
<dc:contributor>équipe LIMC-France</dc:contributor>
```

```

<dc:date>IIe s. av. J.-C.</dc:date>
<dc:date>époque hellénistique récente</dc:date>
<dc:date>2011/08/20</dc:date>
<dc:date>2012/09/27</dc:date>
<dc:type>objet physique, statue</dc:type>
<dc:type>dataset</dc:type>
<dc:format>ressource électronique</dc:format>
<dc:identifiant>http:// http://www.limc-france.fr/objet/14767</
dc:identifiant>
<dc:source>http://www.limc-france.fr</dc:source>
<dc:langue>fr</dc:langue>
<dc:relation>http://www.limc-france.fr/image/1TN0683B</dc:relation>
<dc:couverture>Rome, Fiumicino, Munich, Glyptothèque , Glyptothek,
Antiquité, IIe s. av. J.-C., 2011/08/20</dc:couverture>
<dc:droits>copyright...</dc:droits>

```

Extrait d'une version en CIDOC CRM (prototype IDA)

```

<?xml version="1.0" encoding="UTF-8"?>
...
<crm:E24_Physical_ManMade_Thing> <!-- extrait de l'enregistrement
concernant la statue : -->
...
<crm:P129_is_subject_of> <!-- extrait de la déclaration de
l'enregistrement -->
<crm:E73_Information_Object>
<crm:P48_has_preferred_identifiant>
<crm:E42_Identifier>http://www.limc-france.fr/objet/14767</
crm:E42_Identifier>
</crm:P48_has_preferred_identifiant>
</crm:P129_is_subject_of>
...
<crm:P1_is_identified_by> <!-- extrait des informations concernant
l'objet -->
<crm:E41_Appellation>Aphrodite Braschi</crm:E41_Appellation>
</crm:P1_is_identified_by>
<crm:P45_consists_of> <!-- matériaux -->
<crm:E57_Material>marbre</crm:E57_Material>
<crm:P127_has_broader_term>
<crm:E55_Type>pierre</crm:E55_Type>
</crm:P127_has_broader_term>
</crm:P45_consists_of>
...
<crm:P62_depicts> <!-- mots-clés sur l' iconographie -->
<crm:E28_Conceptual_Object>
<crm:P2_has_type>
<crm:E55_Type>Aphrodite</crm:E55_Type>
<crm:E55_Type>Aphrodite de Cnide</crm:E55_Type>
<crm:P127_has_broader_term>
<crm:E55_Type>déesse</crm:E55_Type>
<crm:E55_Type>divinité</crm:E55_Type>
</crm:P127_has_broader_term>
</crm:P2_has_type>

```

```

</crm:E28_Conceptual_Object>
</crm:P62_depicts>
<crm:P108B_was_produced_by> <!--événement "fabrication de l'objet" -->
<
<crm:E12_Production>
<crm:P32_used_general_technique>
<crm:E55_Type>ronde bosse</crm:E55_Type>
<crm:P127_has_broader_term>
<crm:E55_Type>sculpture</crm:E55_Type>
</crm:P127_has_broader_term>
</crm:P32_used_general_technique>
<crm:P4_has_time-span> <!--datation formulée par fourchette
chronologique ) -->
<crm:E52_Time-Span>
<crm:P82_at_some_time_within>
<crm:E61_Time_Primitive>
<start>-150</start>
<end>-100</end>
</crm:E61_Time_Primitive>
</crm:P82_at_some_time_within>
</crm:E52_Time-Span>
</crm:P4_has_time-span> <!--datation formulée par appellation
traditionnelle -->
<crm:P10_falls_within>
<crm:E4_Period>
<crm:P1_is_identified_by>
<crm:E49_Time_Appellation>époque hellénistique récente</
crm:E49_Time_Appellation>
</crm:P1_is_identified_by>
</crm:E4_Period>
</crm:P10_falls_within>
<crm:P14_carried_out_by> <!-- artiste -->
<crm:E39_Actor>atelier grec</crm:E39_Actor>
</crm:P14_carried_out_by>
</crm:E12_Production>
</crm:P108B_was_produced_by>
...
<crm:P12B_was_present_at><!--informations sur la découverte -->
<crm:E7_Activity>
<crm:P2_has_type>
<crm:E55_Type>découverte</crm:E55_Type>
<crm:P7_took_place_at>
<crm:E53_Place>
<crm:P87_is_identified_by>
<crm:E44_Place_Appellation> Fiumicino</crm:E44_Place_Appellation>
<crm:E47_Spatial_Coordinates>
<latitude>41.7667</latitude>
<longitude>12.2333 </longitude>
</crm:E47_Spatial_Coordinates>
</crm:P87_is_identified_by>
<crm:P89_falls_within>
<crm:E53_Place>Italie</crm:E53_Place>

```

```

</crm:P89_falls_within>
</crm:E53_Place>
</crm:P7_took_place_at>
...
<crm:P130B_features_are_also_found_on><!--reproductions numériques
de l'objet -->
<crm:E38_Image>
<crm:P149_is_identified_by>
<crm:E75_Conceptual_Object_Appellation>URL de l'image (http://
www.limc-france.fr...1DE19367)</
crm:E75_Conceptual_Object_Appellation>
</crm:P149_is_identified_by>
<crm:P104_is_subject_to>
<crm:E30_Right>
<crm:P3_has_note>
<crm:E62_String>notice de copyright</crm:E62_String>
</crm:P3_has_note>
</crm:E30_Right>
</crm:P104_is_subject_to>
</crm:E38_Image>
</crm:P130B_features_are_also_found_on>
...
<crm:P70B_is_documented_in><!-- bibliographie, webographie...
concernant l'objet -->
<crm:E31_Document>Klein, W., Praxiteles (1898) 262-264</
crm:E31_Document>
<crm:E31_Document>Furtwängler, Beschreibung der Glyptothek (1910)
266 n° 258</crm:E31_Document>
<crm:E31_Document>Vierneisel-Schlörb, KatSkulptMünchen II 323-336
n° 31, fig. 158-164</crm:E31_Document>
<crm:E31_Document>LIMC, vol. II, Aphrodite 399*</crm:E31_Document>
<crm:E31_Document>http://...</crm:E31_Document>
</ crm:P70B_is_documented_in>
</crm:E24_Physical_ManMade_Thing>

```

- 14 Dans ce cas, l'avantage du CIDOC CRM par rapport au DC réside dans le fait qu'il permet d'introduire plus de finesse et de sémantique dans la structuration de l'information et donc de traiter davantage d'informations, de données. L'emploi du CRM permet de structurer une information plus riche et plus détaillée : différencier l'objet documentaire (statue) et sa reproduction (photo), différencier/séparer/isoler diverses notions de lieux pertinentes (lieu de fabrication de l'objet, lieu de trouvaille, lieu de conservation actuel, lieux antérieurs de conservations/collections...), de date (date de création de la statue, date d'entrée au musée, date de l'enregistrement). Sa principale différence par rapport au Dublin Core est qu'il permet de distinguer la description de la ressource de celle de l'objet dont parle la ressource.
- 15 La structuration plus fine élargit les possibilités d'utilisation du fichier de données produit. L'utilisation de la donnée a une forte influence sur le modèle de structuration de l'information : si les métadonnées DC peuvent correspondre à une interrogation (type champ Google), un schéma plus fin permet l'élaboration d'une interface Web de recherche différenciant davantage de critères de recherche (facettes...). Un outil complexe servant à recueillir des informations variées et détaillées (une base de

données relationnelle, un fichier RDF...) peut également reposer sur l'emploi de cette norme qui semble adaptée à cette problématique.

- 16 Quelques schémas types fondés sur le CIDOC CRM existent, par exemple LIDO (*Lightweight Information Describing Objects*) destiné aux collections de musées.

Communauté en France

- 17 La documentation mise en ligne sur le site du CIDOC CRM augmente (communications, schémas, liens vers applications, etc.) mais on y constate peu de références françaises. Des travaux ont cependant été amorcés par Pierre Le Bœuf à la Bibliothèque Nationale de France (BnF) et des réflexions sont en cours au Ministère de la Culture. Quant à la communauté scientifique française, elle ne semble pas avoir tellement utilisé ou même travaillé sur ce modèle. Pourtant, il semble interpeller de plus en plus d'acteurs de la recherche, notamment dans les champs disciplinaires liés à l'art et à l'archéologie, comme en témoignent les interrogations (« Vous allez parler du CIDOC CRM ? Justement, je me posais des questions à son propos... ») exprimées lors de rencontres scientifiques, suivies d'exclamations (« Ah ! Mais cela pourrait résoudre mon problème !), vite suivis de nouvelles d'interrogations (« Mais comment ça marche ? »).
- 18 Parmi les initiatives émanant du monde de la recherche, on peut signaler l'élaboration, actuellement, d'un schéma pivot pour un portail dédié aux objets de l'Antiquité (programme IDA avec le TGE Adonis). L'équipe française du LIMC (CNRS UMR7041-ArScAn), impliquée dans ce programme, a également participé à la création du portail international CLAROS dont l'interopérabilité est fondée sur l'emploi du CIDOC CRM et dont le schéma RDF est présenté et commenté sur un site dédié, CLAROSwiki. Sur ce dernier, on peut consulter des exemples de fichiers RDF avec les données/valeurs/instances.
- 19 Si l'on souhaite voir un exemple de site fondé sur l'emploi du CIDOC, on peut aller consulter la réalisation allemande Arachne/iDAI
- 20 On pourrait soumettre l'idée de la création d'une communauté d'intérêt qui fait particulièrement défaut en France. Un accompagnement ou un espace d'échange permettrait de partager des expériences dans ce domaine mais aussi entre spécialistes du domaine, spécialistes du traitement de l'information et informaticiens. Rosemonde Letricot (Institut de recherche et d'histoire des textes, IRHT) a approché le CIDOC dans le cadre du projet IDA. Elle porte actuellement une réflexion autour de l'utilisation d'ontologies pour rendre compte de la complexité du livre médiéval à partir des données de la recherche. Son expérience d'utilisation du CRM l'amène au constat d'un besoin d'un apprentissage de la norme (explications, documents pratiques, schémas, publications de modèles, de schémas incluant les instances/valeurs/données...) et d'accompagnement des utilisateurs. Elle suggère de créer un groupe d'utilisateurs et d'échange pour ne pas démultiplier les créations de modèles ontologiques. Elle souligne l'intérêt de l'ontologie pour des données hétérogènes car celle-ci permet de recontextualiser l'information dans son domaine et de mettre à plat les postulats implicites de ce dernier. Quand on s'attaque à une ontologie, il devient nécessaire d'être critique par rapport aux données car il faut tout formaliser : les concepts, les caractéristiques de ces concepts, leurs relations, leurs rôles, leurs conditions et leurs restrictions.

- 21 Il faut également se demander dans quel but on structure les données (réflexion autour de l'application qui va restituer l'information).
- 22 Il serait aussi intéressant de se positionner par rapport à d'autres propositions, et d'autres normes comme EDM d'Europeana.

Discussion

- 23 La question de la place de l'historien dans la structuration des données et l'implication du chercheur dans la construction d'ontologies a été évoquée.
- 24 **Q :** Le CIDOC CRM peut-il être utilisé dans le cadre d'un projet sur une collection pour créer des liens entre les entités dans les textes et les collections et les outils. Comment au musée de Nuremberg peut-on adapter le CRM, et mélanger des ontologies ?
- 25 **R :** Le Göttingen Centre for Digital Humanities (GCDH) fédère les ontologies produites par différentes institutions².
- 26 **Q :** Dans le cadre d'études architecturales et l'instauration d'un vocabulaire spécifique à la période moderne, le CRM peut-il traiter la gestion des droits de diffusion des prises de vues ? ou répondre à des problématiques concernant l'interopérabilité et la conservation sur le long terme (pérennité), et par rapport aux autres vocabulaires, normes, etc. ?
- 27 **R :** Le CIDOC CRM permet d'insérer de façon précise et détaillée les informations concernant les droits (détenteur des droits, notice de *copyright*... en rattachant précisément ces informations aux documents concernés, par exemple objet documentaire principal [statue], ses reproductions photographiques signalées...). Si on crée des bases de données avec des ontologies spécifiques il est ensuite nécessaire de faire des mappings pour les rendre interopérables.
