



**HAL**  
open science

## Sciences Po. Une histoire de données

Elise Chapoy, Laurent Lesnard, Odile Gaultier-Voituriez, Emilie Groshens,  
Cynthia Pedroja, Anita Beldiman-Moore

► **To cite this version:**

Elise Chapoy, Laurent Lesnard, Odile Gaultier-Voituriez, Emilie Groshens, Cynthia Pedroja, et al..  
Sciences Po. Une histoire de données. [Rapport de recherche] Sciences Po. 2020, pp.91. hal-03612928

**HAL Id: hal-03612928**

**<https://sciencespo.hal.science/hal-03612928>**

Submitted on 18 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The image features a central square with a white background and a thick, golden-brown border. This square is surrounded by several thin, golden-brown lines that radiate outwards from its corners and edges, creating a starburst or network-like effect. The lines vary in length and angle, some extending towards the top and others towards the bottom of the page.

**Sciences Po**

**Une histoire  
de données**

---

Rédaction : Janvier 2016  
Mise à jour : Juin 2020

## Rédacteurs

- Élise Chapoy – DRIS/SAU-ARO – responsable
- Laurent Lesnard – CDSP – directeur
- Odile Gaultier-Voituriez – CEVIPOF – représentante coordination Archives
- Émilie Groshens – CDSP – chargée d'études pour le projet beQuali
- Cynthia Pedroja – DRIS/SAU-ARO – cheffe de projet gestion des données de la recherche
- Anita Beldiman-Moore – DRIS/SAU – responsable (*pour la partie mise à jour*)

## Mots clefs

Données de la recherche, Plan de gestion des données, Entrepôt des données, Cycle de vie des données, Archives de la recherche, *Data Management Plan* (DMP)

# Préambule

Le document ci-après présente un état de l'art en janvier 2016 de la gestion des données de la recherche tant à Sciences Po qu'en France et à l'international. Après deux années de travaux qui ont mobilisé les services et entités concernées au sein de Sciences Po<sup>1</sup> et dont la dernière s'est déroulée sous l'égide d'une cheffe de projet dédiée à la Direction des ressources et de l'information scientifique (DRIS), cette note pose les éléments de contexte et fait le tour des initiatives et des projets déployés à l'époque. A partir de l'éclairage donné par ces exemples, elle a abouti à une proposition de programme fonctionnel, technique et organisationnel au sein de l'établissement.

La mise en oeuvre qui s'en est suivie entre 2016 et 2019 s'est appuyée sur quelques projets pilotes qui ont permis de dégager un début typologie des données et, en s'inspirant du modèle H2020, de rédiger un modèle de plan de gestion des données (DMP) institutionnel en complément des modèles déployés par l'Agence nationale de la recherche (ANR), H2020 et le Conseil européen de la recherche (ERC)<sup>2</sup>, modèle intégré depuis sur OPIDoR. Ces mêmes projets pilotes ont servi à tester l'entrepôt de données Dataverse dans la perspective d'ouvrir un entrepôt institutionnel en lien avec l'archive ouverte de Sciences Po (SPIRE) et les obligations induites par le Règlement général sur la protection des données (RGPD) du 25 mai 2018. Parallèlement, une typologie puis une bibliographie sur les différents aspects du sujet<sup>3</sup> ont été réalisées dès 2016 grâce au travail de Laurène L'Hermite, stagiaire à la DRIS<sup>4</sup>. Cette bibliographie fait depuis l'objet d'une veille partagée avec le Centre pour la communication scientifique directe (CCSD)<sup>5</sup>. Sur le plan organisationnel, en appui à la cheffe de projet, un petit groupe de bibliothécaires a été formé afin de jeter les bases d'une palette de services aux chercheurs dans ce domaine et un guide d'information et de ressources a été publié sur le site de la bibliothèque<sup>6</sup>. Enfin, sous l'égide de Nicolas Sauger, directeur du Centre de données socio-politiques (CDSP), et avec l'appui de la Direction des systèmes d'information (DSI), de la déléguée à la protection des données (DPO) et du médialab de Sciences Po, une formation de deux jours et demi a été proposée aux doctorants, parallèlement à une session de sensibilisation aux chercheurs.

---

<sup>1</sup> A l'époque le CDSP avec Laurent Lesnard, Emilie Grohens, Guillaume Garcia et Sarah Cadorel, la DRIS avec Cynthia Pedroja, Delphine Lereculeur et Goulven Le Brecht et le CEVIPOF et Mission archives avec Odile Gaultier Voituriez.

<sup>2</sup> <https://dmp.opidor.fr/>

<sup>3</sup> Enjeux scientifiques, questions juridiques et éthiques, politiques publiques, enjeux techniques de stockage, référencement et citation et guides et compétences professionnelles

<sup>4</sup> Stage du 11 avril au 14 juin 2016

<sup>5</sup> [https://www.zotero.org/groups/708363/donnes\\_de\\_la\\_recherche\\_-\\_veille\\_collaborative/items/collectionKey/KN6NP7NN](https://www.zotero.org/groups/708363/donnes_de_la_recherche_-_veille_collaborative/items/collectionKey/KN6NP7NN)

<sup>6</sup> <https://sciencespo.libguides.com/donnees-de-la-recherche>

Ces réalisations locales s'inscrivent dans une évolution nationale qui a culminé avec le Plan national pour la science ouverte annoncé le 4 juillet 2018<sup>7</sup>.

Un addendum<sup>8</sup> aux annexes de ce document permet d'illustrer une partie de ces concrétisations et de constater la mesure dans laquelle elles s'inscrivent ou s'écartent du programme envisagé en fonction des évolutions du contexte.

---

<sup>7</sup> <https://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html> (vu le 19/12/2019)

<sup>8</sup> Note "2017, l'expérimentation grandeur nature" ; DMP ; Syllabus de la formation doctorale ; Page d'accueil du Guide sur les données de la recherche ; Note descriptive sur l'entrepôt de données

---

# SciencesPo

**Les données de la recherche à Sciences Po**

**Une année d'exploration : synthèse et perspectives**

**Date de rédaction : janvier 2016**

**[Mise à jour : juin 2020]**

## Introduction

Tout au long de l'année 2015, les membres du groupe de travail Données de la recherche à Sciences Po (GT Données - GTD) ont réalisé une série d'entretiens afin d'étudier le contexte national et européen du champ dénommé "gestion des données de la recherche".

Parallèlement, une série de rencontres a eu lieu avec les laboratoires de Sciences Po afin de pouvoir réaliser une typologie des données produites et utilisées par les chercheurs et d'identifier les besoins en termes de gestion. De ce travail découle la présente note qui constitue l'amorce d'une feuille de route pour l'institution. L'objectif du projet est d'articuler les dispositifs, les initiatives, les outils, les services dans cet environnement tentaculaire. Ainsi, après avoir défini ce que l'on entend par "gestion des données de la recherche", le groupe s'est efforcé de poser les premiers constats et de proposer quelques pistes de réflexion pour le développement d'une infrastructure de services et d'outils à destination de la communauté de recherche de Sciences Po.

Ce projet est piloté par la Direction scientifique (DS), la Direction des ressources documentaires et de l'information scientifique (DRIS) et la Direction des systèmes d'information (DSI). Le Centre de données socio-politiques (CDSP) et la Coordination archives participent également à cette initiative institutionnelle transversale.

Les travaux ont été menés par le groupe de travail données composé d'Élise Chapoy (responsable service Appui à la recherche – ARO à la DRIS<sup>9</sup>), Laurent Lesnard (directeur du Centre de données socio-politiques – CDSP<sup>10</sup>), Odile Gaultier-Voituriez (représentante de la coordination Archives de Sciences Po, responsable du centre de documentation du CEVIPOF), Émilie Groshens (chargée d'études pour le projet beQuali, CDSP) et Cynthia Pedroja (cheffe de projet gestion des données de la recherche au sein du service ARO à la DRIS<sup>11</sup>).

---

<sup>9</sup> Jusqu'en juin 2017

<sup>10</sup> Jusqu'à l'été 2017

<sup>11</sup> Jusqu'en juillet 2019

## Table des matières

<b>I. LES DONNEES DE LA RECHERCHE : QUELQUES ELEMENTS DE CONTEXTE</b>	<b>7</b>
A. Définition et contexte	7
B. Cycle de vie d'un projet de recherche	8
C. Cycle de vie des données de recherche	8
D. Dispositif de gestion des données de la recherche	9
<b>II. 2015, ANNEE DE L'EXPLORATION</b>	<b>12</b>
A. International	12
1. La Data Library de l'université d'Édimbourg, pivot de la gestion des données	13
2. La Research Data Alliance (RDA), une structure internationale fédératrice	18
B. France	20
1. TGIR PROGEDO, au service des enquêtes quantitatives en sciences sociales	21
2. TGIR Huma-Num, un service pour les projets de recherche	21
3. Institut national de l'information scientifique et technique (INIST)	22
4. Projet beQuali, favoriser la réutilisation d'enquêtes qualitatives de sciences sociales	23
5. Consortium Archipolis	23
6. Institut national de recherche agronomique (INRA), un projet à la mesure de l'institution	24
7. Bibliothèque scientifique numérique, segment Données de la recherche (BSN 10)	25
8. Université de Strasbourg (UNISTRA) : les Archives ouvertes de la connaissance (AOC)	25
9. Université Paris-Descartes et Paris-Diderot - Université Paris Sorbonne Cité (USPC)	26
C. Le chercheur, ce champion	27
D. Sciences Po	28
1. Diffusion	29
2. Sensibilisation	29
3. Compétence informatique	29
4. Droit	29
5. Documentation et archivage	30
6. Méthodologie	30
7. Projets pilotes	30
8. Ce qui n'est pas apparu	30

---

---

<b>III.</b>	<b>2016, ANNEE DE L'EXPERIENCE</b>	<b>ERREUR ! LE SIGNET N'EST PAS DEFINI.</b>
<b>A.</b>	<b>Explorer encore</b>	<b>31</b>
1.	Typologie	31
2.	Enquête des besoins	31
3.	Entrepôts	31
4.	Les scénarios pour Sciences Po et les conditions de leur mise en œuvre	32
<b>B.</b>	<b>Mettre en place</b>	<b>32</b>
1.	Formation doctorale	32
2.	Journée d'étude	33
<b>C.</b>	<b>Donner un nouveau cadre au projet</b>	<b>33</b>
1.	Comité de pilotage (COPIL Données)	33
2.	Comité scientifique et technique (CST Données)	33
3.	Groupes de travail	33
4.	Coordinateur	34
<b>IV.</b>	<b>INFORMATIONS GENERALES</b>	<b>35</b>
<b>A.</b>	<b>La carte d'identité du projet Gestion des données de la recherche</b>	<b>35</b>
<b>V.</b>	<b>RESSOURCES</b>	<b>36</b>
<b>VI.</b>	<b>ANNEXES</b>	<b>38</b>

## I. Les données de la recherche : quelques éléments de contexte<sup>12</sup>

### A. Définition et contexte

Une donnée est une information, une description élémentaire. Elle peut prendre plusieurs formes (texte, image, son ou vidéo) et s'inscrire sur des supports de différentes natures (physique ou numérique).

On parle de données personnelles, de données informatiques, de données statistiques, de données patrimoniales, de données administratives, *etc.*, mais également de *données de la recherche*.

Proposition de définition (Edinburgh University, Data Library, 2011) :

“Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.”<sup>13</sup>

Les données de recherche peuvent être disponibles sous forme papier, nativement numériques ou numérisées.

La question de la gestion des données de recherche est aujourd'hui un enjeu majeur pour les sciences de l'homme et de la société. Elle est le fruit de plusieurs mouvements qui accompagnent la recherche en France et à l'international :

- *Open Access* ou libre accès à l'information scientifique et technique. Le programme pilote H2020 se situe dans cette veine.
- administration de la preuve scientifique afin de valider les résultats
- réutilisation des jeux de données
- valorisation et sauvegarde du patrimoine scientifique

---

<sup>12</sup> Les textes de cette partie sont issus de la note de synthèse interne “Données de la recherche” datée du 05/02/2015. Outre les rédacteurs du présent document, ont participé à sa rédaction :

- Sarah Cadorel – CDSP – chargée d'études pour le projet beQuali
- Guillaume Garcia – CDSP – responsable du projet beQuali
- Goulven Le Brech – DRIS/Mission Archives – responsable
- Delphine Lereculeur – DRIS/SAU-ARO – chef de projet appui à la publication scientifique

<sup>13</sup> Cette définition a été retenue par le groupe parmi d'autres, proposées (voir ci-dessous “[Ressources](#)”)

---

## B. Cycle de vie d'un projet de recherche

Les données s'inscrivent dans le cycle de vie d'un projet de recherche que la figure 1 ci-dessous représente schématiquement. De fait, les étapes présentées ici se déroulent souvent en parallèle. Elles se confondent et les rôles occupés par les différents acteurs (chercheurs, ingénieurs, bibliothécaires, documentalistes, chargés d'édition, *etc.*) tendent aujourd'hui à se chevaucher<sup>14</sup> :

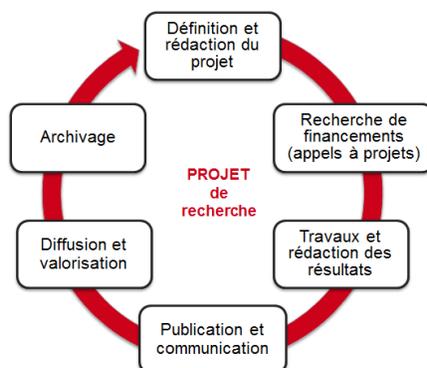


Figure 1 : cycle de vie d'un projet de recherche

## C. Cycle de vie des données de recherche

Si les phases de préparation, de récolte sur le terrain, de traitement et de diffusion des données se déroulent généralement pendant le projet (voir la figure 2. ci-dessous), la gestion des données prise dans son ensemble court tout au long de ce cycle et au-delà. C'est notamment le cas des phases de rédaction du plan de gestion de données (voir ci-dessous) ou des phases d'archivage et de réutilisation. Comme pour le cycle de vie d'un projet de recherche, les différentes étapes du cycle de vie des données sont en fait plus concomitantes que successives.

<sup>14</sup> David Stuard, "Libraries could play key role in managing research data", In *Research Information*, décembre 2014 - janvier 2015 – URL: <https://www.researchinformation.info/feature/libraries-could-play-key-role-managing-research-data> (vu le 19/12/2019)

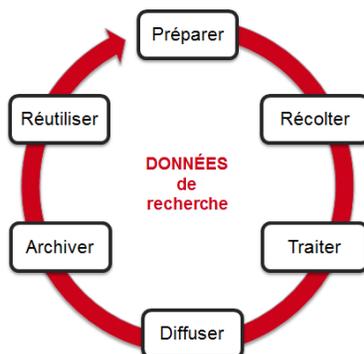


Figure 2 : Cycle de vie des données de la recherche

## D. Dispositif de gestion des données de la recherche

Mettre en place un dispositif de gestion des données au sein d'une institution est un moyen de "s'assurer que les données soient détectables, accessibles et intelligibles pour une réutilisation à long terme"<sup>15</sup>. La mise en place d'un tel dispositif, nécessite de s'interroger et de traiter différentes questions techniques, politiques, humaines. Parmi lesquelles :

- réservoir(s) : où déposer les données ? Quel(s) entrepôt(s) utiliser ? Il s'agit ici d'aborder l'un des aspects techniques de la gestion, à savoir le système de stockage, dans la durée.
- plan de gestion de données : document-cadre "explicitant la manière dont seront obtenues, documentées, analysées, disséminées [archivées] et utilisées les données produites au cours et à l'issue d'une recherche ou d'un projet de recherche"<sup>16</sup>. Ce document doit être rédigé au début du projet puis révisé, parfois plusieurs fois, durant le projet afin de l'adapter aux réalités (auxquelles sont confrontées les équipes). Ce document est obligatoire pour les projets participant au pilote Données de la recherche du programme H2020<sup>17</sup>.

<sup>15</sup> Jones, S., Pryor, G. & Whyte, A. (2013). "How to Develop Research Data Management Services - a guide for HEIs". *DCC How-to Guide*. Edinburgh: Digital Curation Centre. p. 2

<sup>16</sup> Cartier, A., Moysan, M., Reymonet, N., *Réaliser un plan de gestion de données*, 01/12/2014. URL : [http://www.univ-paris-diderot.fr/DocumentsFCK/recherche/Realiser\\_un\\_DMP\\_V1.pdf](http://www.univ-paris-diderot.fr/DocumentsFCK/recherche/Realiser_un_DMP_V1.pdf) (consulté le 30/01/2016)

<sup>17</sup> Lignes directrices pour la gestion des données dans H2020, [http://openaccess.inist.fr/IMG/pdf/14081\\_lignes\\_directrices\\_pgd\\_horizon\\_2020\\_tr\\_fr\\_versionavril2015-2.pdf](http://openaccess.inist.fr/IMG/pdf/14081_lignes_directrices_pgd_horizon_2020_tr_fr_versionavril2015-2.pdf) consulté le 31/01/2016)

- documentation : ensemble des informations servant à décrire le processus de recherche et les différents choix réalisés dans la gestion des données.
  - métadonnées et formats : assurer la compréhension, la lisibilité et la réutilisation des données en utilisant des schémas de description et de format normalisés, reconnus et documentés. Le mouvement du libre accès incite la communauté à utiliser des formats ouverts.
  - droit : à qui appartiennent les données ? Sont-elles communicables ? À qui ? À tous ? À un groupe restreint ? De nombreuses questions de droit, mais également d'éthique sont à résoudre avant que ne débute le projet. Le contexte juridique dans lequel s'inscrit chaque jeu de données conditionne son archivage, sa diffusion, etc.
  - sensibilisation : sensibiliser les différents acteurs de la recherche aux enjeux de la gestion de données en rendant compte de l'actualité, des problématiques soulevées par la question, en alertant sur les points potentiellement sensibles ou les actions à risque de la gestion des données. Il s'agit d'éclairer la communauté sur une question complexe et difficile à appréhender dans son ensemble.
  - formation : former les ingénieurs et chercheurs (rédaction d'un plan de gestion de données, traitement documentaire et dépôt des données, par exemple), et les personnels de l'information scientifique et technique.
  - accompagnement : accompagner les unités et les équipes de recherche dans la mise en place d'un processus de gestion de données à travers différentes actions (réunions, rencontres, documentation papier et en ligne, etc.), notamment lors du montage de projet (DMP) et des phases de dépôt de données.
  - politique : existe-t-il une politique nationale, d'établissement ou d'unité ? Quels sont les projets et donc les données concernées ?
  - traitement : nettoyer, normaliser et structurer les données et documenter le processus de la récolte afin de contextualiser les travaux.
  - diffusion : déposer, rédiger des *Data Paper*, valoriser les jeux par leur signalement dans des catalogues et leur documentation. Faire le lien entre les données et les publications scientifiques.
  - signalement : rendre visibles et accessibles à la communauté les données conservées.  
Communiquer : à qui ? Comment ?
-

- archivage : archivage intermédiaire ? Préservation à long terme ? Traiter ces questions, c'est également aborder la question de la sélection, ce qui doit être conservé, dans quelles et à quelles conditions.

