



N° 4 | 2018

Méthodes et stratégies de gestion de l'information par les organisations : des big data aux thick data

Gestion des données de recherche et thick data : le cas d'une recherche-action sur la médiation numérique du patrimoine

Karel Soumagnac

Maître de conférences

IMS UMR5218 CNRS

University of Bordeaux

Édition électronique :

URL :

<https://revue-cossi.numerev.com/articles/revue-4/1835-gestion-des-donnees-de-recherche-et-thick-data-le-cas-d-une-recherche-action-sur-la-mediation-numerique-du-patrimoine>

DOI : 10.34745/numerev_1607

ISSN : 2495-5906

Date de publication : 04/09/2018

Cette publication est **sous licence CC-BY-NC-ND** (Creative Commons 2.0 - Attribution - Pas d'Utilisation Commerciale - Pas de Modification).

Pour **citer cette publication** : Soumagnac, K. (2018). Gestion des données de recherche et thick data : le cas d'une recherche-action sur la médiation numérique du patrimoine. *Revue COSSI*, (4).

https://doi.org/https://doi.org/10.34745/numerev_1607

La gestion et la mise à disposition des données d'une recherche-action sur la médiation numérique du patrimoine constituent un enjeu en termes de durabilité de l'information (sustainability). Ce travail de gestion rend compte d'une épaisseur socio-culturelle des données (thick data) visant le projet commun d'une science en réseau. Dans une perspective ethnographique et sociale de la recherche nous abordons les ressources numériques patrimoniales, les terrains d'observation et les acteurs selon une dimension systémique. Nous présentons les enjeux actuels des modes de gestion des données de recherche dans le cadre de la recherche-action. Nous revenons sur la méthodologie de la recherche en soulignant les caractéristiques documentaires et communicationnelle de la conservation, du traitement et de la communication des données dans le cadre d'un écosystème informationnel ouvert. Nous terminons par le travail de valorisation, d'information et de médiation documentaire des données pour qu'elles puissent être visibles, compréhensibles et accessibles à tous.

Mots-clefs :

Science ouverte, Données de la recherche, Recherche-action, Patrimoine, Gestion des données, Thick Data

Abstract : The management and the provision of the data of a research-action on the digital mediation of the heritage establish a stake in terms of durability of information (sustainability). This work of management reports a sociocultural thickness of the data (thick data) aiming at the common project of a science in network. In an ethnographical and social standpoint of the research we approach the patrimonial digital resources, the grounds of observation and the actors according to a systematic dimension. We present the current stakes in management modalities of the data of research as part of the research-action. We return on the methodology of research by underlining the documentary characteristics and communicationnelle of the preservation, the treatment and the communication of the data as part of an open informative ecosystem. We end with the work of value, information and documentary mediation of the data so that they can be visible, understandable and accessible to all.

Keywords : Data of research, Data Management, Open Science, Research-action, Thick Data, Resarch-Action, Heritage

INTRODUCTION

En France, il est devenu courant pour les acteurs de la culture de s'investir dans des missions de médiation numérique autour du patrimoine auprès de publics variés et particulièrement auprès des jeunes. Pour comprendre l'effet de l'utilisation des ressources sur un plan éducatif à travers les dispositifs de médiation culturelle, la direction régionale de l'action culturelle en Aquitaine a souhaité mobiliser des enseignants-chercheurs participants à la formation initiale et continue des enseignants pour réaliser un guide de bonnes pratiques en direction des partenaires et structures culturelles et éducatives mobilisés par l'éducation au patrimoine. L'idée a été de lancer un programme de recherche-action qui permette à l'ensemble des acteurs concernés par le projet de travailler ensemble pour évaluer le travail de médiation et de valorisation du patrimoine. Ce projet de recherche-action, intitulé [P@trinum a](#) pour objectifs de qualifier les logiques de médiation du contenu patrimonial, d'identifier les représentations et les usages réels et potentiel des élèves et des enseignants sur le patrimoine numérique, caractériser les communautés professionnelles participant à la médiation des dispositifs numériques du patrimoine à travers sept « expérimentations » de dispositifs de médiation numérique du patrimoine créés par des archivistes, bibliothécaires, universitaires, start-up informatiques, associations culturelles et utilisées en établissements scolaires par les élèves, les enseignants, les intervenants numériques. Dans le cadre de ce projet, nous cherchons à analyser les méthodes de gestion des données de recherche par les acteurs qui y participent. La question de la gestion des données dans les programmes de recherche interroge d'une part l'utilisation des nouvelles technologies pour gérer, traiter communiquer les données issues du terrain ; elle renvoie d'autre part au problème de la conservation des données de recherche à l'issue du projet et à leur valorisation au-delà du projet lui-même. Notre hypothèse de travail est que ce travail de gestion rend compte d'une épaisseur socio-culturelle des données (thick data) servant le projet commun d'une « science en réseau » (Millerand, 2012). Nous nous proposons dans cette communication de partir d'un état des lieux sur le contexte institutionnel et scientifique des données de recherche. Nous abordons ensuite la méthodologie générale du projet puis celle exploratoire que nous venons de mettre en place fondée sur une approche compréhensive des acteurs qui travaillent à la gestion des données de recherche. Enfin nous revenons sur les résultats que nous mettrons en discussion.

