

S'AUTOFORMER  
AUX DONNÉES DE  
LA RECHERCHE :  
GUIDE À  
DESTINATION DES  
PROFESSIONNELS  
DE  
L'INFORMATION  
ET DE LA  
DOCUMENTATION

2020



Une production du gTIGRE (groupe de travail inter-établissements Grand Est)



doi : 10.5281/zenodo.4291896

# Les auteurs

- Laetitia Bracco (Université de Lorraine)
- Mathilde Barthe (Médial)
- Stéphanie Cheviron (Université de Strasbourg)
- Agnès Faller (Université de Reims Champagne-Ardenne)
- Madeleine Hubert (Bibliothèque nationale et universitaire de Strasbourg)
- Sylvie Steffann (Université de Haute-Alsace)
- Noël Thiboud (URFIST de Strasbourg)
- Jean-Baptiste Vu Van (Université de technologie de Troyes)

Le groupe remercie chaleureusement Caroline Sophie Donati (Aix-Marseille Université), Élise Lehoux (Université de Paris) et Delphine Du Pasquier (École des Ponts ParisTech) pour leur relecture attentive et leurs commentaires précieux.

## Introduction

Les formations généralistes sur la gestion et le partage des données de la recherche sont un premier socle indispensable aux professionnels de la documentation pour pouvoir accompagner les chercheurs dans ce domaine émergent. Pour ce faire, l'autoformation est aujourd'hui l'un des meilleurs moyens d'acquérir les compétences nécessaires.

Ce guide n'a pas une portée exhaustive : il existe une pléthore de supports d'autoformation sur les données de la recherche en ligne, de qualité variable. Il a vocation à vous aider à choisir, au sein de cette offre, les supports qui vous seront utiles, quel que soit votre niveau de départ. La durée indicative des différentes formations est indiquée lorsque c'est possible. À l'intérieur des différentes catégories, les formations sont classées par ordre alphabétique.

## Table des matières

<b>1. Niveau « introduction »</b>	2
1.1. Découvrir les données de la recherche (en français)	2
1.2. Découvrir les données de la recherche (en anglais)	3
1.3. Se former en présentiel	4
<b>2. Niveau « avancé »</b>	5
2.1. Approfondir ses connaissances sur les données	5
2.2. Comprendre les enjeux juridiques et éthiques	6
2.3. Monter un service de gestion des données de la recherche	6
2.4. Traiter les données	7
2.5. Former les usagers aux données de la recherche	8
2.6. Se former en présentiel	8
2.7. Parcours spécialisés par disciplines	8
2.8. Découvrir les structures régionales autour des données	10

## 1. Niveau « introduction »

Les formations recensées ici s'adressent aux personnes qui découvrent le monde des données de la recherche.

### 1.1. Découvrir les données de la recherche (en français)

- [CooplST](#) (CIRAD)

**Description** : le module « Gérer des données » aborde tous les aspects de la gestion des données de la recherche de manière précise et pédagogique à travers 10 thématiques incluant des liens pour approfondir la lecture.

**Format** : fiches synthétiques disponibles en format PDF.

- [Datapartage](#) (INRAE)

**Description** : le portail propose des fiches répondant à des questions concrètes sur la gestion des données (Comment FAIR en pratique ? Comment publier ? Comment choisir son entrepôt...).

**Format** : contenu en ligne, FAQ.

- [Dorandum](#) (Inist-CNRS)

**Description** : panorama complet de la gestion et du partage des données en 9 modules thématiques et des exercices pour évaluer ses connaissances.

**Formats** : fiches synthétiques, tutoriels, vidéos, quizz.

- [E-learning Open Data](#) (ODI Open Data Institute, European Data Portal)

**Description** : un total de 16 cours autour de l'*open data*. Évaluation sous forme de questionnaire en fin de cours. Le portail aborde également des aspects pratiques sur les données : leur format, leur nettoyage, leur impact...