Ce dispositif, forme d'écosystème complexe, nécessite l'implication des différents membres de l'équipe de recherche. Le binôme chercheur/ingénieur-technicien forme le socle de ce dispositif. Mais de nombreux autres acteurs au sein de l'institution doivent également y participer : les services de documentation, de bibliothèque et d'archives, la direction scientifique, les services informatiques, les services juridiques, les services de valorisation et les instances de décision politique et scientifique. Nous retiendrons que l'expression "gestion des données" est à comprendre dans ce projet comme le dispositif englobant l'ensemble des problématiques et enjeux liés aux données de la recherche.

## II. 2015, année de l'exploration

On note un grand intérêt actuellement en France et en Europe sur la question des données de la recherche. Les pays et structures institutionnelles sont plus ou moins avancés sur le traitement de ce chantier, certains développant des services nationaux ou locaux dédiés : infrastructures d'hébergement (entrepôts), modules d'(auto-)formation en ligne, documentation fournie, actions de sensibilisation, service d'accompagnement des équipes à la gestion des données, *etc.* font partie des services proposés par les uns et les autres.

Pendant un an, le GT Données a étudié le contexte national et européen de la gestion des données à travers une série de rencontres avec différents acteurs et experts de la question. Parallèlement, il a entrepris de réaliser une typologie des données produites par les chercheurs de Sciences Po ainsi qu'une série de rencontres avec les chercheurs, les ingénieurs et les équipes de direction des unités de recherche afin d'identifier et déterminer les besoins de ces publics.

Dix-sept rencontres se sont tenues entre janvier 2015 et février 2016. Élise Chapoy et Cynthia Pedroja ont fait un voyage d'études (15-16 décembre 2014) pour compléter les entretiens et ont rencontré leurs homologues de la London School of Economics (LSE).

Nous rendons compte ici des constats qui ont été effectués.

### A. International

Les problématiques liées à la gestion des données, et notamment celles de l'accessibilité et de la circulation, n'ont pas de frontières : les projets de recherche sont souvent multi-institutionnels, répartis en France, en Europe et au-delà. Les solutions d'hébergement de données – les entrepôts – ne sont pas forcément nationales, les publications et les données associées sont visibles à l'international.

Certains pays ont pris le virage de la gestion des données depuis des années, voire des décennies. Les infrastructures d'hébergement (entrepôts), les modules d'(auto-)formation en ligne, la documentation fournie, les actions de sensibilisation, les services d'accompagnement des équipes à la gestion des données, *etc.* sont autant de services régionaux et nationaux développés dans d'autres pays.

Il est donc nécessaire de comprendre et d'inscrire ce projet dans sa dimension internationale afin de bénéficier de la capitalisation des expériences d'autres institutions et structures. Nous

---

avons porté notre attention sur le cas du Royaume-Uni, souvent cité en exemple. Un voyage d'étude à l'université d'Édimbourg nous a permis de mieux comprendre les conditions et moyens favorisant la prise en compte de la gestion des données dans une université. Nous nous sommes également intéressés aux travaux menés par la Research Data Alliance (RDA), une infrastructure association internationale réunissant de nombreux acteurs de la question.

## 1. La Data Library de l'université d'Édimbourg, pivot de la gestion des données

web : <http://www.ed.ac.uk/information-services/research-support/data-library>

Interlocuteurs : **Dominic Tate** (Scholarly Communication Manager) - Scholarly Communication and OA, **Pauline Ward** (Data Library Assistant) - Datashare / MANTRA demonstration, **Martin Donnelly** (Senior Institutional Support Officer, DDC) - Digital Curation Centre, **Cuna Ekmekcioglu** (Dr, Senior Research Data Officer) - RDM Training & Awareness, **Stuart McDonald** (RDM Service Coordinator) - RDM Programme, **Anne Donnelly** (Academic Support Librarian) - RDM Librarian training, **Robin Rice** (Data Librarian) - Data Library / EDINA

Le Royaume-Uni est souvent cité comme exemple dans la gestion des données. La documentation et la littérature sont pléthoriques, les événements qu'ils organisent tout comme les participations aux meetings ou organismes internationaux dédiés aux *data* sont nombreux. Nous souhaitons, à travers l'exemple de l'université d'Édimbourg, et de sa Data Library, rendre compte de ce cas particulièrement dynamique au niveau local, national et international.

Un voyage d'études de deux jours, mené par Élise Chapoy et Cynthia Pedroja en décembre 2014 a permis d'échanger avec les membres de la Data Library, d'EDINA et du Digital Curation Center (DCC) situés au sein de l'université d'Édimbourg.

Depuis 1983, l'université d'Édimbourg a développé une Data Library<sup>18</sup> afin d'accompagner sa communauté académique dans l'utilisation et la gestion des données de la recherche, mais également d'en faciliter la découverte et l'accès. A la suite d'un appel d'offres public (open concours) lancé en 1996 par le Jisc<sup>19</sup>, une partie des services développés par la Data Library est devenue un "centre national de données" ; ce centre est alors devenu EDINA<sup>20</sup> (Edinburgh

---

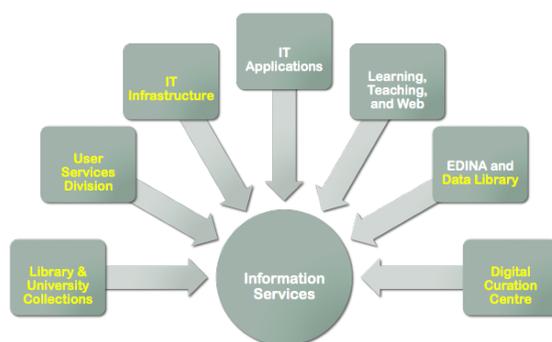
<sup>18</sup> URL : <http://www.ed.ac.uk/information-services/research-support/data-library> (consulté le 31/01/2016)

<sup>19</sup> Le Jisc est un Registered Charity (forme d'association caritative déclarée aux services gouvernementaux) qui soutient depuis 1993 l'utilisation des technologies du numérique dans l'enseignement et la recherche. Ce qui était jusqu'en 2012 un acronyme pour Joint Information Systems Committee est désormais son intitulé officiel. URL : <http://www.jisc.ac.uk> (consulté le 19/12/2019)

<sup>20</sup> URL : <http://edina.ac.uk/> (consulté le 31/01/2016)

---

Data and Information Access). La Data Library et EDINA<sup>21</sup> font partie des Services d'information (SI), un service support de l'université d'Édimbourg. Le SI est piloté par le directeur des systèmes d'information et le directeur de la bibliothèque. Le SI regroupe en plus d'EDINA et de la Data Library (sept personnes), la bibliothèque (collections physiques et numériques), la division des services aux usagers, l'infrastructure informatique, la division des applications/logiciels ou encore la pédagogie numérique.



**Figure 3 : organisation des Services d'information de l'université d'Édimbourg**

À cet enchevêtrement de structures s'ajoute le Digital Curation Center (DCC)<sup>22</sup>. Le DCC est un centre international d'expertise sur l'accessibilité aux ressources, dont les données. Le DCC est une réponse à la problématique selon laquelle une seule institution ou discipline ne peut assurer l'accès et la conservation de l'information. Depuis 2004, il sensibilise, structure et accompagne les institutions du Royaume-Uni dans le développement de politiques et de services de gestion des données en leur sein. Les institutions doivent souscrire au Jisc, dont le DCC est une émanation, pour pouvoir bénéficier de leurs services<sup>23</sup>. La position du DCC leur permet d'avoir un regard sur la situation nationale. Le service est composé de 14 personnes qui sont réparties sur trois sites (Edinburgh, Glasgow et Bath). Il y a parmi ces profils très variés des chefs de

<sup>21</sup> Nommés *EDINA and Data Library*

<sup>22</sup> URL : <http://www.dcc.ac.uk/> (consulté le 31/01/2016)

<sup>23</sup> Dans le cadre d'un programme financé par le Jisc entre 2011 et 2013, le DCC a accompagné pendant soixante jours vingt et une universités dans le développement de leur gestion des données de la recherche. Cette expérience a permis de montrer que l'intégration au sein d'une communauté de praticiens était nécessaire pour le développement de pratiques durables de gestion des données de la recherche.

projets et des bibliothécaires. La formation est l'un des piliers forts de l'action du DCC. L'offre est extrêmement complète et variée : introduction à la gestion de données (RDM), plan de gestion de données (DMP), outils, standards et ressources, licences pour la réutilisation de données, sélection des données à conserver, partage de données et entrepôts, citer les données, etc. Le DCC a également rédigé et diffusé une documentation fort riche autour des aspects méthodologiques du développement de services de gestion des données (comment évaluer et hiérarchiser les besoins de la communauté ? Quelles sont les étapes à suivre et comment les planifier ? Quels sont les enjeux politiques, institutionnels, techniques et financiers auxquels il faudra répondre ? *Etc.*), mais également de la manière dont certains aspects plus techniques doivent être pris en charge (comment évaluer et sélectionner les données à conserver ? Quelles licences utiliser pour les données et à quelles fins ? Quelles compétences pour les Data Librarians ? *Etc.*). Enfin, le DCC a développé une série d'outils pour accompagner les institutions tels que le Collaborative Assessment of Research Data Infrastructure and Objectives (CARDIO)<sup>24</sup>, qui permet de réaliser des audits des infrastructures de gestion des données dans une institution, ou DMPOnline<sup>25</sup>, qui permet de rédiger un plan de gestion des données selon les exigences des financeurs du Royaume-Uni et de l'Union européenne. Même si le DCC cherche avant tout à accompagner les chercheurs, il travaille essentiellement avec des bibliothèques qui sont les points de contact au sein des universités. Il s'est également appliqué à tisser des liens avec les acteurs des secteurs économiques et publics, ainsi qu'avec des organisations soeurs et des groupes internationaux travaillant sur les normes. Dernière brique de ce dispositif, le UK Data Archive<sup>26</sup> a été fondé dans les années soixante par le gouvernement du Royaume-Uni. Cette autorité administrative indépendante est le centre national d'archivage des données en sciences humaines et sociales. Depuis sa création, le UK Data Archive a mis en place de nombreux services pour la recherche en sciences sociales et en économie ainsi que pour l'enseignement : conservation voire achat de données de recherche, publiques ou privées, point d'accès centralisé aux données économiques et de sciences sociales à travers le UK Data Service<sup>27</sup>, expertises techniques (stockage, archivage, documentation, etc.), archivage des archives et données publiques, etc.

---

<sup>24</sup> URL : <http://www.dcc.ac.uk/projects/cardio> (consulté le 31/01/2016)

<sup>25</sup> URL : <https://dmponline.dcc.ac.uk/> (consulté le 31/01/2016)

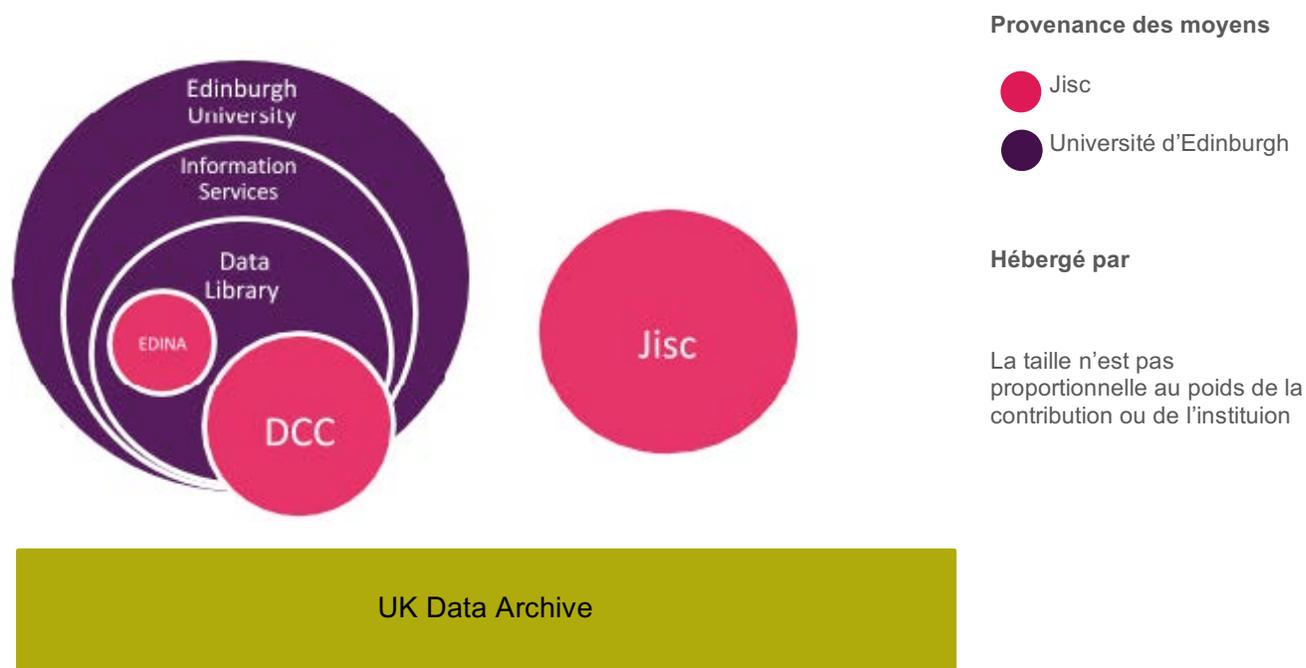
<sup>26</sup> URL : <http://www.data-archive.ac.uk/> (consulté le 31/01/2016)

<sup>27</sup> Données issues des recensements, microdonnées internationales, microdonnées d'entreprises, enquêtes qualitatives, etc.

---

Ces services permettent de répondre aux missions du UK Data Archive : préservation à long terme des données, documentation et validation des formats de données, catalogue de données, potentielle réutilisation et accompagnement des utilisateurs dans cette démarche.

Pour résumer en quelques mots la manière dont s'articulent ces infrastructures autour de la Data Library : les moyens proviennent à la fois de l'université d'Édimbourg, comme entité propre, et du Jisc via EDINA qui est une sous-entité de la Data Library. EDINA et la Data Library appartiennent aux Services d'information qui hébergent également en leur sein l'infrastructure nationale DCC et dont ils ont bénéficié du programme de développement d'infrastructure institutionnelle entre 2010 et 2013. Vient s'ajouter en soutien le service national d'archivage des données, le UK Data Archive. Rencontrer les collègues d'Édimbourg revient donc à pénétrer au cœur du dispositif national des données de la recherche du Royaume-Uni.



**Figure 4 : Les services et moyens à l'université d'Édimbourg**

Les services proposés par la Data Library et EDINA aux membres de la communauté académique de l'université d'Édimbourg sont multiples :

- modules de sensibilisation ou formation en présentiel ou à distance (ex. : Mantra<sup>28</sup>, site de ressources développé par EDINA donc avec des fonds du Jisc),
- entrepôt de données (Edinburgh DataShare<sup>29</sup>) pour les chercheurs et les étudiants,
- accompagnement des membres de la communauté académique dans la gestion de leurs données ;
- catalogue des jeux de données<sup>30</sup> hébergées à l'université d'Édimbourg, au Royaume-Uni et à l'étranger.

L'infrastructure mise en place a pu se développer à la faveur des politiques nationales, des organismes de financement et de l'université. Depuis mai 2011, l'université d'Édimbourg<sup>31</sup> fait partie des trente universités du Royaume-Uni<sup>32</sup> ayant développé une politique en matière de gestion des données. Le DCC a rassemblé<sup>33</sup> les différentes caractéristiques de ces politiques, institution par institution, afin de pouvoir proposer une vision d'ensemble.

Un bémol peut être mis toutefois à ce tableau, peut-être un peu idyllique. Les financeurs du Royaume-Uni exigent aujourd'hui qu'un DMP soit déposé avant le début de tout projet. Pour certains de nos interlocuteurs, une fois ce document rédigé, il n'y a pas de suite concrète. La London School of Economics and Political Science (LSE) a mis en place dernièrement un service de gestion des données. De l'aveu même des membres de ce service, les moyens alloués ne sont pas à la hauteur. Aussi, peu d'actions peuvent être véritablement développées. Nous notons que le DCC ne mentionne pas la LSE comme une institution ayant mis en place une politique.