- 28 Il ne faut pas générer de la différenciation là on cherche l'interopérabilité. Il est nécessaire d'articuler les différents modèles. Par exemple, il faut réfléchir à l'articulation entre ISO 21127 et la norme ISO de gestion des vocabulaires (ISO 25964). La classe E55 Type, utilisée pour porter les valeurs issues de nomenclatures ou de vocabulaires normalisés, doit être rapprochée de la classe ThesaurusConcept d'ISO 25964.
- 29 Au ministère de la Culture et de la communication, un travail est en cours pour articuler ensemble des fonds issus de domaines très divers. Le dialogue passe aussi par le partage de vocabulaires. Un schéma de données (ou une ontologie) décrit bien comment on doit manipuler les concepts (les classes), mais ne dit pas avec quelles valeurs on peut ou on doit les instancier. Il est donc important d'avoir des vocabulaires partagés pour renforcer l'interopérabilité sémantique.
- 30 **Q :** A-t-on connaissance d'ontologies utilisées en SHS ? Le CIDOC CRM permet-il de traiter l'histoire sociale ? Dans le cadre d'une plateforme collaborative et une base de données ouverte³, une équipe lyonnaise a, depuis cinq ans, mis en place une plateforme multi-utilisateurs et les utilisateurs ont créé une ontologie du domaine, en dialogue avec les informaticiens, et ont choisi un modèle ouvert : objets/rôles/informations. Les ontologies sont relationnelles et non RDF car il y a une prédéfinition des prédicats. Ils auraient besoin d'un mapping pour exposer ces données dans une ontologie commune pour partager l'information via le web sémantique (schéma pour l'interopérabilité). Le

CIDOC semble ajouter une couche de description (contextualisation), et avoir des possibilités d'extension.

- 31 **R** : C'est un problème pour la description des textes qui sont très importants dans le travail de l'historien (le CRM propose juste un conteneur).
- 32 Concernant la dimension sociale, les notions d'acteur et d'appartenance à un collectif sont développées dans le CIDOC CRM, mais cela peut-il répondre aux besoins ? La notion d'évènement permet de rendre compte du cycle de vie des objets. Les évènements peuvent être typés (par exemple par la classe E55 Type), comme chaque entité/classe du modèle. Le CRM permet de formuler finement des concepts, notamment pour exprimer l'évènement « fabrication de l'objet » et les informations qui s'y rattachent (lieu de fabrication, artiste, matériaux).
- 33 Le CIDOC CRM est organisé en hiérarchie de classes et de propriétés, avec la notion d'héritage. En particulier, la classe principale (d'entrée) « E1 Entité » porte la propriété « P2 a pour type » (qui s'associe avec la classe E55 Type), qui est donc héritée par toutes les autres classes.
- 34 Si le CRM permet d'exprimer des informations complexes ou riches et d'approfondir leur granularité, certains choix de mise en œuvre peuvent traiter ces mêmes informations en privilégiant l'utilisation, plus facile, d'éléments comme la classe E55 Type (voir le document de la norme) pour catégoriser directement les concepts traités.
- 35 Par exemple, dans le cas du portail CLAROS, le schéma est simple et « plat ». Il repose principalement sur l'utilisation de la classe E55 type. La notion « artiste » est réglée avec la création d'un type E55 = artiste⁴ alors qu'elle aurait pu s'insérer dans une structure réunissant de façon logique plusieurs interactions classe/propriété appropriées (encapsulage d'informations) plaçant, par exemple, l'artiste comme intervenant dans un événement défini comme « processus de création », etc. Cela pose le problème de l'interopérabilité. Si on propose plusieurs schémas pour une même notion, est-ce vraiment interopérable ? Ce choix répond bien au besoin d'interopérabilité du portail : les ressources numériques des partenaires « interopèrent », un nouveau partenaire peut facilement mapper ses données..., mais le nouveau gisement de ressource répond-il vraiment à ce que l'on attend de l'interopérabilité, du web sémantique, etc. ?
- 36 Faut-il privilégier des choix, préconiser des utilisations de classes ou de propriétés ?
- 37 **Q** : Dans le cas d'une plateforme mise en place autour de fichiers son, avec un vocabulaire propre à la discipline, où les notions d'œuvre, de créateurs des ressources, etc. sont importantes, quel format choisir pour les métadonnées ?
- 38 **R** : Dans le cadre du projet européen Caspar sur l'archivage (Jérôme Barthélémy, IRCAM) l'ontologie CIDOC CRM a été utilisée pour l'archivage et la pérennisation, en vue de restituer complètement l'œuvre.
- 39 **Q** : Dans quelle mesure le CIDOC CRM peut-il être souple dans son utilisation et applicable à d'autres types de sources ? Y-a-t-il une norme commune pour les objets hétérogènes ?
- 40 **R** : Il n'y a pas de réponse. La norme prétend couvrir un large spectre de domaines. Il faut explorer la norme pour voir les besoins d'évolution, la confronter à des cas spécifiques. Il y a une nécessité à fédérer les acteurs, qui pourraient alors devenir une force de proposition pour le groupe de réflexion du CRM SIG.

41 **Q** : Y a-t-il une modélisation des ontologies ?

42 **R** : Plusieurs classes peuvent être adaptées à une même notion. Il est parfois difficile de choisir, de percevoir toutes les nuances de la définition. On peut opter pour une classe dont la définition est plus large. Il y a un jeu dans la hiérarchie des classes, et il faut rappeler qu'une sous-classe respecte la définition de ses classes supérieures et hérite de ses propriétés. Cela permet une définition plus ou moins fine des concepts, et aussi des définitions plus complexes.

43 Il est noté l'importance de la collaboration avec les informaticiens. Le travail/dialogue avec eux est essentiel pour la mise en pratique et cet échange profite à l'élaboration de la norme, du schéma, et de la façon de l'appliquer concrètement. C'est un enrichissement mutuel, pour les participants, le projet mais aussi la norme.

44 L'idée de privilégier la simplicité dans la conception de schémas, sans pour autant renoncer à la spécificité et la richesse du CIDOC CRM, pourrait faciliter l'émergence de schémas communs à des besoins, des disciplines (interopérabilité, web sémantique, émergence de besoins et discussion au sein de la communauté d'intérêt...). Il faut trouver un équilibre dans les modèles, car on ne peut pas tout traiter. Même si la norme le permet, faut-il aller vers trop de complexité ? Le sens critique est nécessaire sur des concepts que recouvre l'information à traiter, par exemple les notions de lieux, de datation, d'institutions, de propriété, de localisation, de nommage, etc. Se confronter à la norme oblige à réfléchir davantage et plus précisément aux informations/concepts que l'on souhaite traiter. Cela déclenche l'évaluation critique sur sa propre approche. Une structuration évidente depuis des années peut apparaître inadaptée ou inadéquate lors de ce processus de critique, présenter des distorsions, des lacunes. Cette confrontation avec les définitions du CIDOC CRM est profitable et intéressante à faire même s'il n'y a pas de réalisation pratique prévue. Il se dégage également des notions plus délicates à traiter et à rendre interopérables. Celles-ci apparaissent notamment lorsque l'on élabore un schéma pour un portail visant à rendre interopérables plusieurs ressources numériques semblant pourtant proches dans leur conception : par exemple les datations, les styles, les aires « géo-culturelles » de création, etc.

45 Son utilisation est encore peu répandue. Du point de vue d'un chercheur, on ressent la nécessité de réfléchir sur les normes, car il n'y a pas de norme qui permette de gérer l'ensemble des informations produites par la recherche historique.

46 Il serait intéressant de voir les différents traitements ou schémas proposés et les vocabulaires utilisés. Comment les choisir et comment formuler dans le schéma ?

47 On pourrait fédérer un réseau autour du CIDOC CRM, comme la TEI a su le faire, en associant des spécialistes des sciences de l'information et de l'informatique en plus des spécialistes des champs disciplinaires traités (équipes de recherche).

48 **Q** : Est-ce que le Louvre utilise cette norme ?

49 **R** : Le CIDOC CRM est peu utilisé dans les musées français. Une des explications parfois avancées est que ce modèle émane de la communauté des documentalistes et non de la communauté des conservateurs qui n'y adhèrent pas, peut-être en raison de sa complexité. Dans le cadre de projets dans la nébuleuse Europeana (Athena), on assiste à un rapprochement de plusieurs modèles dans le schéma XML LIDO (CIDOC CRM, Spectrum (gestion des collections), et Museumdat⁵).

50 **Q** : Est-il difficile de s'approprier cet outil, cette norme ?

- 51 **R** : Il y a besoin d'un dialogue pour lever des ambiguïtés possibles et partager les pratiques. À titre d'expérience, les données du LIMC (<http://www.limc-France.fr>) sont dans une base de données dont la première version remonte à 1981 (c'est-à-dire au début du programme d'informatisation). Dans ce cas, la base LIMC n'a pas été remise à plat, mais des mapping ont été créés, d'une part pour préciser les notions traitées – la norme CIDOC CRM devient alors un document de référence pour les définitions, un outil « fort » pour documenter la ressource numérique – et des mapping plus succincts destinés à des extractions pour différents projets (portails IDA et CLAROS, Projet d'OAI-PMH). On peut aussi envisager d'élaborer directement une ressource sur l'emploi de la norme (norme, puis modélisation puis schéma d'implantation, etc.). L'opération de mapping à partir de données existantes n'est peut-être pas aussi facile que les concepteurs du modèle l'ont parfois dit. La communauté aurait besoin d'exemples et de modèles d'utilisation. On rencontre aussi des problèmes pour la traduction d'une norme. Par exemple, on est en attente de la nouvelle version ISO et de sa traduction française, bien qu'une traduction française de l'ancienne version soit déjà disponible.
- 52 **Q** : Dans le contexte de la réalisation d'un programme pédagogique en archéologie où l'on souhaite imbriquer virtuel et réel (terrains, objets) avec un objectif pédagogique et créer des scénarios, est-il possible d'utiliser cette ontologie pour permettre au public de renseigner des informations sur les objets ?
- 53 **R** : La complexité des ontologies impose la mise en place d'interfaces utilisateurs qui masquent cette complexité et qui permettent des restitutions très riches (exemples : timeline, géolocalisation, facettes d'interrogation). Le modèle CIDOC est présenté comme « ouvert » et permet des discours contradictoires sur un même objet.
- 54 **Q** : Comment gérer le modèle dynamique ?
- 55 **R** : C'est une question d'interface : on doit définir les besoins des utilisateurs et se limiter à présenter les « champs », les propriétés qu'ils auront à renseigner.

Conclusions, perspectives

- 56 Il est important de pouvoir présenter et faire connaître cette norme. Il faut aussi structurer un groupe de réflexion sur les modèles d'ontologies ce qui permettrait de faire remonter les usages et les difficultés.
- 57 Il ressort de la discussion qu'il serait bien de faire émerger une communauté d'intérêt en France, mais également ouverte à des collègues étrangers, dans un esprit trans-disciplinaire et inter-institutionnel.
- 58 Pour commencer, on pourrait créer une liste de diffusion et, dans ce but, récupérer les adresses mails des personnes intéressées. Dans un deuxième temps, on pourrait peut-être songer à un carnet de recherche par exemple sur Hypotheses.org pour accompagner cette communauté et lui donner une vitrine dynamique permise par le carnet.

NOTES

1. CIDOC /Comité international pour la documentation de l'Icom.
 2. <http://erlangen-CRM.org/>. Le CRM Erlangen / OWL est une implémentation OWL-DL 1.0 du Modèle Conceptuel de Référence du CIDOC (CIDOC CRM).
 3. (LARHRA - Université de Lyon - Projet SyMoGIH - http://larhra.ish-lyon.cnrs.fr/Pole_Methodes/symogih_accueil_fr.php)
 4. http://explore.clarosnet.org/wiki/index.php?title=CIDOC_CRM_Objects
 5. Harvesting Format for Providing Core Data from Museum Holdings ;LIDO (LIDO is an XML harvesting schema) ; <http://network.icom.museum/CIDOC/working-groups/data-harvesting-and-interchange/what-is-lido/> ; Norme Spectrum http://www.pro.rcip-chin.gc.ca/GetForumRecord.do?type=sd&lang=fr&id=FORUM_23068
-

RÉSUMÉS

L'idée de l'atelier est de présenter et de discuter d'une norme pour structurer et encoder les données sur les œuvres matérielles et immatérielles du patrimoine culturel. Norme qui concerne des documents hétérogène ce qui la rend intéressante pour servir de modèle pivot, mais peu connue en France. Parler de l'émergence d'une communauté d'intérêt et de partage d'expériences.

