A l'épreuve de l'hétérogénéité : données de recherche et interdisciplinarité

Marie Puren

A l'épreuve de l'hétérogénéité : données de recherche et interdisciplinarité. L'exemple du projet européen IPERION-CH

Comme vous pouvez le voir avec le titre de mon intervention, je vais parler ici des données de la recherche dont nous sommes tous producteurs, utilisateurs ou ré-utilisateurs. Et je pense qu'un certain nombre d'entre nous pourraient, au sujet de ces données, reprendre à leur compte le cri que Sir Tim Berners-Lee a lancé lors d'une conférence TED en 2009 : « We want raw data, now ![1] », « Nous voulons des données brutes, maintenant ! ». Nous sommes en effet de plus en plus nombreux à demander aux propriétaires de données, qu'ils soient publics ou privés, d'ouvrir ces dernières. En ce qui concerne les données de la recherche, cela fait plus de dix ans que chercheurs et institutions de recherche réclament leur ouverture. Dès 2003 en effet, la Déclaration de Berlin sur le libre accès à la connaissance demande la mise à disposition en libre accès de la littérature scientifique, et des données brutes à l'origine de celle-ci. Depuis cette date, la volonté d'ouvrir les données de la recherche n'a cessé de grandir. En 2006, l'OCDE insiste sur la nécessité d'élargir l'accès aux données de la recherche financés sur fonds publics. Fin 2007, le Conseil scientifique du Conseil européen de la recherche (ou CER) publie des recommandations appelant à mettre en accès libre les résultats de recherches financés par le CER. En 2012, dans la lignée du rapport *Riding the wave. How Europe can gain from the rising tide of scientific data*[2] publié deux ans plus tôt, la Commission Européenne affirme l'importance d'améliorer l'accès aux données de la recherche, et demande aux Etats européens de définir des politiques de libre accès aux données scientifiques. Ces mesures en faveur du libre accès seront finalement mises en œuvre

[1] Tim Berners-Lee, « Le Web à venir », *TED2009*, février 2009, https://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=fr.

[2] High-Level Group on Scientific Data, *Riding the wave How Europe can gain from the rising tide of scientific data*, Final report of the High level Expert Group on Scientific Data, October 2010, url : http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204

la même année, avec l'*Open Research Data Pilot*[3] qui promeut l'ouverture des données de la recherche générées par les projets financés par le programme européen Horizon 2020.

Afin d'aider à l'ouverture des données de la recherche, l'usage des plans de gestion de données se répand. Ils sont d'ailleurs de plus en plus souvent exigés par les financeurs de la recherche. Ces derniers voient en eux la garantie que les données produites par les projets qu'ils financent seront accessibles par tous et pour tous. Les attentes exprimées sont donc fortes : tous les projets de recherche doivent s'accommoder d'un outil unique pour gérer les données qu'ils génèrent, et s'assurer qu'une véritable politique commune de gestion des données sera ainsi impulsée. On peut se demander si une telle ambition est réalisable. La gestion des données de la recherche au sein de grands projets européens, rassemblant de multiples partenaires, et produisant un nombre croissant de données, peut-elle se satisfaire d'un plan de gestion ? L'exemple du projet interdisciplinaire IPERION-CH, consacré à la conservation et à la restauration du patrimoine, illustre les écueils que peut rencontrer la gestion de données de la recherche, telle qu'elle est pensée aujourd'hui. L'étude du cas IPERION-CH semble en effet souligner l'impossible gestion commune des données hétérogènes que le projet génère, et l'éclatement de la communauté de chercheurs qu'il est censé fédérer.

Afin de bien comprendre les enjeux autour de la gestion des données de la recherche au sein d'IPERION-CH, nous nous intéresserons d'abord aux moyens mis en œuvre pour gérer aujourd'hui les données de la recherche – et plus particulièrement, les plans de gestion de données. Il s'agit en effet de s'assurer de l'ouverture de ces données, afin de développer un nouveau modèle de partage. Toutefois, il semble difficile, pour certains projets, de se contenter d'un plan de gestion de données. C'est pourquoi nous verrons ensuite, en nous appuyant sur l'exemple d'IPERION-CH, que les plans de gestion des données de la recherche

[3] https://www.openaire.eu/opendatapilot

ont leurs limites, et qu'il n'est pas possible de s'en contenter pour tous les projets de recherche. Dans le cas d'IPERION-CH, la création d'un plan de gestion de données est malaisé, et ne fait que souligner l'hétérogénéité intrinsèque des données produites, et la pluralité des chercheurs et institutions que le projet tente de fédérer. Nous verrons donc quelle solution a été adoptée pour gérer les données produites par ce projet.

