



UNIL | Université de Lausanne
Ressources informationnelles et archives
bâtiment Unicentre bureau 214
CH-1015 Lausanne

Université de Lausanne

Archives des savoirs :
De la gestion des données *de* recherche vers une gestion
des données *pour* la recherche

Journées des archivistes des universités et hautes écoles suisses
16 octobre 2014 – UNIL

L'important n'est pas de prévoir l'avenir, mais de le rendre possible
Antoine de Saint-Exupéry

*Sachant nous rencontrer, nous nous donnons des pouvoirs
que nous ne pouvons manifester si nous sommes isolés.*
Albert Jacquard



Table des matières

| | |
|---|-----------|
| Abstract | 4 |
| Préambule | 5 |
| 1. Définitions | 5 |
| Qu'est-ce qu'une donnée ? | 5 |
| Un enregistrement de « faits donnés » | 6 |
| Des données traitées, dérivées, brutes, d'observation, expérimentales, etc. | 6 |
| Des données ouvertes pour un savoir libre | 6 |
| Des données du savoir scientifique | 7 |
| Une définition à huit facettes | 7 |
| 2. Pourquoi gérer c(s)es données de recherche ? | 8 |
| Des raisons contraignantes | 8 |
| Des raisons complémentaires | 9 |
| 3. Le cycle de vie de la donnée de recherche et ses enjeux | 10 |
| Des étapes et des tâches clés | 10 |
| L'emploi de métadonnées | 11 |
| L'élaboration d'un <i>data management plan</i> | 11 |
| 4. Qui sont les acteurs et quels sont leurs rôles ? | 12 |
| Le « bon » chercheur | 12 |
| Une communauté de compétences | 13 |
| 5. Quelles pratiques et expériences au sein du campus lausannois ? | 13 |
| Pas de politique générale explicite | 13 |
| Quelques règles pour les données liées aux inventions | 14 |
| Une Directive sur l'intégrité scientifique dans le domaine de la recherche | 15 |
| Un centre de compétences nationales pour les sciences sociales | 16 |
| Les humanités digitales | 16 |
| 6. Penser globalement et implémenter localement | 17 |
| Un programme de recherche des hautes écoles au niveau suisse | 17 |
| Simple, rapide, efficace | 18 |
| 7. Conclusion : une responsabilité partagée | 18 |
| Quelques références | 20 |
| Glossaire | 22 |

Abstract

The management of research data is vital for research on any kind: Data needs to be managed during the research process to facilitate current research and needs to stay exploitable afterwards to facilitate future research.

The European program *Horizon 2020* requires data management plans to be written for any research project financed by it. This will have a strong impact also on Swiss research, which receives a significant amount of European research grants.

Information professionals specialised in document and data management need to impose themselves as essential partners. They need to make sure that Swiss universities and other institutes of higher education strongly commit to this way of working, because this is where the trend of research information management and open research data is going.

For the University of Lausanne, no general policy has been put into place yet. Some guidelines mention the issue, but only indirectly.

On the campus of Lausanne, certain university partners like the Swiss Foundation of Research in Social Sciences (FORS) and its data services already offer technical and organisational solutions to manage research data. The local implementation of this foundation as well as its openness to federal and international institutions could make it a strategically interesting partner. A partnership with an institution like the EPFL and its Laboratory of Digital Humanities could equally be useful in the creation of a common policy.

Therefore, practical means are required. Furthermore, the policy creation is a shared responsibility and the fruit of collaboration between academic institutions, information and IT specialists, lawyers, etc.



Préambule

« *Must try harder* », tel est l'intitulé d'un article de la revue Nature¹ publié en mars 2012 qui s'étonne du nombre « choquant » d'articles, dans le domaine médical, pour lesquels les conclusions ne pouvaient être reproduites et donc vérifiables par d'autres chercheurs. La faute à des données insuffisamment documentées et gérées. L'article mentionne quelques mauvaises pratiques : références manquantes, contrôles incorrects, ajustements cosmétiques de données non déclarés, méthodes imprécises et incomplètes, mauvaise utilisation de statistiques, etc. Et de terminer par trois questions :

- Qui est responsable ?
- Pourquoi cela arrive-t-il ?
- Comment pouvons-nous stopper cela ?

Ces trois questions seront notre point de départ. Elles nous aideront à cerner la définition des données de recherche, véritables archives des savoirs en devenir, leur cycle de vie et les différents enjeux liés à leur gestion. Nous nous arrêterons sur ce qui est entrepris au sein l'Université de Lausanne et de son campus, sur les acteurs et les moyens nécessaires à mettre en place pour arriver à une (meilleure) prise en compte des données *de* recherche afin de garantir des données *pour* la recherche.

1. Définitions

Doit-on parler de données *de* recherche (*research data*) ou de données de *la* recherche ? Pour simplifier notre propos, on choisira de se « calquer » sur la pratique et la compréhension anglo-saxonne. C'est bien de données *de* recherche dont il est ici question.

Qu'est-ce qu'une donnée ?

De nombreux essais de définitions tentent de circonscrire cette notion. Challenge d'autant plus difficile qu'il convient de trouver un dénominateur commun pour de nombreux types de données (sociologiques, économiques, médicales, biologiques, construites de sons, d'images voire d'odeurs, etc.).