LES DONNÉES DE RECHERCHE : DÉFINITIONS ET ENJEUX DE LEUR GESTION DANS LE CADRE D'UNE RECHERCHE-ACTION

La production de données de recherche est le propre des projets scientifiques quelle que soit leur envergure, les acteurs présents, le budget alloué, les méthodes déployées pour parvenir à donner des réponses aux hypothèses de travail à l'origine de la recherche elle-même. Les recherches de type action qui incluent des professionnels dans la réflexion menée et qui sont très répandues en sciences de l'information et de la

communication (Meyer, 2006) n'échappent pas à ce phénomène. A l'heure du numérique la production de données de recherche entre dans un processus de formalisation appelé « plan de gestion de données » qui peut être l'occasion de discuter de la notion de durabilité, c'est-à-dire de soutenabilité de la gestion des données et de sa valeur-ajoutée pour la communauté de pratiques.

Recherche-action et production de données de recherche : cadre opérationnel des plans de gestion de données

La recherche-action reste « une démarche adaptée pour développer des stratégies permettant de soutenir des professionnels attachés au développement de leur métier » (Meyer, 2006 : 89). Dans cette perspective, elle est le cadre scientifique de production de données de recherche issues de savoirs profanes captés dans une logique opérationnelle. Dans le même temps, les données de recherche représentent une entrée pour comprendre les pratiques professionnelles des chercheurs en termes d'accès au terrain de la recherche et aux acteurs qui le compose.

Selon l'OCDE, les données scientifiques (ou données de la recherche, research data) sont « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche » (OCDE, 2007). Depuis quelques années, les données de recherche font l'objet de préconisations dans de nombreux pays (Données de recherche, Canada, 2011, ADNS, Australie, 2015, Commission européenne, 2013) en termes de plan de gestion de données (Data Management Plan), visant leur inscription dans un écosystème informationnel impliquant des modalités de gestion concertée entre tous les partenaires des projets de recherche. Un plan de gestion de donnée est un « Document rédigé au commencement d'un projet de recherche et qui définit ce que les chercheurs feront de leurs données pendant et après le projet, explicitant notamment la mise à disposition des données » (Deboin, 2014). Cette définition renvoie à une formalisation des procédures à suivre en matière de stratégie documentaire des données.

Dans tous les pays il existe des cadres opérationnels pour la gestion des données. Au Canada, le cadre d'action permettant le développement de la gestion des données de recherche à grande échelle repose sur l'innovation en matière d'économie numérique (Données de recherche, Canada, 2011). En Australie, la collecte de données de recherche doit permettre leur mise en visibilité et leur réutilisation efficaces dans un contexte scientifique international (ADNS, Australie, 2015). En Europe, depuis 2013, et dans le cadre du plan Horizon de la recherche 2020, la commission européenne inscrit les plans de gestion des données dans une perspective d'ouverture des données de recherche scientifique. La nouvelle loi sur la république numérique du 7 octobre 2016 distingue enfin l'accès aux données de recherche et leur réutilisation dans le cadre de

l'ouverture des données à travers l'exploitation et la fouille de texte. Les plans de gestion de données s'inscrivent dans le mouvement de l'ouverture des données de recherche qui « [...] consacr[e] le partage et la réutilisation des productions scientifiques en principe de base, pour une participation de la science à une société plus ouverte, innovante et inclusive » (Bécard et alii, 2016 : préface).

La question de la durabilité des données de recherche dans le domaine de la médiation numérique du patrimoine

Si le consensus autour du développement d'un plan de gestion de données est avéré, leur gestion elle-même reste plus délicate à mettre en place dans une perspective de durabilité (Nolin, 2010). Nous envisageons la question de la durabilité en termes de soutenabilité de leur gestion c'est-à-dire comme une forme de logiques d'actions sur les données qui soit viable par un collectif. Au-delà de l'engagement des structures habilitées à produire des infrastructures pérennes et efficaces pour conserver et rendre accessible les données, la question de la durabilité des données de recherche s'inscrit dans le 4ème paradigme de la recherche pour lequel la théorie, l'expérimentation et la simulation sont intimement liées aux données et dans lequel il faut réfléchir à l'utilisation de leur gestion. Dans ce cadre, les données de recherche participent d'un gain de productivité (mutualisation, reproductibilité) et de création (croisement, visualisation) (Gallezot, 2015). La durabilité implique une négociation autour de d'usages partagés innovants de la donnée visant la production de connaissances (Argote, 2003). Cette dimension est reprise dans le livre blanc du CNRS paru en 2016 : « Les données scientifiques (à financement majoritairement public) doivent devenir un bien commun informationnel. L'objectif est d'autoriser le dépôt en même temps que des articles, des données de base de la recherche. Un tel dépôt faciliterait la reproductibilité de la recherche en même temps qu'il favoriserait l'innovation dans la société civile » (CNRS, 2016 : 111).