---

<sup>28</sup> MANTRA, Research Data Management Training, URL : <http://datalib.edina.ac.uk/mantra/> (consulté le 31/01/2016)

<sup>29</sup> URL : <http://datashare.is.ed.ac.uk/> (consulté le 31/01/2016)

<sup>30</sup> URL : <http://www.ed.ac.uk/information-services/research-support/data-library/finding-data> (consulté le 31/01/2016)

<sup>31</sup> URL : <http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy> (consulté le 31/01/2016)

<sup>32</sup> URL : <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies> (consulté le 31/01/2016)

<sup>33</sup> URL : [http://www.dcc.ac.uk/webfm\\_send/1776](http://www.dcc.ac.uk/webfm_send/1776) (consulté le 31/01/2016)

---

## 2. La Research Data Alliance (RDA), une structure internationale fédératrice

web : <https://rd-alliance.org/>

“Research Data Sharing without barriers”, telle est la raison d’être de la RDA qui rassemble aussi bien des individus que des organisations académiques, économiques et de la sphère publique du monde entier. Née en 2012, la RDA (3610 membres en janvier 2016) est un réseau d’échanges, de partage et d’harmonisation de bonnes pratiques qui a émergé à partir des praticiens, chercheurs, ingénieurs et professionnels de la documentation. Trois pays, à travers leurs agences de financement de la recherche, sont à l’origine de la RDA<sup>34</sup> : l’Australie<sup>35</sup>, la Commission européenne<sup>36</sup> et les États-Unis<sup>37</sup>.

Les livrables attendus sont multiples : établir de nouveaux standards et harmoniser ceux qui existent, favoriser le partage, l’échange, l’interopérabilité, la réutilisation, améliorer l’accessibilité, assurer une bonne gestion, l’administration et la préservation des données<sup>38</sup>.

Les travaux sont menés par trois types de groupes : les *Birds of a Feather (BoF) Groups* (courte durée de vie pour lancer des groupes de travail ou d’intérêt), les groupes de travail (*Working Groups* - 12 à 18 mois d’existence avant dissolution, qui produisent des livrables concrets) et les groupes d’intérêt (*Interest Groups* : qui existent aussi longtemps qu’il le faut, réalisent des études de cas et participent à la coordination et à la communication de RDA). Le travail se fait à distance puis en réunions plénières tous les six mois.

Cette structuration très segmentée et les délais relativement courts (12 à 18 mois) pour la production de livrables des groupes de travail visent à dynamiser les travaux en se focalisant sur une problématique précise et à arriver à des résultats tangibles et applicables rapidement, dans la mesure où les membres de la RDA sont des acteurs effectifs dont une majorité de chercheurs<sup>39</sup>. Nous pouvons déjà en mesurer les résultats au travers des livrables mis à

---

<sup>34</sup> *Research Data Alliance Launch and First Plenary*, 18-20 mars 2013, [https://www.rd-alliance.org/sites/default/files/attachment/RDA\\_1stPlenary\\_Overview.pdf](https://www.rd-alliance.org/sites/default/files/attachment/RDA_1stPlenary_Overview.pdf) (consulté le 31/01/2016)

<sup>35</sup> À travers l’Australian National Data Service ([www.ands.org.au](http://www.ands.org.au)) supporté par le National Collaborative Research Infrastructure Strategy Program et le Education Investment Fund (EIF) Super Science Initiative

<sup>36</sup> RDA Europe Project, 7<sup>e</sup> programme cadre

<sup>37</sup> À travers la National Science Foundation

<sup>38</sup> RDA, Goals and Outcomes, URL : <https://www.rd-alliance.org/working-and-interest-groups/goals-and-outcomes.html> (consulté le 31/01/2016)

<sup>39</sup> 780 en janvier 2016. URL : <https://www.rd-alliance.org/sites/default/files/attachment/RDA-Members%2008January2016.pptx> (consulté le 01/02/2016)

---

disposition sur le site<sup>40</sup>.

Trois membres du GT Données ont participé à la 6<sup>e</sup> plénière qui s'est déroulée à Paris du 23 au 25 septembre 2015<sup>41</sup>. Voici les constats que nous faisons à l'issue de la plénière :

- alors que la 6<sup>e</sup> plénière se passait à Paris, les Français étaient peu représentés, plus que dans les précédentes sessions, mais moins que d'autres pays.  
Certaines institutions ne semblaient pas représentées. Les Nord-Américains sont très présents<sup>42</sup> et ont mis en œuvre de nombreux travaux : une bulle linguistique semble s'être formée et peu d'efforts sont réalisés pour que les membres non anglophones de naissance participent activement. Un groupe d'intérêt de Germaniques s'est constitué, peut-être en réaction à cette situation ou pour faciliter les échanges.
- Les personnes qui ont participé aux éditions précédentes et aux travaux ont appris à se connaître. Des réseaux de travail et d'information se sont déjà constitués. Comment prendre le train en marche ?
- Certains travaux sont déjà bien engagés : il est difficile d'intégrer les groupes. Comment embrasser une littérature très importante et les résultats de travaux qui s'accumulent ?
- Les groupes sont nombreux. En janvier 2016 il y a 21 groupes de travail, 44 groupes d'intérêt et 4 BoF<sup>43</sup>. Les sujets sont variés. Certains groupes thématiques, disciplinaires, sont éloignés des problématiques qui concernent Sciences Po.
- Les travaux menés (métadonnées, entrepôts et terminologies) peuvent être très techniques. Sur certains sujets, nous manquons de compétences au sein de notre GT Données.

Étant donné le nombre et l'avancée des travaux dans RDA, il est difficile d'identifier les groupes qu'il serait pertinent de rejoindre pour le projet institutionnel Sciences Po dont la feuille de route n'est pas encore établie. Comment rattraper, à l'échelle de l'institution, le retard accumulé face à des membres qui travaillent dans des disciplines, des institutions ou au sein de pays dans lesquels la gestion des données est traitée depuis plusieurs décennies ? Participer aux travaux,

---

<sup>40</sup> Chaque groupe de travail bénéficie d'une page dédiée pour exposer les résultats de ses travaux. URL : <https://www.rd-alliance.org/groups/working-groups> (consulté le 31/01/2016). C'est également le cas pour les groupes d'intérêt. URL : <https://www.rd-alliance.org/groups/interest-groups> (consulté le 31/01/2016)

<sup>41</sup> Au CNAM (Paris). URL : <https://www.rd-alliance.org/plenary-meetings/rda-sixth-plenary-meeting.html> (consulté le 31/01/2016)

<sup>42</sup> 35% de Nord-Américains, 50% d'Européens en janvier 2016. URL : <https://www.rd-alliance.org/about.html> (consulté le 01/02/2016)

<sup>43</sup> Les groupes sont créés et disparaissent au fil des plénières.

---

c'est participer aux choix qui sont faits et favoriser la mise en œuvre des résultats dans nos pratiques et le projet. RDA fonctionne d'ailleurs sur un concept simple : créer, adopter, utiliser<sup>44</sup>. C'est le principe du *bottom-up* qui fait loi. Pour bénéficier des résultats des travaux de la RDA et participer à leur élaboration, il faut s'y impliquer dans la durée. L'appropriation des travaux ne saurait s'émanciper d'une participation aux plénières qui sont les seuls moments de rencontre physique des membres de l'Alliance.

Afin de mieux comprendre l'environnement européen actuel en matière d'Open Access, nous recommandons la lecture du rapport final du projet PASTEUR4OA<sup>45</sup>. L'objectif<sup>46</sup> est d'aider au "développement puis au renfort des politiques et des stratégies en matière d'Open Access au niveau national et de faciliter ainsi la coordination entre les membres de l'Union européenne. Pour cela, un programme d'activité collaboratif a été mis en place afin de soutenir les décideurs de niveau national jusqu'aux projets partenaires"<sup>47</sup>. Pendant trois années, les politiques d'Open Access (publications et données de la recherche) de pays (Royaume-Uni, Portugal, Belgique, Danemark, etc.), d'institutions (University College London, université de Liège, université de Turin, etc.) ont été étudiées. Le projet a fourni une riche documentation sur ses travaux. Il permet d'avoir une vision à la fois macro et micro des politiques mises en place. Les données de la recherche

## B. France

Les initiatives en France semblent peu nombreuses, au regard de ce qui se passe au Royaume-Uni, cas évoqué précédemment, aux Pays-Bas, ou encore aux États-Unis. Les politiques nationales ou les exigences des financeurs, la participation à RDA, les services des données de la recherche (Data Libraries) mis en place sont le signe de ce dynamisme. Qu'en est-il en France ? Nous pouvons distinguer deux types de structures : celles, nationales, avec lesquelles Sciences Po a des liens, plus ou moins forts (INIST, TGIR Huma-Num, TGIR Progedo) et aux quelles elle participe (consortium Archipolis, projet beQuali) et celles, nationales ou locales, dont les expériences nous semblent intéressantes. Parmi ces dernières, deux projets ont retenu notre

---

<sup>44</sup> "Create, Adopt, Use"

<sup>45</sup> [http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/PASTEUR4OA/Briefing/Paper\\_FINAL\\_0.pdf](http://www.pasteur4oa.eu/sites/pasteur4oa/files/resource/PASTEUR4OA/Briefing/Paper_FINAL_0.pdf)

<sup>46</sup> URL : <http://www.pasteur4oa.eu/home> (consulté le 31/01/2016)

<sup>47</sup> Liste des partenaires. URL : <http://www.pasteur4oa.eu/partners> (consulté le 31/01/2016)

---

attention : l'université de Strasbourg (AOC) et l'Institut national de la recherche agronomique (INRA). Voici ce que nous avons identifié pour chacune de ces structures :

## 1. TGIR PROGEDO, au service des enquêtes quantitatives en sciences sociales

Interlocuteur : **Pascal Buléon** (directeur)

web : <http://www.progedo.fr>

Née en 2008, la Très grande infrastructure de recherche (TGIR) PROGEDO<sup>48</sup> est un acteur incontournable des données d'enquête quantitatives en sciences sociales. Cette infrastructure nationale articule ses travaux avec d'autres acteurs de la scène européenne<sup>49</sup>. Ses missions consistent à produire et mettre à disposition des données pour les sciences humaines et sociales.

## 2. TGIR Huma-Num, un service pour les projets de recherche

web : <http://www.huma-num.fr/>

Interlocuteurs : **Olivier Baude** (directeur), **Stéphane Pouyllau** (directeur technique)

La TGIR Huma-Num (anciennement ADONIS) propose un certain nombre de services pour la communauté en sciences sociales et humaines. Se fondant sur un "dispositif humain (concertation collective) et technologique (services numériques pérennes)", elle veille à intégrer ses services dans un contexte national et européen<sup>50</sup>. La TGIR a su mettre en place des réseaux d'experts à travers notamment des consortiums dont Archipolis<sup>51</sup> est l'une des émanations. Huma-Num place les données de la recherche au cœur de ses dispositifs en prenant soin de développer des services pour chacun des cycles de vie des données :

---

<sup>48</sup> Très grande infrastructure de PROduction et GEstion des DONnées

<sup>49</sup> European Social Survey (ESS), Survey of Health, Ageing and Retirement in Europe (Share)

<sup>50</sup> URL : <http://www.huma-num.fr/la-tgir-en-bref> (consulté le 31/01/2016)

<sup>51</sup> Voir ci-après [II.B.5. Consortium Archipolis](#)

---

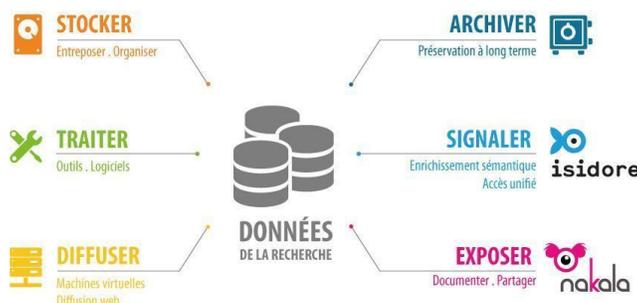


Figure 5 : offre de services et d'outils de la TGIR Huma-Num

Les services proposés par Huma-Num s'adressent aux individus, souvent des chercheurs, dans le cadre de leur projet de recherche, et non aux institutions. Ces dernières peuvent bénéficier de l'expertise technique des membres d'Huma-Num s'ils viennent à développer des infrastructures. Mais aucun service développé à ce jour ne s'adresse directement à elles.

### 3. Institut national de l'information scientifique et technique (INIST), une unité propre nationale ?

web : <http://www.inist.fr/?Donnees-de-la-recherche&lang=fr>

Interlocutrice : **Marie-Christine Jacquemot** (responsable de l'équipe Valorisation des données de la recherche)

Depuis 2011, l'INIST (UPS<sup>52</sup> du CNRS) est en cours de rénovation. En 2012, une étude de faisabilité est menée en interne sur les données de recherche. Une enquête a été réalisée auprès des directeurs d'unité de la délégation Centre-Est<sup>53</sup> ainsi qu'une série d'entretiens avec des chercheurs. L'objectif est d'identifier les services qui pourraient être développés au sein de l'unité. Point d'orgue de ces travaux, une journée d'étude locale a été organisée avec les chercheurs ayant répondu à l'enquête. Les constats sont les suivants : une confusion existe entre les données et les publications<sup>54</sup>, l'intérêt pour des espaces de stockage et des espaces de travail collaboratif est fort, la valorisation des bases de données qui ne sont plus maintenues est nécessaire.

<sup>52</sup> Unité propre de service

<sup>53</sup> L'INIST est situé à Vandœuvre-lès-Nancy (Meurthe-et-Moselle, Alsace-Champagne-Ardenne-Lorraine)

<sup>54</sup> Sur ce point, nous renvoyons également le lecteur au compte rendu d'entretien de Michèle Dassa  
VI. Annexes

## 4. **Projet beQuali, favoriser la réutilisation d'enquêtes qualitatives de sciences sociales**

web : <http://www.bequali.fr>

La banque d'enquêtes qualitatives beQuali, développée au sein du Centre de données socio-politiques (CDSP) de Sciences Po, propose un catalogue d'enquêtes de sciences sociales menées à partir de méthodes qualitatives (par entretiens individuels, focus groups et observations ethnographiques). Le portail met à disposition de la communauté scientifique, sous réserve d'autorisation, des matériaux bruts et une documentation restituant le contexte de leur production. Outil à vocation nationale, il s'inscrit dans la politique patrimoniale de préservation des données de la recherche. Ses objectifs sont également pédagogiques (enseignement et visibilité de la diversité des démarches) et scientifiques (histoire des sciences, revisite et analyse secondaire). Un premier pont va être jeté dans le cadre de la formation doctorale. En effet, l'équipe du projet souhaite développer ses services aux enseignants : il s'agit d'inscrire beQuali dans le catalogue des outils pédagogiques.

## 5. **Consortium Archipolis**

web : <http://archipolis.hypotheses.org/>

ArchiPolis est un consortium national de la TGIR Huma Num ayant pour objectifs l'inventaire, la collecte et l'archivage d'enquêtes qualitatives portant sur l'objet politique dans le domaine des sciences sociales. Labellisé en septembre 2012 pour quatre années, en 2016 il compte neuf laboratoires<sup>55</sup> dont la moitié de Sciences Po. Le travail mené vise à construire, avec les personnels de recherche de ces laboratoires, une méthodologie et des outils communs d'archivage et de mise en visibilité des enquêtes qui y sont produites et conservées. Le consortium a préparé un catalogue commun des enquêtes recensées, avec l'outil Dataverse<sup>56</sup>, accessible au public au second semestre 2016. Cet outil d'inventaire et de description commun rend visibles les productions empiriques qualitatives de chaque laboratoire, en complétant les informations sur les enquêtes (objet de l'enquête problématique et contexte, méthodologie suivie, publications, etc.).

---

<sup>55</sup> CDSP (porteur), CEE, Ceraps (université Lille 2), CSO, Pacte (Sciences Po Grenoble, université Grenoble Alpes), CEVIPOF, Centre Émile-Durkheim (université Bordeaux), OSC, Triangle (ENS Lyon)

<sup>56</sup> URL : <http://dataverse.org/> (consulté le 31/01/2016)

---

## 6. Institut national de recherche agronomique (INRA), un projet à la mesure de l'institution

web : <http://www.inra.fr/>

Interlocutrice : **Odile Hologne** (directrice d'unité de la Délégation information scientifique et technique)

L'Institut national de recherche agronomique (INRA) est un établissement public d'enseignement supérieur (EPST) qui regroupe 8500 titulaires, dont 1850 chercheurs. Trois éléments ont joué en faveur de l'émergence du projet Données de recherche à l'INRA : le besoin exprimé directement par les scientifiques (du fait de la croissance exponentielle du nombre des données issues des séquenceurs et du partage des données en Open Access exigé par certains éditeurs du secteur sciences, techniques et médical (STM) ; le volontarisme de la direction de l'INRA ; l'audit des systèmes d'information de l'INRA qui a montré l'importance du capital scientifique et des nombreuses plateformes existantes dès 2008. Entre 2009 et 2012, une étude sur la gestion et le partage et une enquête comparative des cas du Royaume-Uni, de Canada, des États-Unis et de l'Australie sont menées. Le rapport publié en 2012<sup>57</sup> propose 9 recommandations parmi lesquelles :

- définir la politique de l'établissement,
- mettre en place un comité d'évaluation des données produites par l'INRA,
- s'impliquer dans les comités de standardisation,
- développer un portail d'accès à l'ensemble des ressources (annuaire),
- prendre en compte le cycle de vie dans les projets de recherche (Data Management Plan = plan de gestion de données),
- utiliser des licences CC pour les plateformes (statut des données produites : propriété, réutilisation),
- doter l'INRA d'infrastructures dimensionnées pour le calcul haute performance.