En archivistique, la donnée est souvent comprise comme une représentation formalisée de l'information, adaptée à la communication, l'interprétation ou le traitement, par informatique ou non.

¹ [Nature, « Must try harder », mars 2012, vol.483, p.509.](#)



Selon l'Université de Bristol², en Angleterre, les données de recherche se définissent comme :

« Les données, ou **unités d'information**, qui sont **créées** au **cours d'une recherche**, subventionnées ou non, et qui **sont organisées ou formatées** de telle sorte qu'elles soient **communicables, interprétables et adaptées à un traitement** souvent informatisé. »

On retrouve dans nos deux définitions les notions telles que l'unité d'information, le formalisme nécessaire, l'aspect de traitement et d'interprétation qui doit être possible afin de pouvoir utiliser cette information et la communiquer.

Un enregistrement de « faits donnés »

Certaines définitions mentionnent encore le caractère brut, ou primaire, nécessaire pour en faire une vraie donnée. C'est-à-dire qu'elles « ne sont pas encore le produit d'une analyse ou d'interprétation autres que de calcul »³. Les données ne sont ici que des **enregistrements de « faits donnés »** comme le mentionne Rémi Gaillard⁴.

Des données traitées, dérivées, brutes, d'observation, expérimentales, etc.

Pour ajouter à la nébuleuse, la littérature mentionne également les notions de données « traitées » ou « dérivées » entre lesquelles il est difficile de tracer une ligne claire.

Les données de recherche peuvent donc être des données brutes, des données traitées, des données dérivées, des données d'observation, des données expérimentales ou encore des données computationnelles ou de simulation. Souvent, un des points communs à toutes ces données est leur volumétrie importante (*Big data*). En sciences sociales, mais encore bien plus dans les sciences dites dures (médecine, biologie, géologie, etc.), le **Big data** caractérise les données de recherche de nombreux domaines scientifiques (génomique, astronomique, climat, etc.) qui comprennent des volumétries de l'ordre du téraoctet, voir du pétaoctet⁵.

Des données ouvertes pour un savoir libre

D'autre part, et on le verra avec la notion du cycle de vie des données de recherche, celles-ci devraient être accessibles sous certaines conditions à la communauté (à l'exception des résultats de recherche qui peuvent être protégés selon le principe des patentes, licences ou autres

² Citée dans le mémoire d'étude de Rémi Gaillard intitulé « De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? », ENSSIB, janvier 2014.

³ The Royal Society, Science as an open enterprise : summary report, The Royal Society, juin 2012, p.9, disponible à l'adresse : <https://royalsociety.org>. Cité dans le mémoire d'étude de Rémi Gaillard.

⁴ Rémi Gaillard, idem, p.17.

⁵ 1 To est égal à 1'000 Go et 1 Po vaut 1'000 TO, soit 1 000 000 000 000 000 d'octets.



brevets). Ce caractère d'ouverture se retrouve dans la notion d'**Open data**⁶, notion qui comprend notamment un autre terme anglo-saxon à la mode : l'*Open access* (cf. glossaire).

Les mouvements liés à l'ouverture des données reposent sur de multiples acteurs, principalement non gouvernementaux. Certains d'entre eux prônent notamment un **savoir libre**⁷ qui inclut toutes sortes de données, qu'elles soient scientifiques, historiques, géographiques ou autres.

Des données du savoir scientifique

Certains auteurs, issus des domaines archivistiques ou historiques, ne mentionnent pas explicitement la notion de données de recherche, mais parlent plutôt d'archives des savoirs, soit des **données issues du travail savant**⁸. Un groupe de recherche intitulé Archives des savoirs : enjeux scientifiques, universitaires et patrimoniaux s'est d'ailleurs constitué à l'Université de Genève autour de cette thématique⁹.

Ces archives des savoirs scientifiques, comme par exemple les manuscrits d'un Ferdinand de Saussure¹⁰, d'un Jean Piaget¹¹, d'un d'Albert Einstein¹² ou des cahiers de laboratoires, montrent une pensée en devenir, avec ses hésitations, ses doutes, ses avancées. Elles sont le reflet du travail quotidien du chercheur. Elles matérialisent en quelque sorte la pensée scientifique.

Ces archives sont des documents qui peuvent être transformés en données par numérisation. Elles deviennent ainsi le sujet d'étude de nouveaux domaines de recherche, comme celui des humanités digitales, et requiert une attention particulière car elles dynamisent, comme nous le verrons, la recherche d'aujourd'hui.

Une définition à huit facettes

Ces discussions autour du domaine de la gestion des données de recherche, et nous les avons volontairement abrégées, montrent les difficultés d'obtenir une définition précise et consensuelle, valable pour tous les domaines scientifiques.