La question de la durabilité des données de recherche s'inscrit aussi dans le cadre des humanités numériques « [qui] désignent un dialogue interdisciplinaire sur la dimension numérique des recherches en sciences humaines et sociales, au niveau des outils, des méthodes, des objets d'études et des modes de communication » (Dacos et Mounier, 2004 : 15). Faire en sorte que les données recueillies puissent être gérées de manière durable s'inscrit au-delà des méthodologies déployées autour de la mise à disposition d'infrastructures de conservation et d'archivage des données. Elle renvoie à la patrimonialisation des données de recherche (Fayet-Montagne, 2015). Enfin, la durabilité des connaissances comprise comme production de connaissances construites à partir des données dans une perspective de reproductibilité questionne la communicabilité des données en termes de visibilité et de réutilisabilité.

La gestion des données de recherche pour la communauté de pratiques : quelle valeur ajoutée ?

La diversité des acteurs de la recherche-action représente une communauté de pratiques, tournée vers des pratiques d'information et de communication, qui cherche à se fédérer autour du projet notamment autour de la dimension cognitive de mobilisation

de ressources (Wenger, 1998). Les données de recherches sont le gage à long terme de l'instauration d'un guide de bonnes pratiques autour de la médiation en ligne du patrimoine. Nous savons que la communauté de SHS estime d'une manière générale que le « Partage de données et techniques de "fouille de données et de textes" sont ainsi inégalement répandus selon les types de données, pour des raisons principalement d'obstacles juridiques [données de tiers], et de manque de moyens en personnel pour la production et le maintien de métadonnées de qualité » (CNRS, 2016 : 28). Les pratiques effectives peinent cependant à se mettre en place (Boukacem-Zeghmouri et Délémontez, 2015). Les bibliothèques représentent néanmoins un espace de réflexion sur la veille et la diffusion de l'information autour des projets de recherche et de la valorisation des données. Le travail peut s'organiser avec la communauté de chercheurs, ingénieurs-projets ou archivistes mais aussi avec instances et services dits locaux qui sont directement concernés par ce travail de conservation, de gestion, d'exploitation et de diffusion des données pour permettre le développement de la science en réseau (Millerand, 2012).

MÉTHODOLOGIE DE RECHERCHE : CADRE D'ANALYSE, DE PRODUCTION ET DE TRAITEMENT DES DONNÉES DE RECHERCHE

Les plans de gestion de données ne sont pas encore mis en place dans toutes les universités françaises. Même si la recherche-action [P@trinum](#) s'inscrit dans une situation où la moitié du budget émane d'un acteur public, il a été décidé par les instances partenaires que le travail autour des données de recherche serait artisanal. Nous avons donc établi une méthodologie de travail autour des données de recherche que nous souhaitons valoriser et mettre en discussion pour faire avancer la réflexion sur la question des thick data.

La méthodologie générale de la recherche

Dans le cadre de [P@trinum](#), nous nous situons dans une perspective ethnographique de la recherche-action. A travers les entretiens et les observations, l'entrée par l'action située (Suchman, 2007) nous permet de prendre en compte une réalité de terrain qui est spécifique à chaque expérimentation numérique c'est-à-dire à chaque type de ressources patrimoniales. La perspective de la cognition distribuée (Conein et Thévenot, 2004) est également nécessaire pour comprendre les relations qui se tissent dans les classes entre les différents intervenants et partenaires de la recherche-action qui travaillent auprès d'élèves dans des espaces de travail professionnels du monde de la documentation, des bibliothèques et des archives. La dimension systémique (Engeström, 2001) quant à elle trouve sa véracité dans le fait de cerner comment les activités mises en place font écho à des pratiques réelles et à des représentations que l'on souhaite également capter et mettre en relation entre l'ensemble des acteurs en présence. A partir des éléments que nous avons commencés à recueillir dans le cadre

de cette recherche nous avons choisi d'entamer un travail de réflexion autour de la gestion des données de recherche.