Cette norme est surtout utilisée dans le monde des musées, pas tellement dans la communauté scientifique

Introduction au Modèle Conceptuel de Référence CIDOC (ISO 21127:2006), émanant du monde des musées (ICOM) et utilisé pour structurer, encoder les données sur les œuvres, matérielles ou immatérielles, du patrimoine culturel (*Cultural Heritage*): présentation et discussion sur les domaines d'application (musées, sites archéologiques, gestion de collections et de fouilles, corpus d'objets, etc.), les notions clés (objet/concept, événement/temps, lieu, intervenant/agent de l'événement...) et leurs interactions.

INDEX

Mots-clés : ontologie, patrimoine culturel, code, programmation, collection

Utilisons RDFa pour nos corpus

Proposé par Stéphane Pouyllau

Collectif

Introduction de Stéphane Pouyllau

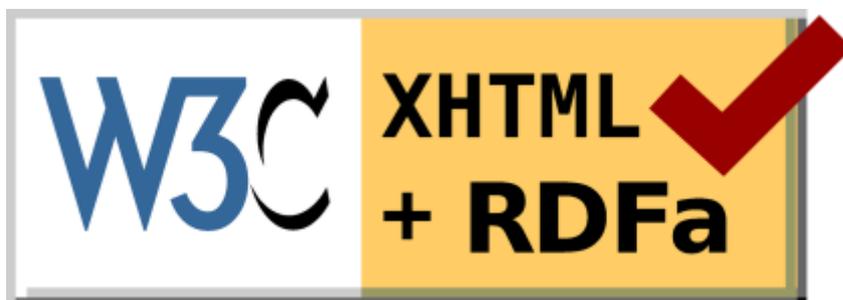
- 1 Il s'agit avec cet atelier d'envisager la question de la publication sur le Web à l'aide du Web de données, de la portabilité du RDFa dans le monde des SHS. C'est-à-dire, de voir comment préparer l'exposition de données numériques complexes dans le Web de données, selon les principes du linked data, et réfléchir ensemble à la manière dont nous pourrions dépasser la phrase habituelle qui consiste à dire « Je vais publier sur le Web ».
- 2 On peut comprendre l'utilisation du RDFa comme quelque chose qui nous offre la capacité, lorsque on fait du traitement sur des données structurées depuis le Web, de faire des traitements augmentés, en reliant l'information à des informations contenues dans des documents distants, ceci en rendant explicite pour des machines, à la fois la structure, mais aussi le contenu sémantique des documents.
- 3 L'idée de cet atelier est de voir comment la mécanique du RDF paraît bien s'appliquer au type de travail des chercheurs en sciences humaines et sociales, au sens où il s'agit de produire de l'information, de la publier et de la rendre explicite. Cela pourrait être en quelque sorte la marche supérieure vers une meilleure interopérabilité des informations et donc des données après l'OAI¹ avec l'utilisation de Dublin Core² : jusqu'à présent on était obligé de coller un petit programme aux bases de données pour exprimer les données, toute modification de la structure de la base de données impliquait donc la réécriture de l'API, RDFa est peut-être la seconde marche vers l'interopérabilité, c'est-à-dire le moment où les chercheurs vont reprendre leurs données et rendre explicite la structuration de toute cette information afin de la publier dans le Web de données.
- 4 Deux grandes notions doivent être posées en introduction afin de mieux débattre, celles des syntaxes RDF et RDFa, et le concept de Web de données et de linked data. Comment faire prendre conscience aux chercheurs de l'importance de la structuration,

y compris dans le Web pour permettre aux moteurs de recherche de mieux retrouver l'information et les données (en passant par des métadonnées exprimées avec plus de richesse sémantique).

L'utilisation des syntaxes RDF et RDFa

- 5 L'utilisation du RDFa est une manière d'exprimer du RDF dans une page web. La norme décrit la manière d'enchâsser de la sémantique à l'aide d'attributs ajoutés dans les balises HTML qui vont pointer vers des URI³. On peut dire que RDFa est une sorte de « micro-formats ». On peut tout à la fois, à l'aide de ce mécanisme, exprimer la structure des pages ou décrire explicitement que telle information désigne le lieu de naissance d'un auteur, la géolocalisation d'un lieu vis-à-vis d'un référentiel géographique présent dans *linked data* (tel que GeoNames.org), etc.

Le RDFa selon le W3C



- 6 Pour décrire cette information, on a recours à des modèles qu'on appelle des ontologies. Divers sites référencent de telles ontologies. Citons par exemple Schemapedia ou encore LOV (Link open vocabulary), un site développé initialement par la société Mondeca et rendu public depuis dans le cadre du projet Datalift. Vouloir rendre explicite de l'information dans une page donnée, et la manière de le faire, relèvent de choix scientifiques. Plusieurs outils permettent ensuite de détecter la présence de RDFa dans les pages web. Toutefois, il n'y a pas encore beaucoup d'outils d'un niveau suffisant pour l'exploitation de cette structuration.
- 7 Au cours de la discussion, Lou Burnard explique qu'il hésiterait à dire qu'il s'agit d'une structuration, il s'agit plutôt pour lui de « labeliser » (au sens d'étiqueter) telle ou telle partie du texte⁴. L'attribut définit tel élément et explique où aller chercher pour obtenir une définition. En revanche, il fait remarquer que cela ne permet pas de savoir ce qu'on peut faire avec. Avec la TEI, on est dans le cadre d'un balisage structurel et informationnel dans un contexte XML. Tandis que dans le cas de l'utilisation du RDF ou de RDFa, on n'est pas forcément toujours dans un environnement XML, et par ailleurs on va pouvoir aller chercher de l'information qui peut être contenue dans des réservoirs qui ne sont pas nécessairement contenus dans la ressource elle-même, mais ailleurs sur le Web. La TEI offre une sémantique, elle est exprimée selon un modèle d'arbre avec XML, le modèle de graphe RDF avec ses relations entre des entités (sujet-prédicat-objet) permet la mise en relation de sources. Ce dernier n'est pas contradictoire avec TEI. On peut dire, en fait, qu'avec RDF on fournit des méta-informations. Si par exemple on désigne un auteur, on informe sur l'endroit où trouver

de l'information sur cet auteur même si les traitements ne sont pas définis par l'encodage⁵.

- 8 L'un des participants souligne la difficulté qu'il y a à identifier des ontologies adaptées ou efficaces dans un contexte de recherche en sciences humaines et sociales. Les répertoires d'ontologies disponibles ne permettent pas toujours d'identifier clairement la portée ou la précision de ces ontologies et leur applicabilité. Les modèles les plus connus comme FOAF (Friend of a Friend), peuvent être d'une portée limitée pour ce qui concerne, par exemple, la description des personnes historiques. Cela constitue sans doute un frein pour la généralisation du recours à RDFa. Toutefois, le modèle conceptuel de référence CIDOC (CIDOC-CRM) offre déjà une ontologie consistante pour les sciences historiques et l'histoire de l'art. Le risque serait de tomber dans l'excès inverse en se lançant pendant 15 ans dans la mise au point d'une ontologie parfaite, en ne faisant rien en attendant.
- 9 Le RDFa permet par exemple d'exprimer une ontologie dans un texte. Un participant fait état de ses expérimentations dans des textes historiques encodés en XML-TEI pour identifier des personnes travaillant dans le même endroit. Le RDFa n'est au fond que le véhicule de ce que fait une ontologie, cela nécessite au préalable de définir (ou de choisir l'ontologie). Un premier travail consiste pour l'historien à savoir quelles sont les ontologies qui peuvent lui permettre d'exprimer toute la richesse de ce qu'il veut dire. Il convient donc au préalable de se mettre d'accord sur la bonne ontologie à utiliser pour exprimer un travail sur un texte. Comment par ailleurs exprimer toute la richesse de l'histoire ; et l'incertitude ?

Identifier les ontologies utiles aux SHS

- 10 La discussion conduit les participants à envisager la mise sur pied d'un site web répertoriant les ontologies disponibles pour les chercheurs en sciences humaines et sociales (historiens, historiens de l'art, sociologues, etc.), en soulignant leurs applicabilité, leur degré de finesse, et en indiquant par qui elles sont maintenues. Un tel site pourrait également, outre des guides pratiques, présenter des exemples de réalisations. Certains participants alertent sur le risque de recloisonnement des disciplines avec des répertoires spécialisés. Même si le modèle conceptuel de référence CIDOC (CIDOC-CRM) a été conçu dans le domaine de l'histoire de l'art, certains éléments qui le composent peuvent être utiles pour d'autres disciplines par exemple. Il y a déjà deux ans, un groupe de travail avait ainsi travaillé sur l'expression de la TEI avec le modèle conceptuel de référence CIDOC (CIDOC-CRM).
- 11 Dans le cadre du projet Athena, et en lien avec Europeana, un travail de recensement avait déjà permis d'identifier récemment un très grand nombre de vocabulaires spécialisés utilisés dans le monde des musées en Europe. Plus de 150 vocabulaires avaient été identifiés, la difficulté consistait à les classer d'après des critères efficaces, par domaines, par disponibilités en *linked data* (ce qui n'avait pas été pensé à l'époque), etc. Le projet Linked heritage qui lui fait suite, permettra de caractériser plus précisément les liens de *mapping* avec une plate-forme fournissant toutes les relations typées, avec des indications sur la finesse des relations, exprimées en SKOS (Simple Knowledge Organization System). Une telle plate-forme devrait rendre possible le jeu entre des silos très spécialisés et devrait éviter toute velléité de produire, *de novo*, une ontologie mondiale hyperspécialisée, ce qui serait tout simplement impossible.

- 12 Si la première marche dans la direction de l'interopérabilité des données scientifiques en sciences humaines et sociale a été la généralisation de l'emploi du Dublin Core via le protocole OAI-PMH, on peut se féliciter déjà de l'emploi de ces six verbes OAI-PMH partout dans les SHS et au niveau mondial. Bien sûr, le Dublin Core simple lissait par trop l'expression de l'information, ainsi certains ont pu avoir recours au Dublin Core étendu (DC Terms). Si l'utilisation du RDFa constitue une deuxième marche vers l'interopérabilité des données scientifiques, le but est peut-être d'abord d'amener les gens à exprimer ce type de structures avec ces modèles. Dans tous les cas, cela passe par des travaux communs et interdisciplinaires pour ne pas retomber dans le travers de l'ultra-spécialisation des ontologies.

Automatiser la sémantisation des corpus

- 13 Pour les corpus existants, deux cas de figure peuvent être envisagés : amener les chercheurs à revenir sur leurs documents pour faire de l'enrichissement sémantique, ce qui peut être une solution lourde, ou bien essayer de produire du balisage automatique. Dans tous les cas, il convient d'être attentif au gain de l'investissement en termes d'utilisation.
- 14 Stéphane Pouyllau ne croit pas trop à la sémantisation automatique, une sémantisation semi-automatique serait en revanche envisageable car il faudra toujours faire des choix. Un participant mentionne, à titre d'exemple, la plate-forme d'annotation sémantique développée par l'Institut de recherche et d'innovation du centre Pompidou (IRI) pour produire une exploration sémantique du portail Histoire des Arts⁶. Même dans un projet comme celui de l'extraction des données structurées des pages de Wikipédia en français, DBpedia en français, à partir des infobox de l'encyclopédie en ligne, les disparités de présentation nécessitent des ajustements manuels.
- 15 Il est peut être aussi souhaitable que les chercheurs soient directement confrontés à l'encodage de ces informations car cela pourrait présenter pour eux un intérêt en terme de questionnements scientifiques. Un participant fait remarquer qu'en 2006 des chercheurs en bio-informatique ont déjà été confrontés à des questions comparables et qu'il pourrait être profitable de prendre contact avec eux⁷.

