

GÉRER et OUVRIR mes DONNÉES de RECHERCHE : les BONNES PRATIQUES

**SEMAINE
DATA/SHS - MSH Dijon**

07 décembre 2022

Léna BOUILLARD,
Maélys LISERON-MONFILS,
Rachel PROST,
Documentalistes
Atelier de la donnée dat@UBFC



Qui sommes-nous ?



dat@UBFC, c'est :

L'atelier de la donnée en Bourgogne Franche-Comté, d'accompagnement à la gestion et la valorisation des données de la recherche

Qui s'appuie sur l'expérience de l'équipe « données de la recherche de l'OSU THETA »

Et s'adresse aux **7 établissements** et **51 laboratoires** répartis sur toute la **région Bourgogne Franche-Comté**

Sylvie Damy, laboratoire Chrono environnement

Léna Bouillard, Maélys Liseron-Monfils, Rachel Prost et Hélène Tisserand, documentalistes

Tom Annebi, Raphaël Melior, Marie Smolinski et Nadir Tabbou, informaticien.ne.s

Sébastien Dassé, chargé d'animation

Un réseau de **référent.e.s** et de membres du **CoPIL** répartis sur tout le territoire de la région BFC

En guise d'INTRODUCTION : les ORIGINES de la SCIENCE OUVERTE

1

**GESTION et OUVERTURE des DONNÉES de RECHERCHE :
CONTEXTE et DÉFINITIONS**

2

**GESTION et OUVERTURE des DONNÉES de la RECHERCHE :
les ENJEUX**

3

Principes **FAIR** et **BONNES PRATIQUES** de **GESTION** des
DONNÉES de **RECHERCHE**

4

STRUCTURATION des acteur.rice.s de l'**OUVERTURE** des
DONNÉES de **RECHERCHE** : l'écosystème
RECHERCHE DATA GOUV et l'**ATELIER** de la **DONNÉE dat@UBFC**



Certains principes de la science ouverte s'appuient sur les « **ethos de la science** ».

Selon Merton, il s'agirait de **normes** et de **valeurs** qui façonnent « la **conscience scientifique** » des chercheur.se.s.

Source : Robert K. Merton, The Normative Structure of Science (1942)

L'universalisme

Le communalisme

Le désintéressement

Le scepticisme organisé

1 GESTION et OUVERTURE des DONNÉES de RECHERCHE : le CONTEXTE

» SCIENCE OUVERTE, définition du PNSO2

La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche.

[...]

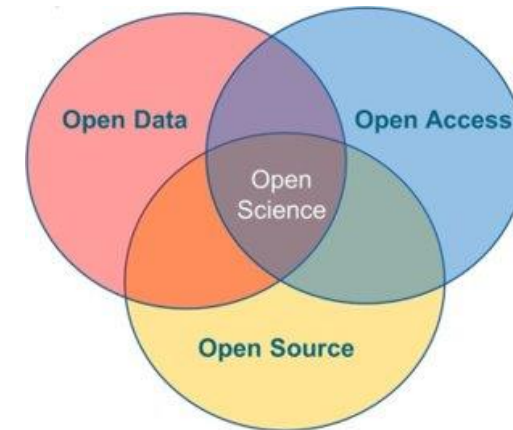
Elle construit un écosystème dans lequel la science est plus transparente, plus solidement étayée et reproductible, plus efficace et cumulative. Elle induit une démocratisation de l'accès aux savoirs, utile à l'enseignement, à la formation, à l'économie, aux politiques publiques, aux citoyens et à la société dans son ensemble. Elle constitue enfin un levier pour l'intégrité scientifique et favorise la confiance des citoyens dans la science.

[Deuxième Plan National pour la Science Ouverte](#)



« Public money, public data »

« Aussi ouvert que possible, aussi fermé que nécessaire »



[UNESCO – Open Science](#)

1 GESTION et OUVERTURE des DONNÉES de RECHERCHE : DÉFINITIONS

» DONNÉES de la RECHERCHE, définition de l'OCDE

« Enregistrements factuels (valeurs numériques, textes, images et sons) issus de la recherche financée en tout ou partie sur fonds publics, utilisés comme sources principales pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. »

Autres objets numériques pertinents au regard de la recherche financés sur fonds publics : **métadonnées, algorithmes, flux de travail, modèles et logiciels** (y compris le code connexe) issus de la recherche financée en tout ou partie sur fonds publics et utilisés dans le cadre de travaux de recherche et de développement.