Cadre d'analyse des données de recherche issues de l'approche qualitative

La notion des « thick data » est un héritage du travail de l'anthropologue de la culture Clifford Geertz qui dans les années 1970 a proposé de réaliser des descriptions fines des observations effectuées dans le cadre du déploiement des études ethnographiques. Cette terminologie de la donnée de terrain fait écho à la notion de « rich data » d'Howard Becker qui de son côté a popularisé le travail de contextualisation de la donnée en préconisant l'observation des phénomènes et la définition du sens de ce qui les organise sans présumer a priori de ce sens. La recherche elle-même gagne en crédibilité et en véracité à partir du moment où la donnée est détaillée, complète et que l'on peut lui donner un sens précis (Onwuegbuzie et Leech, 2007 : 244). A l'heure du big data, certains partisans du thick data revendiquent cette entrée dans la donnée (Wang, 2013), une entrée par la vie professionnelle des gens que l'on accompagne et par laquelle le chercheur entre pour en comprendre les usages, les pratiques et les représentations. Il y a la production d'éléments de contextualisation par le chercheur sous forme de journal de bord mais également la possibilité de tracer les acteurs de terrain suivis à travers les réseaux sociaux. C'est ce que Cardon appelle la « mise en forme du social » (Cardon, 2015) au travers du big data et sur laquelle on peut aussi s'appuyer dans le cadre de la recherche pour étayer et compléter les données de recherche. Les expérimentations suivies ont souvent été tracées par les réseaux sociaux des enseignants de terrain et des commanditaires de la recherche-action, ce sont donc aussi des éléments à prendre en compte lors de l'analyse des données de recherche.

La question des méthodes et des outils pluriels pour capitaliser, traiter et diffuser les données

Nous avons donc réfléchi de manière spécifique à une méthodologie de la recherche exploratoire sur la question des données dans le contexte de la science en réseau et des thick data. On a souhaité dans une perspective ethnographique propre à la recherche-action observer de l'intérieur la gestion des données de recherche par les enseignants-chercheurs et la personne recrutée pour gérer le programme de recherche-action. Ce protocole reste pour l'heure expérimental. Notre hypothèse est que la gestion des données de recherche s'inscrit dans le développement de stratégies et techniques relevant davantage de la boîte à outils du Personal Knowledge Management (Prost et Schöpfel, 2015) que de l'idéal type d'un plan de gestion de données. Nous avons envoyé un questionnaire aux enseignants-chercheurs pour comprendre leur façon de gérer les données. Le gestionnaire réseau a été interrogé de manière semi-directive sur l'organisation de son travail autour des données. L'analyse a également porté sur l'observation des dépôts sur les plateformes d'archivage développées dans le cadre du projet. Le travail s'est enfin nourri des échanges informels avec les enseignants-chercheurs et le chef de projet ainsi que du suivi du projet par le gestionnaire réseau. Le but a été de comprendre les logiques d'actions autour de la gestion des données. Dans le cadre du questionnaire en ligne, on a interrogé les enseignants-chercheurs sur

l'ensemble du cycle de vie des données de recherche (Cirad, 2016), notamment sur les pratiques d'archivage (modalités de dépôt, tailles et types de données, choix des infrastructures), et de traitement des données (principalement autour de la documentation et des métadonnées). La documentation concerne la lisibilité des fichiers par l'être humain, les informations sur le projet (identification du contexte : méthodologie, corpus) et les informations sur les fichiers des données (noms des données, terminologie). Plutôt créées pour être lus par les machines, les métadonnées peuvent aussi émaner de l'outil de capture des données, ou provenir d'informations générées par le capteur des données ou l'utilisateur des données (André, 2016). Enfin on a interrogés les enseignants-chercheurs sur le rôle du gestionnaire de réseau dans la gestion des données, sur la question de la communication des données comme facteur de compréhension d'une science en réseau et sur les liens à établir entre acteurs du projet, institutions et bibliothèques pour mobiliser les données de recherche dans d'autres contextes que celui de la recherche-action.

RÉSULTATS ET DISCUSSION : UNE NÉCESSAIRE ÉPAISSEUR DES DONNÉES POUR COMPRENDRE LA SCIENCE EN ACTION

En termes de résultats, la question de la durabilité des données qui s'appuie sur une politique et des modalités de gestion concertées entre chercheurs, professionnels du patrimoine, enseignants, bibliothécaires est en cours de négociation en l'état actuel du projet. Nous montrons que la constitution du recueil de données est dépendante des outils utilisés et des représentations et pratiques des acteurs. Le traitement des données est également le fait d'un partage de cultures personnelles et professionnelles. C'est le travail autour de la diffusion et la communication qui peut rendre compte de l'épaisseur socio-culturelle de la dimension organisationnelle des données de recherche.