Inciter à l'utilisation d'une structuration de l'information exprimée à l'aide de la syntaxe RDFa

- 16 L'utilisation d'une structuration/étiquetage (labelisation) à l'aide de la syntaxe RDFa dépend très directement des possibilités d'utilisation. Le moteur de recherche Isidore est actuellement l'un des seuls capables de récupérer de l'information disponible sur les pages web en RDFa et de la structurer. Le fait qu'Isidore moissonne les métadonnées est déjà très incitatif⁸. Il a été tenu compte de cette possibilité lors de la conception du moteur dans le cadre d'un travail avec l'Agence bibliographique de l'enseignement supérieur (Abes) qui produisait à l'époque l'application Calames pour le catalogue en ligne des archives et des manuscrits de l'enseignement supérieur et où une structuration RDFa avait été enchâssée dans les pages. Cela avait été réalisé à titre expérimental, mais on a pu constater que le connecteur fonctionnait. Depuis, de nombreux webmasters s'adressent à Isidore pour signifier que le simple moissonnage

RSS ne leur suffit plus en voulant proposer des informations plus précises exprimées à l'aide de RDFa.

- 17 Il est probable que les webmasters se tournent de plus en plus vers RDFa car les moteurs de recherche comme Google s'y intéressent⁹ et que cela peut avoir des répercussions sur les résultats de la recherche. De fait, les possibilités techniques ouvertes par les technologies du web 3.0 (sémantique + LOD Linked Open Data) devraient produire un appel d'air et un cycle d'incitations vertueuses du côté des webmasters.
- 18 Afin d'inciter les chercheurs à l'emploi du RDFa pour la structuration (ou l'étiquetage) de l'information en sciences humaines et sociales publiée sur le Web, on souligne la nécessité d'expliquer les possibilités offertes par une telle description et ses avantages. Il s'agit non seulement de leur donner une idée sur la manière dont cela pourra leur être utile, mais aussi de leur expliquer concrètement ce qu'il faut faire. Aussi, les participants de l'atelier reviennent sur l'idée d'un site web qui recense les ontologies disponibles en sciences humaines et sociales, mais qui fasse peut-être plus : la présentation d'exemples de *mashups* pourrait inciter, par la visualisation de ce que l'on peut faire avec des données, à l'utilisation de cette technologie. Outre des démonstrateurs du potentiel de cet étiquetage, des guides pratiques seraient utiles. La rhétorique consisterait à montrer d'abord que c'est sexy, et ensuite à prouver que cela n'est pas si compliqué à mettre en œuvre.
- 19 Un participant fait aussi remarqué que l'on obtiendra sans doute une adhésion des chercheurs en sciences humaines et sociales et leur participation à partir du moment où l'on pourra arriver à exprimer une fiabilité dans l'information ou à exprimer des notions de débats. C'est sans doute à ce niveau que se joue l'avenir de l'utilisation de RDFa dans le monde des SHS. L'intérêt principal d'une structuration de l'information avec RDFa est sa capacité à introduire une information sur la fiabilité de l'encodage. On a souvent à faire en sciences humaines au problème de l'incertitude. Peut-être que le modèle RDF nous permettra de traiter l'incertitude. Il serait peut-être opportun de créer un centre de ressources pour traiter de cette question.
- 20 S'il s'agit de faciliter l'emploi de RDFa dans le cadre de projets en XML-TEI, il faut fournir au chercheur des propositions d'attributs qu'il puisse utiliser dans son travail. Lorsqu'un chercheur veut bien ajouter des descripteurs, où peut-il trouver des propositions, un mapping des attributs, etc. dont Isidore va se servir ? Comment traduire en RDFa de la sémantique contenu dans les attributs des balises TEI dans le HTML lorsque l'on fait la transformation ? Il manque une documentation et des exemples pour trouver cette information¹⁰.
- 21 Afin qu'il soit bien compris qu'il n'y a pas de contradiction entre « la communauté RDF » et « la communauté TEI », il convient de bien expliquer comment exprimer du RDF dans des documents TEI. Le trésor de guerre des chercheurs se sont les documents XML-TEI sur lesquels ils travaillent souvent depuis plusieurs années. Ils pensent sans doute à raison qu'il serait dangereux de capitaliser uniquement du RDF dans du HTML qui pour eux constitue un format de sortie[[note : L'exemple proposé dans l'atelier, une page HTML avec du RDFa, n'est qu'un exemple et doit être vu comme une « sortie » en ligne, comme une « édition » HTML et pas comme conteneur pérenne de données et d'information.]]. Une des manières d'inciter la communauté des humanités numériques à employer cette structuration de l'information est de leur fournir des moyens sûrs de l'exprimer dans des documents XML-TEI et de leur donner des moyens de l'exporter de

manière fiable. On souhaite pérenniser le travail, et le faire une seule fois, en ce sens la notion de workflow est à prendre en compte.

- 22 D'après Stéphane Pouyllau, le problème n'est pas, alors que l'on passe déjà beaucoup de temps à encoder en TEI, qu'aujourd'hui des informaticiens viennent nous demander d'exprimer les choses en RDF. Quelque soit le langage balisé, on est capable, selon lui, de reconstituer des formats de sortie. De la même manière que de fabriquer des bases de données relationnelles n'empêche pas d'exprimer les choses en RDF, il n'y a pas de problèmes avec la TEI. Afin d'amener une communauté des humanités numériques très largement XML-TEI à inscrire des métadonnées RDFa dans le HTML, il faudrait peut-être alors que la communauté RDF et TEI rédige ensemble une spécification simple exprimant de la TEI en RDF, et proposer des *web services* de génération du HTML/RDFa correspondant pour les documents TEI conformes à cette spécification. Cela existe sans doute.
- 23 Pour ceux qui ne sont pas encore passés au XML-TEI, peut être convient-il également de développer des logiciels performants pour faire de l'annotation RDF. Par ailleurs, il convient de trouver la bonne ontologie et savoir si l'information que l'on décrit existe quelque part. Un des problèmes aujourd'hui est qu'il n'y a pas encore de moteurs de recherche pour savoir si l'information est déjà décrite quelque part. Enfin, il n'existe pas vraiment d'outils accessibles à des non-informaticiens qui permettent une réelle démocratisation du RDFa. L'exploitation des informations en RDFa nécessite souvent des développements à façon.
- 24 L'atelier débouche sur la proposition d'un site web qui :
- 25 - référencerait et documenterait les ontologies disponibles en LOD pouvant intéresser les chercheurs en sciences humaines et sociales, c'est à dire classées, décrites, selon les besoins des disciplines ;
- 26 - comporterait des tutoriels d'utilisation du RDFa ;
- 27 - fournirait une indication concernant les machines en mesure d'exploiter dès à présent ou à terme les attributs RDFa ;
- 28 - offrirait une documentation sur les attributs exploitables par les machines « RDFa-compliant » ;
- 29 - présenterait des démonstrations sur les utilisations possibles et des *mashups* mettant en évidence les possibilités imprévues ;
- 30 - rassemblerait des guides « pas par pas » faisant sentir que le coût d'entrée n'est pas si élevé, ces guides se déclinant en fonction des produits finis que souhaiterait obtenir le chercheur.
- 31 Il pourrait subsidiairement :
- 32 - fournir des applications automatisant un peu l'expression en RDFa des attributs dans le document XML-TEI.

BIBLIOGRAPHIE

Webographie

Stéphane Pouyllau, « Construire le web de données pour les données de la recherche en SHS : comment utiliser RDFa ? », *sp.Blog*, 30 août 2010, <http://blog.stephanepouyllau.org/401>, consulté le 27 septembre 2012.

Got, « Comprendre RDFa en 5 minutes », *Les petites cases*, 6 décembre, 2008, <http://www.lespetitescases.net/comprendre-rdf-en-moins-de-5-minutes>, consulté le 27 septembre 2012.

Got, « Comprendre RDFa en 5 minutes », *Les petites cases*, 6 décembre, 2008, <http://www.lespetitescases.net/comprendre-rdf-en-moins-de-5-minutes>, consulté le 27 septembre 2012.

Got, « RDFaiser votre blog, 1^{ère} partie : la théorie », *Les petites cases*, 24 février, 2008, <http://www.lespetitescases.net/comprendre-rdf-en-moins-de-5-minutes>, consulté le 27 septembre 2012.

Got, « RDFaiser votre blog, 2^{ème} partie : la pratique », *Les petites cases*, 24 février, 2008, <http://www.lespetitescases.net/rdfaisier-votre-blog-2-la-pratique>, consulté le 27 septembre 2012.

Got, « RDFaiser votre blog, 3^{ème} partie : l'exploitation », *Les petites cases*, 24 mai, 2008, <http://www.lespetitescases.net/rdfaisier-votre-blog-3-exploitation>, consulté le 27 septembre 2012.

Adida, Ben and Mark Birbeck, éditeurs, « RDFa 1.1 Primer, Rich Structured Data Markup for Web Documents », W3C Working Group Note, 7 juin 2012, <http://www.w3.org/TR/xhtml-rdfa-primer>, consulté le 27 septembre 2012.

Dan Brickley, « Introducing FOAF », *FOAF Project*, 2000, <http://www.foaf-project.org/original-intro>, consulté le 27 septembre 2012.

The Athena WP4 wiki, <http://www.athenaeurope.org/athenawiki>, consulté le 27 septembre 2012.

Got, « XML vs RDF : logique structurelle contre logique des données », *Les petites cases*, 29 août, 2010, <http://www.lespetitescases.net/xml-vs-rdf>, consulté le 27 septembre 2012.

Normes et standards

Standards RDF sur le site du W3C, <http://www.w3.org/standards/techs/rdf>, consulté le 27 septembre 2012.

Standards RDFa sur le site du W3C, <http://www.w3.org/standards/techs/rdfa>, consulté le 27 septembre 2012.

SKOS Simple Knowledge Organization System, <http://www.w3.org/2004/02/skos>, consulté le 27 septembre 2012.

Lou Burnard et Bauman Syd éditeurs, « P5 : Guidelines for Electronic Text Encoding and Interchange. », *Text Encoding Initiative*, 2007, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html>, consulté le 27 septembre 2012.

ICOM, *CIDOC Conceptual Reference Model*, <http://www.cidoc-crm.org>, consulté le 27 septembre 2012.

NOTES

1. Voir : http://fr.wikipedia.org/wiki/Dublin_Core.
 2. Voir : http://fr.wikipedia.org/wiki/Dublin_C
 3. Voir : http://fr.wikipedia.org/wiki/Uniform_Resource_Identifier
 4. Cette question terminologique est revenue de manière récurrente au cours de la discussion. Nous employons dans ce compte-rendu les termes « structuration de l'information » qui nous semble moins problématique que le mot structuration employé seul, et « étiquetage » à la place de labélisation qui est un anglicisme
 5. Sur la différence de logique entre XML-TEI et RDF, voir ce billet du blog Les petites cases et les commentaires qu'il a suscités.
 6. Voir la démonstration du service web sur le site de l'IRI.
 7. L'équipe mentionnée au cours de l'atelier s'intitule Modélisation conceptuelle des connaissances médicales, unité INSERM U936 de l'université Rennes-1.
 8. Voir la documentation de la plate-forme de recherche Isidore.
 9. Voir à ce sujet l'initiative propre à Google, schema.org mais aussi les « rich snippets » de google.
 10. Le Centre de ressources numériques (MEET) pourrait peut-être proposer cela ?
-

RÉSUMÉS

L'atelier consiste à évoquer la qualité des métadonnées et des données de nos corpus numériques à la suite de la non-conférence de Paul Bertrand. Comment dépasser le plus petit dénominateur d'aujourd'hui, souvent le Dublin Core, pour aller plus loin ? En discutant autour du modèle RDF et plus particulièrement du RDFa pour le web (HTML), il s'agit d'entrevoir ce qu'il est possible de faire en matière de guide de bonnes pratiques, techniques, vecteurs, comment mieux « pousser » des données, dans Isidore par exemple.

L'historien proto-programmeur. Outils et méthodes pour (re)travailler ses données

Proposé par Sylvain Machefert

Collectif

L'autonomie dans la manipulation de ses données

- 1 L'intérêt principal de ce type d'outil est l'autonomie qu'il offre au chercheur dans le traitement de ses données. Le besoin de nettoyage et de modification de données est fréquent, mais il se rencontre particulièrement lors du passage d'un logiciel à l'autre. Ce sont les passages entre deux logiciels, offrant des fonctionnalités différentes, qui constituent souvent un obstacle majeur au traitement heuristique.
- 2 Les données restent stockées localement, mais il est possible de les consolider à partir de données distantes. Ainsi, lors de la constitution d'une base de données prosopographique, il est possible d'utiliser les données issues de DBpedia, puis de les corriger.