[Recommandations de l'OCDE concernant l'accès aux données de la recherche financée sur fonds publics](#)

» DONNÉES de la RECHERCHE, définition de l'UNESCO

« [Les] **données de recherche ouvertes** [...] comprennent, entre autres, les **données numériques et analogiques, tant brutes que traitées**, et les **métadonnées** qui les accompagnent, ainsi que les **scores numériques, les enregistrements textuels, les images et les sons, les protocoles, les codes analytiques et les flux de travail** pouvant être **librement utilisés, réutilisés, conservés et redistribués** par quiconque, moyennant citation de la source. Les **données de recherche ouvertes** sont **mises à disposition en temps utile**, dans un format convivial, **exploitable et lisible par les humains et les machines**, conformément aux principes de **bonne gouvernance** et de **bonne gestion des données**, notamment les **principes FAIR (Facilement trouvable, Accessible, Interopérable et Réutilisable)**, à l'aide d'opérations de traitement et de maintenance régulières. »

[Projet de recommandation de l'UNESCO sur une science ouverte](#)



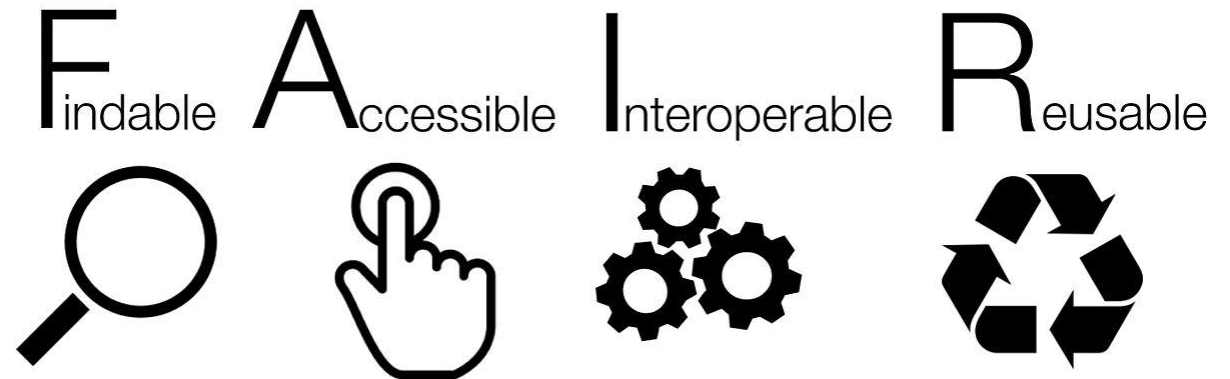
Facile à trouver **A**ccessible **I**nteropérable **R**éutilisable

