

Penser local

Développer une politique de données sur un campus SHS¹

Thinking local

The implementation of an open science policy in a social sciences and humanities university.

Joachim Schöpfel¹

¹ GERiCO, Université de Lille, Domaine Pont de Bois, BP 60149, F-59563 Villeneuve d'Ascq, France, joachim.schopfel@univ-lille.fr

RÉSUMÉ. Dans le cadre du Plan national pour la science ouverte, la structuration et le partage des données de recherche font désormais partie des priorités de la politique scientifique de la France. Chaque établissement et chaque organisme scientifique doit se doter d'une politique de la science ouverte et mettre en place un ensemble de services et dispositifs pour la gestion des données de la recherche.

A partir d'enquêtes sur le terrain, l'article propose une feuille de route pour la mise en œuvre d'une telle politique sur un campus universitaire en sciences humaines et sociales. Dix principes indiquent des pistes pour la gouvernance et le pilotage de cette politique, pour déterminer les priorités de développement et d'investissements, et pour faire le lien avec les infrastructures de recherche, dont notamment Huma-Num.

Il s'agit d'une démarche bottom-up, qui met l'accent sur les pratiques et besoins des chercheurs et qui place les chercheurs au cœur d'une politique institutionnelle dans le domaine des données de recherche.

ABSTRACT. As part of the National Plan for Open Science, structuring and sharing research data is now one of the priorities of French science policy. Each institution and each scientific organization must have an open science policy and set up a set of services and devices for the management of research data.

Based on field surveys, the article proposes a roadmap for the implementation of such a policy on a university campus in the social sciences and humanities. Ten principles indicate perspectives for the governance and steering of this policy, to determine development and investment priorities, and to make the link with research infrastructures, including Huma-Num.

It is a bottom-up approach, which focuses on the practices and needs of researchers and puts researchers at the heart of an institutional policy in the field of research data.

MOTS-CLÉS. données de la recherche, sciences humaines et sociales, science ouverte.

KEYWORDS. research data, social sciences and humanities, open science.

1. Les enjeux des données de la recherche

Avec le développement des données massives, la gestion des données de la recherche est devenue en quelques années l'un des enjeux majeurs des organismes et établissements de l'Enseignement Supérieur et de la Recherche (ESR). Depuis plusieurs années, l'Union européenne a inclus l'ouverture des données de la recherche dans sa politique scientifique, par le biais de ses programmes cadres (H2020), et par la définition d'une feuille de route pour la science ouverte. Ainsi, le *Amsterdam Call for Action on Open Science* de 2016 a adopté le principe selon lequel le

¹ Il s'agit d'une version abrégée et modifiée d'un rapport publié sur HAL (<https://hal.archives-ouvertes.fr/hal-01846849>). L'étude a été réalisée dans le cadre du projet *D4Humanities*, avec un financement de la Maison européenne des sciences humaines et sociales à Lille (MESHS) et du Conseil Régional Hauts-de-France.

partage et la gestion des données constituent l'approche par défaut pour toutes les recherches financées par des fonds publics², un principe confirmé par le Conseil européen en juin 2016.

La politique européenne poursuit plusieurs objectifs, dont l'efficacité et l'efficacités de la recherche publique, le transfert des résultats vers la société civile et l'économie, la transparence de l'action publique et la science citoyenne (cf. Chartron 2018). Le développement des infrastructures de recherche, le changement des modalités et critères d'évaluation et les critères d'attribution des subventions figurent parmi les leviers de cette politique.

En France, la Loi pour une République numérique promulguée en octobre 2016 avait fixé le cadre pour une politique d'ouverture des données et des connaissances. Deux ans plus tard, deux documents stratégiques ont concrétisé cette politique. D'une part, le deuxième Plan d'action national 2018-2020 *Pour une action publique transparente et collaborative*³ précise que la France « soutient la mise en œuvre des principes du gouvernement ouvert pour renforcer (...) l'accès aux matériaux et résultats de la recherche », dans la continuité des chantiers engagés par le gouvernement français sur la transformation numérique de l'Etat et sa modernisation.

D'autre part, le *Plan National pour la Science Ouverte*⁴, présenté en juillet 2018 à Lille, confirme l'engagement de l'Etat français pour une transformation durable de sa recherche publique vers un écosystème de la science ouverte, « dans lequel la science est plus cumulative, plus fortement étayée par des données, plus transparente, plus rapide et d'accès plus universel ». L'objectif est triple :

- faire une science meilleure, plus intègre, plus performante, plus rapide ;
- renforcer la confiance des citoyens dans la science, en particulier dans les domaines sensibles comme le changement climatique, les nouvelles énergies ou les sciences médicales ;
- reprendre le contrôle sur le processus scientifique et sur la diffusion et l'exploitation des résultats de la recherche.

Le Plan national du Ministère propose trois volets pour réaliser cette transformation : une série de mesures et d'actions pour généraliser l'accès libre aux publications issues de la recherche publique ; d'autres mesures pour structurer et ouvrir les données de la recherche ; et plusieurs pistes pour mobiliser l'ensemble des acteurs afin d'intégrer la démarche nationale dans un paysage européen et international. L'approche est double, *top-down* pour déterminer la feuille de route, fixer les objectifs, créer des incitations et coordonner les actions notamment par la mise en place d'un *Comité pour la Science Ouverte*⁵ (CoSO), et *bottom-up* dans la mesure où la politique nationale s'appuie pour l'essentiel sur les initiatives et projets sur le terrain de l'ESR français.

Lors des Premières Journées Nationales de la Science Ouverte à Paris, Borgman (2018) a insisté sur l'importance de placer la gestion des données de la recherche dans le contexte d'une infrastructure de la connaissance (*knowledge infrastructure*), autour d'un écosystème de personnes, pratiques, technologies, institutions etc. Les répertoires des services et entrepôts de données témoignent de la diversité des réalisations⁶. De nombreux dispositifs sont opérationnels depuis de longues années, comme le système des *World Data Centers* ; d'autres en revanche sont en cours de déploiement (comme le réseau des plateformes *DataVerse*) ou seulement à l'étude. La

² <https://www.government.nl/documents/reports/2016/04/04/amsterdam-call-for-action-on-open-science>

³ <https://www.etalab.gouv.fr/plan-daction-national>

⁴ <http://www.bibliothequescientifiquenumerique.fr/plan-national-pour-la-science-ouverte/>

⁵ <https://forum.ouvrirlascience.fr/>

⁶ Cf. re3data <https://www.re3data.org/> et Cat-OPIDoR <https://cat.opidor.fr>

compréhension de l'environnement concret est primordiale pour préparer et mettre en œuvre une politique de données. Il n'y a pas de solution unique.