**Format** : cours en ligne.

- [Projet Train2Dacar](#) (Collaboration suisse : Haute École de Gestion de Genève (HEG-GE) et Fachhochschule Graubünden (FHGR))

**Description** : initiation aux données de la recherche en 9 modules : de base, d'immersion ou didactique. Un des modules fournit des recommandations à destination des formateurs aux données de la recherche.

**Format** : fiches synthétiques disponibles en format PDF.

- [Webinaire « Choisir un dépôt de données de recherche »](#) (Bibliothèque des lettres et sciences humaines de l'Université de Montréal)

**Description** : webinaire réalisé en 2017. Il aborde la question de la donnée, sa typologie, son processus de publication ainsi que sur les entrepôts, sa certification, sa pérennité....

**Format** : webinaire et diaporama de ce même cours sans les commentaires (38 diapos).

**Durée** : le webinaire (captation vidéo) dure 1h07.

- [Webinaire « Citer les données de la recherche »](#) (Bibliothèque des lettres et sciences humaines de l'Université de Montréal)

**Description** : webinaire réalisé en 2018. Il rappelle les enjeux et les principes de la citation des données, donne des exemples de styles bibliographiques...

**Format** : webinaire et diaporama de ce même cours sans les commentaires (16 diapos).

**Durée** : le webinaire (captation vidéo) dure 49 minutes.

## 1.2. Découvrir les données de la recherche (en anglais)

- [DataTree](#) (Natural Environmental Research Council (NERC, UK))

**Description** : cours d'autoformation en ligne complet et gratuit qui aborde toutes les questions relatives à la gestion des données de la recherche, pour les chercheurs et « toutes les personnes amenées à travailler avec les données ». Sur inscription préalable.

**Format** : vidéos, quizz et exemples de gestion de donnée issus de situation concrètes, conseils d'experts en gestion et partage des données. Le cours est conçu et structuré pour un apprentissage progressif des connaissances. Vous pouvez également y piocher des solutions et résoudre des problèmes spécifiques. Un certificat électronique est fourni à la fin de chaque module réussi.

**Durée** : 15 à 20 heures pour la totalité des 8 modules.

- [Open Science basics for KU Leuven library staff](#) (FOSTER)

**Description** : cours en ligne sur les grandes notions de la science ouverte et sur la gestion des données de la recherche. Le portail FOSTER est une plateforme d'e-learning qui rassemble des ressources de formation sur la Science Ouverte. L'inscription sur Foster est nécessaire.

**Format** : contenu en ligne.

- [Research and data management sharing](#) (Université Caroline du Nord Chapel Hill et Université d'Edimbourg)

**Description** : 5 séquences introductives sur les enjeux de la gestion et diffusion des données.

**Format** : MOOC (formation en ligne ouverte à tous).

**Durée** : formation gratuite d'une durée de 16 heures et certification payante. Calendrier variable.

- [Research Data Bootcamp](#) (Université de Bristol)

**Description** : tutoriel (plutôt destiné aux chercheurs) qui aborde les principaux aspects de la gestion des données de la recherche, en anglais. L'approche est générale et généraliste, c'est donc une introduction pour se familiariser avec les données de la recherche, adopter les bonnes pratiques et les appliquer à sa propre recherche.

**Format** : cours en ligne en 8 sections (texte, avec un quizz à la fin de chaque section pour vérifier la compréhension) avec des liens pour aller plus loin.

**Durée** : environ 30 minutes par section, soit 4h en tout.

- [Research Data in context](#) (MANTRA, University of Edinburgh)

**Description** : introduction aux données de la recherche permettant de distinguer les différents types de données, de comprendre l'importance de leur gestion et les grands enjeux.