---

<sup>57</sup> Gaspin, C., Pontier, D., Colinet, L., Dardel, F., Franc, A., Hologne, O., Le Gall, O., Maurin, N., Perrière, G., Pichot, C., Rodolphe, F. (2012). Rapport du groupe de travail sur la gestion et le partage des données. 64 p.  
<https://prodinra.inra.fr/record/206746>

---

## 7. Bibliothèque scientifique numérique, segment Données de la recherche (BSN10)

web : <http://www.bibliothequescientifiquenumerique.fr/>

Interlocuteur : **Francis André** (coordinateur)

Entre le début de la Bibliothèque scientifique numérique (BSN) en 2007-2008 et l'apparition du segment dédié aux données de la recherche (BSN 10) en 2014, plus de cinq ans se sont écoulés. Le ministère de l'Enseignement supérieur et de la Recherche ne considérait pas les données de la recherche comme légitimes pour les bibliothèques. Un long travail a dû être fait avant que BSN 10 ne naisse en 2014. Les travaux menés par ce segment visent à développer une politique nationale en matière de gestion des données<sup>58</sup>. À travers une série d'actions, BSN 10 souhaite "coordonner les questions d'acquisition, de production, de référencement, de protection, de stockage, d'archivage, de diffusion et de réutilisation des données produites dans le cadre des travaux de recherche."<sup>59</sup>. La prise en compte d'un périmètre large incluant toutes les données et les acteurs de la question, la nécessité de définir des stratégies institutionnelles, l'inscription de la France dans un contexte international, la prise en compte des pratiques de gestion et de partage des données dans l'activité scientifique des chercheurs, l'harmonisation de pratique, la rédaction obligatoire d'un DMP pour les projets ANR ou encore la clarification des usages et donc l'adaptation du cadre juridique font partie des recommandations des membres de BSN 10.

## 8. Université de Strasbourg (UNISTRA) : les Archives ouvertes de la connaissance (AOC)

web : <https://www.unistra.fr/index.php?id=accueil>

Interlocutrice : **Adeline Rège** (chef de projet Archives ouvertes)

Le Projet Données de la recherche de l'université de Strasbourg s'inscrit dans un projet institutionnel (contrat de site validé par le MENESR) plus vaste autour du développement de

---

<sup>58</sup> *Propositions des BSN 10 pour la mise en œuvre d'une politique nationale en matière de gestion des données de recherche*, 7 octobre 2014 <http://www.bibliothequescientifiquenumerique.fr/wp-content/uploads/2015/09/BSN10-donn%C3%A9es-de-la-recherche.pdf>

<sup>59</sup> BSN 10 : données de la recherche, <http://www.bibliothequescientifiquenumerique.fr/bsn-10-donnees-de-la-recherche/>

---

l'Open Access ; un volet est consacré à la mise en place d'une archive ouverte<sup>60</sup> appelée les Archives ouvertes de la connaissance (AOC). Au-delà de renforcer la visibilité de la recherche alsacienne en donnant à voir ses pans méconnus, ce projet se veut pionnier dans le partage des données de la recherche.

## 9. Université Paris-Descartes et Paris-Diderot - Université Paris Sorbonne Cité (USPC)

Interlocutrices : **Aurore Cartier** (responsable des Services à la recherche, Service commun de documentation, université Paris-Descartes), **Magalie Moysan** (archiviste, Bureau des archives, université Paris-Diderot), **Valérie Neouze** (directrice du Service commun de documentation, université Paris-Descartes), **Nathalie Reymonet** (documentaliste, Direction d'appui à la recherche et à l'innovation (DARI), université Paris-Diderot).

Une équipe pionnière a produit un plan de gestion de données en français dans le courant de l'année 2014. L'objectif était de pouvoir fournir un formulaire de cadrage pour les projets européens H2020 s'insérant dans le pilote<sup>61</sup>. Ce document a circulé dans de nombreux réseaux d'archivistes, de documentalistes et de bibliothécaires. Cette initiative est venue du service Archives de la cellule de Paris-Diderot et du SCD de Paris-Descartes. Une série de formations a été réalisée pour les ingénieurs Europe en charge du montage des projets de recherche afin qu'ils puissent s'approprier ce document et voir quelles sont les ressources, en IST notamment, à utiliser pour le développer.

Nous avons intégré le groupe de travail ainsi constitué en avril 2015 afin de développer un projet transversal et plus ambitieux. Cinq actions ont été proposées à la COMUE : informatiser le plan de gestion de données, proposer une formation aux doctorants, sensibiliser et former les équipes de recherche (chercheurs et ingénieurs) et les personnels de l'IST et des archives à la

---

<sup>60</sup> “ *Les enjeux actuels en matière de libre accès aux produits de la recherche ont conduit les établissements alsaciens d'enseignement supérieur (Unistra, BNU, UHA et INSA) à s'engager dans le projet Archives ouvertes de la connaissance (AOC), porté dans le cadre du contrat quinquennal de site 2013-2017. Piloté par l'Unistra et l'UHA, il vise à mettre à disposition les publications des chercheurs, toutes disciplines confondues, dans un espace dédié sur Internet. Par leur support numérique, les AOC sont bien référencées par les moteurs de recherche et donc faciles et rapides d'accès.* ”  
(<https://savoirs.unistra.fr/recherche/big-data-revolution-ou-evolution/les-archives-ouvertes-de-la-connaissance/>)

<sup>61</sup> Voir V. Ressources, Commission européenne - H2020

---

gestion de données et enfin réaliser une étude de faisabilité pour un entrepôt de données multi-institutionnel de l'USPC<sup>62</sup>.

En juillet 2015, un accord de principe a été donné pour lancer le projet. La première action mise en place est l'inscription d'un module de formation<sup>63</sup> dans le catalogue du CFDIp<sup>64</sup>. À l'automne 2015, le groupe formé s'est réuni autour des directeurs des trois bibliothèques universitaires partenaires afin de proposer un projet plus ambitieux et un budget prévisionnel.

## C. Le chercheur, ce champion

La connaissance et la pratique du terrain sont des éléments fondamentaux pour accompagner les chercheurs dans la gestion de leurs données. De même, il est nécessaire de pouvoir mobiliser de multiples acteurs qui ne seront efficaces que dans cette proximité (documentalistes, personnels de soutien à la recherche, archivistes, informaticiens et ingénieurs). Pourtant, nous remarquons que la majorité des projets ici mentionnés ont été lancés par des acteurs de la documentation et parfois des archives. Un acteur a souvent été mentionné, sans que son nom apparaisse jusqu'ici : le chercheur "champion".

*"Trouvez votre champion"*<sup>65</sup>. Parmi les préconisations qui nous ont été faites (travailler au plus près des chercheurs, ne pas créer une infrastructure déconnectée des besoins, ne pas reproduire le fameux projet *"top down"*, etc.), celle-ci est revenue maintes fois lors de nos échanges en France, à l'étranger et dans les différents documents dont nous avons pris connaissance. Nous ne sommes pas parvenus à identifier la source de cette injonction qui signifie : identifiez un chercheur confirmé reconnu dans son domaine ou ayant une position stratégique forte au sein de votre institution et sensibilisez-le au bien-fondé de votre démarche.

---

<sup>62</sup> Dans la proposition, cet entrepôt a été abusivement intitulé "Data Center". Il ne s'agit bien évidemment pas de proposer un centre de calcul, mais bien une infrastructure de stockage et de diffusion des données.

<sup>63</sup> Voir III. *2016, année de l'expérience*

<sup>64</sup> Centre de formation des doctorant-e-s aux initiatives professionnelles, *Open access, open data et données de la recherche : quel intérêt pour un jeune chercheur publiant ?*, URL : <http://cfdip.uspc.fr/fr/formations/catalogue-des-formationen/professionnelles-appliquees/valorisation-de-la-recherche/item/222-open-access-open-data-et-donnees-de-la-recherche-quel-interet-pour-un-jeune-chercheur-publiant-2015-2016> (consulté le 31/01/2016)

<sup>65</sup> "Provide a Champion", DCC, *How to Develop RDM Services - a guide for HEIs*, URL : <http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services> (consulté le 31/01/2016)

---

Il pourra ensuite porter la bonne parole à sa communauté scientifique. D'autres projets pourraient bénéficier de cette préconisation. Toutefois, en la confrontant au constat que nous avons pu faire sur la prédominance des métiers de l'IST en France sur ces questions, il apparaît que la présence d'un champion est nécessaire pour assurer l'adhésion de la communauté académique à ce projet.

D'une manière générale, nous avons constaté une bonne réception du Projet données par nos différents interlocuteurs qui posent un regard attentif sur nos travaux afin de voir ce que nous allons faire et la manière dont nous allons procéder. L'intérêt porte également sur le fait que ce projet soit mené par trois directions (DS, DSI et DRIS).

À la lumière des travaux menés, nous avons identifié ce que pourraient être deux schémas de développement de services. Le premier consiste à prendre en charge les données à travers le plan de gestion de données. Les chercheurs sont invités à s'emparer de ce document-cadre afin d'annoncer la manière dont ils vont gérer ou ont géré leurs données. Un service d'accompagnement, souvent situé au sein des bibliothèques, implique alors les différents acteurs nécessaires pour remplir ce document. C'est ce que nous nommerons la voie 1. La deuxième voie consiste à proposer un service de valorisation documentaire, c'est-à-dire accompagner et distribuer les forces nécessaires pour réaliser un site web qui permettrait d'exposer les jeux de données et offrirait au visiteur une expérience d'appropriation grâce aux outils développés autour des données (visualisation, simulation, etc.). La réalisation d'un tel produit nécessite, comme pour la voie 1, que l'on sache comment gérer les données. Mais cela se fait en toute transparence pour le chercheur qui n'est pas confronté à la dimension administrative un peu rébarbative du plan de gestion de données.

## D. Sciences Po

Dans le but de réaliser une typologie des données produites par les chercheurs des laboratoires de Sciences Po, le GT Données a pris contact avec chacune des unités pour présenter le projet et évoquer les modalités du recueil. Les rencontres se sont tenues soit avec l'équipe de direction, soit dans le cadre de réunions du laboratoire<sup>66</sup>. Le GT Données était représenté au

---

<sup>66</sup> Consultez la liste des rencontres au sein des centres de recherche [VI. Annexes](#)

minimum par Laurent Lesnard et Cynthia Pedroja. Un ou plusieurs autres membres pouvaient être présents.

Le groupe a constaté une adhésion générale de principe au projet. Nos interlocuteurs comprennent la plupart des enjeux de cette problématique tout en ne mesurant pas toujours l'envergure et la complexité. Parmi les besoins, les constats et autres interrogations de nos interlocuteurs, sept ont retenu notre attention<sup>67</sup> :

## 1. Diffusion

En accès libre, restreint, sans droit d'accès, la question de la diffusion a été certainement le premier élément mis en avant. Les pratiques et les envies divergent selon les communautés et les laboratoires. Pour l'ensemble des personnes rencontrées, la mise à disposition (Open Access, avec accès restreint ou encore sous embargo) est la finalité de la gestion des données.

## 2. Sensibilisation

Il y a un besoin général d'informations sur les méthodes de travail, les enjeux, les questions juridiques, etc. Les doctorants et les étudiants de master constituent un public cible à privilégier, mais l'ensemble des personnels de Sciences Po est concerné, même ceux qui semblent plus éloignés de cette question comme les personnels administratifs des directions qui produisent des données.

## 3. Compétence informatique

Certaines équipes manquent de compétences informatiques pour le traitement et la diffusion des données. Il ne s'agit ici pas de stockage, de gestion des flux, des formats ou des métadonnées, mais de l'accompagnement pour rendre accessibles en ligne les données traitées dans des dispositifs outillés. Un besoin qui rejoint les constats faits par un certain nombre de nos interlocuteurs (UNISTRA, INIST, BSN 10).

## 4. Droit

L'absence de soutien juridique de l'institution pour savoir ce que l'on peut faire des données (diffusion, stockage, archivage) ou encore comment l'on peut (ré)utiliser des données produites par d'autres est perçu comme un frein à la pratique.

---

<sup>67</sup> Retrouvez l'ensemble des comptes en [VI. Annexes](#)

## 5. Documentation et archivage

Seul le médialab a formulé ces besoins qui se cachent pourtant derrière tous les autres points évoqués. Soit parce qu'ils étaient induits, soit parce qu'ils relèvent des métiers de l'IST et des archives qui ne sont pas forcément représentés dans les centres de recherche, sauf au CEVIPOF et au CHSP. C'est la lecture que nous en faisons.

## 6. Méthodologie

Il existe un questionnement général sur les méthodes de traitement des données (éthique, archivage, conservation des sources, manipulation des données, etc.) et un besoin de partager avec la communauté, peut-être au-delà du centre.

## 7. Projets pilotes

« Il est nécessaire d'identifier les projets susceptibles de devenir pilotes pour les actions qui seront menées dans le cadre du Projet Données ». Cette phase a été lancée lors des rencontres menées dans les centres de recherche. Il s'agit de confirmer et d'affiner la liste des besoins afin de pouvoir, le moment venu, travailler avec l'équipe du projet sur tel ou tel aspect que le GT devrait mettre en place.

## 8. Ce qui n'est pas apparu

Certaines actions ou étapes constitutives de la gestion des données n'ont pas ou peu été mentionnées : le catalogue des données pour les rendre visibles, le lien entre publications et données, l'archivage, le stockage et le travail sur les métadonnées, comme nous le disions. Soit parce que le besoin ne semble pas nécessaire de prime abord, soit parce que certaines actions relèvent de pratiques documentaires et ne sont pas constitutives de pratiques de recherche. Ces actions sont en fait induites par les besoins mentionnés précédemment.

Le GT Données trouve pertinent la poursuite des travaux visant le développement d'une infrastructure de services et d'outils pour la gestion des données de la recherche à Sciences Po. Les raisons sont de trois ordres : Sciences Po est propice au développement d'une telle infrastructure. Le nombre de chercheurs et de disciplines représentés est restreint et nous bénéficions de l'expérience et de l'expertise du CDSP et du consortium Archipolis. Le moment est favorable : le contexte national n'est pas figé et l'expérimentation est possible ; les retours d'expérience et les outils développés dans les autres pays constituent un socle solide. Il y a un véritable enjeu stratégique à la mise en place d'une infrastructure qui mènerait Sciences Po à devenir un des moteurs de la question pour les SHS en France.

---

## III. 2016, année de l'expérience

Le GT Données a mis en place un guide thématique<sup>68</sup> Données de la recherche, à la manière de ceux déjà mis en place par la DRIS sur l'Open Access<sup>69</sup> ou la science politique<sup>70</sup> et sur le modèle des guides développés par d'autres institutions<sup>71</sup>.

Ce guide<sup>72</sup> propose une sélection de ressources pour les personnes s'intéressant à la question des données de la recherche et notamment leur traitement au sein de l'institution.

Les constats établis à l'issue de l'année 2015 doivent être confortés ou infirmés. C'est pourquoi nous proposons de prolonger notre étude afin de pouvoir poser les bases d'une feuille de route en travaillant sur les points suivants :

### A. Explorer encore

#### 1. Typologie

Poursuivre la réalisation de la typologie dans la lignée de ce qui a été commencé afin d'affiner et de quantifier les objets produits. Il n'y a pas de méthodologie globale à développer, mais il faut s'adapter aux pratiques ou aux souhaits des directeurs d'unité et de leurs équipes, tel que cela a été mentionné dans les entretiens.

#### 2. Enquête des besoins

En parallèle, une enquête de besoins des équipes et des directions des laboratoires sera menée. Il s'agit d'identifier, qualifier et mesurer les besoins sur la base des premiers constats réalisés lors des rencontres avec les laboratoires.

#### 3. Entrepôts

Étudier finement les entrepôts existants afin d'avoir une cartographie complète des possibilités proposées aux chercheurs. Une étude des politiques des éditeurs en SHS vis-à-vis des données

---

<sup>68</sup> Liste des guides thématiques disponibles : <http://sciencespo.libguides.com/>

<sup>69</sup> <http://sciencespo.libguides.com/openaccess>

<sup>70</sup> <http://sciencespo.libguides.com/science-politique>

<sup>71</sup> Liste des guides thématiques consacrés aux données VI. Annexes

<sup>72</sup> <http://sciencespo.libguides.com/donneesdelarecherche>

---

des auteurs et des dispositifs techniques qu'ils ont pu mettre en œuvre devra également être menée.

#### 4. Les scénarios pour Sciences Po et les conditions de leur mise en œuvre

Le groupe propose de réaliser des scénarios de mise en œuvre de ces voies afin de pouvoir estimer les moyens nécessaires à leur développement. Le groupe s'appuiera sur les projets potentiellement pilotes identifiés lors du tour des centres de recherche<sup>73</sup>.

## B. Mettre en place

### 1. Formation doctorale

Il s'agit de rassembler des acteurs issus de corps professionnels variés qui ont en commun un intérêt pour les données de la recherche. Académiques, archivistes, bibliothécaires et documentalistes s'intéressent de plus en plus à cette question. Cet intérêt est cependant marqué, ce qui est tout à fait logique et légitime, par le point de vue de chacun de ces corps professionnels: les archivistes s'intéressent à la préservation des données de la recherche, les bibliothécaires veulent faire le lien entre publications et données, les chercheurs veulent accéder aux données pour faire de l'analyse secondaire ou pour l'enseignement et les documentalistes regardent de près les questions de formats et métadonnées. L'objectif de la journée d'étude est d'ouvrir le dialogue entre ces corps professionnels pour développer une approche qui intègre ces points de vue et qui décloisonne les initiatives qui sont en train d'émerger sur cette question.