Aussi, les mesures et actions du Plan national pour « structurer et ouvrir les données de la recherche » constituent les nouvelles conditions auxquelles les acteurs de l'ESR doivent trouver des réponses (obligation de la diffusion ouverte des données de la recherche, généralisation des plans de gestion, certification, création d'une fonction administrateur des données etc.) sans pour autant imposer un modèle de référence pour les initiatives sur le terrain. Le Collège Données de la recherche du nouveau CoSO⁷ contribuera au développement de ces initiatives par le biais d'un « kit d'accompagnement », mais n'a pas vocation à proposer une solution clé en main.

Dans cet état d'esprit, deux enquêtes ont été menées sur le campus Pont de Bois de l'Université de Lille, dans le domaine des Sciences Humaines et Sociales : la première enquête a été réalisée en 2015 auprès de 270 enseignants-chercheurs, chercheurs, doctorants pour évaluer les pratiques, besoins et attentes en matière de gestion des données de la recherche (Prost et Schöpfel 2015) ; la deuxième enquête a pris la forme d'un audit auprès de 51 chefs de projets, directeurs de laboratoires, chargés de missions afin de mieux comprendre les incitations et verrous sur le terrain et proposer quelques perspectives de développement (Schöpfel 2018).

L'objectif de cet article est de faire une synthèse des recommandations issues de ces deux enquêtes, sous forme d'une liste de dix principes pour la mise en place d'une politique de données sur un campus universitaire en sciences humaines et sociales. Il s'agit donc d'une approche *bottom-up* qui place les pratiques et attentes des acteurs du terrain – chercheurs, enseignants-chercheurs, ingénieurs, doctorants – au cœur de l'action, dans une démarche que Chartron (2018) a caractérisée comme un « processus de co-construction (par les communautés scientifiques) avec les politiques et les corps intermédiaires ».

2. Le terrain

L'étude empirique des pratiques individuelles et collectives de gestion des données de recherche est nécessaire au développement de nouveaux services et dispositifs, comme par exemple des entrepôts, plates-formes, formations etc., pour une meilleure gestion. Ce qui est plus important encore, pour une politique de données institutionnelle, c'est la compréhension des priorités propres des chercheurs concernant les données de la recherche, les incitations et les verrous. Quel est leur agenda ? Quelles sont les opportunités et les menaces du développement d'une culture de données ? Voici un aperçu des principaux résultats de nos enquêtes qui sont convergents avec d'autres analyses, dont notamment l'enquête sur le campus de l'Université de Rennes 2 (Serres *et al.* 2017) et une autre enquête à Bordeaux-Montaigne (Duprat 2019).

Tout d'abord, que veut dire « données de recherche » ? L'acceptation du terme varie selon les disciplines, les approches, les contextes, révélant aussi bien une complexité épistémologique et sémantique qu'une construction communautaire et politique (Malingré *et al.* 2018). Certes, il est possible de distinguer plusieurs catégories de données (typologie), de déterminer leurs principales fonctions et finalités (données en tant que « valeur »), et de faire le lien entre données et processus scientifique (données primaires/sources et données secondaires/résultats) ; il reste cependant le constat d'une relation forte entre la compréhension exacte et spécifique de ce terme et les thématiques, disciplines, méthodes et équipements (Schöpfel *et al.* 2017).

⁷ Un ensemble de groupes de travail pour accompagner la politique de données de recherche du Ministère <https://www.ouvrirlascience.fr/college-donnees-de-la-recherche/>

En ce qui concerne leurs données, la priorité de la plupart des chercheurs interrogés n'est pas la gestion en tant que telle, la préservation ou le partage, mais bien la sécurité des données et, au sens large, la sécurité des dispositifs utilisés pour le stockage et l'analyse des données. Cette observation rejoint la conclusion de l'enquête rennaise, selon laquelle le manque de sécurité des données est l'un des points les plus cruciaux mis en évidence. La deuxième préoccupation des chercheurs interrogés est la communication de « données chaudes » (en cours d'analyse) tout au long du projet, au sein de l'équipe scientifique. Il ne s'agit pas de partager des données dans le sens d'une ouverture ou d'une diffusion vers un public plus large, mais bien d'un échange ou d'un transfert de données dans le cadre du cahier des charges d'un projet. La question cruciale est l'opposition entre ce besoin de communication et les contraintes d'une protection imposée par la réglementation.

Quels facteurs favorisent les bonnes pratiques en matière de données ? Quelles sont les raisons pour lesquelles les chercheurs mettent en œuvre une gestion réfléchie de leurs données ? Selon nos résultats, six facteurs peuvent se chevaucher :

- les critères des appels à projets du programme cadre H2020,
- les appels à projets des programmes de l'Agence nationale de recherche (ANR),
- les exigences d'une bonne gestion de projet, en particulier dans les projets d'envergure,
- les contraintes juridiques liées à des données spécifiques (vie privée, santé, mineurs...),
- les protocoles et normes éthiques et
- la politique éditoriale d'un nombre croissant de revues académiques qui demandent la disponibilité des jeux de données associés aux articles.

Plusieurs mesures du Plan national du Ministère vont créer de nouvelles incitations, comme le lancement d'un « appel ANR Flash » pour accélérer l'adoption des principes FAIR⁸ (Wilkinson *et al.* 2016) et l'ouverture des données, la généralisation de la mise en place de plans de gestion dans les appels à projets de l'ANR et d'autres agences à partir de 2019, la création d'un prix des données de la recherche pour récompenser les équipes et projets exemplaires, la construction d'un réseau de correspondants données sur le terrain, et l'adoption d'une politique de données ouvertes associées aux articles de revue.