**Format** : cours en ligne avec vidéos et exercices sur une vingtaine de diaporamas.

### 1.3. Se former en présentiel

Outre les contenus en ligne, les Unités Régionales de Formation à l'Information Scientifique et Technique (URFIST) et les Centres Régionaux de Formation aux Carrières des Bibliothèques (CRFCB) proposent également des formations introductives aux données de la recherche en présentiel qui sont gratuitement accessibles aux personnels des bibliothèques universitaires, mais aussi aux doctorants et chercheurs. En voici des exemples :

- CRFCB Médial : « Kit de survie dans le monde des données de la recherche »
- URFIST : « Gérer et diffuser ses données de recherche : introduction aux enjeux, méthodes et pratiques »

## 2. Niveau « avancé »

Ces formations sont destinées à un niveau plus avancé et permettent d'approfondir des aspects plus spécifiques des données.

### 2.1. Approfondir ses connaissances sur les données

- [Managing and sharing research data](#) (FOSTER)

**Description** : cours en ligne sur l'ouverture des données, le plan de gestion de données, le choix des entrepôts. L'inscription sur Foster est nécessaire.

**Format** : contenu en ligne.

- [Research Data Management Librarian Academy](#) (Elsevier)

**Description** : parcours de formation en 8 modules abordant le cycle de vie de la donnée, les partenariats à nouer autour des données, les premières étapes nécessaires pour monter une offre de services, les plans de gestion de données, la visualisation de données avec R, Python et les notebooks Jupyter.

**Format** : cours en ligne.

**Durée** : environ 2 heures par module, soit 16 heures en tout.

- [UK Data service](#)

**Description** : plateforme proposant guides, tutoriels vidéos et ressources de formation. Les supports sont classés par domaine, et impliquent une connaissance préalable du sujet.

**Format** : guides spécifiques et vidéos généralistes. Au total [29 courtes vidéos](#) en anglais (sous-titrées en anglais) de 1'30 à 9 minutes chacune en fonction des sujets :

- *Accessing Data* (descriptif des types de micro et macro données mises à disposition sur le site UK Data Service)
- *Analysis Tools* (outil propre au site pour la recherche et l'exploitation des données)
- *Why cite Data?* (vidéos généralistes)
- *Deposing Data* (pour le dépôt sur UK Data Service uniquement)
- *Managing Data* (*Write a Data Management Plan*, indications propres au contexte britannique)
- *Teaching with Data* (appliqué à UK Data Service).

Les modules de 2 heures (*Data Skills Modules*) présentent différents cas d'utilisation des données (données d'enquête, données longitudinales et données agrégées).

## 2.2. Comprendre les enjeux juridiques et éthiques

- [Data protection and ethics](#) (FOSTER)

**Description** : cours en ligne en anglais sur la gestion des données personnelles et l'anonymisation des données. L'inscription sur Foster est nécessaire.

**Format** : cours en ligne.

- [Ethiquedroit](#)

**Description** : élaboré par un groupe de travail d'universitaires, c'est un recueil de bonnes pratiques relatives aux différentes étapes du cycle de vie des données.

**Format** : blog.

## 2.3. Monter un service de gestion des données de la recherche

- [23 Things: Libraries for Research Data](#) (Research Data Alliance (RDA))

**Description** : synthèse (en anglais) sur les ressources gratuites et en ligne permettant aux professionnels de l'information de s'engager dans la gestion des données de la recherche.

**Format** : fiches en format PDF.

- [Building Blocks: Laying the Foundation for a Research Data Management Program](#) (OCLC)

**Description** : conseils en anglais pour mettre en place un service de gestion des données de la recherche à partir de zéro.

**Format** : rapport en format PDF.

- [Delivering research data management services](#) (Future Learn)

**Description** : formation en anglais pour aider à la configuration et au développement de services autour des données de la recherche.

**Format** : MOOC.

**Durée** : 5 semaines à raison de 3 heures hebdomadaires. Calendrier variable.

## 2.4. Traiter les données

- [Assessing the FAIRness of data](#) (FOSTER)

**Description** : cours en ligne pour apprendre à évaluer le respect des principes FAIR au sein d'un jeu de données à l'aide d'outils et de ressources librement accessibles. L'inscription sur Foster est nécessaire.