I. La gestion des données de la recherche au sein de grands projets européens : objectifs, méthodes et limites

a. *Le partage et la diffusion des données de recherche : un nouveau modèle*

Actuellement, on assiste à un mouvement croissant en faveur de la Science ouverte, qui prône plus de partage et plus de transparence dans la recherche[4], et cherche ainsi à imposer un nouveau modèle d'ouverture. Généralement, on considère que la Science ouverte repose sur six piliers principaux : les données ouvertes ou « Open data » ; l'Open source ; la méthodologie ouverte ou « Open methodology » ; l'évaluation transparente par les pairs ou « Open Peer Review » ; l'accès ouvert ou « Open access » ; et les ressources ouvertes pour l'enseignement ou « Open Educational Resources ». En ce qui nous concerne, nous allons nous concentrer sur l'ouverture des données. Pour reprendre la définition d'Open Knowledge International, « Open means anyone can freely access, use, modify, and share for any purpose […][5] ». Ce modèle d'ouverture a donc créé un nouveau cadre, dans lequel les données de la

[4] "A movement to promote greater sharing and transparency", University of Oregon Libraries, "Best practices", *Research Data Management*, url: https://library.uoregon.edu/datamanagement/sharingdata.html#one
[5] *Open Knowledge International*, "The Open definition", url: http://opendefinition.org/

recherche sont considérées comme des produits de la recherche, au même titre que des articles scientifiques ou des actes de conférence[6].

Avant d'aller plus loin, il paraît nécessaire de nous arrêter sur ce que recouvre le terme « données de la recherche ». Car la définition même de ces données spécifiques pose problèmes, tant elles prennent des formes variées. Il n'y a en effet pas de véritable consensus autour de la définition : cette dernière varie en fonction de la discipline et en fonction de ceux qui les manipulent. C'est ce que souligne Agnès Magron sur le site du Centre pour la Communication Scientifique directe (CCSD) lorsqu'elle explique que « Chaque discipline, chaque communauté scientifique et peut-être même chaque chercheur établira une définition différente des données de la recherche, en fonction de sa culture disciplinaire ou de ses objets de recherche[7] ». Dans son *Livre blanc – une science ouverte dans une république numérique*, le CNRS adopte une perspective large, et propose de définir ces données de la manière suivante : « Ensemble des données ayant servi à l'établissement d'un résultat scientifique[8] ». Tout en rejoignant cette perspective, les bibliothèques de l'Université de Boston précisent également que les données de recherche peuvent être recueillies, observées ou créées à des fins d'analyse, pour produire des résultats de recherche originaux[9].

On peut également catégoriser ces données en fonction de la méthode qui a permis de les obtenir. Celles-ci peuvent en effet être des données :

- d'observation,
- expérimentales,
- générées à partir de modèles (comme la simulation),

---

[6] Felicia LeClere, "Too Many Researchers Are Reluctant to Share Their Data", *The Chronicle of Higher Education*, August 03 2010, url: http://www.chronicle.com/article/Too-Many-Researchers-Are/123749/

[7] Agnès Magron, « Comment diffuser mes données de recherche ? (1/2) », https://www.ccsd.cnrs.fr/fr/2016/11/comment-diffuser-mes-donnees-de-recherche-1/

[8] DIRECTION DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE – CNRS, « Glossaire » in *Livre blanc — Une Science ouverte dans une République numérique* [en ligne]. Marseille : OpenEdition Press, 2016. Disponible sur Internet : http://books.openedition.org/oep/1554. DOI : 10.4000/books.oep.1554.