Quant aux verrous, les études sur les pratiques scientifiques donnent parfois l'impression que le principal obstacle à une bonne gestion est le manque de motivation et/ou de compétences des chercheurs eux-mêmes. Les interviews sur le campus des sciences humaines et sociales (SHS) de l'Université de Lille dessinent une autre image. Le problème principal n'est pas la psychologie ou le savoir-faire, mais l'absence d'informatique dédiée (stockage, communication, bases de données, etc.) et de ressources humaines (département informatique et juridique, bibliothèque universitaire, laboratoires, etc.). Les chercheurs et les autres membres du personnel concernés sont généralement conscients de ce qui devrait être fait pour assurer un minimum de gestion efficace des données de la recherche, mais ils sont souvent incapables de le faire correctement en raison du manque de ressources et de la lourde charge de travail.

Parmi les problèmes identifiés figurent le manque d'outils et d'assistance pour les doctorants, des programmes de formation insuffisants, des problèmes de conservation à long terme sur les serveurs de laboratoire, un manque de suivi des projets à grande échelle et un manque de procédures (cf. Serres *et al.* 2017 et Duprat 2019). A ceci s'ajoute le vécu d'« injonctions contradictoires » (Corvol 2018) entre pratiques communautaires, dispositifs de valorisation, appels à l'ouverture et au partage et les contraintes réglementaires et techniques.

⁸ *Findable, accessible, interoperable, reusable* : des principes de bonnes pratiques qui cherchent à rendre les données de recherche plus facilement trouvables et utilisables, notamment par des machines

Les chercheurs interrogés constatent un manque de ressources informatiques et de personnel spécialisé dans les laboratoires et dans les services communs et centraux. Cette observation ne minimise pas la réalité de ce que Serres *et al.* (2017) appellent l'impact des écosystèmes et pratiques de recherche dans certaines disciplines. S'agissant de bonnes pratiques et de contraintes liées à la législation ou aux programmes de recherche, le principal verrou n'est pas psychologique ou sociologique, mais matériel, faute de moyens.

Par rapport aux ressources, le Plan national préconise le développement des centres de données thématiques et disciplinaires, ainsi que la création d'un « service générique d'accueil et de diffusion de données simples », à l'instar des grands entrepôts multidisciplinaires comme Zenodo⁹, DANS¹⁰, DRYAD¹¹ ou Figshare¹² ou des dispositifs comme Dataverse.nl ; dans ce sens, le CoSO a décidé de fédérer le déploiement des plateformes Dataverse au niveau des institutions et structures de recherche (Sciences Po Paris, CIRAD, INRA, IRD, INSERM etc.).

3. Des stratégies convergentes

A l'heure actuelle, un nombre croissant d'universités déploient une politique de données, d'une manière pragmatique et à partir d'un socle de dispositifs et de services opérationnels. Certes, chaque établissement compose avec ses particularités et son environnement, et il faut admettre, avec la Ligue des Universités de Recherche Européennes, que « Open Science represents a complex and multi-dimensional process of transition, different for every university » (LERU 2018). Néanmoins et malgré ces différences, on peut observer une sorte de convergence empirique d'approches et de modèles, que ce soit au niveau des bibliothèques universitaires ou au niveau des institutions.

Ainsi, la Ligue des Bibliothèques Européennes de Recherche a formulé dix recommandations pour inciter les bibliothèques universitaires à développer de nouveaux services et contribuer à la mise en œuvre d'une politique de données institutionnelle (Liber 2012). Cette contribution est décrite comme un ensemble cohérent de prestations, dont l'information et la formation des personnels scientifiques, l'assistance et le conseil personnalisé des chercheurs et la mise en place d'outils et de dispositifs pour la gestion des données, tout cela en étroite collaboration avec les communautés scientifiques (cf. Reznik-Zellen *et al.* 2012, Pryor *et al.* 2014). Parmi les recommandations spécifiques aux bibliothèques universitaires figurent l'attribution d'identifiants, l'indexation des données (métadonnées), la contribution à la normalisation des formats de données et de métadonnées, le signalement des données dans des catalogues et bases de données, l'assistance à la préparation des plans de gestion puis l'administration et l'alimentation des entrepôts de données (cf. Liber 2012).

L'importance de la curation des données de recherche est primordiale (Neuroth *et al.* 2013) et la définition même du concept des données de recherche doit inclure cette dimension, au même titre que leur finalité et leur caractère communautaire (Schöpfel *et al.* 2017). La curation des données mobilise des compétences cœur des métiers de l'information, telles que l'indexation et le catalogage. Les bibliothèques universitaires sont de ce fait un acteur-clé sur le campus, avec leur savoir-faire et leurs outils en matière de curation des métadonnées. Cependant, une politique de la donnée contient d'autres dimensions (informatiques, juridiques, éthiques, scientifiques) qui

⁹ <https://zenodo.org/>

¹⁰ <https://dans.knaw.nl/en>

¹¹ <https://datadryad.org/>

¹² <https://figshare.com/>

dépassent le périmètre des bibliothèques et qui constituent un challenge pour l'ensemble de l'institution.

Sur le plan institutionnel, l'Université de Bielefeld propose une stratégie qui fait le lien entre la gouvernance, le pilotage et la communication, les infrastructures techniques et les services opérationnels pour accompagner et faciliter le travail des chercheurs. Cette stratégie s'articule autour d'une structure interdisciplinaire de formation et de recherche, le *Bielefeld Center for Data Science*, et s'oriente vers un pôle de compétence où seront regroupés tous les services et acteurs sur le campus qui contribuent à la gestion des données de la recherche (curation, droit, dépôt, publication, formation etc.).

Aux Pays Bas, l'Université de Wageningen a préparé sa politique institutionnelle à partir d'une analyse des bonnes pratiques, en prenant soin de valoriser la diversité des réalisations et initiatives pour crédibiliser et illustrer les actions de communication, d'information et de formation (van Zeeland et Ringersma 2017). Une autre particularité de l'approche de Wageningen est le choix de huit cas d'application dans cinq domaines différents, dont l'alimentation (*consumer science*), la biologie végétale (*genetics*) et l'environnement (analyse de risque) ; cette approche met en avant l'exemplarité et l'excellence scientifique.