Le recueil des données : la question des outils comme espace de représentations et de pratiques des acteurs du projet

Le choix des outils pour héberger et qualifier les données a été effectué par le gestionnaire réseau. Son travail participe d'un « travail d'articulation » (Strauss, 1985) nécessaire à la compréhension de ce qu'est une donnée de recherche. Plusieurs types de données ont été recueillis, des vidéos, des photos, et des enregistrements sonores. Le gestionnaire réseau a fait une offre de proposition d'outils pour travailler sur les différents types de données, établir la volumétrie et établir la démarche qualité des données recueillies. Pour les vidéos Vimeo a été choisi en fonction de son coût, de son utilisabilité, de la rapidité avec lequel le chef de projet a pu contracter l'abonnement. Les données plus légères comme les photos, les enregistrements audio ont été recueillies sur Google Drive (Figure 1) que la majorité des enseignants-chercheurs de l'équipe utilise au quotidien.

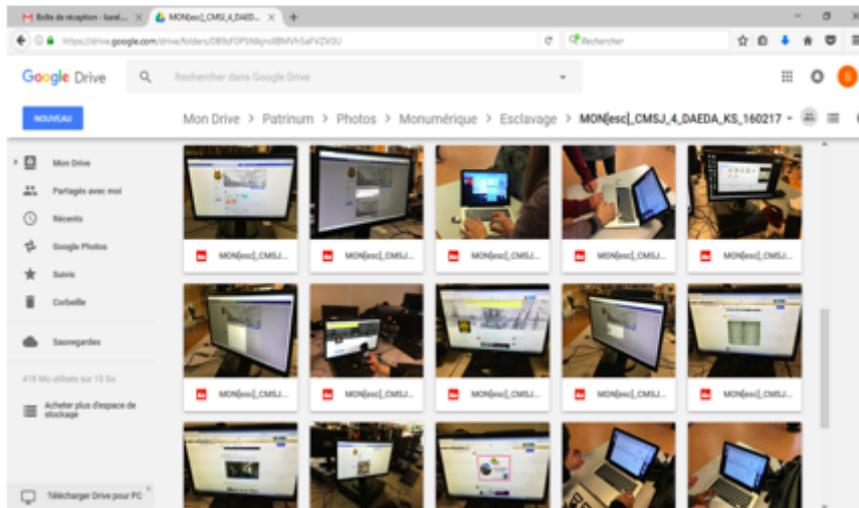


Figure 1 : recueil de données : création d'une base de données dans GoogleDrive

Du côté des enseignants-chercheurs, chacun a ses méthodes et outils de travail qui dépendent de ses pratiques informationnelles personnelles (« ensemble de procédures apprises dans le cadre d'une formation spécifique ou acquise par tâtonnements susceptibles de s'enrichir par l'échange et avec le temps » Gardiès, Favre, Couzinet, 2010 : 128). Le recueil des données de terrain a posé un certain nombre de difficultés ne serait-ce que sur le choix des outils de capture de l'information sur le terrain pour l'observation et les entretiens (GoPro, enregistrement numérique, smartphone, I-pad). Elles dépendent en même temps de la façon de percevoir ce qui relève de la qualité d'une donnée car la donnée elle-même relève d'une forme de publication (Beaudry, 2010). Concernant le traitement des données, certains chercheurs de l'équipe valorisent la documentation de type technique autour de la donnée pour qu'elle soit consultable par n'importe quel public. Pour d'autres, il s'agit simplement de déposer la donnée de recherche sur les outils définis dans le protocole de recherche et par lesquels le gestionnaire des données va pouvoir leur attribuer justement cette épaisseur documentaire. On voit ici que la représentation des enjeux du recueil et de la mise à disposition des données est variable d'une personne à l'autre. Les représentations « constructions sociales de la réalité » (Berger et Luckmann, 1986), dépendent souvent de l'interprétation de la valeur de la donnée en tant qu'elle apporte des résultats significatifs à la problématique de la recherche.

Le traitement des données : classer, indexer, évaluer : quel partage entre acteurs du projet, utilisateurs et institution ?

Suite aux choix des infrastructures de recueil, le gestionnaire réseau a proposé une organisation des données de recherche suivant un classement relatif aux expérimentations de la médiation numérique du patrimoine. A la suite de (Foucault, 1966) la notion de classement peut être comprise comme le fait « d'accorder une

priorité à un objet, à un terme, d'établir une hiérarchie, ce qui exprime un pouvoir sur les choses » (Maury, 2013 : 24). L'entrée dans les données repose tout d'abord sur le classement par expérimentations (Figure 2).