[9] "Research data is data that is collected, observed, or created, for purposes of analysis to produce original research results". Boston University Libraries, "What is 'Research Data'?", *Research Data Management*, url: https://www.bu.edu/datamanagement/background/whatisdata/

- dérivées ou compilées (par exemple avec la fouille de textes et de données),
- de référence (par exemple, les banques de données de séquençage génétiques).

Les données de recherche peuvent donc prendre des formes multiples comme des documents en .txt ou en .doc, des feuilles de calcul, des cahiers de manipulation, des questionnaires, des vidéos, des présentations, des échantillons, des bases de données, ou encore des sorties de logiciels d'analyse… Il faut également noter que les données de recherche peuvent être diffusées sous la forme d'un fichier unique ou bien sous celui d'un jeu de données (traduction du « dataset » anglais). Un jeu de données rassemble des données (brutes ou dérivées) qui présentent une certaine unité et qui sont réunis sous la forme d'un ou plusieurs fichiers électroniques de manière à former un ensemble cohérent[10].

Du côté de l'ouverture des données de la recherche, on peut souligner le soutien récent offert par l'Union Européenne. Avec la publication en 2013 des *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, le pilote de libre accès aux données de la recherche a annoncé l'engagement de l'Union Européenne à ouvrir les données produites par des projets Horizon 2020 préalablement sélectionnés. Le pilote a pour but d '« élargir et […] améliorer l'accès aux données générées par les projets, et leur réutilisation, au profit de la société et de l'économie[11] ». Deux types de données sont concernés par le pilote :

- les données nécessaires pour valider des résultats dans des publications scientifiques ;
- et, plus largement, toutes les données considérées comme ayant de la valeur. La définition de ces données est par contre laissée à l'appréciation des projets de recherche.

[10] Gaillard Rémi, *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?*, mémoire d'études sous la direction de Monique Joly pour l'obtention du diplôme de conservateur de bibliothèque, Lyon, Enssib, janvier 2014, p.19. http://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf

[11] Commission Européenne, « La Commission lance un projet pilote de libre accès aux données issues de la recherche financée sur fonds publics », *Base de données des communiqués de presse*, Bruxelles, 16 décembre 2013, url : http://europa.eu/rapid/press-release_IP-13-1257_fr.htm

Depuis le 25 juillet 2016, ce pilote a été étendu : il impose à tous les bénéficiaires du programme européen Horizon 2020 de rendre leurs données de recherche « repérables, accessibles, interopérables et réutilisables[12] ».

b. Des plans de gestion de données de plus en plus répandus…

Dans ce but, il leur est demandé de fournir un plan de gestion de données ou « Data Management Plan (DMP)[13] ». Ce document formalisé constitue une feuille de route régulièrement mise à jour, destinée à exposer le cycle de gestion des données générées, collectées, disséminées et utilisées au cours d'un projet de recherche, et une fois ce projet terminé. Les plans de gestion de données ont pour mission d'assurer que les données seront préservées, mais aussi utilisables à la fois dans le présent et dans le futur, par leurs créateurs et d'autres ré-utilisateurs. Par ailleurs, ils encouragent et aident à générer des données de recherche FAIR – c'est-à-dire « **F**indable », « **A**cessible », « **I**nteroperable » et « **R**e-usable »[14], c'est-à-dire « repérables, accessibles, interopérables et réutilisables ». L'utilisation d'un plan de gestion, dont l'objectif est d'ouvrir les données décrites, facilite en effet l'application des principes FAIR, définis pour la première fois en janvier 2014 par le comité FORCE 11[15]. En incitant les projets de recherche à réfléchir sur la manière dont ils vont rendre accessibles les données produites pendant le projet et à la fin de celui-ci, les plans de gestion de données invitent en effet à utiliser ces principes.

---

[12] "This document helps Horizon 2020 beneficiaries make their research data findable, accessible, interoperable and reusable (FAIR)". European Commission (Directorate-General for Research & Innovation), *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, 26 July 2016, p.3, url: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[13] European Commission, "Horizon 2020: Work Programme update supports competitiveness through open science", *Base de données des communiqués de presse*, Brussels, July 2016, url: http://europa.eu/rapid/press-release_MEMO-16-2604_en.htm.