2014 : Naissance des principes FAIR

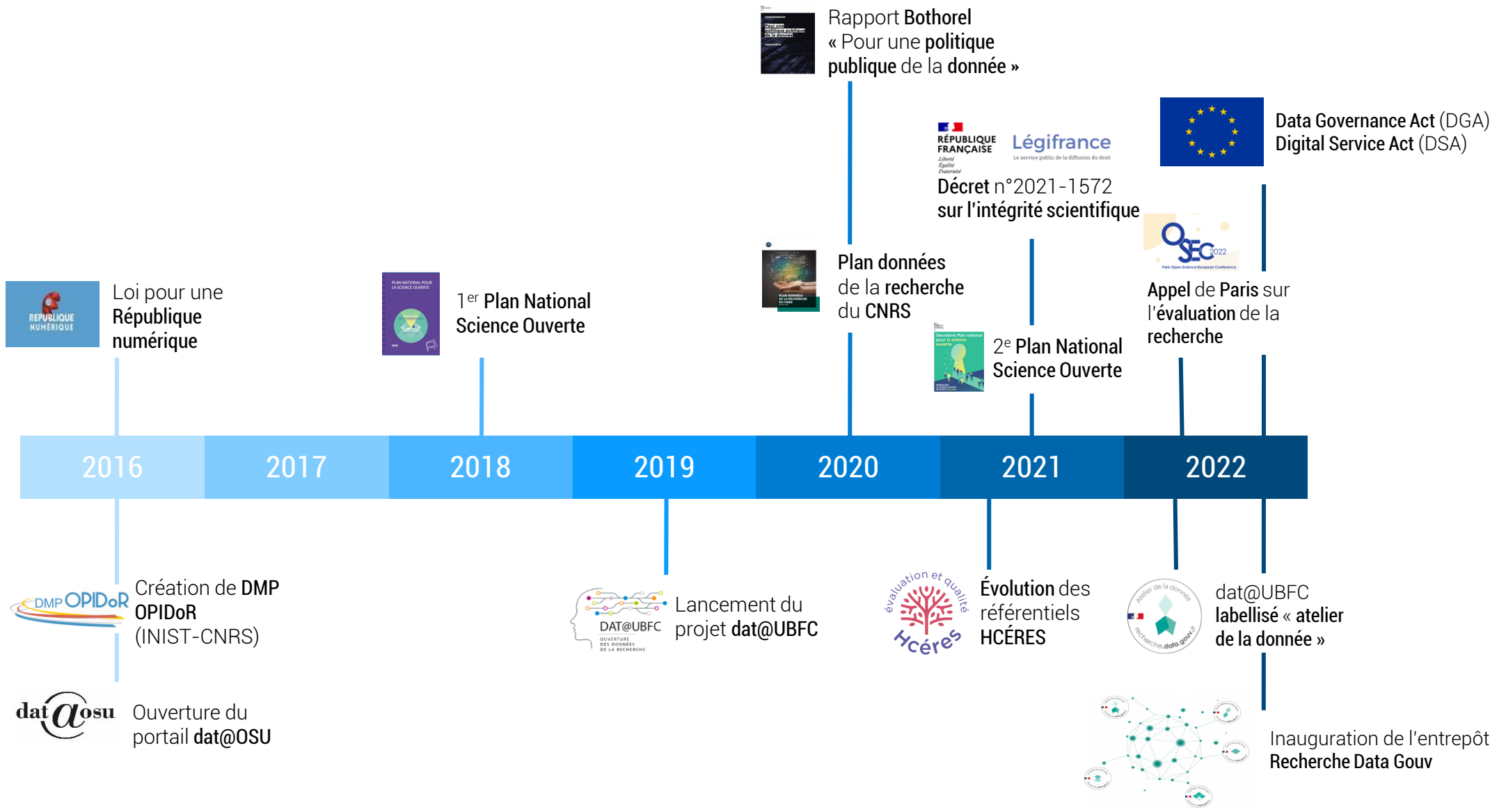
2016 : Publication fondatrice des [principes FAIR](#) dans *Nature*

2022 : Publication dans *Nature* des principes FAIR appliqués aux logiciels et codes source de la recherche (FAIR4RS Principles)

→ [Introducing the FAIR Principles for research Software](#)



1 GESTION et OUVERTURE des DONNÉES de RECHERCHE : l'ACCÉLÉRATION





VISIBILITÉ

- Lien entre données et publications
= **augmentation du taux de citation**



QUALITÉ et ÉVALUATION de la RECHERCHE

- **Monter en compétence** sur la gestion des données
- Faciliter l'**évaluation par les pairs**, la **reproduction** du résultat
- **Conserver/archiver** les données importantes



ÉTHIQUE et INTÉGRITÉ scientifique

- **Transparence**
- **Lutte** contre le **plagiat**, la **fraude scientifique**
- **Efficacité** des **investissements publics**
- **Confiance** des **citoyens** dans la **science**



RESPECT des EXIGENCES des FINANCEURS

- **Horizon Europe**
- **ANR** et autres **agences de financement** françaises
- **Décret** du 03/12/2021 (art. 6) sur l'**intégrité scientifique**

ENJEU #1 : la CULTURE des SHS

La QUANTITÉ de DONNÉES :

- Moins de données produites qu'en sciences dites « dures » (ex : les climatologues)
- De moins en moins vrai tout de même : l'exemple des **humanités numériques**



Impact principal : le **STOCKAGE**

Les PUBLICATIONS :