L'Université d'Iowa représente un troisième cas de figure car elle s'appuie sur les plans de gestion imposés par les appels à projets de la *National Science Foundation* (NSF) et des *National Institutes of Health* (NIH), notamment dans le domaine de la santé ; Averkamp *et al.* (2014) ont analysé comment cette approche institutionnelle a modifié les pratiques, les attitudes et les services sur le campus : le fait d'introduire les plans de gestion comme nouvel outil a contribué à une prise de conscience des enjeux, à une amélioration des pratiques en matière de stockage, de description (curation) et de conservation, et à davantage de motivation pour ouvrir et partager les données.

Ces trois exemples de politique d'établissement ne sont pas incompatibles ; un cadre général a été formulé par Cox et Verbaan (2018), en termes de cinq axes pour le développement d'un ensemble de services de données au sein d'une université. Ces cinq axes contiennent aussi bien la définition d'une politique au plus haut niveau que le développement de nouveaux services à partir d'une analyse détaillée des pratiques et besoins, l'accompagnement par une stratégie de communication et de formation, et le développement d'infrastructures, d'outils et de procédures pour la gestion des données tout au long de leur cycle de vie (figure 1).

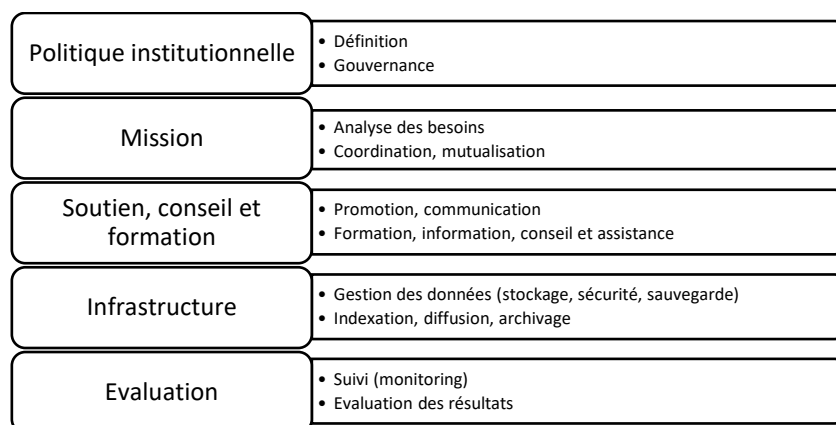


Figure 1. Les cinq axes de Cox et Verbaan (2018)

Cette stratégie paraît intéressante pour deux raisons. D'une part, elle place la définition d'une politique de données et la gouvernance institutionnelle au sommet des cinq axes ; il est tout à fait possible d'envisager un ensemble de services et de dispositifs sans coordination ou avec une coordination minimale à un niveau subalterne, mais cela ne s'appelle pas une politique institutionnelle. D'autre part, l'approche de Cox et Verbaan montre la complexité des enjeux, allant

d'une stratégie marketing à l'articulation avec les infrastructures et à la mise en place d'un dispositif de suivi, tout cela sur un terrain peu exploré, dans un environnement dont les conditions technologiques et juridiques changent rapidement.

Le dernier aspect – l'axe « évaluation » - introduit un élément évoqué par la Ministre Frédérique Vidal lors des premières Journées Nationales de la Science Ouverte (JNSO) début décembre 2018 : la nécessité de développer un ensemble d'indicateurs pour mesurer les résultats de la politique pour la science ouverte dans le domaine des données de recherche, et l'intérêt d'accompagner la mise en œuvre de cette politique par des études dédiées.

4. Dix principes pour co-construire une politique de données en SHS

Force est de constater qu'il n'existe pas de modèle de référence pour la mise en œuvre d'une politique de données sur un campus universitaire. Chaque établissement doit faire face au défi de développer une stratégie adaptée aux projets et dispositifs existants qui tient compte à la fois du cadre national et international et des pratiques et besoins de ses personnels scientifiques sur le terrain. Certains établissements disposent déjà de services de données bien identifiés et opérationnels, d'autres sont dans une phase de montée en charge ou réfléchissent à la meilleure manière de préparer et coordonner un projet collectif.

A partir des résultats des deux enquêtes de Lille, dix principes ont été établis pour une politique de données sur un campus SHS, en cohérence avec le Plan national et en capitalisant les initiatives et expériences d'autres établissements.

Regroupés en trois ensembles – gouvernance, dispositifs campus et infrastructures – ces dix principes sont formulés d'une manière pragmatique et réaliste pour réussir une politique de données, avec un souci de faisabilité et d'acceptabilité par les différentes communautés, composantes et services sur le campus. L'acceptabilité signifie ici qu'une nouvelle politique de données sera d'autant plus efficace si elle répond aux besoins et priorités des chercheurs, sans discours d'injonction idéologique sur la question de l'ouverture des données (« partage ») et sans éparpillement des efforts et ressources.

4.1. Gouvernance

La bonne gouvernance est l'un des facteurs clés du succès d'un projet de données de la recherche sur un campus. Il s'agit d'une question sensible avec plusieurs dimensions, qui mobilise des décisions politiques aussi bien que des choix stratégiques dans les domaines du développement des services (marketing) et de la communication.

4.1.1. Mettre en place un pilotage scientifique

Développer une politique de données en SHS nécessite avant tout un pilotage scientifique, par un comité de pilotage et de coordination rattaché à la direction de la recherche, légitime et accepté par les chercheurs, avec des compétences politiques et scientifiques. Ce comité réunirait les Vice-Présidents en charge de la recherche et du numérique, des représentants de laboratoires et projets scientifiques, des représentants des directions de la recherche (valorisation, ingénierie et management de projets), des systèmes d'information et des affaires juridiques, le responsable de la sécurité des systèmes d'information (RSSI) et le délégué à la protection des données (DPD), le président du comité d'éthique et le directeur du service commun de la documentation (SCD).