- 3 Google Refine est adapté pour de petites tâches quotidiennes du chercheur sur ses données. Par exemple, si les villes ne sont pas dans le champ « ville », il est possible de les y déplacer automatiquement. Il est également possible de créer un morceau de code permettant de répéter une opération autant de fois que nécessaire. Ce code permet de créer un bouton automatisant une tâche répétitive.
- 4 Google Refine n'est pas le seul outil de ce type. On peut par exemple utiliser RelFinder (Interactive Relationship Discovery in RDF Data).

La structuration des données

- 5 Alors que le Web de données manque encore de maturité, il est nécessaire d'y contribuer pour le faire progresser. Or, les outils utilisés le plus largement sont assez

frustrés. C'est le cas de Microsoft Word ou de Google Books, qui font partie de la panoplie quotidienne de l'historien, sans proposer aucune structure forte des données qui sont collectées. On trouve des dispositifs permettant de passer de fichiers bureautiques à une TEI structurée de façon minimal en utilisation Lodel ou Oxgarage. D'autres utilisations structurantes sont possibles, comme le passage d'un traitement de texte vers un tableur, voire, idéalement, vers une base de données.

- 6 On bloquera souvent sur le problème de l'identification unique des objets manipulés. Une des pistes permettant de relier les données entre elles est l'utilisation de l'International Standard Name Identifier (ISNI), un identifiant unique pour les « identités publiques », qui joue un rôle comparable à celui joué par l'ISBN dans le monde du livre.
- 7 *The International Standard Name Identifier (ISNI) is an ISO standard (ISO 27729) that identifies public identities of parties. Essentially, this means the identities used publicly by individuals or organizations involved in creating, producing, managing and distributing content. The ISNI system uniquely and authoritatively identifies Public Identities across multiple fields of creative activity. ISNI provides a tool for disambiguating Public Identities that might otherwise be confused (such as Michele Smith the female singer, and Michele (Italian form of "Michael") Smith the male author). The ISNI is not intended to provide direct access to comprehensive information about a Public Identity – instead, it's a tool that can provide links to other systems where such information is held.*

De la ligne de commande aux interfaces visuelles

- 8 Reste à convaincre les chercheurs de faire évoluer leurs pratiques. Trop de chercheurs perdent du temps à construire à la main l'index de personnes de leur ouvrage, alors qu'il existe des outils bureautiques simples pour automatiser cela. La formation est un levier majeur pour résoudre cette difficulté. Des exemples édifiants peuvent aussi nourrir l'argumentaire. Une enquête historique mal montée pourra être partiellement récupérée avec une utilisation raisonnée de Google Refine, par exemple. Les interfaces en ligne de commande permettent depuis longtemps de réaliser ce type de traitement. Ce qui change avec un outil de la famille de Google Refine, c'est l'interface, très proche de celle d'un tableur, qui rend l'outil à la fois abordable et visible, c'est-à-dire rassurant. Il ne faut pas sous-estimer la capacité à l'auto-formation et respecter la diversité des usages, pour ne pas introduire un rapport contraint à la technique. En sociologie des médias, les figures du nul en informatique ou du super-geek permettent d'évacuer certains problèmes.
- 9 Montrer concrètement comment on peut insérer une référence bibliographique structurée dans Word avec Zotero remplace beaucoup de discours sur l'utilité des bibliographies structurées. De même, les tutoriels de Google Refine sont d'excellents ambassadeurs pour favoriser l'appropriation. Il faut aussi miser sur un continuum temporel, avant et après la formation, pour initier, construire et consolider les méthodes transmises en formation, afin de s'assurer que la greffe prenne, selon le rythme et les besoins de l'utilisateur. Enfin, le rapport à la documentation et à la recherche de solution (forums, listes de discussion, etc.) doit évoluer, pour devenir plus maîtrisé et moins timoré, afin de permettre une plus grande autonomie.

Conclusion

- 10 Si Google Refine a servi de fil conducteur à cette réflexion sur l'historien proto-programmeur, il n'est qu'un exemple marquant, et il faut éviter de s'arrêter à un outil particulier. Les logiciels proposant de structurer les données de l'historien avec plus de simplicité qu'avec des instructions saisies en ligne de commande se multiplient, et c'est un atout pour l'émergence de compétences plus fortes dans la recherche en sciences humaines et sociales. La condition de la réussite de leur implantation dans les usages quotidiens est l'existence d'une formation bienveillante, autant que la possibilité de recourir à l'auto-formation, à condition que les usages potentiels soient promus et illustrés de façon concrète et convaincante.

Les fonctionnalités de Google Refine

Google Refine est un logiciel libre écrit en JAVA, sous licence BSD. Il fonctionne comme une application web, mais s'utilise de façon locale, sur un poste personnel.

Il permet :

- L'analyse des données, de manière simple : facettes sur occurrence, sur valeurs numériques et options d'analyse ;
- De gérer des volumes assez importants (centaines de milliers de lignes) ;
- De nettoyer ses données, notamment les erreurs de saisie ou les données non normalisées : rapprochement « sonore », distance de Levenstein, unification de la casse, normalisation des informations ;
- D'enrichir ses données depuis des données du web (web services et linked data) :
 - Connexion à des services web : par exemple récupérer les coordonnées géographiques via geocode de Yahoo,
 - Connexion au LinkedData (LinkedGeoData par exemple) ;
- De transformer ses données grâce à des expressions régulières, à l'aide du langage Google Refine Expression Language (GREL et de Jython (variante Python pour java) ;
- D'importer / d'exporter ses données :
 - Assez ouvert en termes de format : CSV, XML, Triplets RDF, Json, xls, etc.
 - En entrée : CSV, Excel, HTML, RDF,
 - Des templates permettent de générer des sorties à façon,
 - Export en RDF (une sorte de première étape d'Open Data structuré dans une norme du web de données).

Ressources

- 11 Google Refine: <http://code.google.com/p/google-refine/>
 Didacticiels vidéos sur Google Refine, notamment dans la perspective de publier des données dans le Linked Data : <http://freemetadata.org/>
 OxGarage : <http://oxgarage.oucs.ox.ac.uk:8080/ege-webclient/>
 RelFinder: <http://www.visualdataweb.org/relfinder.php/>
 ISNI: <http://isni.org/>
 Freebase: <http://www.freebase.com/>
 Zotero : <http://www.zotero.org/>

RÉSUMÉS

Entre l'historien programmeur et l'historien « wordiste », uniquement capable de manipuler sommairement un traitement de texte, il faut aussi accorder une place à l'historien « bidouilleur », que l'on préférera appeler l'historien proto-programmeur. Celui-ci est capable d'exploiter au mieux ses données sans s'arrêter à un tableur, sans pour autant maîtriser un langage de programmation ou attendre qu'un informaticien traite ses données. Cette figure intermédiaire est essentielle au développement des humanités numériques. Le logiciel Google Refine est un bon exemple d'instrument accessible sans apprentissage de la programmation, proposant pourtant des fonctionnalités assez avancées de traitement de l'information. Google Refine (GR) est issu d'un logiciel libre, Freebase gridworks, que Google a acheté en 2010. GR est un logiciel libre et gratuit, multiplateforme (Linux/Windows/Mac) qui permet, selon le propos de la page d'accueil du site : « Google refine is a power tool for working with messy data » (« Google Refine est un outil de productivité pour travailler avec des données désordonnées ».) Tout un programme, propre à séduire un historien dont les données initiales sont souvent d'inégale qualité !

INDEX

Mots-clés : données, programmation, histoire, sciences humaines

Sommes-nous en train de perdre la mémoire ? Mémoire et archivage du web

Proposé par Frédéric Clavert et Clément Oury

Collectif

Introduction

- 1 En 2003, dans son article « *Scarcity or Abundance ? Preserving the Past in a Digital Era* »¹, Roy Rosenzweig attirait l'attention sur les enjeux de la préservation des traces numériques du passé. Malgré l'abondance des sources désormais disponibles sur le web, il existe un risque réel de perdurer dans une rareté de l'information qui a toujours été, et de ne pouvoir, à l'avenir, que construire une histoire fragmentaire de notre époque. Insistant sur l'urgence de trouver les moyens de pérenniser les sources numériques primaires, il montrait les limites de certaines techniques, notamment l'impression des sources numériques (l'*Auswärtiges Amt* à Berlin procède de cette manière pour les courriers électroniques), la conservation d'un lecteur pour les supports obsolètes (pratiqué par l'Institut national de l'audiovisuel en France, INA), la migration régulière d'un support ou d'un logiciel à l'autre ou la virtualisation. Pour Rosenzweig, une des solutions est la mobilisation commune des historiens et archivistes, comme celle qui a permis, en 1934, la création de la *National Archives and Records Administration* aux États-Unis avant l'éloignement de leurs organisations professionnelles respectives². Ce dernier point rappelle l'une des conclusions de THATCamp Paris 2010 et l'une des particularités des *Digital Humanities* : la nécessité de l'interdisciplinarité et du travail en commun de multiples corps de métier.
- 2 L'article de Roy Rosenzweig date de 2003. Depuis, le web a évolué vers plus de « Web 2.0 » et, notamment, vers les réseaux sociaux comme Twitter ou Facebook, pour nommer les plus connus. La question de la préservation des données publiées sur les réseaux sociaux se pose avec autant d'acuité que celle du site satirique *Bert is evil* pris en exemple par Rosenzweig, où il souligne que la préservation de ces données pose deux

questions principales. Doit-elle être laissée à des organismes privés – comme Internet Archive et surtout les entreprises possédant les réseaux sociaux dont l'intérêt est, avant tout, l'exploitation commerciale de ces réseaux et, certainement pas, l'archivage à des fins d'analyse historique ? Qui peut accéder à ces données et quelle législation va encadrer cet accès, entre la loi des trente ans s'appliquant aux archives publiques et un « droit à l'oubli » en émergence³ ?

- 3 Un article récent⁴ démontre que les ressources partagées sur les réseaux sociaux disparaissent au rythme de 0,02 % par jour. Depuis la fin des révolutions libyenne, égyptienne et tunisienne du printemps 2011, environ 30 % des sites partagés via les réseaux sociaux ont tout bonnement disparu. En clair, la situation décrite par R. Rosenzweig ne s'est que peu améliorée. De plus en plus de bibliothèques nationales archivent des portions du web, au titre du dépôt légal. Cet archivage vient s'ajouter à celui d'entreprises privées comme Internet Archive et sa *wayback machine* toutefois quelque peu négligée ces derniers temps. Des logiciels pour la préservation de collections numériques, appelés *repository* (Fedora, DSpace, Alfresco⁵) permettent, bien utilisés, de gérer des millions de documents de manière pérenne (si le financement est bien entendu aussi pérenne). Enfin, des identifiants uniques ou permanents apparaissent : le DOI (*Digital Object Identifier*), devenu une norme ISO en 2012, implémente le système Handle, les adresses web permanentes de type purl, etc.⁶.
- 4 Dans une perspective plus historique, la question de la gestion des archives a toujours été double. D'un côté, se pose celle de la sélection des archives, un problème important dans le domaine numérique à l'ère de « l'infobésité », de l'autre, celle de leur préservation. La perte des archives n'est pas un phénomène nouveau. En Égypte, par exemple, nous disposons des papyri qui ont été préservés dans les zones sèches et désertiques. Ceux qui ont été conservés dans des endroits trop proches du Nil, en zone humide, n'existent plus. Les guerres ont toujours fait disparaître des archives, soit de manière directe à cause des combats, soit de manière indirecte en raison de leurs conséquences politiques. Certaines disparitions furent temporaires. Ainsi, la division est/ouest pendant la Guerre froide a bloqué pour les historiens de l'Ouest l'accès à de nombreuses archives du régime nazi qui avaient été emportées par les Soviétiques. Mais, la multiplication des sources numériques donne une dimension nouvelle par sa masse à ces problèmes classiques pour les archivistes et les historiens.
- 5 À travers plusieurs questions principales, cet atelier se concentre sur la spécificité de cette problématique de l'archivage et de la perte à l'ère du numérique, notamment à propos de la mémoire du Web. Peut-on mesurer l'ampleur des pertes d'archives, et les comparer à celles que nous avons déjà connues auparavant ? Quelles solutions de préservation et d'archivage émergent ? Si nous devons sélectionner les archives numériques à conserver, quels critères vont être adoptés ? Vont-ils différer de ceux utilisés pour la sélection des archives de l'ère « papier » ? Pouvons-nous faire le choix de l'abondance, c'est-à-dire de tout garder ? On rejoint là un questionnement traditionnel, mais qui a sans doute une résonance spécifique dans le cas des sources numériques. Enfin, un atelier animé par G. Poupault avait déjà abordé la question lors du THATCamp de Paris en 2010, soulignant qu'il fallait « archiver le temps » selon l'expression de Frédéric Clavert, puisque le document numérique, fluide, change de forme et de contenu. Il faudrait donc conserver non seulement le document, mais aussi tout son historique. Pouvons-nous le faire ?