Figure 2 : Le classement des données de recherche par expérimentation sur Vimeo

Ensuite, le gestionnaire réseau a rédigé un document sur le nommage des fichiers présentant le protocole d'identification des données. Par rapport à l'utilisation de Vimeo, on s'aperçoit que dans le cadre de notre projet, l'utilisation de cet outil ne semble pas adaptée aux procédures de dépôt des enseignants-chercheurs qui pour des raisons souvent de droit à l'image conservent les données sur leur ordinateur personnel. Géré par le gestionnaire réseau du projet, l'outil de partage de vidéos propose également, par le biais d'étiquettes, une forme d'indexation ouverte à une communauté d'utilisateurs qu'il s'agit de mobiliser autour d'usages partagés des données (Figure 3). Les enseignants-chercheurs n'ont pas un usage des tags, des like ou d'annotation servant la réutilisation des données. Le commentaire associé à la donnée peut pourtant servir d'évaluation, expliquer le contexte de captation, offrir un cadre d'interprétation de la ressource. Ce n'est pas encore le cas dans ce projet de recherche-action qui conforte le fait que l'évaluation des données de recherche passe encore par le prisme de la publication plutôt que par l'infrastructure les recueillant (Beaudry, 2010).

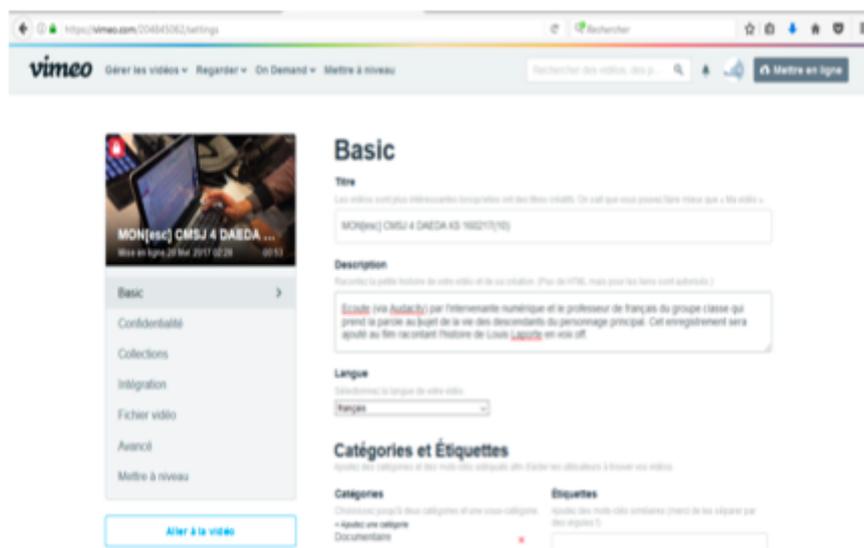


Figure 3 : L'espace « description » sur Vimeo : contexte de description et cadre d'interprétation de la ressource

Les résultats de l'enquête en ligne auprès des enseignants-chercheurs montrent enfin qu'ils ne souhaitent pas forcément diffuser les données brutes auprès des acteurs de la recherche-action. Pour les uns « Une donnée n'est pas a priori une information communicable en dehors de la recherche » (EC 1) car elle engage les acteurs qui y ont participé sur le plan du droit à l'image. Pour d'autres, les données sont « susceptibles de présenter de l'intérêt pour tous » (EC 2). La donnée n'est pas exploitable brute sans métadonnées contextuelles. Les métadonnées peuvent émaner de l'outil de capture des données, ou provenir d'informations générées par le capteur des données ou l'utilisateur ré-utilisateur des données (André, 2016). L'identification du contexte et de la provenance des données (par des mots-clés ou compte-rendu de la séance observée) permet d'avoir confiance dans la donnée produite, et l'utilisateur peut l'évaluer et la rendre reproductible. L'absence de guide ou d'aide a été déplorée parfois alors que le gestionnaire réseau (Millerand, 2012) a mis en place des protocoles de nommage des fichiers et de description des vidéos. Enfin pour le partage des données, les textes institutionnels invitent au dépôt des données sur les dispositifs de communication autour des projets. Dans le cadre la recherche-action les enseignant-chercheurs proposent la création de capsules vidéo pour faire-valoir la recherche en train de se faire, le dépôt des données pour illustrer et documenter le travail de recherche, tandis que le gestionnaire réseau revendique la création de comptes partagés avec les acteurs du terrain pour accéder aux données brutes.