[14] FORCE 11, "The FAIR data principles", *FORCE 11*, url : https://www.force11.org/group/fairgroup/fairprinciples

[15] Force 11, "GUIDING PRINCIPLES FOR FINDABLE, ACCESSIBLE, INTEROPERABLE AND RE-USABLE DATA PUBLISHING VERSION B1.0", *Force 11*, url: https://www.force11.org/node/6062/#Annex1.

Le programme Horizon 2020 met lui aussi l'accent sur ces principes, et sur le fait que les plans de gestions de données facilitent la production de données FAIR[16]. D'ailleurs, l'utilisation d'un DMP fait en général suite à une demande des agences de financement public : elles souhaitent en effet s'assurer de l'accès libre aux données produites par les projets de recherche qu'elles financent, espérant ainsi stimuler d'autres recherches sur ces mêmes données, et *in fine* augmenter l'impact économique des projets financés[17].

c. *...Mais qui ont du mal à faire face à des situations plus complexes*

Les plans de gestion de données sont des outils puissants, qui permettent aux chercheurs d'avoir une vision d'ensemble des données qu'ils vont générer tout au long d'un projet de recherche. Les DMP les aident également à mettre en place des stratégies de gestion efficaces de leurs données et à s'assurer de la préservation de ces dernières.

Moins le projet produit des jeux de données différents, plus il est facile de créer un DMP. Il est en effet bien plus aisé de réfléchir sur des jeux de données peu nombreux, issus d'un nombre restreint de sources. On peut en effet écrire un plan de gestion de données pour un seul jeu de données, produit lors d'une expérience unique par exemple[18]. Mais il est tout à fait possible de créer un DMP pour un projet qui produit plusieurs jeux de données différents – comme c'est le cas lorsque l'on passe à l'échelle européenne. C'est la situation à laquelle est confronté le projet Horizon 2020 *ESiWACE. Centre of Excellence in Simulation of Weather and Climate in Europe*. Dans son Data Management Plan publié en février 2016[19], *ESiWACE* a ainsi identifié quatre jeux de données expérimentales. Les projets Horizon 2020

[16] European Commission (Directorate-General for Research & Innovation), *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*…, p.3,

[17] OCDE (Organisation de Coopération et de Développement économiques), *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds public*, OCDE, 2007: http://www.oecd.org/fr/sti/sci-tech/38500823.pdf. Consulté le 6 septembre 2016.

[18] Un exemple ici : http://www.dcc.ac.uk/sites/default/files/documents/adocs/DMTpsych-example-DMP.pdf

[19] ESiWACE, *Data Management Plan Deliverable 5.4*, 28 février 2016, url : https://www.esiwace.eu/results/deliverables/Data_Management_Plan_Annex_final.pdf.

ne sont toutefois pas les seuls à produire plusieurs jeux de données : c'est en fait le cas de la plupart (voire la totalité) des projets de recherche. Mais à la lecture de ces plans de gestion de données, on remarque cependant qu'au sein d'un même projet, ces jeux de données, peu nombreux, partagent des points communs, en termes d'accès aux données et/ou de la nature des données collectées. Ces jeux de données peuvent également participer du même processus ou être issus de la même expérience. Par exemple, des données primaires (des données d'observation) deviennent ensuite des données secondaires (ou des données d'analyse)[20].

Mais il en va tout autrement lorsque l'on s'intéresse à des projets vastes, qui rassemblent de nombreux partenaires et qui produisent un grand nombre de jeux de données différents. C'est le cas du projet Horizon 2020 *FREME* : *Open Framework of E-Services for Multilingual and Semantic Enrichment of Digital Content*[21]. Le but de FREME consiste à créer un ensemble de services en ligne destinés à faciliter l'enrichissement multilingue et sémantique de contenus numériques. Le projet va utiliser des jeux de données existants pour conversion et enrichissement. C'est donc un projet ambitieux qui mobilise potentiellement de nombreux jeux de données différents. Ceux-ci ont d'ailleurs été listés dans un autre deliverable[22] et neuf sont répertoriés sur le site du projet[23]. Les dimensions du projet et la vaste de gamme de jeux de données disponibles rendent plus difficile la création d'un plan de gestion de données. En témoignent d'ailleurs les auteurs de la première version du DMP livré en juillet 2015, qui se sont contentés de décrire un seul jeu de données exemple[24].