**Format** : cours en ligne.

- [Data carpentry](#)

**Description** : plateforme qui aide à acquérir diverses compétences orientées gestion de données, notamment en visualisation de données (R, Python...). Les formations peuvent avoir lieu en ligne. Sous l'onglet « *lessons* » : les supports qui ont servi aux cours donnés. Très spécifique, ce cours demande un grand niveau d'autonomie technique et permet des approfondissements disciplinaires (écologie, génomique, sciences sociales, données géospatiales...).

**Format** : contenu en ligne, présentations PowerPoint, vidéos.

- [Effectively Managing & Sharing Research Data in Spreadsheets](#) (Johns Hopkins Libraries Data Services)

**Description** : cours pour apprendre à utiliser les tableurs afin d'augmenter les possibilités de réutilisation des données qu'ils contiennent et réduire les risques d'erreur. Formation en anglais.

**Format** : modules et présentations d'outils.

**Durée** : prévoir deux heures : 60 minutes pour les modules et 53 minutes pour la présentation des outils.

- [Planning for software reproducibility reuse](#) (Johns Hopkins Libraries Data Services)

**Description** : cours pour apprendre à créer des logiciels et scripts réutilisables. 6 modules : bonnes pratiques pour écrire et documenter du code ; utiliser une version de contrôle ; lier le code aux résultats de la recherche (crédits) ; rendre le code publiable et citable ; les licences ; l'archivage. Formation en anglais.

**Format** : 6 webinaires courts.

**Durée** : 22 minutes en tout.

- [School of Data](#)

**Description** : plateforme d'autoformation sur la curation de données, pour apprendre à nettoyer des jeux de données, à collecter des données... Certains modules sont très poussés. Le contenu est surtout orienté sur le traitement des données (extraction, collecte, nettoyage...).

**Format** : contenu en ligne.



## 2.5. Former les usagers aux données de la recherche

- [Do-It-Yourself Research Data Management Training Kit for Librarians](#) (MANTRA, University of Edinburgh)

**Description** : boîte à outils pour les formateurs en bibliothèques. Nécessite des connaissances préalables assez poussées.

**Format** : présentations (.ppt ou .pdf), cours en ligne avec vidéos et transcriptions, podcast (15 minutes environ), exercices et formulaires d'évaluation. La formation était conçue à l'origine pour des groupes de doctorants et/ou bibliothécaires avec une séance de formation présentielle de 2 heures pour chaque session et des travaux à réaliser en dehors des sessions. Thèmes des sessions : *Data management planning, Organizing & documenting data, Data Storage & Security, Ethics & copyright, Data Sharing*.

**Durée** : prévoir environ 1h30 à 2 heures pour chaque session.

- [Research management open training materials](#)

**Description** : liste de documents pédagogiques pouvant être réutilisés pour construire des formations à destination de doctorants, de chercheurs... Les documents sont en anglais, ils nécessitent donc à la fois une bonne maîtrise de la langue et une bonne maîtrise du sujet pour pouvoir faire le tri.

**Format** : ensemble de documents déposés sur Zenodo.

## 2.6. Se former en présentiel

Les formations en ligne peuvent être complétées par des formations en présentiel. Voici quelques exemples de formations proposées par le réseau des URFIST et par l'Inist-CNRS :

- Inist-CNRS : [OpidorTour](#), présentation de l'outil DMP OPIDoR qui permet de créer des plans de gestion de données collaboratifs en ligne.
- URFIST : [Gestion et diffusion des données de la recherche](#)
- URFIST : [Données de la recherche : outils et méthodes](#)

Différentes journées d'étude sont organisées régulièrement sur le sujet par divers établissements d'enseignement supérieur et organismes de recherche.

## 2.7. Parcours spécialisés par disciplines

- [Best practices for biomedical research data management](#) (Harvard Medical School)

**Description** : parcours complet de formation aux données de la recherche spécialisé dans le domaine de la biomédecine. Le suivi du parcours donne lieu à certification. En anglais uniquement.