La mission d'un tel comité serait la préparation d'une politique de données institutionnelle à décider par les conseils centraux, et la coordination de sa mise en œuvre, y compris du suivi et de l'évaluation. Sa légitimité repose essentiellement sur le pilotage par les communautés scientifiques elles-mêmes. Ses valeurs sont à définir à partir de la collégialité, la scientificité, l'intégrité, la

cohérence et la transparence de la démarche, et de la subsidiarité en ce qui concerne la mise en œuvre. Il faudra prendre au sérieux ce qui a été décrit comme une nouvelle forme de contrat social entre l'ensemble des utilisateurs à chaque étape de la chaîne des données, de la production et collecte au traitement et stockage et à la conservation et la réutilisation par d'autres (Reyes 2017) ; il s'agit d'un lien fort entre l'institution et les chercheurs, pas seulement de quelques nouvelles prestations ou contraintes administratives. Par ailleurs, un pilotage scientifique paraît le meilleur garant à même d'initier et d'assurer le suivi scientifique mentionné par la Ministre lors des JNSO 2018.

4.1.2. Investir d'une manière ciblée

Vouloir mener une « politique générale, unique et identique pour tous » (Serres *et al.* 2017) est voué à l'échec, non approprié à cause de la diversité des pratiques et besoins, et irréaliste du fait des ressources humaines, informatiques et financières limitées. D'emblée il faudrait abandonner l'idée de trouver des solutions à tous les problèmes et besoins en matière de données, au bénéfice d'une stratégie d'acculturation à la donnée à partir de la politique scientifique, avec des domaines d'action prioritaires, comme par exemple :

- Les grands programmes de recherche (H2020, ANR etc.), en particulier avec des partenariats internationaux.
- La formation et le suivi des doctorants, pas seulement pour les plans de gestion, mais également pour les espaces de stockage sécurisés et la formation à la gestion des données personnelles et sensibles ; à faire avec les écoles doctorales et d'autres acteurs, comme les Urfist.
- Des services spécifiques à certains outils, méthodes, équipements ou domaines particulièrement exposés.

Investir n'est pas arroser. Concentrer des ressources sur certains domaines d'action prioritaires permet de mettre en œuvre une démarche efficace et efficiente. En plus, si cette démarche est accompagnée d'une communication appropriée, on peut compter sur un effet de marketing viral parmi les communautés scientifiques sur le campus.

4.1.3. Viser les projets

A l'instar de l'observation d'Averkamp *et al.* (2014) que les principaux enjeux de la gestion des données pointent vers des services orientés projets, la troisième recommandation serait de choisir les projets comme principal vecteur et courroie de transmission. Viser prioritairement les projets de recherche et non pas les laboratoires semble plus pragmatique et prometteur, pour plusieurs raisons dont :

- Des besoins précis et immédiats (montage, suivi),
- des contraintes imposées par les agences de financement,
- des obligations légales et réglementaires (données personnelles etc.),
- une gouvernance plus simple,
- l'expérience des pratiques collaboratives au sein des équipes de projet.

Les enjeux juridiques, éthiques, techniques mais aussi politiques concernent avant tout les projets de recherche, où ils se manifestent non pas comme sujets à débattre mais comme problèmes en attente d'une solution. Cette approche mobiliserait en priorité la Direction Recherche et le RSSI, aussi bien pour l'accompagnement (« kit données ») et l'assistance (sécurisation des données) que pour le suivi des dossiers.

4.1.4. Utiliser les plans de gestion comme levier

Pour développer une culture de la donnée et mettre en œuvre de bonnes pratiques scientifiques, l'expérience de l'Université d'Iowa a montré que les plans de gestion représentent un levier efficace. Ils sont devenus obligatoires pour les projets européens du programme H2020, ils figurent parmi les actions prioritaires de la politique du Ministère, et l'ANR a annoncé vouloir les rendre obligatoires à partir des appels à projets en 2019.

Dans la mesure où ils décrivent l'ensemble du cycle de vie de la gestion des données, de la collecte au traitement et à la génération de nouvelles données, les plans de gestion des données constituent un élément clé de la bonne pratique en matière de données. L'idée est donc d'utiliser ces plans comme levier, en s'appuyant sur une contrainte forte externe, liée aux projets les plus prestigieux et les mieux dotés. Il faudrait faire le lien avec le protocole éthique et son volet « traitement de données » pour éviter un double travail, et il faudrait essayer d'assurer un suivi au moins ponctuel, dans la mesure où les projets H2020 (et demain sans doute aussi ANR) demandent des plans de gestion à mi-parcours et à la fin de projet.

4.1.5. Soutenir les bonnes pratiques

A l'exemple du modèle de l'Université de Wageningen, une proposition complémentaire est de valoriser les bonnes pratiques sur le campus, à titre d'exemple et de modèle, pour la communication, pour la formation, aussi pour la promotion et le marketing de nouveaux services et outils. Cette valorisation pourrait également permettre de faire émerger certaines initiatives éligibles au prix récompensant les équipes et projets exemplaires annoncé par le Plan d'action du Ministère. Une telle stratégie impliquera plusieurs étapes :

- Identifier les cas de bonnes pratiques, par discipline, laboratoire, équipement, méthodologie ou type de projet.
- Décrire ces cas dans leur contexte, avec leurs facteurs-clés de succès, leurs retombés etc. (« story-telling »).
- Créer des vitrines virtuelles pour les rendre visibles au plus grand nombre (« showcases »).
- Construire du matériel de communication (vidéos, sites, plaquettes...) et de formation (recommandations, guides, procédures, modélisation, « Kit données ») à partir des exemples de bonnes pratiques.

Cette valorisation pourrait cibler, du moins au début, certains domaines et cas de figures prioritaires, comme les recherches en santé, les projets européens, les enquêtes, les enregistrements vidéo, l'utilisation d'un entrepôt comme Zenodo etc.

4.2. Dispositifs campus

D'après les résultats de nos enquêtes, les chercheurs expriment en priorité le besoin de disposer d'outils performants pour sécuriser leurs données et pour communiquer ces données à d'autres chercheurs partenaires de leurs projets de recherche. Il s'agit souvent d'obligations et de contraintes auxquelles ils ont du mal à répondre avec les ressources du campus - la protection des données personnelles, la confidentialité des données dans les projets du domaine de la santé, les exigences de partenaires industriels ou étrangers etc. Une politique de données qui met l'accent sur l'ouverture et le partage des résultats sans faire de la sécurité du stockage et de la communication une priorité manquera de crédibilité. Quant à l'ouverture des données et à leur gestion en conformité avec les principes FAIR, il faut apporter des réponses pragmatiques et réalistes.