Nous proposerons de considérer la donnée de recherche comme :

- Une unité d'information ;
- Créée, collectée ou rassemblée ;

⁶ Voir la philosophie d'ouverture des données de la Fondation nationale suisse pour la gestion des données de recherche (FORS) : <http://forscenter.ch/fr/about-us/philosophy/open-data-policy-at-fors/>

⁷ <http://opendefinition.org/od/francais/>

⁸ Jean-François Bert, Qu'est-ce qu'une archives de chercheur, OpenEdition Press, Encyclopédie numérique, Marseille, 2014. Disponible en OpenBook à l'adresse <http://books.openedition.org/oepp/438>

⁹ <http://www.unige.ch/rectorat/maison-histoire/recherche/archives-1/>

¹⁰ <https://www.ville-ge.ch/bge/evenements/expositions-virtuelles/saussure/introduction.html>

¹¹ <http://www.fondationjeanpiaget.ch>

¹² <http://www.alberteinstein.info/>

- Issu du travail savant ;
- Qui représente un enregistrement de faits objectifs ;
- Qui est organisée ou formatée pour être rendue communicable et interprétable ;
- Qui est susceptible de subir un traitement, informatisé ou non ;
- Dont la volumétrie peut être importante (*Big data*) ;
- Et qui est accessible, à terme, à la communauté (*Open data*).

2. Pourquoi gérer c(s)es données de recherche ?

Les nombreux guides, bonnes pratiques, plans et autres lignes directrices disponibles sur internet donnent des pistes sur le « *comment ?* » gérer c(s)es données, mais plus rarement sur le « *pourquoi ?* » les gérer.

Plusieurs raisons peuvent être évoquées pour tenter d'expliquer pourquoi la gestion des données de recherche est une nécessité.

Des raisons contraignantes

- C'est **scientifiquement incontournable**, car la recherche repose sur la capacité de vérifier et de reproduire¹³. La production, l'exploitation et l'archivage de données font parties d'une recherche intègre et responsable¹⁴.
- C'est **financièrement obligatoire**, car l'octroi de fonds, tant publics¹⁵ que privés, est de plus en plus souvent conditionné à la mise en place d'une gestion des données du projet de recherche.
- C'est **techniquement indispensable**, car la recherche d'aujourd'hui nécessite des technologies computationnelles et informatiques. De plus, la gestion de volume de données toujours plus important nécessite une réponse organisationnelle autre qu'une simple accumulation quantitative d'espaces de stockage.
- C'est **démocratiquement essentiel**, car les citoyens, les groupements civils ou les organismes non gouvernementaux, demandent plus de transparence et un contrôle sur les informations (données) qui soutiennent les conclusions scientifiques.
- C'est **juridiquement nécessaire**, car certaines lois, notamment dans le domaine archivistique et de l'information, interdisent les destructions sauvages d'informations

¹³ <https://royalsociety.org/policy/projects/science-public-enterprise/Report/>

¹⁴ Voir à ce sujet le guide édité par le CNRS : <http://responsable.unige.ch/main/directives-aux-chercheurs/directives-du-cnrs-france.html>

¹⁵ Voir le projet européen H2020 : <http://ec.europa.eu/programmes/horizon2020> et notamment les *Guidelines on Data Management in Horizon 2020* du 16 décembre 2013 : http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

produites tant par des services étatiques que par des organismes privés et obligent les producteurs à les organiser et à les gérer¹⁶.

Des raisons complémentaires

- C'est **stratégiquement populaire**, car c'est dans l'air du temps et que le partage de données de recherche ramène en popularité et en notoriété au chercheur et à son institution de rattachement. C'est également un important critère pour l'avancement d'une carrière et la reconnaissance par ses pairs.
- C'est **éditorialement recommandable**, car les journaux scientifiques, revues et autres maisons d'édition ont intérêt à rendre accessibles, utilisables et vérifiables les données qui ont permis d'aboutir à la publication en question. Cette pratique renforce la garantie scientifique et donc le sérieux professionnel de l'étude et de la maison d'édition qui la publie.
- C'est **politiquement souhaitable**, car c'est reconnaître le potentiel de l'*open research data* et l'*open science* comme moyen de renforcer les bases scientifiques et l'innovation d'une nation.
- C'est **commercialement intéressant**, car l'accès à certains types de données par d'autres publics peut déboucher sur le développement de nouvelles prestations.
- C'est **intellectuellement stimulant**, car le croisement de données provenant de champs d'études différents ouvre des perspectives infinies de nouvelles recherches¹⁷. De même, une bonne gestion offre une meilleure exploitation des données existantes et une réduction des besoins en matière de récoltes de nouvelles données.
- C'est **historiquement responsable**, car les données d'hier peuvent devenir des données essentielles pour la science de demain. Elles servent également à tracer l'histoire d'une recherche ou d'un domaine de recherche.

La gestion des données de recherche et leur accessibilité offre donc de multiples avantages à différents niveaux, tant à l'ensemble des acteurs qu'à la société en général. On peut alors se poser la question de savoir pourquoi cela ne se fait pas naturellement puisque l'intérêt commun de tous est de le faire ? L'évidence n'est pas forcément source d'action et il faut bien reconnaître que la mise en place d'une telle politique n'est pas chose aisée, d'autant plus qu'il convient de prendre en considération l'entier du cycle de vie des données.

¹⁶ La quasi totalité des cantons suisses disposent de législation en matière de gestion des archives et d'accès à l'information, bien souvent couplée avec la protection des données. Il en est de même au niveau fédéral. Voir les principales références légales qui s'appliquent pour l'UNIL : <http://www.unil.ch/uniris/home/menueinst/references/lois-reglements-et-normes.html>

¹⁷ Croiser des données médicales avec des données sociologiques ou économiques peut offrir une nouvelle cartographie des quartiers d'habitation où l'obésité est la plus importante. Voir notamment l'article : <http://www.arcinfo.ch/fr/suisse/geneve-correlation-entre-le-poids-et-le-lieu-de-domicile-566-1269446>

3. Le cycle de vie de la donnée de recherche et ses enjeux

Schématiquement, le cycle de vie d'un document ou d'une donnée peut être représenté par la période qui s'étend de la conception du document ou de la donnée, à son utilisation jusqu'à sa destruction ou sa conservation pour raison historique ou scientifique.