Archiver le Web : la politique de la BNF

- 6 Clément Oury présente tout d'abord le dépôt légal du Web dont la BNF est chargée depuis la loi relative au droit d'auteur et aux droits voisins dans la société de l'information (DADVSI) de 2006, parallèlement à l'INA qui s'occupe de collecter les sites de la communication audiovisuelle⁷. La loi étend le dépôt légal, dont l'origine remonte à 1537 sous François I^{er} et qui a d'abord été appliquée aux livres publiés en France, à tous les « signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique⁸. » La radio, la télévision et les logiciels (dont les jeux vidéo) étaient, de leur côté, soumis au dépôt légal, depuis 1992, et le besoin d'archiver le nouveau support qu'est le Web s'est fait sentir à partir du début des années 2000. La loi permet, en outre, l'archivage des sites sans qu'une autorisation soit à demander aux personnes qui éditent et produisent des sites sur le domaine national français. En contrepartie, la consultation des archives du Web s'effectue uniquement sur place.
- 7 Ce cadre juridique nouveau est très important : il signifie que les sources Web sont considérées comme des contenus patrimoniaux au même titre que les autres documents collectés par la BNF. Du point de vue de cette institution publique nationale, l'objectif est donc d'organiser la mémoire du Web en fonction de son savoir-faire et de sa mission patrimoniale traditionnelle. Mais, pour y parvenir, elle a aussi besoin aussi des regards externes et, notamment, de ceux des chercheurs qui sont les principaux utilisateurs de ces collections.
- 8 Cette mission est cependant difficile à mener sur un terrain aussi vaste et mouvant que le Web. D'une part, si le dépôt légal s'applique en théorie à ce qui est diffusé sur le plan national, cela n'a pas grand sens de parler de territorialité sur Internet. En pratique, il s'agit donc de collecter des sites produits et édités sur le « Web français » (le domaine national français en « .fr » et les sites des départements et des territoires d'outre-mer), ou les sites de personnes physiques ou morales domiciliées en France (sites en « .org », « .net », « .com », etc.). Ce n'est pas la langue qui définit le dépôt légal. Ainsi, un site s'exprimant dans une autre langue mais hébergé en « .fr » sera également concerné par l'archivage du Web français. D'autre part, le dépôt légal est censé être exhaustif. Mais, archiver le Web de façon exhaustive est bien sûr un objectif impossible. La BNF remplace donc l'exhaustivité par la « représentativité », notion qui suscite beaucoup de discussions lors de l'atelier.
- 9 Enfin, une autre distinction cruciale est apparue dans les échanges sur la définition du périmètre à archiver. Comment distinguer entre la nature privée et publique d'une source Web ? En effet, le dépôt légal s'applique à ce qui est publié, non aux échanges à caractère privé. Du point de vue de la BNF, un blog est une publication et non une conversation. Il est donc concerné par le dépôt légal. En revanche, un forum ne l'est pas et donc n'est pas conservé. Ceci peut certainement se discuter. Une correspondance privée, à l'ère du numérique, ne sera pas non plus archivée par la BNF et il faudra recourir à des services en ligne pour récupérer des boîtes mails. De même, les espaces privés des réseaux sociaux ou des sites intranet sont exclus. *A priori* la limite entre le public et le privé est posée par l'internaute lui-même. Mais il est vrai que ces notions sont très floues quand elles s'appliquent à Internet, et sont en voie de redéfinition.

Médiévizmes @medieviz

@inactinique #tcp2012 on expliquera sur le pad pourquoi ne pas conserver un forum ?

- 10 Une contrainte forte influence ces choix et ces définitions : la méthodologie et les outils utilisés pour la collecte. Concrètement, celle-là est effectuée par des robots qui « moissonnent » le Web français, c'est-à-dire 2 millions de sites en « .fr », avec une profondeur de capture moyenne. Tout ce qui est collecté est mis sur un serveur. Certains sites sont « photographiés » une fois par jour si leur importance patrimoniale le justifie. C'est le cas, par exemple, des sites des grands quotidiens d'actualité. La discussion porte aussi sur d'autres types de collectes plus ciblées et liées à un grand événement ou à un thème précis. Dans ce cas, la moisson peut aller plus en profondeur. Les campagnes électorales en ligne, régionales, nationales ou européennes, sont ainsi préservées par la BnF. L'an dernier une collecte spécifique a été effectuée dans l'urgence sur les révolutions du « printemps arabe » entre décembre 2010 et mars 2011.

Jean-Pierre Masse @jpmasse

#tcp2012 la BnF et l'INA ont "des robots qui se promènent" et qui explorent le web

- 11 L'exemple est discuté assez longuement car il pose bien le problème de la constitution d'une archive du Web. Les sites sont majoritairement en français et non en arabe. Du point de vue de l'historien, il s'agit d'une source européenocentrée qui introduit un biais sur les événements. Clément Oury est conscient des limites du corpus, mais souligne qu'il faut bien avoir en tête les conditions matérielles de collecte du côté des archivistes. La BnF, tout comme les gouvernements, a été prise au dépourvu et a organisé une collecte dans l'urgence en ayant conscience de la fragilité de ces sites. Le personnel étant majoritairement francophone, il s'est en effet attaché en priorité à des sources en français. Cependant, le service des collections orientales a pu leur indiquer des sites en arabe qui ont ainsi été moissonnés. Une coopération internationale a pu être construite pour ajouter d'autres sources et renforcer les contenus de cette collection Web⁹. Sur Internet comme pour les archives non numériques, souligne-t-il, ce qui est important est d'expliquer clairement comment la collection a été constituée. Les historiens peuvent ensuite prendre en compte ces biais dans leur analyse des sources.

Des problèmes classiques qui se posent de façon plus aiguë

- 12 Il est intéressant de constater que les discussions de l'atelier sur la mémoire du Web abordent de grandes questions classiques sur la constitution des archives et le rapport des historiens à leurs sources. C'est le cas, tout d'abord, de la distinction entre privé et public. L'hébergeur d'un forum qui n'est pas pris en compte par le dépôt légal du Web va décider s'il archive ou non son site en recourant à des services privés. L'archivage des réseaux sociaux est avant tout constitué par les sociétés privées qui les développent. Or les entreprises privées ont toujours eu une perspective différente de

celles des pouvoirs publics dans la constitution de leurs archives. Elles gardent ce qui leur est utile, et non ce qui va être utile aux historiens. C'est le cas de BNP-Paribas, par exemple, qui voulait récemment jeter toutes ses archives liées à l'Indochine (puisque la Banque d'Indochine faisait partie d'une partie de son histoire sans intérêt pour la banque d'aujourd'hui). De la même façon, risquons-nous un jour de devoir aller voir chaque hébergeur (par exemple Twitter, qui ferme son API) pour accéder à des archives qui auront le biais de l'entreprise privée ? La nature numérique ou non de ces archives ne change pas fondamentalement le problème.

- 13 Sur ce plan, les collections Web de la BnF ont les limites de tout dépôt légal. Certains sites en sont exclus et, même pour ceux qui sont concernés par la loi, il ne s'agit pas d'une solution de pérennisation en tant que telle. Certes, si un site disparaît par accident et qu'il a été collecté, la BnF peut lui fournir une copie de secours, à charge pour lui de remettre en ligne le contenu. Mais ce n'est pas la fonction principale de l'archive du Web de l'institution. Des services commerciaux d'archivage et de sauvegarde spécifiques existent pour ce type de besoins. Enfin, certains participants à l'atelier évoquent les risques de « privatisation », ou du moins le caractère limité de l'accès à ces collections numériques puisqu'il faut aller sur place pour les consulter, en vertu de la loi DADVSI¹⁰. Cependant, c'est une contrepartie à la possibilité qui est donnée à la BnF de collecter les sites sans demander d'autorisation préalable à leurs auteurs – une exception à la règle habituelle –, ce qui limiterait beaucoup le nombre de sites effectivement collectés, et donc la qualité de l'archive finale¹¹.
- 14 La sélection de sources au moment de la constitution de l'archive est un autre problème récurrent. La position de la BnF présentée au début de l'atelier est de privilégier la « représentativité » des sources et non une impossible exhaustivité. Clément Oury souligne aussi que la perspective patrimoniale de l'institution est « d'anticiper sur les besoins du chercheur dans 500 ans ». Ces expressions, bien sûr, font réagir les participants. La représentativité est vue avant tout par la BnF comme une question de fréquence de collecte. L'indexation est automatique et se fait sur l'ensemble des sites pour les collectes larges. Il y a intervention humaine pour la sélection des sites de base utilisés pour les collectes thématiques ciblées, par exemple le Web politique en 2012. Dans ce cas, les sites sont fournis par une vingtaine de bibliothèques. En outre, les choix de sélection et d'indexation sont documentés et accessibles. La liste des URL des sites archivés depuis 2002 pendant les périodes électorales est ainsi accessible sur www.data.gouv.fr¹². Un effort important est donc fait pour que le processus soit transparent.

Frédéric Clavert @inactinique

@medieviz échantillonnage? On sélectionne une partie des sources pour conservation ou on sélectionne un échantillon représentatif? #tcp2012

- 15 On sait bien, toutefois, que les sources considérées intéressantes à analyser par les historiens ne se limitent pas à ce que les administrations gouvernementales jugent ou ont jugé, à une certaine période, nécessaire ou pertinent de conserver. Les historiens des représentations ou de l'intime, entre autres, sont familiers de ces débats sur ce qui peut devenir source, même les éléments considérés comme insignifiants à un moment donné, ou du point de vue des institutions politiques. Ceci recoupe également une différence familière de perspective et de culture professionnelle entre les historiens,

dont le premier mouvement serait de tout conserver, et les archivistes qui sont conscients de la nécessité de faire des choix pour pouvoir matériellement conserver des collections d'archives dans de bonnes conditions. Bien sûr, « anticiper sur les besoins du chercheur dans 500 ans » peut sembler un objectif d'une « arrogance intellectuelle incroyable », comme le notent certains participants, mais c'est aussi le mandat attendu d'une institution patrimoniale nationale qui a pour devoir de faire des choix afin d'être opérationnelle. Cette formule est certainement propre à provoquer la réflexion et traduit bien le fait que, dans le domaine du numérique comme dans des fonctions patrimoniales plus classiques, on récupère ce qu'on choisit de garder. Il y a donc bien, à la base et dans toute politique publique d'archivage, un choix politique.

- 16 En fin de compte, sur ces deux points, les problèmes ne sont pas nouveaux mais ils acquièrent une dimension particulière en raison de la masse des données collectées sur le Web et de la fluidité des évolutions en cours. Cependant, d'autres enjeux débattus pendant l'atelier montrent qu'il existe une spécificité de l'archivage du Web qui nécessite de repenser la notion de préservation.

Repenser la notion de préservation

- 17 Un premier enjeu évoqué par les participants concerne l'accessibilité des données. Le stockage est une chose, mais la préservation des collections suppose aussi de mettre à disposition des outils pour les rendre utilisables par les chercheurs. Dans un contexte numérique, la question est particulièrement importante. Il faut fournir des outils de fouille des données, et la BNF réfléchit à une navigation dans les versions conservées, à une indexation de meilleure qualité, ou à des cartographies de sites avec Gephi, par exemple¹³. Mais ceci n'est pas encore disponible. Avec 17 milliards de fichiers, il est difficile d'avancer vite.
- 18 En outre, il faut assurer la lisibilité des archives numériques. Se pose donc le problème de l'évolution des formats et des logiciels, qui deviennent rapidement obsolètes. Il faut conserver à la fois la source informatique et la source logicielle. Pour les archives du Web, la migration vers un format nouveau fonctionne mal. On privilégie les logiques d'émulation pour lire le Web d'il y a cinq ans avec des outils reproduisant les technologies de l'époque. Avec un navigateur et quelques *plug-ins*, on peut lire 99 % des archives actuellement conservées. Cependant, toutes les procédures d'émulation ne sont pas encore bien établies. Il faut aussi faire des répertoires de formats, les documenter et stocker des exemples de formats et d'outils de lecture. La BNF travaille avec Microsoft sur les formats et les émulations, mais cela pose des problèmes juridiques considérables pour les logiciels propriétaires.