4.2.1. Apporter des réponses aux contraintes de sécurité

La sécurité est la première préoccupation des chercheurs, et l'établissement doit proposer un environnement de travail à la hauteur des enjeux et des obligations réglementaires, sur ses propres

serveurs et/ou « in the cloud », avec des partenaires publics ou privés. Souvent, les équipes n'ont pas les moyens pour réaliser une analyse des risques en amont d'un projet et pour assurer une protection suffisante des données contre le piratage et d'autres risques (incendie, vol, crash...), pendant la durée du projet. Les projets de recherche en SHS dans le domaine de la santé et avec les hôpitaux, notamment en psychologie, sont particulièrement exposés et sensibles. Il faudra sensibiliser sur les risques de services tiers non affiliés à l'ESR, dans le cadre nouveau du règlement général sur la protection des données (RGPD) et en partenariat étroit avec le délégué à la protection des données (DPO) de l'établissement.

Les enquêtes sur la gestion des données en SHS décrivent des pratiques centrées sur l'ordinateur personnel, privé et/ou professionnel. Le cryptage des disques des personnels pourrait être une solution, tout comme la mise à disposition d'une solution de connexion sécurisée (réseau privé virtuel). Cependant, une démarche sécurité pourrait changer la situation, en mettant le cloud au cœur de la gestion et de la protection des données, à la place de l'ordinateur personnel.

4.2.2. Apporter des réponses aux besoins de communication

La communication sécurisée des données au sein d'une équipe projet et avec d'autres établissements et collègues est l'autre préoccupation des chercheurs. Une politique de données doit faire de ce besoin une deuxième priorité.

Il peut y avoir plusieurs solutions dont certaines ne se trouvent pas nécessairement sur le campus. Néanmoins, dans le cadre d'un dispositif de données, l'option la plus réaliste semble aujourd'hui un serveur de stockage temporaire avec des fonctionnalités de partage, du genre *Invenio* ou *Dataverse*, un système qui, comme évoqué plus haut, fait actuellement l'objet d'une attention particulière de la part des organismes de recherche et du CoSO.

4.2.3. Apporter des réponses aux besoins de curation

Ouvrir les données rend une curation professionnelle indispensable (Neuroth *et al.* 2013), surtout quand il s'agit d'appliquer les principes FAIR préconisés par l'Union européenne et mis en avant dans le Plan national pour la science ouverte. Cela veut dire en bref :

- Contribuer à la normalisation des métadonnées (formats génériques ou disciplinaires),
- contribuer à l'utilisation (attribution) d'identifiants uniques (DOI¹³, ORCID iD¹⁴ etc.),
- contribuer à la création de liens avec les publications associées (archive ouverte institutionnelle ou autre).

Les chercheurs n'ont généralement ni les outils ni les compétences ou le temps pour faire ce travail correctement. Comme il s'agit d'activités et de compétences qui relèvent des métiers de l'information, la curation des données nécessite une étroite coopération entre bibliothécaires, documentalistes, chercheurs et, le cas échéant, ingénieurs de données. Une telle offre de service doit rester flexible et s'adapter aux infrastructures, politiques et pratiques des chercheurs, en proposant aussi bien de la médiation, de la formation et de l'assistance, mais aussi de la « sous-traitance », via la prise en charge de la curation par les professionnels. Aussi, une partie de la curation demande une expertise de spécialistes d'un domaine ou d'un équipement, qui se trouve davantage dans les laboratoires que dans les services communs.

¹³ Digital object identifier <https://www.doi.org/>

¹⁴ Identifiant unique pour des chercheurs <https://orcid.org/>

4.3. Infrastructures

D'autres besoins nécessitent une réponse dont la réalisation se trouvera partiellement ou complètement en dehors du campus. Pour la mise en œuvre d'une politique de données institutionnelle, cela implique nécessairement des partenariats avec d'autres établissements, organismes ou réseaux, notamment dans le domaine de l'archivage pérenne mais aussi pour la diffusion et l'exposition.

4.3.1. Proposer plusieurs solutions pour la conservation des données

Le terme de conservation englobe, du point de vue fonctionnel mais aussi informatique, plusieurs dispositifs, de la conservation à court terme (archives vivantes), pour la durée d'un projet ou de sa valorisation (publications etc.), de la conservation à plus long terme (archives définitives), et enfin de la conservation à long terme sans diffusion (« dark archive »). Proposer des solutions pour tous ces besoins ne veut pas dire, développer toute une gamme de dispositifs sur le campus ; ce serait irréaliste. Mais il faudrait pouvoir proposer des espaces appropriées (en termes de volumétrie, sécurité, accessibilité) pour les besoins des chercheurs, soit sur les serveurs de l'université, soit en partenariat avec des prestataires externes, y compris le Centre Informatique National de l'Enseignement Supérieur (CINES). Le projet d'un entrepôt national générique et transdisciplinaire tel qu'évoqué par le Plan national prend ici tout son sens. Pour le reste, il s'agira de conseiller, d'orienter, le cas échéant de faire le lien avec des infrastructures nationales, internationales et/ou disciplinaires, y compris les services de Renater et les data centers régionaux.

4.3.2. Institutionnaliser le lien avec la TGIR Huma-Num

La TGIR Huma-Num¹⁵, seule Très Grande Infrastructure de Recherche dédiée aux lettres, sciences humaines et sociales et aux humanités numériques, propose des services numériques pour les programmes de recherche, dont une gamme de prestations pour faciliter l'accès, le signalement, la conservation et l'archivage à long terme des données de la recherche.

Huma-Num a mis notamment en œuvre un service d'exposition de données appelé NAKALA avec deux types de services : des services d'accès aux données elles-mêmes et des services de présentation des métadonnées. Les données hébergées par NAKALA peuvent être éditorialisées sur le web à l'aide d'un « pack » associant une bibliothèque numérique (Oméka) et NAKALA.

Afin de promouvoir le dispositif et de coordonner les contacts avec les laboratoires, il faudrait désigner un correspondant local TGIR Huma-Num au sein du service commun de la documentation, à l'instar de l'Université de Nice Sophia Antipolis.