Des étapes et des tâches clés

Selon le *UK Data Archives*¹⁸, spécialisé dans les données de recherche en sciences sociales, le cycle de vie des données de recherche comporte les étapes suivantes :

- La **création** des données
 - Définit le référentiel de la recherche
 - Mise en place du *Data management plan*
 - Localise les données
 - Collecte les données
 - Décrit les données

- Le **traitement** des données
 - Saisit les données, les digitalise, les traduit, les transcrit
 - Contrôle, valide, nettoye les données
 - Anonymise et décrit les données
 - Gère et stocke les données

- L'**analyse** des données
 - Interprète
 - Produit des données dérivées
 - Produit des *output* de recherche
 - Prépare les données pour la préservation

- La **préservation** des données
 - Migration dans un format pérenne
 - Création des métadonnées
 - Documentation des données
 - Archivage des données

- L'**accessibilité** et le **partage** des données
 - Distribution/partage des données
 - Contrôle des accès (ou non)
 - Etablissement des protections (copyright vs. copyleft)
 - Promotion des données via une plateforme internet ouverte

¹⁸ <http://data-archive.ac.uk/create-manage/life-cycle>

- La **réutilisation** des données
 - Suivi et revues des recherches
 - Nouvelles recherches à partir des données
 - Croisement des données avec d'autres données issues d'autres domaines

La gestion des données de recherche doit prendre en compte l'entier du cycle de vie de la donnée et non pas uniquement l'une ou l'autre de ses phases. La préservation et l'accessibilité des données ne se décident pas au moment de l'archivage, mais elles sont affaire d'anticipation. Pour pouvoir gérer les données tout au long de leur cycle de vie, celles-ci doivent être décrites dès leur création au moyen de métadonnées.

L'emploi de métadonnées

A chacune de ces étapes, correspondent donc des tâches ou actions nécessaires à la bonne réalisation de l'étape en question. L'ajout de métadonnées, c'est-à-dire de données sur les données, est ici essentiel. Les métadonnées renseignent la donnée numérique au moyen de descripteurs. Ceux-ci peuvent se répartir selon quatre domaines : le **contenu**, le **contenant**, le **contexte** de création et l'**administration technique** de la donnée. La norme ISO 23081 (*Dublin core*) peut offrir les bases nécessaires à cette description¹⁹.

L'élaboration d'un data management plan

La délimitation des corpus de données prévus pour un archivage définitif peut varier selon les domaines concernés et les habitudes disciplinaires. Le plan de gestion des données (*data management plan*) documente formellement comment les données seront gérées dans le cadre d'un projet scientifique. Il tient compte du domaine de recherche dans lequel le projet s'inscrit et des particularités dudit projet.

Un plan de gestion des données²⁰ est essentiel. Il est même parfois obligatoire pour s'assurer le financement de certains projets²¹. Il permet de sensibiliser le chercheur à cette problématique tout en lui fournissant un guide de gestion de ses données. Celui-ci doit comprendre :

- Des informations sur les données et leurs formats (type de données, logiciels utilisés, modèles, etc.) ;
- Les métadonnées du projet et des données collectées ;
- La politique d'accès, de partage et de réutilisation ;
- La préservation (ou non) à long terme de ces données, selon quel format et modalités ;
- Le budget disponible pour gérer les données ainsi que les composantes techniques ; etc.

¹⁹ Voir notamment le document [Métadonnées des documents et répertoires électroniques](#) développé par UNIRIS.

²⁰ A titre d'exemple : UK Data Archives, [Managing and sharing data, Best practice for researchers](#), may 2011.

²¹ Voir le Programme européen Horizon 2020 pour la recherche et l'innovation qui exige notamment un plan de gestion des données pour toutes recherches financées par le programme : <http://ec.europa.eu/programmes/horizon2020/>. Voir notamment les [Guidelines on Data Management in Horizon 2020](#), du 16 décembre 2013

4. Qui sont les acteurs et quels sont leurs rôles ?

Une saine gestion des données de recherche est le fruit d'un assemblage d'intérêts, communs ou non, autour d'un projet susceptible d'amener une plus-value à la société prise dans son ensemble. Le défi est de donner corps à cette communauté d'intérêts multiples et variés, et de rassembler ces différents acteurs autour de cette gestion.

Le « bon » chercheur

Pour une bonne gestion des données de recherche et leur partage, le premier concerné est le chercheur lui-même, responsable et maître d'œuvre de la recherche. Comme producteur de données, lui seul a la capacité de déterminer la réelle valeur des données collectées. De nombreuses publications ou autres codes de déontologie font état des bonnes pratiques du chercheur dans le cadre de sa recherche.