Fiche projet : [VI. Annexes](#)

---

<sup>73</sup> Projets de recherche pilotes :

- CREATE (entretiens individuels (audio), entretiens groupes (audio), rapports)
- Politique de la vulnérabilité épistémique à l'ère du nucléaire (Chaire d'excellence junior en études de sécurité USPC) (archives)
- Ricardo (archives)
- Travaux sur la politique étrangère américaine sous la présidence Obama (articles de presse)
- Choosing Faith and Facing Race: White Converts to Islam in France and the United States (entretiens - transcription)

## 2. Journée d'étude

Dans l'écosystème actuel de la recherche scientifique, un chercheur se distingue autant par la qualité de son travail que par la visibilité qu'il parvient à lui donner. Cette visibilité est en grande partie le fruit de ses choix de publications, de diffusion et de préservation des résultats comme des données qu'il met en œuvre. Cette formation a pour objectif de transmettre aux jeunes chercheurs l'ensemble des éléments nécessaires à la compréhension des stratégies de valorisation documentaire des résultats (publications et données) qui serviront leur carrière.

Fiche projet : [VI. Annexes](#)

## C. Donner un nouveau cadre au projet

Afin d'impliquer au mieux la communauté académique de Sciences Po et pour dynamiser les espaces de réflexion et de décision, le GT Données préconise la mise en place d'une nouvelle organisation des groupes de travail. Proposition : 3 instances :

### 1. Comité de pilotage (COPIL Données)

Composé des représentants des trois directions, du coordinateur du projet et d'invités, il veille au fonctionnement du projet, analyse et arbitre les propositions stratégiques sur la base des propositions du comité stratégique. Il fixe les grandes orientations, les calendriers et identifie les investissements nécessaires. C'est l'organe de décision. Il pourrait se réunir tous les deux mois.

### 2. Comité scientifique et technique (CST Données)

Ce comité est le représentant de la communauté scientifique<sup>74</sup>, ses membres réfléchissent aux grandes questions, parfois techniques, proposent des solutions et des orientations au comité de pilotage. Il pourrait se réunir deux fois par an.

### 3. Groupes de travail

Il s'agit du groupe opérationnel actuel. Les groupes de travail seront mouvants : il s'agit de mobiliser des membres du comité stratégique pour effectuer les travaux. Le GT Données a lancé ce fonctionnement en multigroupes autour de l'organisation de la journée d'étude, de la formation doctorale et de la production d'une documentation. Ces groupes de travail se

---

<sup>74</sup> Statuts, type d'expertise, genre, unité de rattachement, communauté, etc.

réunissent une à trois fois par mois selon les échéances et les livrables. Sa composition peut être éventuellement remaniée.

#### 4. **Coordinateur**

Anime, organise, pilote et coordonne le projet.

## IV. Informations générales

### A. La carte d'identité du projet Gestion des données de la recherche

#### Début du projet

Janvier 2015.

#### Objectif

Étudier l'opportunité de créer une infrastructure de services et d'outils pour les équipes de recherche de Sciences Po.

#### Missions 2015

Réalisation d'un panorama des projets et initiatives de gestion des données en France et en Europe.

Etablissement d'une typologie et d'un recueil de besoins au sein des laboratoires de Sciences Po.

#### Pilotage du projet : groupe exploratoire

Stéphane Auzanneau (DSI), Bénédicte Barbé (DS), Alexandre Biotteau (chercheur, LIEPP), François Cavalier (DRIS), Élise Chapoy (DRIS), Odile Gaultier-Voituriez (Coordination archives), Olivier Godechot (chercheur, OSC), Émilie Groshens (CDSP), Laurent Lesnard (CDSP), Christine Musselin (DS), Cynthia Pedroja (DRIS), Mirna Safi (chercheuse, OSC).

#### Exécution des travaux : groupe opérationnel

Élise Chapoy, Odile Gaultier-Voituriez, Émilie Groshens, Laurent Lesnard et Cynthia Pedroja.

#### Coordination

Cynthia Pedroja.

## V. Ressources

### Débuter

- Contat, Odile, « Gestion des données de la recherche : quelques pistes pour démarrer, comprendre et se former », CORIST-SHS, 2014. [http://corist-shs.cnrs.fr/gestion\\_donnees\\_recherche\\_guideetformation](http://corist-shs.cnrs.fr/gestion_donnees_recherche_guideetformation) [consulté le 2 décembre 2014]
- « Gestion des données de la recherche : quelques pistes pour démarrer, comprendre et se former - Episode 2 », CORIST-SHS, 2014. [http://corist-shs.cnrs.fr/gestion\\_donnees\\_recherche\\_guideetformation\\_2](http://corist-shs.cnrs.fr/gestion_donnees_recherche_guideetformation_2) [consulté le 2 décembre 2014]

### Texte de référence

- *Déclaration de Berlin sur le libre accès à la connaissance* (22 octobre 2003) <http://openaccess.mpg.de/Berlin-Declaration>

### Formation

- Research Data Management Training, University of Edinburgh (section bibliothécaires) <http://datalib.edina.ac.uk/mantra/>
- *Sensibilisation aux enjeux et bonnes pratiques en matière de gestion, partage et réutilisation des données de la recherche sous forme d'interventions en présentiel ou à distance*. INIST, septembre 2014. [http://www.inist.fr/donnees/co/Donnees\\_recherche\\_web.html](http://www.inist.fr/donnees/co/Donnees_recherche_web.html)

### Définitions "données de recherche"

- Pomart, Julien, (18/07/2014) AAF / Section Aurore : Un groupe de travail sur les données de la recherche (billet de blog), sur *Archives de la FMSH* <http://archivesfms.hypotheses.org/1209> (consulté le 14/01/2015)
- *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*. OCDE, avril 2007. <http://www.oecd.org/sti/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>

### BSN10 - Données de la recherche

- André, Francis BSN10 : Données de recherche (Journées DIRBUIST, 2-4 juin 2014) <http://www.enseignementsup-recherche.gouv.fr/cid80646/journees-dirbuist-2014.htm>
- Genova, Françoise, Le partage des données de la recherche : un changement de paradigme (Journées DIRBUIST, 2-4 juin 2014) <http://www.enseignementsup-recherche.gouv.fr/cid80646/journees-dirbuist-2014.htm>

## Commission européenne - H2020

- Lignes directrices pour la gestion des données dans Horizon 2020  
[http://openaccess.inist.fr/IMG/pdf/lignes\\_directrices\\_pgd\\_horizon\\_2020\\_tr\\_fr.pdf](http://openaccess.inist.fr/IMG/pdf/lignes_directrices_pgd_horizon_2020_tr_fr.pdf)
- Lignes directrices pour le libre accès aux publications scientifiques et aux données de recherche dans Horizon 2020  
[http://openaccess.inist.fr/IMG/pdf/14086\\_lignes\\_directrices\\_la\\_horizon\\_2020\\_tr\\_fr\\_version-oct2014-2.pdf](http://openaccess.inist.fr/IMG/pdf/14086_lignes_directrices_la_horizon_2020_tr_fr_version-oct2014-2.pdf)
- *Open access to research data: the Open Research Data Pilot (H2020)*  
<https://www.openaire.eu/oa-in-h2020/h2020/h2020-oa-data-pilot>

## Archivage

- Massol, Marion, Larrousse, Nicolas, Jacobson, Michel, « La question de l'archivage des données de la recherche en SHS ». <https://halshs.archives-ouvertes.fr/halshs-01025106/document> (consulté le 02/12/2014)

## Acteurs et veille

- Digital Curation Center (DCC) <http://www.dcc.ac.uk/>
- Données de la recherche <http://www.donneesdelarecherche.fr>
- European Data Infrastructure (EUDAT) <http://www.eudat.eu/what-eudat>
- Research Data Alliance (RDA) <https://rd-alliance.org/>

## VI. Annexes

- 1) Fiche communication
- 2) Liste des entretiens menés hors de Sciences Po
- 3) Liste et résumé des entretiens menés au sein de l'établissement
- 4) Programme prévisionnel de la formation doctorale
- 5) Programme prévisionnel de la journée d'étude
- 6) Liste des LibGuides sur les données de la recherche
- 7) Note "2017, l'expérimentation grandeur nature"
- 8) DMP
- 9) Syllabus de la formation doctorale
- 10) Page d'accueil du Guide sur les données de la recherche
- 11) Note descriptive de l'entrepôt de données de Sciences Po

## DONNÉES DE LA RECHERCHE UN PROJET INSTITUTIONNEL

La question de la gestion des données de recherche est aujourd'hui un enjeu majeur pour les sciences de l'homme et de la société. Elle est le fruit de plusieurs mouvements qui accompagnent la recherche en France et à l'international parmi lesquels : l'*Open Access* ou le libre accès à l'information scientifique et technique, l'administration de la preuve scientifique afin de valider les résultats, la réutilisation des jeux de données, la valorisation et la sauvegarde du patrimoine scientifique.

La Direction scientifique (DS), la Direction des systèmes d'information (DSI) et la Direction des ressources et de l'information scientifique (DRIS) ont lancé en 2015 un projet sur les données de la recherche afin d'étudier l'opportunité de créer une infrastructure de services et d'outils pour les équipes de recherche.

### ACTIONS

#### 2015 : explorer

- Etude du contexte national et européen à travers une série de rencontres avec différents acteurs et experts de la question.
- Réalisation d'un panorama des types de données produites et des pratiques en matière de gestion au sein des laboratoires de Sciences Po (rencontres avec les directeurs d'unités, les chercheurs et les ingénieurs).
- Étude des entrepôts de données disciplinaires, institutionnels et thématiques mis à disposition des chercheurs afin d'en évaluer la pertinence et l'adéquation avec les besoins des équipes.

#### 2016 : sensibiliser

##### Journée d'étude

Il s'agit de rassembler des acteurs issus de corps professionnels variés qui ont en commun un intérêt pour les données de la recherche.

##### Formation doctorale (module 3h)

La formation sera centrée sur les différentes modalités, stratégies et conditions préalables à la diffusion des résultats de la recherche avec un focus particulier sur la gestion, la valorisation et la conservation des données de la recherche au cours et à l'issue des processus de recherche et de publication. Les enjeux scientifiques, bibliométriques et socio-économiques du libre accès aux résultats scientifiques, pour le chercheur, seront au cœur de la formation. Ils seront complétés par une approche plus pratique des modalités de dépôts et outils disponibles.

## GROUPES DE TRAVAIL

### Groupe exploratoire

#### Missions

Définir des objectifs et des travaux, établir les priorités, valider les travaux du groupe opérationnel

#### Directions et structures de Sciences Po impliquées

Direction scientifique, Direction des systèmes d'information et la Direction des ressources et de l'information scientifique, Centre de données socio politiques (CDSP), Coordination archives

#### Fonctionnement

Réunions semestrielles.

### Groupe Opérationnel

#### Missions

Réalisation des travaux et production de livrables.

#### Fonctionnement

Réunions mensuelles.

## CONTACT

#### Cynthia Pedroja (coordinatrice)

DRIS - Appui à la recherche et outils numériques  
cynthia.pedroja@sciencespo.fr / 01 45 49 59 98

## 2) Liste des entretiens menés hors de Sciences Po

Nom	Contacts	Fonction	Date RV	P1	P2	Type d'acteur
BSN10	Francis André	coordinateur	31/03/2015	Laurent	Cynthia	TGIR
CNRS-INIST	Marie-Christine Jacquemot	responsable Valorisation des données	16/11/2015	Élise Odile	Cynthia	Institut
CNRS-INSHS	Didier Torny	DAS section 36, 40, pôle IST	03/03/2015	Laurent	Cynthia	Institut
CNRS-INSHS	Michèle Dassa	responsable IST et RIBAC	27/03/2015	Odile	Cynthia	Institut
INRA	Odile Hologne	directrice d'unité Délégation IST	24/03/2015	Élise	Cynthia	EPST
Isore	Catherine Fauveaud	coordinatrice	16/04/2015	Odile	Cynthia	Réseau
Sciences Po Grenoble	Marc Sabin	directeur Documentation et recherche	11/03/2015	Élise	Cynthia	IEP
TGIR Huma-Num	Olivier Beaude Stéphane Pouyllau	directeur directeur technique	29/01/2016	Émilie	Cynthia	TGIR
TGIR PROGEDO	Pascal Buléon	président	01/04/2015	Laurent	Cynthia	TGIR
Université Paris Descartes Université Paris Diderot	Aurore Cartier Valérie Neouze Magalie Moysan Nathalie Reymonet	responsable des Services à la recherche, SCD directrice du SCD archiviste, Bureau des archives documentaliste, DARI	15/04/2015	Élise	Cynthia	Université
Sciences Po (USPC)	Amélie Antoine-Audo Marie Ferrazzini	déléguée générale pour Sciences Po déléguée de la documentation	14/04/2015	Laurent	Cynthia	COMUE
INED	Chantal Cases	directrice	07/04/2015	Laurent	Cynthia	Autre
MESR	Jacques Dubucs	directeur scientifique du dept. SHS de la DGRI	25/03/2015	Laurent	Cynthia	Autre
USPC	Serge Bauin	chargé de mission Documentation scientifique	12/05/2015	Élise	Cynthia	COMUE
UNISTRA	Adeline Rège	chef de projet Archives ouvertes	14/10/2015	Élise Odile	Cynthia	Université

## 3) Liste et résumés des entretiens menés au sein de l'établissement

### Centre d'histoire - 01/10/2015

**Présents :**

Centre d'histoire :

- Béatrice Dedinger, ingénieur d'étude
- Marc Lazar, directeur
- Paul-André Rosental, directeur adjoint à la recherche

GT Données : Odile Gaultier-Voituriez, Laurent Lesnard, Cynthia Pedroja

**Lieu :** Centre d'histoire

#### **Ce que l'on peut retenir**

les données ne sont pas seulement statistiques.

Il est important de valoriser le travail de recherche, mais la contextualisation des données, l'accès aux données souvent construites sont complexes.

Comment gérer les questions de droit ? Il faut des compétences dans l'institution.

Il faut investir dans les publics de doctorants et de masters.

---

## CERI - 21/09/2015

### Présents :

#### CERI :

- Alain Dieckhoff, directeur
- Ewa Kulesza (EK), directrice exécutive
- Stefan Seiler, secrétaire général
- Dorian Ryser (DR), documentaliste et cartographe, en charge des ressources

numériques

GT Données : Elise Chapoy, Laurent Lesnard, Cynthia Pedroja

**Lieu** : CERI

### **Ce que l'on peut retenir :**

le plus important est d'informer et de sensibiliser à la gestion de données

> problématique largement méconnue des chercheurs, voire des ingénieurs ;

les doctorants - plus sensibles aux outils et au maniement des données - sont une cible importante > DR informera les doctorants de la prochaine offre de formation CFDIP lors de la réunion de rentrée du 6 octobre ;

les projets financés du CERI (20 à 30) feront l'objet d'une recension orientée données > EK et DR produisent ce bilan pour début décembre 2015 (21/12/2016 : 5 projets identifiés) ;

lors de l'AG du CERI de mi-janvier 2016, la gestion de données et les travaux du groupe REDA seront à l'ordre du jour pour 1 h d'intervention (12/01/2016 : intervention de 20 minutes) ;

repérage d'un deux projets pour expérimentation avec le groupe de travail (projet de Stéphanie Balme évoqué + 1 projet débutant) ;

Attirer les chercheurs vers la gestion de données (a minima pour l'archivage, au mieux pour la diffusion) en les outillant pour assurer la pérennité de la conservation. Le plus important est l'offre de services.

## CEVIPOF - 16/10/2015

**Présents :**

CEVIPOF :

- Martial Foucault, directeur
- Madani Cheurfa, secrétaire général

GT Données : Odile Gaultier-Voituriez, Laurent Lesnard, Cynthia Pedroja

**Lieu :** CEVIPOF

### Ce que l'on peut retenir

Les données doivent être mises à disposition, dans le respect des cadres légaux et des embargos. Une politique institutionnelle serait un levier pour les directeurs de laboratoire.

Seules les données servant le laboratoire doivent être mises à disposition. Il faut donc choisir, par la voie du consensus, quelles sont ces données.

Les doctorants et étudiants de master sont un public à privilégier ; les outils

---

## École de Droit - 26/10/2015

**Présents :**

École de droit :

- Frédéric Audren,
- Christophe Jamin,
- Anne-Solenne de Roux

GT Données : Elise Chapoy, Laurent Lesnard, Cynthia Pedroja

**Lieu :** École de droit

### Ce que l'on peut retenir

La question des données est hors champ en droit si on parle de données produites et non utilisées.

## médialab - 03/11//2015

### Présents :

médialab : Diego Antolinos Basso, Audrey Baneyx, Babara Bender, Davy Braun, Paul Girard, Biljana Jankovic, Christophe Leclercq, Vincent Lépinay, Benjamin Ooghe, Guillaume Plique, Benoît Verjat, autres membres du laboratoires dont des étudiants

GT données : Élise Chapoy, Odile Gaultier-Voituriez, Émile Groshens, Laurent Lesnard, Cynthia Pedroja

**Lieu** : médialab, 13 U

### Ce que l'on peut retenir

Importante production de données (médialab et autres centre de recherche).  
Travail sur la valorisation et la réutilisation de données.  
Intérêt pour le développement d'une offre de services pour l'archivage, la documentation, la diffusion (droits, CNIL) ; ces compétences leur manquent.

Le médialab peut soutenir le projet pour les questions d'exploration des données et l'Open Data.

---

## OFCE - 14/12/2015

### Présents

#### OFCE :

Amel Falah : responsable des bases de données à l'OFCE

Xavier Timbeau : directeur principal de l'OFCE (=directeur scientifique)

#### GT Données :

Elise Chapoy, Odile Gaultier-Voituriez, Laurent Lesnard, Cynthia Pedroja

**Lieu :** OFCE

### **Ce que l'on peut retenir**

L'OFCE est consommateur avant d'être producteur de données. Les données produites sont des agrégats et l'OFCE ne possède pas forcément les droits pour une gestion (stockage, archivage, diffusion).