- 19 Même dans le cas où les formats ne sont pas devenus obsolètes, ils constituent un obstacle de taille dans l'accès aux archives du Web. En effet, les robots moissonneurs maîtrisent certains formats et pas d'autres¹⁴. Les données incluses dans des pages auxquelles ils ne peuvent accéder ou dans des formats illisibles pour eux ne seront donc pas conservées. Les chercheurs qui travaillent sur ces collections ont donc l'habitude d'utiliser des archives « à trous ».
- 20 La question technique et la masse même des données en jeu sur le Web posent de façon brutale la question des moyens et de leur affectation. Archiver le Web demande de gros équipements et des budgets conséquents qu'il faut prévoir bien en amont. Les sources

numériques sont pérennes à condition qu'un investissement suffisant soit prévu pour fournir une redondance du stockage en interne, et une réplication dans des sites secondaires. En pratique, il s'agit actuellement d'une annexe en banlieue, selon un système partagé avec le Centre informatique national de l'enseignement supérieur (CINES)¹⁵. Les projets ont été lancés récemment dans des collectes sophistiquées, où le budget nécessaire au stockage n'a pas été inclus, courent le risque de perdre leurs données (cela s'est déjà produit).

Dépôt légal Web BnF @DLWebBnF

La réplication est capitale ds le domaine de la conservation du numérique #tcp2012

Jean-Pierre Masse @jpmasse

#tcp2012 la BnF redonde en interne ;-)

- 21 De plus, les institutions comme la BnF doivent faire un second arbitrage : affecter plus de ressources à la collecte (plus d'archives disponibles plus tard aux chercheurs) ou à la valorisation des données actuellement conservées (ce qui permet un meilleur accès immédiat des chercheurs, mais au détriment de la quantité de données collectées, et donc des archives numériques disponibles plus tard). C'est un problème de financement et de choix, toujours difficile, entre le court et le long terme. La BnF a une position institutionnelle qui la pousse à accorder autant d'intérêt aux besoins du chercheur dans 500 ans qu'à celui d'aujourd'hui. On ne peut pas connaître les futurs utilisateurs des sources, mais on conserve pour eux. L'archivage du Web est donc nécessairement pris en charge par une institution capable mettre en place une organisation importante, ce qui introduit une distorsion inévitable pour les futurs historiens.
- 22 La BnF et le dépôt légal du Web ne peuvent couvrir qu'une partie des archives Internet. La plate-forme HAL conserve les productions scientifiques déposées dans le cadre des archives ouvertes, notamment la littérature grise¹⁶. Le CINES offre des solutions aux bibliothèques universitaires¹⁷. Les responsables des diverses institutions sont censés le faire pour leurs données. L'entrepôt SPAR (Système de préservation et d'archivage réparti) de la BnF propose également un service tiers-archivage¹⁸. Le TGE-Adonis est également un acteur important dans ce domaine. Son offre d'archivage, lancée en 2009, va être redéployée très prochainement, toujours avec le CINES. Des coopérations devraient être développées entre tous ces acteurs, mais il n'est pas toujours simple de les mettre en place, essentiellement pour des raisons institutionnelles (ils dépendent de ministères différents).
- 23 Enfin, même dans le cas de projets de taille bien plus modeste lancés par des chercheurs, la collecte et l'archivage de sources numériques nécessitent un soutien institutionnel fort. C'est le cas, par exemple, du projet *Bracero History Archive* qui collecte et présente des sources diverses sur l'histoire d'un programme bilatéral qui a, entre 1942 et 1964, organisé l'immigration de travailleurs mexicains aux États-Unis¹⁹.
- 24 Au-delà de ces questions fondamentales de moyens et de taille institutionnelle de ces projets, la spécificité du web comme support, dominé par les flux de données et les fluctuations dans le temps, nous oblige tous, historiens comme archivistes, à repenser ce que nous entendons par « préservation ». À chaque lecture, l'archive doit être réactivée et recontextualisée. Il faut pouvoir documenter l'évolution de son contexte –

à la fois celui des choix de conservation, mais aussi celui des attentes et de *l'habitus* des lecteurs des sources. Ceci est particulièrement difficile à faire pour des sources du Web. Il manque des modèles de gestion des données permettant d'archiver aussi le temps du support de l'archive et le temps de sa réception. L'enjeu est donc de réinventer un modèle conceptuel autour de cette notion de préservation appliquée aux archives du Web.

Dépôt légal Web BnF @DLWebBnF

De la difficulté de pérenniser tout en continuant à donner accès et à documenter le contexte de consultation #tcp2012

Conclusion

- 25 Quatre points principaux se dégagent de l'atelier. Il est d'abord urgent de définir un modèle conceptuel de la préservation par l'accès (et de l'accès) aux archives du Web. C'est en fonction de ce modèle qu'il faudrait, dans un second temps, définir les missions à assurer au profit de l'utilisateur/lecteur. Le choix des moyens techniques nécessaires et des politiques de conservation devrait en découler. Enfin, les possibilités institutionnelles d'organisation de cet archivage devraient être développées dans un cadre de mutualisation au profit de tous les acteurs de l'enseignement supérieur et de la recherche. Aussi bien ces conclusions que le ton animé des discussions sont révélateurs du vif intérêt suscité par la problématique de la mémoire et de l'oubli sur le Web, mais aussi des divergences de perspectives qui subsistent entre archivistes et historiens.
- 26 Les suggestions de Rosenzweig en 2003 sont donc toujours aussi pertinentes. Le paradigme de la surabondance des sources introduit par l'ère numérique, même s'il a tout pour réjouir les historiens, constitue en réalité pour eux un redoutable défi. S'ils ne s'impliquent pas maintenant dans la réflexion sur la préservation des archives du Web et dans le lobbying en faveur de politiques d'archivage adéquates, en alliance avec les archivistes, ils se préparent des lendemains douloureux, paradoxalement dominés par la rareté plus que par l'abondance.

NOTES

1. *American Historical Review*, 108 (3), June 2003, pp. 735-762, reproduit dans le *Read/Write Book 2*, <http://press.openedition.org/265>.
2. Rosenzweig, *op. cit.*, pp. 759-761.
3. Cf. les débats en France en 2009-2010 autour de la « Charte du droit à l'oubli » proposée par le gouvernement, le nombre croissant de plaintes auprès de la CNIL en vertu de ce même droit à l'oubli sur Internet encore mal défini (voir le rapport d'activité 2011 de la CNIL, en ligne sur son site <http://www.cnil.fr>), et la proposition de règlement de la Commission européenne en janvier

2012 (COM(2012) 11 final, http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_fr.pdf).

4. H. M. Salah Eldeen et M. L. Nelson, « Losing My Revolution : How Many Resources Shared on Social Media Have Been Lost? », arXiv :1209.3026, septembre 2012.

5. Fedora Project (<http://fedoraproject.org/>), DSpace (<http://www.dspace.org/>) et Alfresco (<http://www.alfresco.com/>).

6. Voir le site de la DOI Foundation (<http://www.doi.org/>), celui du système Handle (<http://www.handle.net/>) et de purl (<http://purl.oclc.org/docs/index.html>).

7. Loi n° 2006-961 du 1^{er} août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information, <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000266350>. Voir aussi le Code du patrimoine, Titre III, dépôt légal (partie législative) <http://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006074236&idArticle=LEGIARTI000006845516> et la partie réglementaire (Article R132-23) <http://legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006074236&idArticle=LEGIARTI000025004800>, qui date seulement de décembre 2011.

8. Loi n° 2006-961 du 1^{er} août 2006, Titre IV, article 39.

9. Deux cents sites, blogs ou pages Facebook ont été sauvegardés en coopération avec Internet Archive, la Bibliothèque du Congrès, l'Université de Stanford, l'Université américaine du Caire, la British Library et la Bibliothèque d'Alexandrie. Voir <http://blog.bnf.fr/lecteurs/index.php/2012/07/02/la-revolution-du-jasmin-sur-la-toile/>

10. Pour respecter le droit d'auteur, aucune diffusion en ligne de ces archives du Web n'est autorisée. On ne peut pas non plus copier librement les sources numériques consultées sur place. Il est possible d'imprimer les pages consultées, de prendre des notes ou de copier des extraits de texte, mais les captures d'écran ne sont pas autorisées.

11. Le dépôt légal du Web en Grande-Bretagne, par exemple, fonctionne sur un modèle différent qui permet des collectes bien moins riches. Le cadre législatif britannique protège les droits d'auteurs même dans le cas des publications numériques comme des sites Web. La British Library ne peut donc qu'encourager les propriétaires et les producteurs à lui donner la permission d'archiver leur site. Voir par exemple le résumé sur le site de la British Library : <http://www.bl.uk/aboutus/stratpolprog/legaldep/>. Du coup, seules 22 000 captures de 6 000 sites web sont disponibles pour la recherche, un chiffre à comparer aux 2 millions de sites mentionnés par la BNF pour son échantillon représentatif du Web français en 2011, ou même des 30 000 sites concernés par les collectes ciblées, le tout représentant un total de 16,5 milliards de fichiers dans les archives de l'Internet de la BNF fin 2011.

12. Collectes du Web électoral par la BNF, données CSV accessibles en licence ouverte, <http://www.data.gouv.fr/donnees/view/Collectes-du-Web-électoral-par-la-BnF-551866>

13. <http://gephi.org/>

14. C'est le cas des accès par mot-clé ou payants et des liens non HTML, c'est-à-dire notamment les menus déroulants, des liens inclus dans des animations, des liens ouvrant vers des *popups*, de ceux qui n'apparaissent pas explicitement dans la barre d'état du navigateur et des vidéos et sons diffusés en flux plutôt que sous forme d'un fichier à télécharger.

15. Voir par exemple O. Rouchon (CINES) et L. Duploux (BNF), « *Collaboration over datasets replication* », un résumé disponible en PDF (http://www.mops1.com/oracle/event/pasig/downloads/Collaboration_over_datasets_replication_CINES_BnF.pdf)

16. <http://hal.archives-ouvertes.fr/>. HAL est pilotée par le Centre pour la documentation scientifique directe (<http://www.ccsd.cnrs.fr>). Voir aussi l'atelier *Évolution de l'archive ouverte HAL-SHS* du THATCamp 2012 animé par Laurent Capelli sur HAL_SHS

17. Il fournit aussi à HAL un archivage pérenne de ses collections. <http://www.cines.fr/>

18. SPAR, projet majeur piloté par la BNF, a été lancé en 2005 et mis en production à partir de 2010 pour permettre à l'institution patrimoniale de rationaliser et mieux gérer son stockage de données numériques. Il aura l'avantage, à terme, de fournir à la BNF un parc de stockage homogène et des procédures respectant les bonnes pratiques en matière d'archivage de données, tout en offrant une masse critique suffisante pour faire baisser les coûts et mutualiser le stockage entre plusieurs institutions. Voir la présentation de SPAR sur le site de la BNF : http://www.bnf.fr/fr/professionnels/conserver_spar.html

19. <http://braceroarchive.org/>

RÉSUMÉS

Allons-nous vers une période d'abondance gérée de l'information ou resterons-nous dans la rareté ? L'atelier essaiera d'analyser quelles sont les conséquences de ces pertes de données pour les sciences humaines et sociales, de resituer ces pertes dans un contexte historique (ce n'est pas la première fois que nous perdons des archives) et de dresser un éventail des solutions disponibles, notamment les politiques d'archivage du web actuellement mises en place.

INDEX

Mots-clés : mémoire, archivage, archives, perte de données

Rôle et enjeux du design graphique

Proposé par : Nicolas Thély, David-Olivier Lartigaud et Gilles Rouffineau

Collectif

Les enjeux du design graphique

- 1 L'idée de cet atelier est d'ouvrir un nouveau chantier. Un atelier précédent intitulé *l'historien programmeur*, a posé le couple chercheur-programmeur. On se propose d'ajouter un troisième personnage troisième, le designer.