Comme nous l'avons déjà mentionné, il existe des listes qui sont censées permettre au chercheur de gérer ses données. Une de ces listes²² mentionne par exemple :

- Que le fait de **partager les données pousse la science en avant** et qu'il convient donc de rendre accessibles des données organisées et de les partager sur des plateformes existantes. Cet « impératif volontaire » répond aux institutions internationales de financement, poussées notamment par les lignes directrices de l'OCDE²³ ou encore par les programmes Européens tel que H2020²⁴. La gestion des données répond également aux financeurs privés et aux journaux responsables des publications, qui estiment que leur financement achète le résultat, mais également la recette ayant permis d'atteindre ce résultat.
- Que toute recherche doit prévoir un **plan de gestion des données** pensé à l'avance afin de créer des données de haute qualité, durables qui peuvent (et doivent) être partagées.
- Que les données doivent être **documentées**, faciles à comprendre, facile à utiliser et bien formatées et organisées.
- Que les données doivent également être **conservées** de façon **sécurisée** et **réutilisable**. Qu'elles doivent pouvoir être transmises de façon cryptée, au besoin.
- Que le chercheur doit être attentif à la transmission des données sensibles et confidentielles. Il doit se doter de **règles éthiques**, en accord avec la législation en vigueur et tenir compte du consentement des intéressés.
- Que les notions de copyright doivent être prises en compte, sans pour autant limiter le partage des données.

²² UK Data Archives, [Managing and sharing data, Best practice for researchers](#), may 2011.

²³ Rapport général de l'OCDE de 2007 : [Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics](#).

²⁴ Programme européen pour la recherche et l'innovation : <http://ec.europa.eu/programmes/horizon2020/>. Voir notamment les [Guidelines on Data Management in Horizon 2020](#), du 16 décembre 2013

On le constate aisément, les tâches décrites dans ce genre de liste s'avèrent difficile à mettre en place pour un chercheur tant il doit disposer de compétences en matière juridique, informatique, de communication, de préservation, de stratégie voir d'éthique pour être qualifié non pas de bon chercheur, mais de chercheur « tout court ».

Une communauté de compétences

On l'aura compris, une bonne gestion des données de recherche passe par la mise en place d'une communauté de compétences au service d'un même but. Il s'agit donc de réunir les métiers :

- De l'**archiviste** ou du **records manager** pour les aspects organisationnels, de description, de sélection et de préservation des données ;
- De l'**informaticien** pour les questions technologiques liées aux plateformes de stockage et de diffusion ;
- Du **juriste** pour les aspects légaux de copyrights, de distributions et d'accessibilité des données ;
- Du responsable de **communication** pour les questions liés à la mise en valeur des données et de la recherche ;
- Des **organes de direction** de l'institution pour offrir un cadre général, une stratégie et le code déontologique qui s'applique aux recherches menées dans l'institution en question ;
- De **formateurs** pour former et sensibiliser les chercheurs, nouveaux et anciens, notamment via les programmes doctoraux²⁵.

5. Quelles pratiques et expériences au sein du campus lausannois ?

Pas de politique générale explicite

Il n'existe pas de politique générale concernant explicitement la gestion des données de recherche au sein de l'UNIL, ni d'ailleurs dans les hautes écoles et universités suisses, à notre connaissance.

Le service des ressources informationnelles et archives (UNIRIS) de l'UNIL ne conserve que quelques rares fonds d'archives qui proviennent de professeurs. Ces documents issus de la recherche sont traités comme des fonds privés et font l'objet de conventions de don.

A noter qu'il existe une politique de records management et d'archivage pour une gouvernance informationnelle, validée par la Direction de l'UNIL le 30 juin 2014²⁶. Cette politique vise la

²⁵ <http://www.unil.ch/researcher/>

²⁶ <http://www.unil.ch/uniris/home/menuguid/a-telecharger/documents-de-reference.html>

gestion des données administratives de l'institution. Elle rappelle le cadre normatif et réglementaire en vigueur. Elle explicite les objectifs, les moyens opérationnels et les rôles des différents acteurs de l'UNIL dans la gestion de leurs documents et de leurs archives.

Concernant les données issues de la recherche, cette nouvelle politique mentionne que :

« La gestion des données issues de l'activité académique (recherche et contenu des enseignements) n'est pas prise en compte dans ce document, mais devra l'être à terme. »

Cette phrase donne deux indications : d'une part, la problématique des données de recherche est officiellement mentionnée et ce, pour la première fois, au niveau d'une politique validée par la Direction de l'UNIL. D'autre part, ces données ne sont pas, pour l'heure, gérées, mais devront l'être à terme.

Quelques règles pour les données liées aux inventions

La gestion de certains types de données de recherche est toutefois mentionnée dans le cadre de la Directive du Conseil de Direction UNIL-CHUV du 2 décembre 2009 relative aux contrats et à la valorisation de la recherche²⁷. Celle-ci fixe les procédures applicables en matière de valorisation financière des résultats de la recherche à l'UNIL et au CHUV, quant à la conclusion de contrats liés à la recherche et quant à l'obligation d'annoncer les inventions et tout autre résultat de la recherche potentiellement valorisable. Est soumis à cette Directive, tout collaborateur employé par l'UNIL ou par le CHUV, avec ou sans charge d'enseignement, quelle que soit la source de son salaire ou son taux d'activité.