Achat de données coûte cher : possibilité d'une politique institutionnelle (mutualisation) ?

Absence d'une compétence informatique pour traiter les problèmes de fond (par exemple : connexion aux données statistiques).

Besoin d'une aide juridique pour mesurer les risques (réutilisation, diffusion, etc.) et savoir quelles sont les procédures administratives.

---

## OSC - 29/09/2015

### Présents:

#### OSC :

- Solenne Brun : doctorante
- Bernard Corminboeuf : ingénieur de recherche documentation
- Marta Dominguez : associate professor (quantitatif et qualitatif, famille, genre)
- Marie Ferrazzini, secrétaire générale
- Juliette Galonnier : doctorante (qualitatif, ethnologie)
- Ilam Madji : assistante ingénieure statisticienne
- Marco Oberti : directeur OSC, professeur (inégalités urbaines, quanti/quali)
- Ettore Recchi : professeur (unités spatiales, inégalités)
- Anne Revillard, OSC-LIEPP : associate professor (qualitatif, handicap, genre, inégalités)
- Yannick Savina : ingénieur d'études (données quantitatives, cartographie)
- Louis-André Vallet : directeur de recherche

GT Données : Emilie Groshens, Laurent Lesnard, Cynthia Pedroja

**Lieu** : OSC

### **Ce que l'on peut retenir**

Les données produites ne sont pas uniquement des données d'enquêtes qualitatives et quantitatives.

Il faut former les doctorants, mais également les Masters, voire les personnels administratifs de Sciences Po (pour réutiliser des données de l'institution).

4 personnes souhaitent participer aux premiers travaux pour la typologie.

Il y a une urgence à rassurer les chercheurs sur ce qu'ils peuvent ou ne peuvent pas faire dans la gestion de leurs données (archivage, diffusion, etc.).

## 4) Programme prévisionnel de la formation doctorale

Formation doctorale - juin 2016

Document de travail - janvier 2016

### Accroche

La gestion des données de la recherche : quel(s) intérêt(s) pour un jeune chercheur ? L'objectif de cette formation est de proposer des outils concrets et des clefs de réflexion pour la gestion des données dans le cadre d'une thèse, et au-delà.

### Présentation

Cette formation propose de s'intéresser à la gestion et la valorisation des données de la recherche au cours et à l'issue des processus de recherche et de publication. En nous appuyant sur le plan de gestion de données (Data Management Plan) adopté par l'USPC, nous veillerons à expliciter les différentes étapes qui constituent le cycle de vie des données (préparation, récolte, traitement, diffusion, archivage et réutilisation). L'objectif est de donner à comprendre les enjeux scientifiques, documentaires et techniques de la gestion de données en les réintroduisant dans des processus de recherche. L'angle d'attaque envisagé est celui de la documentation (démarche scientifique, métadonnées, formats pérennes, etc.) des données afin d'en assurer la préservation et la réutilisation pour leur usage personnel dans un premier temps. Les enjeux sociétaux, scientifiques, financiers et pédagogiques seront également évoqués.

### Comités

#### Organisation

Laurent Lesnard, Marie-Sophie Dumont, Cynthia Pedroja

Intervenants pressentis

Laurent Lesnard, Marie-Sophie Dumont, Odile Gaultier-Voituriez, Claire Lemerrier, membre beQuali

### Programme

#### Orientations envisagées :

1. enjeux : être efficace pour publier, des outils pédagogiques pour l'enseignement, administration de la preuve, réutilisation, etc.
2. bonnes pratiques : documenter la production et l'analyse par l'enrichissement (métadonnées, expliquer les choix, etc.), utiliser les formats pérennes
3. valoriser, exposer, accéder)

ou

A. moi (le doctorant) :

B. la communauté (scientifique, société, pédagogie, financeurs) : enjeux pour la réutilisation, la réanalyse, l'administration de la preuve, outils d'enseignement, etc.

C. limites : juridiques, risques, pourquoi (ne pas) partager ?

### Dates envisagées

Juin 2016

### Budget

Nul

---

## 5) Programme prévisionnel de la journée d'étude

### Données de la recherche - Journée d'étude 2016

Document de travail - janvier 2016

#### Présentation (Laurent Lesnard)

Il s'agit de rassembler des acteurs issus de corps professionnels variés qui ont en commun un intérêt pour les données de la recherche. Académiques, archivistes, bibliothécaires et documentalistes s'intéressent de plus en plus à cette question. Cet intérêt est cependant marqué, ce qui est tout à fait logique et légitime, par le point de vue de chacun de ces corps professionnels : les archivistes s'intéressent à la préservation des données de la recherche, les bibliothécaires veulent faire le lien entre publications et données, les chercheurs veulent accéder aux données pour faire de l'analyse secondaire ou pour l'enseignement et les documentalistes regardent de près les questions de formats et métadonnées. L'objectif de la journée d'étude est d'ouvrir le dialogue entre ces corps professionnels pour développer autour des données de la recherche une approche qui intègre ces points de vue et qui décloisonne les initiatives qui sont en train d'émerger sur cette question. Cette journée est la première d'une série.

#### 1. *Accroche*

Poser les bases d'une réflexion au sein de l'institution et au-delà.  
Susciter l'intérêt pour le projet Données de la recherche, commencer la formation des membres de la communauté scientifique.

#### 2. *Comités*

#### Pilotage

Groupe de travail exploratoire du projet Données de la recherche : Christine Musselin, Bénédicte Barbé, François Cavalier, Stéphane Auzanneau, Olivier Godechot, Mirna Safi, Alexandre Biotteau.

#### Organisation

Claire Lemerrier, Laurent Lesnard, Cynthia Pedroja, édition Sciences Po (CERI), et un informaticien externe (Paris) profil traitement des données

#### 3. *Programme*

#### Matin (ateliers, 5 à 6 personnes)

Les ateliers sont l'occasion de faire travailler séparément les différentes communautés professionnelles sur :

- notions de gestion de données (diffusion, traitement, archivage, documentation, archivage, DMP, etc.) ;
  - enjeux (personnels, institutionnels, sociétaux) ;
  - types de données traitées (typologie) ;
  - leurs pratiques et la manière dont ils perçoivent, comprennent les pratiques et les attentes des autres communautés.
-

L'objectif de ces travaux est de pouvoir identifier l'état des connaissances et des pratiques des différentes communautés, mais également de pouvoir mettre en exergue *a priori* des communautés les unes vis-à-vis des autres. Les groupes identifiés à ce jour sont : chercheurs, doctorants, archivistes, informaticiens, chargés d'édition, documentalistes / bibliothécaires.

Veiller à la représentativité des disciplines chez les chercheurs et doctorants.

Pour chaque atelier, une personne animera les débats (étudiants Master journalisme ? Doctorants ? autres ?). Ces ateliers seront préparés en avance par les organisateurs avec les animateurs.

## Après-midi (plénière)

- restitution des ateliers métiers
- restitution de la typologie des données réalisée par le groupe de travail
- conférence(s) / tables rondes : réintroduire la question des données dans une perspective disciplinaire plus large et internationale. Ouvrir à d'autres thèmes dont le data-journalisme et l'utilisation des données de recherche dans un contexte journalistique, Open Data. À cette occasion, les questions d'éthique et de droit pourront être évoquées.

### 4. *Restitution*

Delphine Lereculeur, chef de projet appui à la publication scientifique à la DRIS, est responsable de la restitution web qui prendra une forme plus dynamique que des actes uniquement textuels.

Idée à creuser : impliquer des doctorants (rémunération ?) pour des restitutions des débats.

### 5. *Budget*

## Gestionnaires :

DS : Vincent Morandi / DRIS : Flavien Graziani

## Prévisionnel :

50% DS / 50% DRIS

<b>Journée d'étude</b>		<b>4 200€</b>
transport (5 pers.)		600€
hébergement (5 pers.)		600€
repas (pause café, buffet debout, 60 pers.)		2 500€
communication		500€

## Calendrier (2016)

février	- présentation projet au GT exploratoire - réunion 1 comité d'organisation
mars	- réunion budget
avril	- réunion 2 comité d'organisation
juin	- réunion 3 comité d'organisation - pré-communication (date, intentions)
septembre	- communication (programme, diffusion)
décembre	Journée d'études

## 6) Liste des LibGuides sur les données de la recherche

- **About - Research Data Management - Subject Guides at John Jay College of Criminal Justice, CUNY**

<http://jjay.v1.libguides.com/content.php?pid=447569>

"Resources and tools for managing your research data. (For faculty and graduate students.)"

- **About this guide - Research data across the lifecycle - The Library's Subject Guides! at American University**

<http://american.v1.libguides.com/content.php?pid=550921>

"An overview of considerations, resources and tools for working with data in your research projects"

- **Data Management Defined - Research Data Management - Guides at Texas A&M University**

<http://guides.library.tamu.edu/content.php?pid=320939>

"This guide covers principles of data management and data management planning, along with summaries of various agency requirements, links to example data management plans, and pointers to the best tools and resources around."

- **Data Repositories - Research Data Services - WSU Libraries Guides at Wayne State University**

<http://wayne.v1.libguides.com/content.php?pid=327769>

"Information for discovering, sharing, and managing research data."

- **Enquête sur les pratiques et les besoins à l'égard de la gestion des données de recherche - Guides de recherche / Research guides at University of Ottawa**

<http://uottawa.v1.libguides.com/content.php?pid=435308&sid=3562006>

- **Get Started - Research Data Management - UQ Library Guides at University of Queensland Library**

<http://uq.v1.libguides.com/content.php?pid=547615>

"Research Data Management covers the planning, collecting, organising, managing, storage, security, backing up, preserving, and sharing your data. "

- **Getting Started - Research Data Services - Research Guides at University of Kentucky Libraries**

<http://uky.v1.libguides.com/content.php?pid=608982>

"Tools and information for researchers to help with data management, data preservation, and data sharing"

- **Home - Research Data Management Curriculum Guide - Research Guides at University of Washington Libraries**

---

<http://guides.lib.washington.edu/content.php?pid=637322>

"A guide to resources to support research data management (RDM) instruction, including curricula, mailing lists, conferences and more."

- **Home - Research Data Services - Library Guides at Arizona State University**

<http://asu.v1.libguides.com/content.php?pid=359860>

"Designed for the ASU Faculty and other researchers to become familiar with the growing literature on research data services generally and what we have at ASU."

- **Home - Research Data Services - Subject Guides at Syracuse University Library**

<http://syr.v1.libguides.com/content.php?pid=340558>

"SU Libraries offers a broad range of research data services related to the identification, collection, management, analysis, and curation of quantitative and qualitative research data."

- **Home - Research Data: Finding it and Using it - LibGuides Sandbox for Library Schools at Springshare**

<http://libraryschool.libguidescms.com/content.php?pid=464498>

"This guide will identify popular Canadian microdata resources available to the members of the Carleton University Social Science community. As well, resources to help with data analysis are identified. "

- **Introduction - Research Data Sources - Research Guides at Singapore Management University**

<http://researchguides.smu.edu.sg/content.php?pid=400946>

"This guide highlights the various types of data available in databases, and reduces the time navigating within the database. "

- **Open Repositories and Research Data - Claremont Research Guides at Claremont Colleges Library**

<http://claremont.v1.libguides.com/content.php?pid=90265>

"Resources to find, manage, use, and share data and research. Please send me suggestions of resources to add!"

- **Overview - Research Data Management - Research Guides at Florida Institute of Technology**

<http://libguides.lib.fit.edu/content.php?pid=402634>

"Help with data management plans"

- **Overview - Sensitive Research Data Management - LibGuides at Purdue University Libraries**

<http://purdue.v1.libguides.com/content.php?pid=219836>

"A guide for addressing issues with sharing research data involving human subjects or other sensitive data sets."

- **Research Data Management - Library Guides at University of the Sunshine Coast**

<http://libguides.usc.edu.au/content.php?pid=345268>

---

"Managing and sharing the data that fuels research."

- **Research Data Management - Subject Guides at University of Victoria Libraries**

<http://uvic-ca.v1.libguides.com/content.php?pid=567688>

This guide is intended to support researchers who want to effectively manage their data. It is equally important for researchers to understand new requirements from grant funding agencies for managing research data and the research data life-cycle.

UVic Libraries can help researchers with data management. For guidance in writing a data management plan or with data curation, please contact your subject specialist.

- **Research Data Management Toolkit - Library Guides at Bond University**

<http://bond.libguides.com/content.php?pid=430932>

"This guide is designed as an introduction to research data and the management of data for the research community at Bond University."

- **The basics - Manage Your Created Research Data - Research Guides at University of Waterloo Library**

<http://subjectguides.uwaterloo.ca/content.php?pid=662042>

"Curate and store the research data you create."

- **Why Manage Research Data? - Research Data Management - LibGuides at UCD Library**

[http://libguides.ucd.ie/data/why\\_manage](http://libguides.ucd.ie/data/why_manage)

"Bringing together University resources and services to facilitate researchers in the production of high quality data"

### Gestion des données de la recherche 2017, l'expérimentation grandeur nature

D'un projet initié en 2015 à la définition de services et d'outils en 2016, la DRIS propose de réaliser en 2017 une expérimentation grandeur nature afin d'éprouver le programme Gestion des données de la recherche. Ce document définit les grandes lignes de cette expérimentation qui se fera en coordination étroite avec l'ensemble des acteurs de la question à Sciences Po : DS et DSI en premier lieu en tant que copilotes, mais également les centres de recherche producteurs de données, les chercheurs et ingénieurs ou encore les projets et infrastructures tels le CDSP et Archipolis, etc.



Début 2018, un bilan de cette expérimentation sera proposé dans le cadre d'une journée d'étude interne à l'institution. Les échanges permettront à la fois d'ajuster le programme au plus près des besoins de la communauté de recherche et d'initier une discussion collective.

#### Gouvernance et communication institutionnelle

Depuis deux ans les groupes de travail opérationnel et exploratoire ont mené les travaux d'exploration et défini le programme Gestion des données de la recherche.

En 2017, la restructuration des instances permettra de dynamiser les espaces de réflexion et de décision en impliquant au mieux les centres de recherche.

- COPIL : réunissant les directeurs de la DRIS, la DS, la DSI et du CDSP, ainsi que le coordinateur du programme, il veille au fonctionnement du programme, analyse et arbitre les propositions stratégiques sur la base des propositions du CST. Il fixe les grandes orientations, les calendriers et identifie les investissements nécessaires. C'est l'organe de décision.
- CST (comité scientifique et technique) : représentant la communauté scientifique, ses membres réfléchissent aux grandes questions, parfois techniques, proposent des solutions et des orientations au comité de pilotage.

Divers supports de communication seront produits (guide thématique, bibliographie sélective, fiches de présentation, etc.)

#### Actions mises en place dès janvier 2017

##### Entrepôts de données

Installation d'un entrepôt Sciences Po<sup>1</sup> pour stocker et diffuser les données. Fruit d'une collaboration entre la DRIS et la DSI et nous appuyerons sur l'expérience menée par le CDSP dans le cadre du consortium Archipolis.

<sup>1</sup> Dataverse (<http://dataverse.org/>), solution développée par Harvard

## SciencesPo

DONNÉES DE LA RECHERCHE - 2017, l'expérimentation grandeur nature

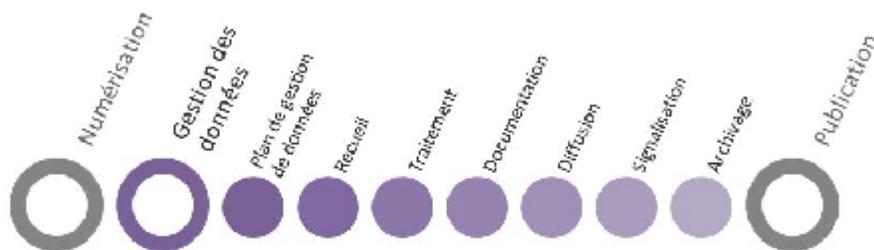
### Formation et sensibilisation

Développement de l'offre de formation et d'actions de sensibilisation avec, par exemple l'accompagnement à la gestion des données pour les chercheurs et doctorants, la montée en compétence des personnels DRIS, ou encore l'établissement de bonnes pratiques entre ingénieurs.

### 10 projets pilotes

De la numérisation, lorsque les supports sont nativement papier, à la publication, la DRIS propose d'intégrer la gestion des données dans une chaîne de traitement de contenus de formes variés.

L'atelier données de la recherche proposera une gamme de services, dès le début de la recherche :



Chaîne de services proposée par la DRIS

Afin d'éprouver les services proposés par l'atelier, 10 au moins projets seraient pris en charge, dont :

Contact	Centre	Projet	Type de données	Service(s) rendu(s)	Début
Charlotte Halpern	CEE	CREATE	entretiens (audio) rapports	Traitement Documentation Diffusion Signalisation Archivage	février
Benoît Pelopidas	CERI	Politique de la vulnérabilité épistémique à l'ère du nucléaire	archives	Numérisation Traitement Documentation Diffusion Signalisation Archivage	janvier
Béatrice Dedinger	CHSP	Ricardo	archives	Documentation Diffusion Signalisation Archivage	janvier
Mario Del Pero	CHSP	(travaux sur la politique étrangère américaine sous la présidence Obama)	articles de presse	Documentation Diffusion Signalisation Archivage	mars
Juliette Galonnier	OSC	Choosing Faith and Facing Race: White Converts to Islam in France and the United States	entretiens (transcription)	Traitement Documentation Diffusion Signalisation Archivage	septembre

Ces projets ont identifié lors d'entretiens menés pour établir les besoins des chercheurs.

## SciencesPo

DONNÉES DE LA RECHERCHE - 2017, l'expérimentation grandeur nature

---

Sur 6 entretiens, 5 ont donné lieu à une programmation de prise en charge de données.