La difficulté des DMP à s'adapter à des projets plus vastes, ayant à gérer de nombreux jeux de données, s'explique sans doute par le fait que la responsabilité de la gestion des données repose, dans la grande majorité des cas, sur les épaules des producteurs des données

[20] Par exemple : http://library.umassmed.edu/necdmc/dmp
[21] http://www.freme-project.eu/
[22] D1.1. « Initial requirement analysis and specification for the FREME framework and the e-Services » (cf. : http://www.freme-project.eu/resources/deliverables/)
[23] https://freme-project.github.io/api-doc/list-datasets.html
[24] FREME (Open Framework of E-Services Multilingual and Semantic Enrichment of Digital Content), *Data management plan*, version 1, 24 July 2015, url : http://www.freme-project.eu/resources/FREME_Deliverable_D7-4.pdf

eux-mêmes, c'est-à-dire les chercheurs. La création des plans de gestion de données est, la plupart du temps, le résultat d'une demande des agences de financement, qui « responsabilise » avant tout le chercheur [25]. Si ce genre d'approche « top-down » s'accommode facilement des projets à petite échelle, ce n'est pas le cas des grands projets européens qui rassemblent de multiples partenaires.

II. La gestion des données au sein d'IPERION-CH : créer une culture du partage

Peut-on alors parler d'un échec des plans de gestion de données lorsqu'il s'agit de gérer les données de recherche d'un projet à très grande échelle ? C'est en effet une question cruciale lorsque l'on s'intéresse à la gestion des données du projet Horizon 2020 IPERION-CH qui réunit 27 partenaires et doit s'assurer de la gestion de 56 jeux de données différents.

a. IPERION-CH : un projet européen et interdisciplinaire

IPERION-CH[26] (*Integrated Platform for the European Research Infrastructure ON Cultural Heritage*) est un projet Horizon 2020 qui préfigure la mise en place d'une infrastructure de recherche unique pour la restauration et la conservation du patrimoine. Il s'agit de l'infrastructure E-RIHS (*European Research Infrastructure for Heritage Science*), dédiée à l'étude des matériaux anciens du patrimoine culturel et naturel. Celle-ci a été présentée en mai 2016 au ministère de l'Education nationale, mais l'infrastructure sera créée en 2022.

[25] Gaillard Rémi, *De l'Open data à l'Open research data…*, p.32.
[26] http://www.iperionch.eu/

Par la nature des recherches menées en son sein, IPERION-CH embrasse un vaste ensemble de disciplines s'intéressant à la conservation, à l'interprétation et à la gestion du patrimoine. De ce fait, IPERION-CH rassemble des chercheurs issus à la fois des humanités (sciences humaines et sociales, arts, archéologie) et des sciences naturelles (physique, chimie, biologie). Le projet revendique la création d'une communauté de recherche interdisciplinaire, avec la mise en réseau de chercheurs travaillant sur des sujets variés et avec des méthodes très différentes. Il rassemble aussi des instituts de recherche, des centres de conservation, des laboratoires, des archives, des musées : en tout, IPERION-CH compte 28 partenaires européens.

L'interdisciplinarité est donc au centre de ce projet, qui se retrouve à devoir gérer des données de recherche générées par 19 institutions dans 11 pays européens. On a d'un côté de la documentation : archives et catalogues de musée, collections de photographies, rapports de restauration, articles scientifiques. De l'autre, des résultats d'expérience : des « raw data » issus du bombardement d'un artefact ancien par un faisceau d'électrons, des analyses de pigments, des radiographies d'œuvres d'art, des chromatographies en phase gazeuse. La nature interdisciplinaire du projet va donc de pair avec des données de recherche très hétérogènes. Il est clair en effet que l'hétérogénéité ne cesse d'être à l'œuvre au sein d'IPERION-CH :

- Hétérogénéité des objets étudiés. Les chercheurs peuvent s'intéresser à tous types de matériaux anciens : tissus, fossiles, pigments, poteries, tableaux, sculptures… de la préhistoire à nos jours.
- Hétérogénéité des disciplines. Un tesson de poterie peut passer entre les mains d'un chimiste qui s'intéressera à la composition de la peinture qui a été appliquée à sa surface ; un historien de l'art cherchera à déchiffrer le sens des motifs qui sont

représentés sur ce fragment ; un archéologue cherchera à déterminer quel peuple l'a façonné.