L'article 5 de ladite Directive, qui traite de la procédure d'annonce d'une invention, précise que les collaborateurs de l'UNIL ou du CHUV qui ont développé une invention doivent l'annoncer en prenant toutes les mesures nécessaires pour la protéger, notamment :

qu'« ils tiennent des cahiers de laboratoires et documentent leurs expériences selon les règles de l'art ».

On retrouve ce souci de conserver et documenter la recherche dans le modèle 2009 de convention de transfert (*Material Transfer Agreement - MTA*)²⁸ qui stipule que :

« RECIPIENT shall, in accordance with its established practice, keep complete and accurate accounts, notes, data and records of the Research. ».

Cette précision concernant les documents et les données de recherche ne figure toutefois plus dans la version 2014 du modèle *MTA*²⁹.

²⁷ Disponible sur le [site internet du PACTT](#), le bureau de valorisation de la recherche.

²⁸ http://www.pactt.ch/files/template_mta_unil_academic_out.doc

²⁹ http://www.pactt.ch/pactt_home/pactt-contracts-and-agreements/pactt-documents.htm

Cette Directive sur la valorisation de la recherche à l'UNIL et au CHUV ne touche à la problématique des données de recherche que de façon très limitée.

Une Directive sur l'intégrité scientifique dans le domaine de la recherche

Outre sa charte³⁰ qui mentionne le savoir critique comme l'une de ses valeurs fondamentales et un code de conduite pour les doctorants³¹, l'Université de Lausanne a édicté une Directive en matière d'intégrité scientifique³². Son point 2.4 traite de l'accès aux données et précise qu'il :

« convient de documenter les données relatives à l'avancement de la recherche et les résultats des expériences originelles ("données de base") d'une manière claire, complète et précise, afin d'exclure autant que possible tout dommage, toute perte ainsi que toute manipulation ciblée, selon les règles établies dans chaque discipline. Il en va de même pour les données électroniques (sauvegardes des données sur CDROM, etc.) ainsi que pour la documentation originale des projets de recherche indiquée dans le protocole de recherche. »

La formule est bien trop générale pour assurer la gestion de l'ensemble de la problématique. De même, une « sauvegarde des données sur CDROM » n'est pas suffisante et ne saurait répondre aux besoins de conservation à long terme et à l'archivage.

La Directive précise encore que c'est le chef de projet :

« qui doit veiller à ce que les données de base produites dans le cadre de la recherche soient conservées en sécurité pendant au moins dix ans après l'achèvement de l'étude. En cas de départ de l'institution, il doit s'assurer d'une conservation appropriée des données. »

Sans pour autant fournir les moyens techniques, ni la recette pour assurer une « conservation appropriée ». Les protocoles de laboratoire ainsi que les publications qui peuvent en résulter doivent « être suffisamment documentés pour que d'autres chercheurs puissent reproduire les résultats obtenus ».

Au chapitre des manquements à l'intégrité, sont susceptibles de constituer des infractions :

« La falsification intentionnelle de données de base (...) ; La suppression de données de base consignées, avant l'expiration du délai de conservation prescrit (...) ; La dissimulation de données ; (...) »

L'absence d'une bonne gestion des données peut très bien arriver au même résultat, sans pour autant être intentionnelle. La Directive, qui se veut générale, manque de précisions et aucun moyen n'est mis à disposition des chercheurs pour leur permettre de gérer adéquatement leurs données.

³⁰ <http://www.unil.ch/central/home/menust/organisation/les-documents-officiels/charte-unil.html>

³¹ Ce *code of practice for the doctorate* renvoi les futurs doctorants vers les éventuels codes éthiques et autres régulations qui existent concernant notamment les problématiques de plagiat ou de *data collection*. Il ne mentionne pas les données de recherche.

³² Voir la directive de la Direction 4.2 : [Intégrité scientifique dans le domaine de la recherche et procédure à suivre en cas de manquement à l'intégrité.](#)



A noter que le CNRS, via son Comité d'éthique, a publié en juillet 2014 un guide³³ destiné à encadrer l'activité de recherche intitulé : « Promouvoir une recherche intègre et responsable ». Un chapitre traitant de la production, de l'exploitation et de l'archivage des données y a été notamment intégré. Un modèle similaire pourrait être suivi à Lausanne.

Un centre de compétences nationales pour les sciences sociales

Depuis quelques années, l'UNIL abrite et finance en partie la Fondation suisse pour la recherche en science sociale (FORS)³⁴, centre de compétence nationale dont le mandat est de :

- produire des données dans le cadre d'enquêtes nationales et internationales ;
- distribuer des données destinées à l'analyse secondaire ;
- développer la recherche empirique, en particulier les méthodologies d'enquête ;
- conseiller les scientifiques et les milieux intéressés en Suisse et à l'étranger.

Son service de données et d'information sur la recherche (DARIS) a comme objectif de promouvoir une utilisation plus large et plus efficace des données. Il développe notamment des outils, des bonnes pratiques et des compétences pour aider les chercheurs à gérer leurs données³⁵. DARIS met à disposition un certain nombre de données issues de la recherche et permet aux chercheurs de déposer leurs données, de documenter facilement leurs projets de recherche et d'assurer ainsi leur partage selon diverses modalités tout en les conservant à long terme.