5. Discussion

Ces principes forment un cadre global pour une politique de données qui se résume en trois axes :

- Mettre en place un pilotage scientifique, pour une coordination des actions et services de l'ensemble des acteurs (services centraux et communs, structures de recherche, composantes etc.).
- Concentrer la politique sur certaines actions ciblées, en tenant compte des priorités des chercheurs et en mettant l'accent sur les projets de recherche (H2020, ANR etc.).
- Positionner la démarche clairement au sein des infrastructures nationales et européennes en SHS, en particulier par une institutionnalisation des liens avec la TGIR Huma-Num.

Ce cadre s'appuie sur une approche *bottom-up*, dans la mesure où il trouve sa légitimité et sa justification dans les résultats empiriques d'enquêtes de terrain et dans l'analyse des conditions du

¹⁵ <https://www.huma-num.fr/>

travail scientifique sur un campus universitaire en SHS. Il est tout à fait compatible avec les initiatives de Bielefeld ou de Cox et Verbaan (2018), même si l'accent est mis sur les besoins réels des chercheurs pour la gestion de leurs données, et non pas sur l'objectif d'ouverture des données. Mais il est certain aussi que cette approche doit être accompagnée par une forte implication de l'encadrement scientifique, à commencer par les responsables des équipes et structures scientifiques dont le rôle crucial a été souligné par une enquête auprès de 400 directeurs de laboratoires du CNRS (Schöpfel *et al.* 2018).

Dans le contexte de la science ouverte, l'adossement des principes à une analyse empirique des pratiques scientifiques limite cette approche car elle inverse en quelque sorte les priorités, plaçant les besoins de chercheurs en matière de sécurité, de communication et de curation des données avant les objectifs politiques de partage et d'ouverture des données.

Cet inversement est assumé selon deux explications. D'une manière pragmatique et opérationnelle, au vu des retours d'expérience il paraît illusoire de vouloir développer un ensemble de services de données en mode top-down, par des mandats et injonctions, sans proposer des solutions aux problèmes de la gestion quotidienne des données de recherche par les chercheurs. Il est non seulement impossible de mener une politique générale, unique et identique pour tous mais en plus, une stratégie faite de « discours d'injonction, de pression moralisatrice ou culpabilisante sur la question du partage » (Serres *et al.* 2017) est condamnée à l'échec ; la priorité doit être accordée aux besoins et contraintes auxquels les chercheurs et ingénieurs sont confrontés tous les jours. En d'autres termes, se contenter d'appels au partage ne remplacera jamais l'investissement nécessaire dans les services, dispositifs et infrastructures, pour sécuriser la gestion des données, permettre une communication au sein des équipes de recherche et assurer la fonction de curation. Des discours sans ressources ne mènent nulle part.

La deuxième raison est la conviction qu'avant de parler d'ouverture et de partage, il faudra développer davantage la culture de la donnée, la « data literacy » ; celle-ci se définit par un ensemble de compétences et de connaissances spécifiques pour transformer des données en information et en connaissance pratique ; elle se définit aussi par la capacité de savoir accéder, interpréter, évaluer, gérer et utiliser de manière éthique les données de recherche (cf. Koltay 2016). Les enquêtes dans les universités et laboratoires de recherche reflètent une grande diversité et des compétences et connaissances inégales. Du point de vue des communautés scientifiques, pour optimiser les pratiques et pour rendre la performance scientifique plus efficace, il paraît plus urgent de renforcer et consolider cette culture de la donnée, plutôt que se focaliser uniquement sur l'ouverture et le partage.

Ce besoin de développer une culture de la donnée scientifique est peut-être caractéristique pour un campus universitaire en SHS, avec une multitude de disciplines, domaines, équipes et projets, mais sans grand équipement ou infrastructure de recherche d'envergure capable, à l'instar d'un observatoire ou d'un synchrotron, de fédérer les compétences et les pratiques des chercheurs. Aussi, les enquêtes à Lille ou à Rennes laissent penser que le développement des grandes infrastructures numériques, avec des data centers etc., semble assez éloigné et peu connu par la plupart des chercheurs dans les laboratoires en SHS. Ceci est certes une autre limitation de cette étude qui n'a pas vocation à proposer un modèle pour tous les établissements, organismes ou disciplines. C'est justement pour cette raison que nous suggérons un partenariat fort et « politique » avec la TGIR Huma-Num, afin de rapprocher les communautés SHS de cet environnement en évolution rapide.

Un dernier point est peut-être également lié à la limitation de l'analyse aux SHS. Il s'agit de s'appuyer d'abord sur les projets et leurs porteurs ; ceux-ci, à l'inverse des laboratoires, sont considérés comme principal vecteur d'une politique institutionnelle de données. La raison n'est pas un manque d'appréciation du rôle des laboratoires sur le campus ou une volonté d'affaiblir leur importance dans l'écosystème de la science ouverte. Mais force est de constater qu'en matière de

données, les besoins et demandes s'articulent essentiellement autour des projets et moins au niveau des laboratoires (par exemple quand il y a une base de données centrale) ; par conséquent il paraît plus pragmatique de construire un ensemble de services autour des projets de recherche, sachant que les laboratoires continueront à jouer leur rôle d'intermédiaire entre administration et équipes de recherche, pour assurer une cohérence globale des différents projets et pour la formation des doctorants.

Pour résumer, il faut éviter tout discours d'injonction idéologique sur la question de l'ouverture des données, tout comme il faut éviter l'éparpillement des efforts et ressources. Il n'est pas possible de donner une réponse à toutes les demandes, d'autant qu'une partie des solutions se trouvent à l'extérieur du campus, dans les projets et communautés de recherche, dans les infrastructures et services au plan national, et dans les réseaux internationaux. La politique à mener devrait appliquer des principes de subsidiarité et de complémentarité, ce qui implique une très bonne connaissance du terrain de la recherche et des dispositifs de données.

6. Conclusion

La structuration et l'ouverture des données de la recherche figurent parmi les objectifs du Plan national pour la science ouverte du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Le Ministère a mis en place un comité pour coordonner les différentes mesures et actions et pour accompagner les établissements et organismes de recherche sur le terrain, par des incitations, des recommandations, des forums de discussion et d'échange, mais aussi par le financement de certaines infrastructures etc., sachant qu'il n'existe pas de modèle unique ou de stratégie clé en main pour mettre en œuvre une politique de données au niveau local. Tout n'est pas à inventer puisqu'il existe un nombre croissant d'initiatives et de projets, de retours d'expériences et d'études de cas ; cependant chaque établissement et organisme doit décider de sa propre stratégie, à partir des services et dispositifs existants et en fonction de ses compétences, profils et ressources.