Discussion : La diffusion et la communication autour des données : les thick data comme espace de réutilisation des données

Pour les acteurs enseignants, les archivistes, les médiateurs et les institutionnels du patrimoine, le site web outil de communication du projet^[1] pourrait faire office de porte d'entrée dans les données de recherche via une rubrique consacrée à la recherche en cours qui explique pour les non spécialistes les processus de collecte, de traitement et de diffusion des données. Il est cependant déconnecté de l'ensemble de l'activité de

gestion des données de recherche qui participe d'une classification des objets étudiés, des personnes rencontrées et de leurs discours sur les pratiques et les représentations de la médiation numérique du patrimoine. Selon (Foucault, 1966) le travail de classification est un indice de la visibilité de l'organisation des savoirs. En ouvrant Viméo à d'autres recherches sur les humanités numériques, on offre la possibilité d'accéder, de réutiliser les données « travaillées » par l'être humain, les thick data. Ces données « documentarisées » (Zacklad, 2007), - la documentarisation est « le travail consistant à équiper un support pérenne des attributs qui faciliteront sa circulation dans l'espace, le temps et les communautés d'interprétation » (Zacklad, 2007 : 23) - deviennent des données « augmentées » ou big data, car cette fois ce ne sont plus les acteurs de la recherche-action qui la documentent par un journal de bord où sont consignés les éléments de contexte et d'environnement dans lequel s'exercent et se déploient des pratiques (d'information, de médiation, de patrimonialisation) mais des communautés plus larges, grand public qui comme les acteurs que nous avons observés dans leurs pratiques professionnelles souhaitent connaître et identifier les savoirs, les compétences à l'œuvre dans les métiers traversés par la recherche-action, et les indexer et les faire circuler.

Finalement, il est difficile de s'assurer que l'accès aux données de recherche est garanti pour tous car le comportement informationnel (Taylor, 1991) des chercheurs et des utilisateurs grand public reste variable et la mise en sens des données peut impliquer des réutilisations plurielles (Heaton, 2004, Chabaud et Germain, 2006). Le travail des chercheurs dans l'apport des données et leur qualification auprès de l'ensemble des acteurs participant à la mise en place des plans de gestion de données est primordial pour que la gestion et la diffusion des données puissent s'inscrire dans ce courant amorcé de l'open research data. Il reste à savoir si dans le cadre modélisant de plan de gestion de données et d'ouverture d'infrastructures adaptées au dépôt et au traitement des données de recherche, leur épaisseur ne perde pas ce qui fait leur originalité et leur particularité. Les entrées croisées sur les terrains d'observation, le prisme d'intérêt scientifique de chaque chercheur font aussi partie des éléments de ce qu'on nomme la thick data (cette épaisseur que l'on attribue à la donnée de recherche).

CONCLUSION

Favoriser la dynamique du thick data à travers le processus de gestion des données est primordial pour développer une vraie gouvernance des données (Maurel, 2012). Les plans de gestion de données peuvent être un moyen pour une communauté de pratiques d'instaurer un processus de gestion qui soit identifiable et compréhensible pour tous tout en rendant compte des compétences de tous les acteurs du projet à gérer efficacement les données de recherche. Pour que la question de la durabilité soit réelle et inscrite dans le travail de la gestion des données, le développement d'une gouvernance des données de recherche mériterait un rapprochement avec la culture de la donnée ou data literacy. L'idée de curation est enfin au cœur de la politique de

développement des données de recherche en promouvant la culture du partage des connaissances au sein d'une société ouverte et innovante.

BIBLIOGRAPHIE

André, F. (2016). Gestion des données de la recherche dans le contexte de l'Open Science. *Action nationale de formation RENATIS Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données*. Disponible à : <https://anfdonnees2016.sciencesconf.org/resource/page/id/7>

ANDS. (2011). *Research Data Australia Guide*. Disponible à : <http://ands.org.au/guides/research-data-australia.pdf>

Argote, L. (2003). *Organizational Learning : Creating, Retaining, and Transferring Knowledge*. Berlin : Springer.

Bécard, N., Castets-Renard, C., Chassang, G., Courtois, M.-A., Dantant, M., Gandon, N., Martin, C., Martelletti, A., Mendoza-Caminade, A., Morcrette, N., Neirac, C. (2016). *Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France*. Disponible à : <http://prodinra.inra.fr/ft?id={97224C30-C56B-4CC2-8F78-41C7E1AF5148}>

Beaudry, G. (2010). *La communication scientifique directe : un nouveau champ éditorial*. Hermès, 2(57), pp. 51-57.

Berger, P. et Luckmann, T. (1986). *La Construction sociale de la réalité*. Paris : Méridiens Klincksieck.

Cardon, D. (2012). Regarder les données. *Multitudes*, (49), pp. 138-142.

Chabaud, D. et Germain, O. (2006). La réutilisation de données qualitatives en sciences de gestion : un second choix ? *M@n@gement*, 9(6), pp. 199-221.

CIRAD. (2016). *Le cycle de vie des données. Intégrer la gestion de données scientifiques aux activités de recherche*. Disponible à : <https://coop-ist.cirad.fr/content/download/5922/43494/version/1/file/Cycle-vie-donnees-Poster-Cirad-2016.pdf>

Conein, B. et Thévenot, L. (dirs). (1997). *Cognition et information en société*. Paris : École des hautes études en sciences sociales.