- 2 Si l'on veut définir le design, on peut dire que c'est de la conception d'interfaces. C'est d'ailleurs la définition donnée ici depuis deux jours. Or, le design n'est-il pas plutôt la conception d'une structure et la définition d'une fonction ? La définition admise est la conception d'outils, de corpus, et la restitution des résultats, (design par wikipedia).
- 3 Nous devons réfléchir à de nouveaux outils et méthodes de travail, être inventifs et innovants. Nous utilisons des programmes conçus ou des petites machines de calculs dont nous devons discuter le statut. Ces programmes sont standards ou non, et d'ailleurs on peut se poser la question de tous les standardiser. Quelles visualisations données aux nouveaux objets produits par ces programmes ?
- 4 Quelques questions :
- 5 Les artefacts qui sont produits : quelle visualisation, quel schéma ?
- 6 Les nouvelles « formes de sensibilité » : identifier, authentifier, commenter ?
- 7 Quels médiums de véracité ?

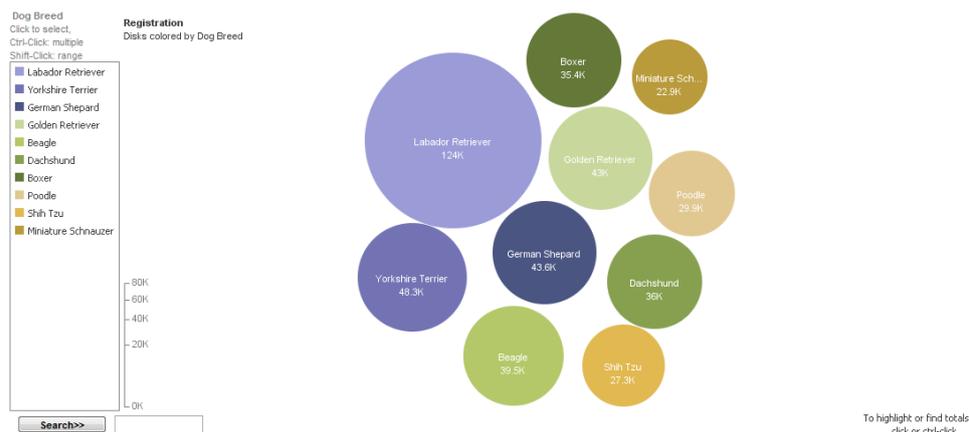
Quelques références et éléments de réflexion

- 8 What is visualization de Lev Manovich (2009) : Lev Manovich distingue le « design » d'information c'est-à-dire le questionnement sur la représentation de la donnée et la visualisation de l'information (extraction des données et *processing*) comme deux éléments différents.

- 9 DIRT (Digital Research Tool) est une boîte à outils très riche inventariant des outils des humanités numériques pour des utilisations universitaires.

- 10 DIRT propose ainsi des programmes libres ou payants. Parmi eux, on peut trouver le logiciel GEPHI (libre) qui est une plate-forme interactive de visualisation et d'exploration de toutes sortes de réseaux et de systèmes complexes et/ou dynamiques. Un autre exemple d'outil proposé par la plateforme DIRT mais développé cette fois par IBM est ManyEyes. Manyeyes est un outil de visualisation de données, sans caractère privatif, dont l'accès est public.

- 11 Manyeyes permet notamment d'expérimenter les représentations graphiques de données (voir la copie d'écran ci-dessous).



Autres références intéressantes :

- 12 John Maeda au MIT : Design By Numbers (1999)
- 13 processing.org et David-Olivier Lartigaud
- 14 Aaron Koblin qui a travaillé au CreativeLab Google
- 15 visualcomplexity.com et Manuel Lima
- 16 Fernanda B. Viégas et Martin Wattenberg, qui depuis la fin des années 1990, ont essaimé des exemples de visualisation : cas de 2003 et Historyflow
- 17 Trois livres circulent dans la salle :
- 18 Form+Codes édité par Casey Reas et Chandler McWilliams
- 19 Software Studies. A Lexicon édité par Matthew Fuller
- 20 Data Flow 2. Visualizing Information in Graphic Design édité par R. Klanten, S. Ehmann, T. Tissot, N. Bourquin

Démonstration à partir d'un outil de visualisation

- 21 Des données vont être visualisées de différentes façons sur un logiciel plutôt orienté graphistes 2D/3D et lié au prototypage d'interface (Arduino).
- 22 Cette démonstration est réalisée à partir d'une masse de données issues de la captation du déplacement du doigt de l'utilisateur sur l'écran d'une tablette. Il est possible de visualiser ce déplacement de plusieurs manières différentes : avec des points, des lignes, des couleurs différentes, des nuages de zone de contact qui s'amplifient en fonction du nombre de fois où le doigt passe sur une partie de la tablette. La représentation est aussi possible par points en fonction de la chronologie des événements. Une information temporelle peut être transmise en plus d'une information spatiale. En comparant avec l'action d'un autre utilisateur de cette même tablette, on voit que le toucher n'est pas le même grâce à une visualisation appropriée. Cette démonstration pose déjà une question de design : quelle est l'option de visualisation la plus appropriée pour ce que l'on souhaite montrer ?

- 23 Il existe donc différents types de visualisation pour les mêmes données. Ce n'est pas seulement un effet décoratif car ce que dit le graphisme est signifiant pour celui qui reçoit l'information.
- 24 La question du design peut se faire aussi en amont dans la conception de la base de donnée, de l'interface des logiciels utilisés, etc.
- 25 Exemple : Carte de métro du trafic internet ou des éléments de la vie quotidienne
- 26 <http://store.informationarchitects.net/product/web-trend-map-4>



- 27 Icepick.com : visualisation liées à la vie quotidienne, fréquence d'ouverture de la porte du frigo, chasse d'eau, éléments culinaires, etc.



- 28 La mise en forme des données pose la question des outils à utiliser (Python, Processing ?). Afin de choisir l'outil adéquat, une étude sur les logiciels « Software Studies » rédigé par M. Fuller est disponible à l'url <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11476>. Ce *software studies* permet la critique de logiciels en 40 courtes critiques.
- 29 Un autre exemple intéressant de mise en forme de données est l'application Pirate Pad, utilisé pour prendre des notes collaboratives lors de ce THATCamp Paris 2012. Pirate Pad modifie la pratique d'écriture de par les fonctionnalités qu'il propose et son ergonomie.
- 30 La question de la critique logicielle, au sens de M. Fuller, est donc un domaine qui doit entrer dans le cadre des « bonnes pratiques » des humanités numériques.
- 31 Pour Norman Potter (*What is a designer*), « Chaque être humain est un designer » (p. 1). Les trois notions clefs du design sont la spéculation, le rapport à la technique et la responsabilité.
- 32 Le bon design oppose une solution élégante à une solution productive (p. 50) c'est à dire une économie de moyen (l'élégance) contre une ambition plus grande (lié à la production) où l'on propose une nouvelle intégration du sens (par déplacement).
- 33 Un autre exemple d'interface est le site UbuWeb. UbuWeb est une ressource indépendante dédiée à tous les courants avant-gardistes d'art... La composition du site est très structurée et laisse à penser qu'on est dans une base de données. L'approche graphique ressemble à un programme, les pages sont combinées comme un tableau à double entrée.
- 34 Au-delà de la séduction qu'offre le design, chaque position des auteurs et titres-thèmes dans les pages respecte une grille-code graphique homogène. Il s'agit d'une rigueur programmatique, une pensée visuelle et plastique.

Le public visé par le design : le cas des chercheurs

- 35 À la suite de la présentation des différentes interfaces, la question que l'on pose est celle de la réception des informations proposées, c'est-à-dire le public visé. Quel est le meilleur design et pour quel public et quels objectifs ? Faut-il former ce public à lire des designs particuliers ?
- 36 En effet, l'efficacité diffère d'un mode de visualisation à un autre suivant l'objectif visé. La frontière entre l'art et le design, entre la visée expressive et un moyen plus « pragmatique » de transmettre, de communiquer un certain sens précis, est ténue. Il est, de ce fait, important de faire une distinction entre l'artiste et le designer : le designer n'est pas l'artiste.
- 37 La question de la réception se pose en plusieurs temps. Elle se pose d'abord au moment de la production, en fonction des potentiels destinataires et des formes produites. Dans le cadre de cet atelier, la cible c'est d'abord le chercheur. L'habitus du chercheur peut jouer contre une nouvelle ergonomie proposée. La transmission d'information implicite se réalise également dans un design achevé. Une des limites cependant est le design qui va très loin et qui empêche parfois l'interprétation correcte des données.
- 38 Dans la question du design, la conception même de l'outil est importante. Quelles sont les frontières réelles entre les compétences du programmeur, du chercheur et du

designer ? C'est une triade de savoir faire, triade mouvante où les compétences circulent. Certaines personnes sont les trois à la fois mais ce sont des personnes rares ! La conception débute dans la base de donnée, on est déjà dans le design, c'est-à-dire une pensée structurante.

- 39 La question de la formation en design se pose. Les chercheurs ont une appréhension du rôle du graphiste. Le travail sur la visualisation de la donnée est souvent sous-évalué par les chercheurs et les doctorants. On peut se demander ce qui fait obstacle à cette prise de conscience de l'importance du design. Les temps d'apprentissage du design sont compliqués. La difficulté réside également dans l'accoutumance à certains codes préexistants. Cependant, au moment de la production des données, une des dimensions importantes pour le chercheur est de prendre en compte la visualisation des données et d'anticiper la conception.
- 40 Existe-t-il des outils préférables à d'autres pour les SHS ? Et pourquoi ?
- 41 Il est difficile de généraliser, de dire quel outil est préférable pour quoi. Il est tout à fait possible de construire son propre outil mais on ne devient pas programmeur « comme ça ». L'étape de programmation d'un outil ne peut pas se cumuler avec la fonction de chercheur. Un des risques est la perte des données : il y a un enjeu de qualité et d'exigence quand on programme. Le programmeur du dimanche le fait à ses risques et périls. Faire appel à un développeur pour programmer un logiciel est une garantie.
- 42 Cependant, un des avantages à ce que le chercheur se frotte à la programmation est que cela lui permet de prendre un certain recul et d'avoir un esprit critique sur « la boîte noire de l'algorithme qu'il utilise pour ses données ». S'interroger systématiquement sur les outils utilisés, c'est aussi une partie du travail du chercheur. Il faut un dialogue chercheur-designer mais il est important également qu'un contact se crée entre le chercheur, la plasticité de l'information et l'algorithme. Un exemple concret du lien design/chercheur est le carnet de recherche d'Hypothèses : comment cet outil détermine-t-il un type d'expression lié à son format ?
- 43 Est-ce que c'est le déficit de formation qui fait que l'historien doit faire appel au développeur ?
- 44 Il est important d'organiser du dialogue entre les compétences. Dans le cas du design, cela se passe en dehors du temps identifié de la recherche. Concernant la question de la pensée de l'objet final tel qu'il se présente dans sa forme visuelle, il faut que dans les formations, les chercheurs soient sensibilisés à ces traces de conception aussi.
- 45 Aux Éditions de la Maison des sciences de l'homme, l'équipe a travaillé avec des graphistes qui sont aussi développeurs et à qui il a été imposé d'entendre la parole de l'auteur-chercheur. L'objectif était de pouvoir fabriquer quelque chose qui rendait compte (permettait de visualiser sa démonstration argumentée et longue) mais qui faisait un peu perdre le fil aux autres. Cette initiative a été bien reçue par tous et le chercheur a en plus aperçu (à la visualisation) de nouveaux éléments en regardant le schéma travaillé (animé) par le graphiste. Le médium : l'éditeur qui veut un propos, le graphiste qui rend visible et le chercheur qui rajoute finalement un chapitre car il a « vu » de nouvelles choses ! On peut ainsi imaginer ouvrir le triangle chercheur/designer/programmeur avec l'éditeur. D'autres acteurs sont, bien entendu, à ajouter.
- 46 Dans les *software studies*, on est obligé d'avoir un rapport critique au logiciel utilisé.
- 47 Marie-Haude Caraës et Nicole Marchand-Zanartu, *Images de pensée*, RMN, 2011.

48 <http://strabic.fr/La-pensee-a-l-oeuvre.html>

RÉSUMÉS

À partir de cas de figures et de situations concrètes, cet atelier propose de discuter la manière dont le design graphique est convoqué dans l'agencement et la production de savoir.

INDEX

Mots-clés : graphisme, programmation, design