De manière générale, FORS défend l'idée selon laquelle les données de sciences sociales issues de projets de recherche financés par les fonds publics devraient être³⁶ :

- mises à disposition des chercheurs n'appartenant pas à l'équipe de recherche originale ;
- faciles à trouver et gratuitement accessibles ;
- archivées au sein d'une institution spécialisée et en suivant des standards internationaux.

La riche expérience et les compétences éprouvées en matière de gestion de données de recherche de FORS gagneraient à être (*re*)connues au sein des institutions universitaires suisses.

Les humanités digitales

Egalement sur le campus lausannois et partenaire de l'UNIL, le Collège des humanités³⁷ de l'École polytechnique fédérale (EPFL), se profile dans le domaine de gestion des données numériques de masse. Son laboratoire des humanités digitales³⁸ développe de nouvelles

³³ http://www.cnrs.fr/comets/IMG/pdf/guide_promouvoir_une_recherche_inte_gre_et_responsable_8septembre2014-2.pdf

³⁴ <http://forscenter.ch/fr/>

³⁵ <http://forscenter.ch/fr/service-de-donnees-et-d-information-sur-la-recherche/gestion-de-donnees/>

³⁶ <http://forscenter.ch/fr/about-us/philosophy/open-data-policy-at-fors/>

³⁷ <http://cdh.epfl.ch/digital>

³⁸ <http://dhlabs.epfl.ch/>



approches computationnelles. Il innove notamment sur la mise en commun des données issues de la dématérialisation à grande échelle d'archives et leur gestion³⁹. A noter que l'UNIL dispose également de son laboratoire de cultures et humanités digitales (LADHUL)⁴⁰.

L'ensemble de ce savoir-faire pourrait servir d'incubateur au développement d'une politique concertée sur la gestion des données de recherche.

6. Penser globalement et implémenter localement

Seul un cadre global développé en partenariat au niveau des hautes écoles et des universités suisses, peut permettre la mise en place d'une réelle politique commune en matière de gestion des données de recherche.

Pensée globalement, cette politique devrait être implémentée localement par les différents acteurs en présence : universités, laboratoires de recherche, instituts, écoles, chercheurs. Les moyens de mise en place d'une telle politique doivent être fournis en adéquation avec les besoins des chercheurs. Ces derniers doivent être étroitement associés à la constitution de la politique générale dont la mise en application reposera principalement sur leurs épaules.

Un programme de recherche des hautes écoles au niveau suisse

Une solution sortira peut-être du programme CUS 2013-2016 P2 « Information scientifique : accès, traitement et sauvegarde »⁴¹ lancé par la Conférence des Recteurs des universités suisse en 2014. Son objectif, d'ici à 2020, est que les enseignants, les chercheurs ainsi que les étudiants de la place scientifique suisse disposent d'instruments adéquats pour accéder et gérer de manière optimale l'information scientifique numérique.

La vision est de regrouper et de développer les efforts que les hautes écoles déploient actuellement de manière dispersée pour mettre à disposition et traiter des informations scientifiques.

L'un des axes du programme vise à établir des processus d'accès aux données de recherche unifiés par-delà les frontières institutionnelles ainsi que des normes pour leur traitement et leur archivage. Certains projets devraient donc se pencher sur la problématique de gestion des données de recherche⁴².

³⁹ Voir leur projet de numérisation des archives de Venise : <http://actu.epfl.ch/news/traveling-to-the-venice-of-the-doges-with-your-f-5/>

⁴⁰ <http://www.unil.ch/ladhul/home.html>

⁴¹ <http://www.crus.ch/information-programmes/projets-programmes/isci.html?L=1>

⁴² Le projet *Research Data Life-Cycle Management : From Pilot Implementations to National Service Concept*, lancé par l'Université de Genève en partenariat avec l'EPFL, l'ETHZ, la HEG-GE, Switch, l'Université de Bâle et de Zurich a malheureusement été rejeté en juin 2014. Voir : <http://www.crus.ch/dms.php?id=29371>



Simple, rapide, efficace

Des lignes directrices simples, une mise en place rapide et un système informatique de gestion efficace sont le garant d’une appropriation effective des nouvelles pratiques par les chercheurs. Dans tous les cas, la recherche scientifique ne doit pas être entravée ou freinée par cette nouvelle politique et les pratiques de gestion qu’elle suppose.

7. Conclusion : une responsabilité partagée

Aux niveaux anglo-saxon et européen, la gestion des données de recherche est un passage obligé si le chercheur veut pouvoir mener sa recherche. L’obligation de rédiger des plans de gestion des données dans le cadre des recherches financées par le programme H2020⁴³ aura des conséquences décisives en la matière, notamment pour la recherche suisse qui obtient de nombreux financements européens.

Les métiers de l’information (records manager, archivistes, bibliothécaires), spécialisés dans la gestion et la documentation des données et documents, doivent s’imposer progressivement comme des partenaires indispensables au service de la formation et de la recherche⁴⁴. Les professionnels de l’information doivent presser les universités et les hautes écoles suisses pour qu’elles s’engagent résolument dans cette voie, car c’est à leur échelle que se prépare la gestion et l’ouverture des données de recherche.

Pour l’UNIL, il n’existe pas actuellement de politique générale explicite en la matière. Certaines directives abordent toutefois cette problématique de manière détournée.