4 entretiens sont programmés avant la fin 2016 (Jérôme Aust, Marie-Emmanuelle Chessel, Richard Banégas, Catherine Cavalin). Ce liste est amenée à évoluer. Dernière liste à jour disponible :

<https://docs.google.com/a/sciencespo.fr/spreadsheets/d/19iCQvHdvn3zIPZWmSSMHRIS46pC3C1tBpIDX-4ivoyw/edit?usp=sharing>

7) Note "2017, l'expérimentation grandeur nature"

## 8) DMP

### Sciences Po: Sciences-Po-template (fr) - Informations générales

#### INFORMATIONS GÉNÉRALES SUR LE DMP

Historique des mises à jour du DMP

*Recommandations :*

Aide à la saisie : par ordre antéchronologique.

Responsable(s) du DMP

*Recommandations :*

Aide à la saisie : personne à contacter.

Modèle : Nom, Prénom, Institution, Laboratoire, Ville, Pays

Exemple : Goguel, François, CNRS, Centre de recherches politiques (CEVIPOF, UMR 7048), Paris, France

Rédacteurs et relecteurs du DMP

*Recommandations :*

Aide à la saisie : ensemble des personnes ayant participé à la rédaction ou relecture du plan de gestion de données.

Modèle : Nom, Prénom, Institution, Laboratoire, Ville, Pays

Exemple : Goguel, François, CNRS, Centre de recherches politiques (CEVIPOF, UMR 7048), Paris, France

Date d'échéance pour la remise du DMP

*Recommandations :*

---

Aide à la saisie : dans le cadre d'un projet dans lequel le DMP est un livrable, la date d'échéance est celle figurant dans le calendrier du projet. À titre d'exemple, pour les projets H2020, le DMP qui doit être envoyé dans les 6 mois après le début effectif du projet.

Date de soumission effective

*Recommandations :*

Aide à la saisie : dans le cadre d'un projet dans lequel le DMP est un livrable, mentionner la date d'envoi effectif au destinataire.

Niveau de confidentialité du DMP

*Recommandations :*

Exemple : Confidentiel, seulement les membres de la Commission européenne.

## INFORMATIONS GÉNÉRALES SUPPLÉMENTAIRES SUR LE DMP POUR LES PROJETS STRUCTURÉS EN WORKPACKAGE

Numéro du workpackage (WP)

*Recommandations :*

Exemple : WP8

Titre du workpackage

*Recommandations :*

Exemple : Coordination and Management

Responsables du workpackage

*Recommandations :*

Modèle : Nom, Prénom, Institution, Laboratoire, Ville, Pays

---

Exemple : Goguel, François, CNRS, Centre de recherches politiques (CEVIPOF, UMR 7048), Paris, France

Numéro de livrable

*Recommandations :*

Exemple : WP8.5

## PRÉSENTATION DU PROJET DE RECHERCHE

Description et objectifs du projet de recherche

*Recommandations :*

Aide à la saisie : résumé du projet.

Organisme financeur du projet

*Recommandations :*

Aide à la saisie : pour les projets avec financement.

Exemples : Agence nationale pour la recherche (ANR), Commission européenne

Référence de la convention de financement

*Recommandations :*

Aide à la saisie : si la convention est numérotée, ajouter la référence de la convention.

Exemple : 741718

Date de début du projet

*Recommandations :*

---

Aide à la saisie : date de début du financement ou date de début des travaux pour les projets sans financement

Date de fin du projet

*Recommandations :*

Aide à la saisie : date de fin du financement ou des travaux pour les projets sans financement

Coordinateur(s) scientifique(s) du projet

*Recommandations :*

Modèle : Nom, Prénom, Institution, Laboratoire, Ville, Pays

Exemple : Goguel, François, CNRS, Centre de recherches politiques (CEVIPOF, UMR 7048), Paris, France

Partenaires du projet

*Recommandations :*

Aide à la saisie : structures partenaires du projet (universités, entreprises, associations, etc.).

Modèle : Institution, Laboratoire, Ville, Pays

Exemple : CNRS, Centre de recherches politiques (CEVIPOF, UMR 7048), Paris, France

Site web du projet

*Recommandations :*

Modèle : <http://www.siteprojet.fr>

## Sciences Po: Sciences-Po-template (fr) - Gestion des données

### DESCRIPTION GÉNÉRALE

Description de l'ensemble des données

*Recommandations :*

Aide à la saisie : expliquer la contribution des données produites et utilisées aux objectifs du projet de recherche.

### DONNÉES PRODUITES DANS LE PROJET

Mode de collecte et type de données

*Recommandations :*

Exemples :

comme type de données : données du recensement, enregistrements audio et transcriptions, photographies d'archives, etc. ;

comme mode de collecte : questionnaire, entretien semi-directifs, note d'observation, etc.

Fréquence de collecte des données

*Recommandations :*

Exemples :

données collectées mensuellement, annuellement, etc. ; données collectées deux fois dans le cadre du projet.

Volume de l'ensemble données produites

*Recommandations :*

---

Aide à la saisie : indiquer au début de projet, un volume prévisionnel de données produites au cours du projet. En fin de projet, reporter le volume réel à la place du volume prévisionnel préalablement saisi.

Exemples : [X] Mégaoctets, [n] Téraoctets, d'autres unités de grandeur peuvent être utilisées comme le nombre de fichiers, la durée des enregistrements, etc.

Formats de fichiers utilisés

*Recommandations :*

Aide à la saisie : pour connaître les formats de fichiers à privilégier, consulter le guide Données de la recherche. Exemples : image (JPEG, PNG, TIFF), texte (doc, plain text (TXT), HTML, XML, PDF), audio (AIFF, WAVE), base de données (XML, CSV)

Logiciels et matériel utilisés pour la production de données

*Recommandations :*

Exemples : LimeSurvey, Qualtrics, Hyphe

## DONNÉES TIERCES UTILISÉES DANS LE PROJET

Données et autres sources auxquelles vous avez recours

*Recommandations :*

Aide à la saisie : si vous utilisez des données déjà existantes, indiquez lesquelles et leur origine.

Exemples :

1. Musselin, Christine, Friedberg, Erhard (2018) : Comparaison des ministères de l'Enseignement supérieur de France et d'Allemagne 1987-1990 (cdsp\_bq\_s4) [data set], Centre de données socio-politiques, beQuali. [https://doi.org/10.21410/dshs\\_2016/p9kv-f227](https://doi.org/10.21410/dshs_2016/p9kv-f227)
2. Déclaration annuelle des données sociales (DADS) : postes et salariés au 1/25e - 1996, INSEE (producteur), ADISP (diffuseur)
3. Archives départementales du Gard, série F125

## 4. Revue de presse disponible sur Zotero

## STOCKAGE ET PROTECTION DES DONNÉES

Lieu et mode de stockage des données le temps du projet

*Recommandations :*

Exemples :

les données sont stockées dans le Google Drive Sciences Po (Google Apps for Education, sous contrat institutionnel) ;

les données sont stockées sur un disque dur externe dans mon bureau.

Règles de nommage et de classement des données

*Recommandations :*

Aide à la saisie : indiquer les règles de nommage et d'organisation des données mises en place dès le début du projet afin de faciliter la recherche puis la mise à disposition des données.

Pour plus d'informations, consulter la page Organiser du guide [Données de la recherche](#)

Utilisateurs des données durant le projet

*Recommandations :*

Aide à la saisie : indiquer la liste des personnes qui ont accès aux données mobilisés pendant toute la durée du projet.

Modèle : Nom, Prénom, Institution, Laboratoire, Ville, Pays

Exemple : Goguel, François, CNRS, Centre de recherches politiques (CEVIPOF, UMR 7048), Paris, France

Propriétaire(s) des données

*Recommandations :*

---

Pour en savoir plus, contacter Rémi Pignal de la DAJAM

Mesures et solutions mises en œuvre pour garantir la protection des données, notamment en matière de confidentialité

*Recommandations :*

Informations importantes :

dans le cas où vous produisez ou utilisez des données à caractère personnel, fournir le numéro de déclaration de traitement des données (accès au formulaire) ;

pour les autres données, en particulier celles présentant un risque pour la sûreté de l'État, la sécurité publique, la protection du secret de la défense nationale, la sécurité de l'établissement, le secret professionnel, etc. indiquer tous les dispositifs de protection mis en œuvre.

## ACCÈS AUX DONNÉES

Accessibilité des données

*Recommandations :*

Aide à la saisie : pour chaque type de données, indiquer les conditions d'accès pour les personnes extérieures au projet de recherche.

Données	Condition d'accès	Modalités	Temporalité	Contractualisation
	Restrictions			

La liste est établie dans le paragraphe C.1

Libre/restreint/privé	Comment les données sont-elles accessibles
?	Quand les données sont-elles accessibles
?	Licence pour faciliter la réutilisation ou contrat de protection des données Expliciter les restrictions et les embargos

Exemple :

Données	Conditions d'accès	Modalités	Temporalité	Contractualisation
	Restrictions			

---

## Notes d'observation

### Restreint

Entrepôt de données institutionnel Sciences Po - À la fin du projet

- Après un embargo de... mois
- Jamais

Contractualisation au cas par cas.

Données personnelles et sensibles ne pouvant faire l'objet d'une anonymisation complète. Le porteur du projet octroiera les accès au cas par cas, sur justificatif de recherche.

Utilisation de standards d'identification de type identifiant unique

### *Recommandations :*

Aide à la saisie : indiquer si les données rendues accessibles peuvent être trouvées grâce à un standard d'identification.

Exemple : Chaque jeu de données est identifié grâce à un DOI

Pour en savoir plus, consulter la page « Héberger et publier » du guide [Données de la recherche](#)

## Versionning des jeux de données

### *Recommandations :*

Aide à la saisie : les données rendues accessibles sont-elles versionnées ? En cas de mise à jour des données (correction d'une erreur dans les métadonnées, ajout de nouvelles données), leur numéro de version sera-t-il incrémenté afin de

pouvoir retrouver les modifications ?

Thèmes ou mots-clés

*Recommandations :*

Aide à la saisie : mots décrivant le projet et permettant d'optimiser la visibilité des données et séparer par des points virgules.

Métadonnées permettant de décrire vos données

*Recommandations :*

Aide à la saisie : indiquer la manière dont les données seront décrites. L'utilisation d'un standard de métadonnées est recommandée.

Les données produites par les centres de recherche de Sciences Po peuvent être décrites par le schéma Data Documentation Initiative (DDI). Pour en savoir plus, consulter le guide [Données de la recherche](#)

## RESSOURCES POUR LA GESTION DES DONNÉES

Acquisition et installation de matériel spécifique

*Recommandations :*

Exemples : serveurs, machines virtuelles, logiciels dédiés, licences supplémentaires, etc

Recrutement et formation

*Recommandations :*

Aide à la saisie : préciser les recrutements de personnel et les formations envisagées pour assurer la gestion des données du projet.

Exemple : Un ETP de [n]% est budgété pour couvrir la gestion, la documentation et la curation des données

Coûts

*Recommandations :*

Aide à la saisie : sur la base des informations ci-dessus, préciser le surcoût lié à la gestion et la diffusion des données en précisant les modalités de prise en compte de ce coût supplémentaire. Dans certains projets financés, les coûts sont éligibles au financement.

Pour en savoir plus, contacter la Mission d'appui aux projets (MAPS) de la Direction scientifique

## 9) Syllabus de la formation doctorale

### Data management

#### A short / intensive course for PhD students

Principal instructors: Cynthia Pedroja & Nicolas Sauger

*Description* : If your PhD project involves working and even more collecting data, data management is key for success. This course offers a review of most pressing issues when dealing with data. Organization is key for efficient work. But data management goes way beyond. A number of demands are going to guide your use of data. Journals and funders request more and more often to make data accessible and data processing available for review. Regulations organize what kind of regulation applies to research data, how it can be collected, stored and shared. But data is now an asset in its own right. It is a way to become cited and recognized. It is a way to achieve important research results.

This course covers the critical areas related to research data management: licensing and intellectual property, metadata and contextual description, obtaining informed consent for participation in research, anonymising research data for reuse, data storage, and data archiving. It introduces key contemporary concepts such as open science and FAIR principles. It also guides participants in the discovery of available tools for better data management.

This course is not a course about the methods of data collection or analysis.

This course is proposed by Sciences Po's library and the CDSP (Center for Social-Political Data). Cynthia Pedroja and Nicolas Sauger will be the principal instructors for this course, attending all sessions. A variety of instructors from the library, CDSP, medialab, IT and Legal Department will participate in animating sessions over the course of this programme. Cynthia Pedroja is project manager in research data management at the Library. Nicolas Sauger is Professor of political science at Sciences Po. He is the director of the CDSP and has participated in large scale projects of data collection such as the European Social Survey and the Comparative Study of Electoral Systems. He has a long experience with quantitative and experimental data.

*Format*: This course is organized in 5 sessions in 2 days. Its combines lectures, presentations, exercises and discussions. Participants will have begun their *data management plan* for their thesis during this course.

*Sessions:*

## **A. Session 1 : introduction to data management and DMPs**

This session introduces the importance of data management. Key focus are: open science and FAIR principles; the added value of managing efficiently data; data lifecycles. It will discuss the role of good data management in terms of scientific integrity and re-use of the data. It will also show how to access existing data. It will also cover the Data management plan principles and tools. It will introduce the main sections of DMP and how to fill them. These tools will serve as general guidelines for the rest of the course.

## **B. Session 2: organizing and cleaning data**

This session will include a theoretical part followed by a practical one built around concrete case studies to experience good practices and pitfalls in terms of data management. We will discuss the issues of versioning, how to aggregate data, file formatting, data transformation, anonymisation, how to structure your data according to your research question, *etc.*

Instructors:

- Audrey Baneyx, Data manager (medialab)
- Selma Bendjaballah, Research engineer (CDSP)
- Simon Le Corgne, Data manager (CDSP)

## **C. Session 3: Security, regulations and ethics**

This session raises the question of how to deal with data collection and data use in the right way. It will review requirements and policies, as set especially in GDPR. It will introduce the issue of licensing of data and intellectual property (“IP”) rights. The ethics of data collection and data storage will be also thoroughly covered, from informed consent to anonymisation. It will question data safety and integrity. It will finally consider the issue of the security of data and for instance introduce best practices in terms of encryption.

Instructors:

- Marion Lehmans - Data protection officer
- Laurence Gagelin - Information Systems Security Officer (DSI - IT Department)
- Rémi Pignal Legal council - Intellectual property (DAJAM - The legal department)

## **D. Session 4: sharing, storing, and preserving research data**

This session will have a closer look at how to preserve and distribute data. It will cover metadata and their role. It will also review the question of physical and digital preservation of data. It will set out the difference between preservation and long-term archiving. It will introduce key archival standards and trusted repositories. We will do practical exercise to depose a dataset in the future Sciences Po Repository.

Instructors:

- Alina Danciu, Data manager (CDSP)
- Cynthia Pedroja, Project manager in research data management (Library)

## **E. Session 5 : wrap-up and conclusion**

This session concludes the course and provides a final review of data management strategy of participants.

### *References and sources:*

Henderson, M.E. 2016 *Data management*, New York, Rowman & Littlefield.

Plotkin, D. 2013 *Data stewardship*. Burlington, M. Kaufman.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3. Queensland University of Technology. (2013). Management of Research data. Retrieved from

[http://www.mopp.qut.edu.au/D/D\\_02\\_08.jsp](http://www.mopp.qut.edu.au/D/D_02_08.jsp)

EPSRC, Engineering and Physical Sciences Research Council. Scope. Retrieved from

[www.epsrc.ac.uk/about/standards/researchdata/scope/](http://www.epsrc.ac.uk/about/standards/researchdata/scope/)

Crossley, J.; Stewart, J. (2013). Managing research data. Retrieved from

<http://www.scribd.com/doc/138079216/Managing-Research-Data>

University of Sheffield. (2013, September 2). Research data management: what is research data? [blog].

Retrieved from <http://researchdatamanagement.blogspot.co.uk/2013/09/what-is-research-data.html?m=1>

Scott, M. (2013). Introducing Research Data, University of Southampton. Retrieved from

[https://eprints.soton.ac.uk/403440/1/introducing\\_research\\_data.pdf](https://eprints.soton.ac.uk/403440/1/introducing_research_data.pdf)

Burnham, A. (2013). An introduction to managing research data. University of Leicester. Retrieved from

<http://www2.le.ac.uk/services/research-data/documents/an-introduction-to-managing-research-data>

University of Leicester. (2012). Research data: definitions. Retrieved from

[http://www2.le.ac.uk/services/research-data/documents/UoL\\_ResearchDataDefinitions\\_20120904.pdf](http://www2.le.ac.uk/services/research-data/documents/UoL_ResearchDataDefinitions_20120904.pdf)

[https://www.CESSDA.eu/content/download/3845/35038/file/20170707\\_How\\_FAIR\\_are\\_your\\_data\\_Jones.pdf](https://www.CESSDA.eu/content/download/3845/35038/file/20170707_How_FAIR_are_your_data_Jones.pdf)

<http://www.dcc.ac.uk/resources/how-guides/license-research-data>

<https://www.cnil.fr/fr/ce-que-change-la-loi-pour-une-republique-numerique-pour-la-protection-des-donnees-personnelles>

<https://www.CESSDA.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/3.-Process/Data-authenticity>

<https://www.CESSDA.eu/Research-Infrastructure/Training/Expert-Tour-Guide-on-Data-Management/3.-Process/Data-entry-and-integrity>

<http://openaccess.couperin.org/comment-deposerpublier-des-donnees/>

[https://eprints.soton.ac.uk/403440/1/introducing\\_research\\_data.pdf](https://eprints.soton.ac.uk/403440/1/introducing_research_data.pdf)

---

## 10) Page d'accueil du Guide sur les données de la recherche

### DONNÉES DE LA RECHERCHE

BIENVENUE | LES DMP | GÉRER SES DONNÉES - RÉGLEMENTATION ET ÉTHIQUE | HÉBERGER ET PARTAGER - FAQ | ALLER PLUS LOIN | CONTACTS

English version

## DONNÉES DE LA RECHERCHE

### CHERCHEURS, INGÉNIEURS, SUIVEZ LE GUIDE !

Au cœur de la production scientifique, les **données de recherche** sont variées par leur forme (texte, image, son, vidéo), leur support (papier, analogique ou numérique), leur format (carnet, cassette, fichier, base de données structurées, etc.) et leur état (brutes, traitées ou dérivées).