**Format** : MOOC.

- [Big Data to Knowledge](#) (Oregon Health and Science University)

**Description** : autre parcours de biomédecine conçu dans le cadre du programme [Big Data to Knowledge](#) des National Institutes of Health, surtout pour étudiants et chercheurs. Aborde des thèmes allant du Big Data dans les sciences biomédicales à la question des normes d'interopérabilité et à l'annotation ou à la curation des données cliniques en passant par les aspects éthiques et réglementaires, les ontologies, l'utilisation de données secondaires, le partage de données. En anglais uniquement.

**Format** : le cours complet comporte 22 modules qui se déclinent chacun en trois vidéos de 5 à 30 minutes (en fonction du sujet abordé).

**Durée** : compter 45 minutes en moyenne pour chaque module, soit environ 16h30 au total.

- [Dataacc](#) (Universités Claude Bernard Lyon 1 & Grenoble Alpes)

**Description** : plateforme d'accompagnement à la gestion des données de recherche en physique et en chimie. L'onglet « bonnes pratiques » fournit des conseils utiles dans le domaine disciplinaire, avec un focus sur les carnets de laboratoires. En français uniquement pour le moment.

**Format** : fiches pratiques, *serious game*.

- [Data Train](#) (Archaeology Data Service)

**Description** : parcours de formation aux données de la recherche dédié à l'archéologie et réalisé en 2011. Il est composé de 8 modules. En anglais uniquement.

**Format** : cours en format PDF.

- [vads4R](#) (Visual arts data skills for researchers)

**Description** : parcours complet de formation aux données de la recherche destiné aux chercheurs en arts visuels. En anglais uniquement.

**Format** : parcours de formation dynamique en ligne.

## 2.8. Découvrir les structures régionales autour des données

Développer sa culture dans le domaine des données de la recherche, c'est aussi connaître les acteurs de terrain qui fournissent des infrastructures techniques et proposent un accompagnement aux chercheurs.

Dans les établissements d'enseignement supérieur, de plus en plus de services de soutien à la gestion des données sont mis en place, comme on peut le constater en consultant le répertoire [SOS-PGD](#).

De nombreuses plateformes de calcul intensif, ou des plateformes liées à de grands instruments, sont des partenaires incontournables autour des données qu'il convient de connaître. Quelques exemples : le mésocentre [EXPLOR](#) en Lorraine ou le [centre de calcul](#) de l'Université de Strasbourg.

Les Très Grandes Infrastructures de Recherche (TGIR) [Progedo](#) et [Huma-Num](#) proposent des services aux chercheurs en Sciences Humaines et Sociales. Certaines universités disposent d'antennes locales de ces TGIR, localisées dans les Maisons des Sciences de l'Homme (MSH). Les relais locaux de Progedo, par exemple, sont les [plateformes universitaires de données \(PUD\)](#). Sous la direction d'un ou d'une responsable scientifique, un ingénieur statisticien accompagne les chercheurs dans l'utilisation des bases de données quantitatives en sciences sociales et dans leur formation à l'utilisation d'outils statistiques.

Deux PUD exercent en-dehors d'une MSH à Paris :

- Au sein du [Centre de Données Socio-politiques](#) à Sciences Po ;
- Au cœur des Grands Moulins, dans le cadre d'un projet plus large d'accompagnement aux Humanités Numériques porté par l'Université de Paris.

Huma-Num, quant à elle, est en cours d'implantation dans les MSH et fournit des services couvrant le cycle de vie des données, depuis le traitement des données jusqu'à leur diffusion et leur archivage.

Pour trouver l'interlocuteur susceptible de renseigner les chercheurs sur ces services dans son établissement, un bon point de départ est de vérifier si celui-ci dispose d'une MSH ou d'un pôle Humanités Numériques, localisé ou non dans une MSH. Si ce n'est pas le cas, on peut toujours les adresser au service d'accompagnement à la recherche.

Source image d'illustration : <https://stories.freepik.com/illustration/blog-post/bro> > Graphique vecteur créé par stories - fr.freepik.com