Sur le campus lausannois, certains partenaires de l’Université comme la Fondation suisse pour la recherche en sciences sociales et son service de données offrent déjà des solutions techniques et organisationnelles pour gérer les données de recherche. L’implémentation locale, au sein d’une université comme l’UNIL, et son caractère fédéral ouvert à la communauté scientifique suisse et internationale, place la fondation à un carrefour stratégique. De même, le développement de partenariats avec les laboratoires des humanités digitales permettrait de forger une politique commune.

La mise en place d’une politique de gestion des données scientifiques doit s’accompagner des moyens pratiques de la réaliser. Elle est une **responsabilité partagée** et le fruit d’une collaboration entre institutions académiques et d’une complémentarité de compétences entre

⁴³ Voir note 12.

⁴⁴ En Suisse, dans le domaine de l’*Open government data* (OGD), ce sont par exemple les Archives fédérales et leurs partenaires qui exploitent le portail pilote visant à donner accès, de manière centralisée, aux données ouvertes de l’administration (<http://opendata.admin.ch/fr>).



chercheurs, spécialistes de l'information, informaticiens, juristes, etc. Une telle politique doit se mettre au service de la recherche pour que la gestion des données *de* recherche deviennent une gestion des données *pour* la recherche.



Quelques références

- Rémi Gaillard, De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? », Mémoire d'étude, ENSSIB, janvier 2014. Document qui offre un bon panorama des enjeux et questions sous-jacentes ainsi que de nombreux exemples d'application.
- Rapport de 2012 de la Royal Society de Londres intitulé *Science as an open entreprise*. Plaidoyer pour l'ouverture des données de recherches sous forme de 10 recommandations. Intéressant comme cadre général à mettre en place pour gérer les données. L'objectif avoué : que toute la littérature soit en ligne, que l'ensemble des données le soit également, et que les deux soient interopérables : <http://royalsociety.org/policy/projects/science-public-enterprise/report/>
- Rapport général de l'OCDE de 2007 : Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics : <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>
- Un site d'information très riche sur la problématique des données de recherche en général : <http://www.donneesdelarecherche.fr/>
- Trois exemples de sites de données de recherche et d'outils : <http://datadryad.org/> ; <http://dataverse.scholarsportal.info/dvn/> ; <http://zenodo.org/>
- Le CINES - Centre Informatique National de l'Enseignement Supérieur en France (www.cines.fr) ou le JISC en Grande-Bretagne (<http://www.jisc.ac.uk/>)
Présentation du projet : Archivage intermédiaire de données scientifiques : les fonctionnalités ISAAC - information scientifique archivée au CINES. Très intéressant en matière d'organisation de la gestion des données (étapes de la vie d'un projet d'archivage) : <http://www.cines.fr/spip.php?rubrique369>. Voir notamment la présentation : <http://www.cines.fr/spip.php?rubrique384>
- Au niveau européen, voir notamment :
 - Le projet EUDAT (<http://www.eudat.eu/>), dédié à la construction d'une infrastructure de préservation des données ou *OpenAIREplus* (<http://www.openaire.eu/>) dont l'objectif est de donner de la visibilité aux données en les reliant à des publications et à des informations sur les projets de recherche dont ils sont issus.
 - Le groupe de travail sur les données de recherche du *Science Europe*, association de l'*European Research Funding Organisation* (RFO) : <http://www.scienceurope.org/policy/working-groups/Research-Data>

Glossaire

Archivage

Démarche d'organisation qui a pour objectif d'identifier, de mettre en sécurité et de maintenir disponibles l'ensemble des informations qui engagent une institution ou un organisme vis-à-vis de tiers ou de son activité future et dont le défaut représenterait un risque. Il ne s'agit pas ici du backup ou de sauvegarde qui caractérisent uniquement une copie temporaire de fichiers dans le but d'en prévenir la perte ou la destruction accidentelle.

Conservation/pérennisation

Ensemble des opérations techniques qui permettent de maintenir dans le temps des objets documentaires (quel que soit leur support), de préserver leur intégrité et de garantir l'accès à leur contenu.

Cycle de vie

Période qui s'étend de la conception d'un document ou d'une donnée, à son utilisation jusqu'à sa destruction ou sa conservation pour raison historique ou autres. Ainsi, tout document ou donnée passe par une ou plusieurs périodes caractérisées par la fréquence et le type d'utilisation qui en est faite.

Donnée

Représentation formalisée de l'information, adaptée à la communication, l'interprétation ou le traitement.

Données ouvertes (open data)

Données numériques d'origine publique ou privée, mise à la disposition du plus grand nombre selon une méthodologie bien spécifique. Les caractéristiques des données ouvertes publiques (*open government data*) sont l'intégrité, l'originalité (données originales), l'immédiateté, l'accessibilité, la lisibilité par machine, la non discrimination (disponibles sans enregistrement préalable), non propriétaires, disponible en licence libre, disponible de manière permanente et gratuite⁴⁵.

Libre accès (open access)

Le libre accès comprend la mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers un texte, un document ou une donnée et s'en servir à toute fin légale, sans barrière financière ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution, et le seul rôle du copyright dans ce domaine devrait être de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités⁴⁶.

⁴⁵ <http://fr.opendata.ch/manifeste/>

⁴⁶ Budapest open access initiative (2002), premier rassemblement historique, fondateur du mouvement « libre accès » : <http://www.budapestopenaccessinitiative.org/fr/index>