- Hétérogénéité des données produites. Chaque discipline possède ses propres méthodes de recherche, et produit donc des données très variées, en termes de contenus, de formats, de conservation et de gestion.

b. Une impossible gestion commune des données ?

En ce qui concerne la gestion de ces données, les pratiques sont très hétérogènes. Les formats utilisés varient en fonction de la discipline, de l'outil qui a permis de produire les données, et de l'institution concernée. La plupart du temps, il s'agit de formats fermés, générés par des instruments spécialisés et gérés par des logiciels propriétaires. Et souvent, il y a peu de documentation disponible sur ces formats très spécifiques. Les données sont conservées sur des supports très différents : serveurs, disques durs externes, et même ordinateurs personnels. Les durées de conservation ne sont pas indiquées et laissées à l'appréciation du producteur des données, ou bien elles ne dépassent pas quelques mois. De manière générale, les données ne sont donc pas conservées de manière pérenne. Par ailleurs, les conditions d'accès restent très floues, et sont soumises au bon vouloir du producteur des données : l'accès est autorisé au cas par cas, sur demande expresse d'un utilisateur extérieur.

Ces 19 institutions, productrices d'un large éventail de données de recherche, utilisent donc des outils scientifiques très variés pour les générer, et pour la plupart d'entre elles, gèrent ces données avec des solutions « maison ». Nous faisons donc face à une situation très complexe en termes de gestion de données, due à l'hétérogénéité intrinsèque des jeux de données produits. Ces obstacles scientifiques et humains ne facilitent donc pas la création d'un plan de gestion de données pour IPERION-CH. C'est d'abord l'aspect interdisciplinaire du projet qui remet en cause l'usage d'un DMP. Dès 2006, l'OCDE avait ainsi insisté sur « la

nécessité de privilégier une approche disciplinaire[27] » dans la gestion des données de la recherche, de façon à respecter les pratiques observées par chaque communauté. Dans notre cas, la diversité des formats et des outils utilisés reflète assez bien la pluralité des habitudes disciplinaires. S'y ajoute un certain goût pour le secret, ou du moins une absence de transparence en ce qui concerne les règles d'accès et de réutilisation. La responsabilité des producteurs des données est en effet primordiale dans la diffusion, l'archivage et la durée de conservation des données de recherche, comme le souligne dès 2005 un rapport de la *National Science Foundation*[28].

Dans ses conditions, il est très difficile de mettre en place un plan de gestion de données. Plutôt que de contribuer à l'unité du projet et de rassembler autour d'elles, les données de recherche générées par le projet IPERION-CH semblent avant tout souligner l'éclatement de la communauté de recherche qui les produit. Serait-ce alors le signe d'une impossible collaboration entre les différents acteurs impliqués ?

c. *Créer une culture du partage des données avec la « Cultural Heritage Data Reuse Charter »*

Comment faire face à cette situation ? Est-ce finalement la marque d'un échec du projet, qui serait incapable d'assurer la cohésion des différents acteurs qui en font partie ? Plutôt que de parler d'échec, parlons de défi. Il s'agit ici de prendre conscience de la nécessité de développer une nouvelle politique en matière de données de recherche, afin de gérer ces dernières au sein de grands projets de ce type. Dans ce contexte, le but n'a pas été de produire un plan de gestion standard; on a plutôt choisi de rassembler les elements-clés qui pourraient aider à définir des recommandations en termes de politique de données. La nécessité de