Comment assurer la lisibilité de vos données dans le temps, d'un projet à l'autre ? Comment répondre aux exigences des financeurs pour la livraison d'un plan de gestion de données ? Comment partager vos données afin que vos pairs puissent comprendre au mieux vos recherches ? Autrement dit, comment rendre visibles vos travaux ?

En optimisant leur gestion ! Chercheurs, ingénieurs, voici les premières réponses à vos interrogations.

#### SOMMAIRE

- DMP Sciences Po
- Gérer ses données
  - Traiter ses données
  - Organiser ses fichiers
  - Documenter les processus
- Réglementation et éthique
- Héberger et partager
  - Pendant le projet
  - Après le projet
- FAQ
- Aller plus loin
- Contacts

#### OUTILS ET MODÈLES

- DMP Sciences Po
  - FR
  - UK
- Engagement de confidentialité
  - FR
- Analyse d'impact sur la vie privée
  - FR

#### SCIENCES PO, UNE HISTOIRE DE DONNÉES

Depuis 2015, Sciences Po développe une gamme de services et d'outils pour la gestion des données de la recherche, dans la continuité de l'archive ouverte institutionnelle SPIRE.

Le projet est mené par la Direction des ressources et de l'information scientifique (DRIS), en collaboration avec la Direction scientifique (DS), la Direction des systèmes d'information (DSI) et le Centre de données socio-politiques (CDSP) et les centres de recherche de Sciences

#### ACTUALITÉS

##### Ressource

**Feuille de route du CNRS pour la science ouverte**

À l'occasion des deuxièmes journées nationales de la science ouverte qui se sont tenues les 18 et 19 novembre dernier à Paris, Antoine Petit, président-directeur général du CNRS, a annoncé la feuille de route de l'organisme dans ce domaine.

Quatre axes : l'accès ouvert aux publications scientifiques, le partage des données scientifiques, le développement et la promotion d'outils pour la fouille des contenus scientifiques, et enfin, l'évaluation individuelle des chercheurs.

Données personnelles : formalités éthiques

- FR

Déclarer un traitement de données

- FR

Formulaire de consentement

- FR
- UK

2019 — Développer

2018 — Ouvrir

2017 — Expérimenter

## 11) Note descriptive l'entrepôt de données de Sciences Po

# NOTE DE PRÉSENTATION DE *DATA.SCIENCESPO*, L'ENTREPÔT DES DONNÉES DE LA RECHERCHE DE SCIENCES PO

### Contexte

Depuis plusieurs années, la Direction des ressources et de l'information scientifique (DRIS) travaille conjointement avec la Direction scientifique (DS) et le Centre des données socio-politiques de Sciences Po (CDSP) à l'élaboration d'une politique de gestion des données de la recherche et à la mise en place d'un entrepôt visant à préserver et à valoriser les données issues de recherches en sciences humaines et sociales, menées notamment à Sciences Po.

Ce projet de fournir un accès gratuit et universel aux produits de la recherche s'inscrit dans le mouvement de la science ouverte dont une étape fondamentale a été franchie, au niveau national, en juillet 2018 avec le lancement du [plan national pour la science ouverte](#) par Frédérique Vidal, Ministre de l'Enseignement supérieur, de la Recherche et de l'Innovation.

Ce contexte national de progressive mise en œuvre de principes présidant en 2016 à la rédaction de la [Loi pour une République numérique](#), qui sont aujourd'hui repris et déclinés par des acteurs nationaux de la recherche comme le CNRS ([feuille de route science ouverte, novembre 2019](#)) et l'ANR ([obligation de fourniture de DMP dans le cadre de sa politique de science ouverte](#), mars 2019), est particulièrement favorable au déploiement de l'entrepôt des données de la recherche de Sciences Po prévu pour le début de l'année 2020.

### Objectifs de *data.sciencespo*

Les principaux objectifs de l'entrepôt des données de la recherche, nommé en 2019 [data.sciencespo](#), sont les suivants :

- valoriser en ouvrant, autant que possible dans le respect des lois en vigueur, les données produites à Sciences Po dans le cadre de projets de recherche ;
- valoriser en ouvrant, autant que possible, les données produites dans le cadre de projets de recherche en SHS ;
- permettre la consultation, voire l'accès et la réutilisation des données diffusées, entre pairs comme au plus grand nombre ;
- fournir un moyen de préserver les données de la recherche (sans pour autant constituer un système d'archivage pérenne) ;
- constituer un entrepôt de référence pour les données en SHS à l'échelle de l'institution mais également à l'échelon national et européen.

Ces objectifs, qui relient la mission locale de Sciences Po et l'ambition nationale du CDSP, se traduisent dans une organisation de l'entrepôt par collections :

- Collection réservée aux membres de la communauté scientifique de Sciences Po (sous la forme d'auto-dépôt), administrée par la DRIS ;
- Collection réservée aux enquêtes documentées et diffusées par le CDSP (données provenant de l'entrepôt Nesstar et du site beQuali) ;
- Collection ouverte aux données en SHS (lancement prévu dans un second temps).

## Public de *data.sciencespo*

L'application *data.sciencespo* vise à être consultable en ligne par le plus grand nombre.

Néanmoins, le dépôt de jeux de données dans l'entrepôt est réservé aux :

- membres de la communauté académique de Sciences Po (pour ce qui concerne la collection de l'entrepôt destinée aux personnels de Sciences Po et administrée par la DRIS) ;
- ingénieurs du CDSP (pour ce qui concerne la collection de l'entrepôt destinée aux enquêtes dans le domaine des SHS, produites principalement par des institutions françaises et confiées au CDSP pour documentation et diffusion) ;
- aux chercheurs en SHS souhaitant valoriser leurs données via *data.sciencespo* (prochainement).

## Solution logicielle

La solution logicielle libre [Dataverse](#), développée et maintenue par l'université de Harvard, qui a été testée par la DRIS et le CDSP pour les projets ArchiPolis (catalogue signalant des archives collectées et au service de la science politique) et DataSPIRE (plateforme expérimentale préfigurant *data.sciencespo*), a été identifiée comme solution adaptée pour la mise en place de *data.sciencespo*.

## Équipe projet

- Genevieve Michaud (responsable de l'équipe informatique du CDSP), cheffe de projet ;
- Eleonora Moiraghi (cheffe de projets numériques, DRIS), responsable du déploiement de la collection destinée à la communauté académique de Sciences Po ;
- Alina Danciu (chargée d'études, CDSP), responsable de la collection réservée au CDSP ;
- Rémi Pignal (juriste droit privé, DAJAM), aide à la rédaction des [conditions générales d'utilisation et du contrat de dépôt](#) ;

Le travail de cette équipe projet est supervisé par deux niveaux de gouvernance :

- les Directeurs François Cavalier (DRIS) et Nicolas Sauger (CDSP) ;
- le comité pour le traitement des données et la science ouverte réunissant les directeurs de la DRIS, du CDSP, de la DS et de la DSI ainsi que d'autres acteurs majeurs de la science ouverte au sein de l'établissement.

## Maintenance et alimentation

Le CDSP, en collaboration avec la DSI de Sciences Po, est en charge de l'implémentation et de la maintenance technique de *data.sciencespo*.

La DSI assure également l'hébergement *in situ* de l'application.

La DRIS est pour sa part responsable de la gestion de la collection destinée à la communauté académique de Sciences Po et de l'accompagnement des usagers de cette collection.

## Contenus et fonctionnalités principales de *data.sciencespo*

### ▪ Types de ressources

L'application *data.sciencespo* a vocation à accueillir des jeux de données produits dans le cadre de projets de recherche. Par exemple des corpus de transcriptions d'entretiens, des statistiques, des corpus d'images de terrain.

Un jeu de données peut être constitué de plusieurs fichiers (ex. fichiers tabulaires, images, textes, code informatique, etc.) dans des formats différents (ex. .png, .jpg, .docx, .rtf, etc.). L'application accepte une variété de formats considérable. Néanmoins, les formats ouverts et documentés sont recommandés.

### ▪ Recherche

L'application *data.sciencespo* permet une recherche simple et avancée dans toutes les collections et jeux de données ainsi qu'un affinage des résultats via des facettes et des options de tri.

### ▪ Téléchargement

L'application *data.sciencespo* permet le téléchargement des fichiers en accès ouvert rattachés à un jeu de données.

### ▪ Partage

L'application *data.sciencespo* permet le partage d'une notice d'un jeu de données sur les réseaux sociaux.

---

## ▪ **Authentification**

La connexion au back-office de l'application *data.sciencespo* se fait via la fédération d'annuaires de RENATER.

## ▪ **Gestion des utilisateurs**

L'attribution des profils de droits aux utilisateurs est réservée aux administrateurs de collections et peut se faire au niveau de chaque collection ou sous-collection.

Un profil spécifique a été créé pour la collection gérée par la DRIS. Ce profil permet à un utilisateur de publier un jeu de données sans passer par l'étape de validation afin de fluidifier le dépôt.

## ▪ **Gestion du dépôt**

L'application *data.sciencespo* permet aux usagers de déposer des jeux de données dans les collections pour lesquelles ils ont les droits.

Déposer un jeu de données signifie essentiellement :

- documenter un jeu de données en remplissant le formulaire de métadonnées proposé ;
- ajouter les fichiers composant le jeu de données décrit ;
- fixer les conditions d'accès et d'utilisation des fichiers ajoutés.

Un modèle de formulaire de métadonnées est proposé pour la collection gérée par la DRIS. Il comprend les métadonnées suivantes qui reposent sur le schéma standard [DDI](#), version 2:

MÉTADONNÉES	DÉFINITIONS
<b>Title*</b> / Titre	Le titre du jeu de données
<b>Subtitle</b> / Sous-titre	Le sous-titre du jeu de données
<b>Other URL</b> / Autre URL	Une autre URL ou d'autres URL donnant accès au jeu de données ; par exemple la page d'un site personnel ou d'un projet
<b>Author*</b> / Auteur	La ou les personnes physiques ou morales qui ont créé le jeu de données
<b>Contact*</b> / Personne-ressource	La ou les personnes à contacter pour avoir des informations sur le jeu de données, voire pour y avoir accès

---

<b>Description*</b>	Un résumé indiquant le contexte de production du jeu de données et la nature des données
<b>Subject* / Sujet</b>	Le domaine dans lequel le jeu de données peut être classifié
<b>Keyword* / Mot-clé</b> (vocabulaire contrôlé <a href="#">ELSST</a> )	Le ou les mots-clés qui décrivent brièvement les contenus du jeu de données
<b>Topic Classification / Classification par sujets</b> (vocabulaire contrôlé <a href="#">CESSDA Topic Classification</a> )	Les grands thèmes couverts par le jeu de données

<b>Related Publication / Publication connexe</b>	La ou les publications liées au jeu de données
--	--

<b>Notes / Remarques</b>	Toute information sur le jeu de données ne pouvant pas entrer dans un des champs proposés
<b>Language* / Langue</b> (ISO 639-1)	La langue du jeu de données
<b>Producer / Producteur</b>	La ou les personnes physiques ou morales qui détiennent la responsabilité financière ou administrative du jeu de données
<b>Contributor / Collaborateur</b>	La ou les personnes physiques ou morales ayant contribué à la constitution du jeu de données
<b>Grant Information / Renseignements sur la subvention</b>	Informations sur le financement (financeur.s et financement.s)
<b>Distributor* / Diffuseur</b>	La personne morale désignée par l'auteur ou le producteur pour la mise à disposition de copies du jeu de données y compris sous la forme de révision ou édition

---

<b>Depositor / Déposant</b>	La personne physique ou morale qui a fait l'acte de dépôt dans <i>data.sciencespo</i> (elle peut coïncider avec l'auteur mais peut aussi être différente comme dans le cas d'un assistant de recherche ou documentaliste)
<b>Deposit Date / Date de dépôt (ISO 8601)</b>	La date du dépôt dans <i>data.sciencespo</i>
<b>Time Period Covered / Période couverte (ISO 8601)</b>	Période couverte par les données à distinguer des dates de collecte et de traitements
<b>Date of Collection / Date de collecte (ISO 8601)</b>	Dates entre lesquelles les données ont été collectées
<b>Kind of Data / Type de données (vocabulaire contrôlé <a href="#">DDI Alliance General Data Format</a>)</b>	Type.s des données présent.s dans le jeu de données
<b>Software / Logiciel</b>	Solution.s logicielle.s utilisée.s pour créer ou traiter le jeu de données

<b>Related Material / Documentation connexe</b>	Tout document lié au jeu de données
<b>Related Datasets / Ensembles de données connexes</b>	Tout jeu de données ayant un lien avec le jeu de données
<b>Data Sources / Sources de données</b>	Tout document qui a servi de source pour la constitution du jeu de données
<b>Origin of Sources / Origine des sources</b>	Réservé aux documents historiques pour indiquer le contexte de production des sources
<b>Geographic Coverage / Couverture géographique (ISO3166 2-letter)</b>	La couverture géographique du jeu de données
<b>Universe / Univers</b>	La/ les population.s couverte.s par le jeu de données

---

---

<b>Time method / Méthode temporelle</b> (vocabulaire contrôlé <a href="#">DDI Alliance Time method</a> )	La méthode temporelle adoptée
<b>Sampling procedure / Méthode d'échantillonnage</b> (vocabulaire contrôlé <a href="#">DDI Alliance Sampling procedure</a> )	La procédure adoptée pour la constitution de l'échantillon
<b>Collection mode / Mode de collecte</b> (vocabulaire contrôlé <a href="#">DDI Alliance Mode of collection</a> )	La méthode de collecte des données
<b>Research instrument / Type d'instrument de recherche</b> (vocabulaire contrôlé <a href="#">DDI Alliance Type of instrument</a> )	Le type d'instrument utilisé pour la collecte des données

La normalisation des champs a été principalement traitée par Eleonora Moiraghi et Alina Danciu en collaboration avec Alexandre Mairot, ingénieur expert des métadonnées du CDSP, et Clément Siberchicot, bibliothécaire chargé des métadonnées rattaché à la DRIS. Le modèle de métadonnées de référence utilisé pour la normalisation des champs est le suivant :

[Borschewski, Kerrin, Förster, André, Friedrich, Tanja, Zenk-Möltgen, Wolfgang,](#)

[Miranda, Patrícia, Moura Ferreira, Pedro, ... Bolton, Sharon. \(2019\). CMM CESSDA](#)

[Metadata Model \(Version 1.0, DOI:10.5281/zenodo.3236171\)](#)

Cette liste de métadonnées et les vocabulaires contrôlés associés ont fait l'objet de tests utilisateurs conduits par Eleonora Moiraghi avec la collaboration de deux membres du groupe Databibs (Muriel Dennefeld, bibliothécaire référente en histoire à la DRIS, et Anna Egea, ingénieure d'études au Centre de Sociologie des Organisations, CSO) entre octobre et novembre 2019.

Concernant les conditions d'accès aux fichiers, trois choix sont possibles dans l'application :

- accès ouvert (en cas de présence de données à caractère personnel, ce type d'accès suppose une étape préalable d'anonymisation des données) ;
- accès sur demande ;
- accès privé (avec possibilité de partage d'URL privée).

La licence [CC-BY](#), proposée par défaut, fixe les conditions d'utilisation. Néanmoins, il est possible d'apporter des informations complémentaires ou d'attribuer une autre licence, voire une licence *ad hoc*.

La DRIS a pour mission d'assurer l'accompagnement des usagers affiliés à Sciences Po notamment dans le processus de dépôt brièvement décrit.

## ▪ **Gestion de la publication**

Un dépôt peut être modifié à plusieurs reprises. Une fois le dépôt finalisé, il est possible, pour un utilisateur de la collection gérée par la DRIS, de le publier directement sur l'interface publique de l'application sans passer par une étape de validation.

Comme l'action de publication génère automatiquement un identifiant unique et pérenne DOI, les modifications après publication sont affichées sous forme de versions. Toutes les versions, antérieures et postérieures à des modifications, restent visibles après publication.

## ▪ **Moissonnage et visibilité**

Les métadonnées renseignées dans *data.sciencespo* sont moissonnables via le protocole OAI-PMH par d'autres entrepôts ou services. Outre la visibilité permise grâce au moissonnage (ex. catalogue de données [CESSDA](#)), *data.sciencespo* a vocation à être visible dans le répertoire international d'entrepôts de données [re3data](#) et dans le répertoire national [Cat OPIDoR](#).

## **Communication autour de *data.sciencespo***

La communication envisagée autour de l'ensemble de l'entrepôt *data.sciencespo* est à distinguer de la communication autour des collections que l'entrepôt contient.

Pour ce qui concerne la communication autour de l'entrepôt, une communication via email de la Direction scientifique de Sciences Po et des présentations dans des centres de recherche ainsi que lors de conférences, sont envisagées.

Concernant la collection gérée par la DRIS, un événement de lancement est envisagé ainsi qu'une communication via les canaux utilisés habituellement par la DRIS : communication interne, actualités sur le site web, guide thématique sur les données de la recherche, fiches imprimées, réseaux sociaux.