Dacos, M., et Mounier, P. (2014). *Humanités Numériques : État Des Lieux et Positionnement de La Recherche Française Dans Le Contexte International*. Institut français. Disponible à : <http://www.enssib.fr/bibliotheque-numerique/documents/65357-humanites-numeriques-etat-des-lieux-et-positionnement-de-la-recherche-francaise-dans-le-contexte-international.pdf>

Deboin, M-C. (2014). *Découvrir des plans de gestion des données de la recherche, en 4 points*. Montpellier : CIRAD, 6 p. <http://coop-ist.cirad.fr/content/download/5435/40362/version/4/file/CoopIST-plan-gestion-donnees-recherche-20140717.pdf>

Délémontez, R. et Boukacem-Zeghmouri, C. (2015). Données de la recherche : entre discours, réalités et valeur. *I2D - Information, données & documents*, 53(4), pp. 56-57.

Direction de l'Information Scientifique et Technique - CNRS. (2016). *Livre blanc - Une Science ouverte dans une République numérique*. Marseille : Ed OpenEditionPress: 195 p. Disponible à : <http://books.openedition.org/oep/1548>

Données de recherche Canada. (2011). Portrait de la situation des données. *Rapport sur le Sommet 2011 sur les données de recherche canadiennes*. 49 p. Disponible à : <http://www.rdc-drc.ca/wp-content/uploads/f-Report-of-the-Canadian-Research-Data-Summit.pdf>.

European Commission. (2013). *Guidelines on Data Management in Horizon 2020*. [Traduction française]. Disponible à : http://openaccess.inist.fr/IMG/pdf/14081_lignes_directrices_pgd_horizon_2020_tr_fr_versionavril2015-2.pdf

Engeström, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, 7(43), pp. 960-974.

Fayet-Montagne, C. (2015). *Patrimonialisation et réutilisation des données de la recherche en sciences humaines et sociales*. Villeurbanne : ENSSIB. 107 p.

Foucault, M. (1966). *Les mots et les choses*. Paris : Gallimard.

Gallezot, G. (2016). *Les enjeux des données de la recherche*. URFIST Nice, 23 juin. Disponible à : http://urfist-apps.unice.fr/documents/160623_JE_DR/160623_gallezot_enjeux.pdf

Gardiès, C., Fabre, I. et Couzinet, V. (2010). Re-questionner les pratiques informationnelles. *Études de communication*, (35), pp. 121-132.

Heaton, J. (2004). *Reworking Qualitative Data*. London : Sage.

Maurel, D. (2012). Chapitre 8 : Gouvernance informationnelle et perspective stratégique. Dans C. Paganelli (dir.), *L'information professionnelle*. Paris : Hermès sciences publication, pp. 175-197.

Maury, Y. (2013). Classements et classifications comme problème anthropologique : entre savoir, pouvoir et ordre. *Hermès*, 2(66), pp. 23-29.

Meyer, V. (2006). De l'utilité des recherches-actions en SIC. *Communication et*

organisation, (30). Disponible à : <http://communicationorganisation.revues.org/3455>

Millerand, F. (2012). La science en réseau. *Revue d'anthropologie des connaissances*, 6(1), pp. 163-190.

Nolin, J. (2010). Sustainable information and information science. *Information Research*, 15(2). Disponible à : <http://InformationR.net/ir/15-2/paper431.html>

OCDE. (2007). Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics. Disponible à : <https://www.oecd.org/fr/sti/sci-tech/38500823.pdf>

Onwuegbuzie, A. et Leech, N. (2007). A call for qualitative power analyses. *Quality & Quantity: International Journal of Methodology*, (41), pp. 105-121.

Prost, H. et Schöpfel, J. (2015). *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3 : Rapport final*. [Rapport de recherche] Lille 3. Disponible à : <http://hal.univ-lille3.fr/hal-01198379/document>

Strauss, A. (1985). Work and the Division of Labor. *The Sociological Quarterly*, (26), pp. 1-19.

Suchman, L. (2007). *Human-Machine reconfigurations. Plans and situated action*. New York : Cambridge University. 328 p.

Taylor, R. (1991). Information use environments. *Progress in Communication Sciences*, (10), pp. 217-255.

Wang, T. (2013). *Big data needs thick data*. Disponible à : <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/>

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Zacklad, M. (2007). Une théorisation communicationnelle et documentaire des TIC. Dans C Brossaud., et B. Rebert (dirs), *Humanités numérique 2 Socio-informatique et démocratie cognitive*. Paris : Hermès Science Publications, pp. 20-35.

[1] Le site web du projet est accessible à l'adresse suivante : <http://patrinum.espe-aquitaine.fr/>