Dans ce contexte, l'étude propose un cadre global autour de dix principes pour la mise en place d'une politique de données sur un campus SHS, à partir de plusieurs enquêtes à Lille, Rennes et ailleurs. Il s'agit d'une feuille de route faisant le lien entre les besoins, les attentes et les pratiques identifiés sur le terrain et les modèles et stratégies dans l'environnement de la science ouverte. L'objectif principal est de (re)mettre les chercheurs, enseignants-chercheurs, ingénieurs et doctorants au cœur de la politique de données sur le campus, non pas (uniquement) comme clients et usagers d'un nouvel ensemble de dispositifs et de services de données, mais (surtout) comme partie prenantes, acteurs et pilotes.

Ce rôle actif des communautés scientifiques paraît indispensable pour construire l'écosystème de la science ouverte qui figure dans le Plan national du Ministère, tout en évitant les écueils des injonctions contradictoires mentionnées plus haut. Par ailleurs, l'approche *bottom-up* sera également garante d'une évaluation indépendante et scientifique des différentes politiques, services, dispositifs et infrastructures dans cet écosystème, à partir d'indicateurs à déterminer ; cette approche mesurera l'impact réel de cet écosystème non seulement sur les comportements, méthodes et performances des chercheurs, mais aussi sur le développement d'une culture de la donnée, en termes de compétences, pratiques et valeurs sur le terrain des sciences humaines et sociales.

Bibliographie

- Averkamp S. *et al.* (2014). *Data Management at the University of Iowa: A University Libraries Report on Campus Research Data Needs*. University of Iowa. http://ir.uiowa.edu/lib_pubs/153/
- Borgman C. L. (2018). Big Science, Little Science, and Open Science: Sustainability, Stewardship, and Knowledge Infrastructures. *JNSO 2018. Premières Journées Nationales de la Science Ouverte "De la stratégie à l'action"*, 4-6 décembre 2018, Paris.

- Chartron G. (2018). L'Open science au prisme de la Commission européenne. *Education et sociétés*, vol. 41, n° 1, p. 177-193. doi:10.3917/es.041.0177
- Corvol P. (2018). Le chercheur et la science ouverte. Au cœur d'injonctions contradictoires ? *JNSO 2018. Premières Journées Nationales de la Science Ouverte "De la stratégie à l'action"*, 4-6 décembre 2018, Paris.
- Cox A. et Verbaan E. (2018). *Exploring research data management*, Facet, London.
- Duprat J. (2019). *Les données de la recherche à l'Université Bordeaux Montaigne : Synthèse d'une enquête qualitative auprès des chercheurs*. Rapport de recherché, Université Bordeaux Montaigne. 2019. <https://hal.archives-ouvertes.fr/hal-02020141>
- Koltay T. (2016). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, vol. 49, n° 1, p. 3-14. doi:10.1177/0961000615616450
- LERU (2018). *Open science and its role in universities: A roadmap for cultural change*. League of European Research Universities, Leuven, Belgium. <https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change>
- Liber (2012). *Ten recommendations for libraries to get started with research data management*. Final report. LIBER working group on E-Science / Research Data Management, The Hague. <https://libereurope.eu/strategy/research-infrastructures/rdm/>
- Malingre M.L. et al. (2019). Construction(s) et contradictions des données de recherche en SHS. *Recherche d'information, document et web sémantique*, vol. 2, n° 1. doi :10.21494/ISTE.OP.2019.0336
- Neuroth H. et al. (dir.) (2013). *Digital curation of research data. Experiences of a baseline study in Germany*. vwh, Glückstadt.
- Prost H. et Schöpfel J. (2015). *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3*. Rapport final. Université de Lille 3, Villeneuve d'Ascq. <http://hal.univ-lille3.fr/hal-01198379v1>
- Pryor G. et al. (dir.) (2014). *Delivering research data management services: fundamentals of good practice*. Facet, London.
- Reyes A. (2017). La data literacy ou la "culture de la donnée", le prochain enjeu de nos sociétés. *Les Echos*, 20 février 2017 <https://www.lesechos.fr/idees-debats/cercle/cercle-166504-la-data-literacy-ou-la-culture-de-la-donnee-le-prochain-enjeu-de-nos-societes-2066367.php>
- Reznik-Zellen R. et al. (2012). Tiers of Research Data Support Services. *Journal of eScience Librarianship*, vol. 1, n° 1, p. 27-35. <http://escholarship.umassmed.edu/jeslib/vol1/iss1/5/>
- Schirrwagen J. et al. (2018). Expanding the research data management service portfolio at Bielefeld University according to the three-pillar principle towards data FAIRness. *CODATA RDM Symposium 2018*, 18-20 mars 2018, Göttingen. <https://pub.uni-bielefeld.de/publication/2919659>
- Schöpfel J. et al. (2017). « Pour commencer, pourriez-vous définir 'données de la recherche' ? » *CEUR Workshop Proceedings 1860, CEUR-WS.org 2017*. <https://dblp.uni-trier.de/db/conf/inforsid/vador2017.html>
- Schöpfel J. (2018). *Vers une culture de la donnée en SHS. Une étude à l'Université de Lille*. Rapport final. Université de Lille, Villeneuve d'Ascq. <https://hal.archives-ouvertes.fr/hal-01846849>
- Schöpfel J. et al. (2018). Research data management in the French National Research Center (CNRS). *Data Technologies and Applications*, vol. 52, n° 2, p.248-265. doi:10.1108/DTA-01-2017-0005
- Serres A. et al. (2017). *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2*. Rapport final. Université Rennes 2. <https://hal.archives-ouvertes.fr/hal-01635186>
- Wilkinson M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, vol. 3, n° 160018. doi:10.1038/sdata.2016.18
- van Zeeland H. et Ringersma J. (2017). The development of a research data policy at Wageningen University & Research: best practices as a framework. *LIBER Quarterly*, vol. 27, n° 1, p. 153-170. Doi :10.18352/lq.10215/