[27] Gaillard Rémi, *De l'Open data à l'Open research data…*, p.20.
[28] National Science Foudation, *Long-Lived Digital Data Collections : Enabling Research and Education in the 21st Century*, septembre 2005, 87 p. https://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf

développer une culture de partage des données entre les différents acteurs d'IPERION-CH - les institutions patrimoniales, les laboratoires et les chercheurs – est clairement apparue. L'une des difficultés, pour chaque acteur impliqué, consiste manifestement à fournir les conditions d'utilisation et de réutilisation des données produites, et d'indiquer les éventuelles contraintes liées à leur reutilisation. Pour dépasser ces difficultés, IPERION-CH souhaite travailler main dans la main avec d'autres infrastructures européennes s'intéressant à la recherche sur les humanités et le patrimoine. Il s'agit de définir ainsi les contours d'une future plateforme numérique, destinée à rassembler et à diffuser des informations sur les jeux de données générés par IPERION-CH, mais aussi par ses partenaires. Actuellement, cet environnement, ou « Cultural Heritage Data Reuse Charter », est développé par DARIAH-EU[29], avec les projets européens HaS[30] et IPERION-CH. On passe ici à un autre niveau : du niveau technique du plan de gestion de données, à un niveau politique avec la Charte de réutilisation des données patrimoniales.

Après un développement s'étalant sur plusieurs mois entre 2016 et 2017, la Charte verra le jour à l'été 2017. Elle prendra la forme d'un environnement en ligne, destiné à soutenir tous les acteurs intéressés par la réutilisation des données, qui ont été générées par les institutions patrimoniales, et par les laboratoires et chercheurs en sciences du patrimoine. En pratique, les utilisateurs de la Charte s'enregistreront au sein de cet environnement, avant de déclarer toutes les informations nécessaires à la réutilisation des jeux de données qu'ils ont produits. En partageant ces informations, la Charte aidera également à documenter le processus de génération, de diffusion et de réutilisation des données. Par conséquent elle aidera aussi à améliorer la qualité des données et des métadonnées qui y sont liées. Plus largement, la Charte offre un cadre concret d'implémentation pour les principes FAIR. Celle-

[29] http://dariah.eu/
[30] http://has.dariah.eu/

ci invite en effet les utilisateurs à rendre leurs données FAIR en mettant l'accent sur les points suivants :

- Pour trouver ces données, celles-ci doivent avoir été décrites avec des métadonnées et identifiées grâce à un identifiant pérenne.
- Pour que ces données soient interopérables, il est nécessaire d'utiliser des formats ouverts, des standards communs, des spécifications de données documentées, et des vocabulaires ou des ontologies cohérents.
- Pour faciliter la réutilisation de leurs données, les droits d'accès doivent être clairement définis et des licences appropriées doivent avoir été choisies.

La Charte n'offre pas une alternative aux plans de gestion de données, mais une autre perspective concernant la gestion des données de la recherche. Il est clair que les plans de gestion de données sont extrêmement utiles, et qu'ils suffisent pour gérer les données de la majorité des projets de recherche. Mais pour certains, ce document formalisé, qui met l'accent sur l'aspect technique de la gestion des données, ne permet pas de développer une véritable politique commune en matière de gestion de données de la recherche.

Pour IPERION-CH, il n'était pas possible de réunir, ou d'implémenter, des habitudes de bonne gestion qui n'existaient pas encore. Il était donc nécessaire d'inciter plutôt que d'imposer, en offrant un environnement qui permettra à chaque producteur de données de réfléchir sur les conditions de partage et de réutilisation de ses « datasets ». Là encore, il s'agit de responsabiliser le producteur des données, qui a un rôle primordial à jouer dans l'ouverture des données de la recherche. La Charte ouvre ainsi une réflexion sur la gestion des données de la recherche ; et sans imposer en amont des règles de gestion – ce qui serait difficile voire impossible au vue du nombre de jeux de données produits et des partenaires impliqués -, elle invite *in fine* à améliorer la qualité des données produites, facilite la diffusion

et l'ouverture des données de la recherche, tout en essayant de stimuler l'émergence d'une culture du partage. Une telle initiative rejoint donc pleinement l'espoir exprimé par Aurélien Berra au début de cette conférence : celui que les humanités numériques arrivent à développer des espaces de dialogue et d'échange, un « intérêt au désintéressement »[31].

[31] Aurélien Berra, Pierre Mounier, « Humanités numériques : quelle(s) critique(s) ? », *DHNord 2016*, MESHS de Lille, 21 novembre 206, url : http://publi.meshs.fr/ressources/humanites_numeriques_quelles_critiques/@@video