- Publication d'ouvrages ou de chapitres d'ouvrages (en STM, on publie plutôt des articles dans des revues)
- **Embargos plus longs** : 12 mois pour les SHS contre 6 mois pour les STM



Impact principal : **FREIN** à l'**OUVERTURE** des **DONNÉES**



RESSOURCES :

PROGEDO et les PUD

<https://www.progedo.fr/>

Réseau des MSH

<https://www.msh-reseau.fr/>

InSHS

<https://www.inshs.cnrs.fr/fr>

ENJEU #2 : LA MÉTHODOLOGIE D'ENQUÊTE

Les SOURCES :

- Sources textuelles
- Images, audio, vidéo
- Réutilisation de données existantes : portails SHS comme Nakala, Quetelet-Progedo-Diffusion, Data.sciencespo, etc



Impact principal : **TYPES, QUANTITÉ** et **FORMATS** de **DONNÉES** différents selon les cas

Les MÉTHODES :

- **Méthodes d'enquête** : entretien, questionnaire, observation, analyse de statistiques, collecte de données web, etc
- **Nouvelles méthodes de collecte et de traitement des données** : travail avec des APIs, création de bases de données, data visualisation



RESSOURCES :

PROGEDO et les PUD

<https://www.progedo.fr/>

Huma-Num

<https://www.huma-num.fr/>

Réseau Mate-SHS

<https://mate-shs.cnrs.fr/>

Nakala

<https://nakala.fr/>

ENJEU #3 : LES ENJEUX ÉTHIQUES

ENQUÊTER sur l'HUMAIN :

- En SHS, on s'intéresse aux **PERSONNES**
- Nécessité de protéger ces personnes : **RESPECT** de la **LOI (RGPD)**, mais aussi **ENJEUX ÉTHIQUES**



Impact principal : la question des **DONNÉES PERSONNELLES**



FOCUS : Les **DONNÉES COLLECTÉES** sur le **WEB**

« Ce qu'il faut éviter : collecter des données personnelles accessibles publiquement sur internet sans prévoir une information des personnes ni respecter leurs autres droits » (CNIL)



RESSOURCES :

[Recherche scientifique \(hors santé\) : les mesures de sécurité et de confidentialité \(CNIL\)](#)

[Recherche scientifique \(hors santé\) : comment assurer le respect des droits des personnes ? \(CNIL\)](#)

PLANIFIER la recherche



Rédiger un Plan de Gestion de Données (PGD) ou Data Management Plan (DMP)

- Anticiper les difficultés liées à la gestion des données
- Synthétiser la description des jeux de données du projet de recherche
- Préparer partage, réutilisation et pérennisation des données

Répondre aux exigences des financeurs

- Élaboration obligatoire d'un DMP
- Ouverture obligatoire des publications et données



Bénéficier d'un accompagnement par l'atelier de la donnée dat@UBFC

- **Formation** à l'utilisation de l'outil **DMP OPIDoR** (école doctorale + formation permanente / continue)
- **Accompagnement** à la rédaction du DMP pour les porteur.se.s de projet **ANR/Horizon Europe** ou les équipes de recherche souhaitant réfléchir à la gestion de leurs données

COLLECTER les données



Bien documenter la collecte

- Recenser l'**identité** des contributeur.rice.s
- Décrire les **protocoles** et unités de mesures utilisé.e.s
- Utiliser des **vocabulaires standardisés**



Pour trouver des vocabulaires standardisés :

[FAIRsharing](#) et [Ontobee](#)

Organiser le stockage intermédiaire

- **Lieu de stockage** des données
- Sécurisation de l'**accès** aux données
- **Partage** des données avec l'équipe du projet



Règle du **3-2-1** :
3 copies de vos données sur **2** supports différents et au moins **1** copie à distance

Prendre en compte les aspects éthiques et juridiques

- **Données personnelles, données sensibles** (RGPD, recueil du consentement → voir le [Guide CNRS InSHS](#) : *Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte*)
- **Réutilisation de données existantes** (accords, conventions, droit d'auteur)
- **Éthique et déontologie**
[Charte nationale de déontologie des métiers de la recherche](#)
[Code de conduite européen pour l'intégrité en recherche](#)



Voir une nouvelle ressource sur DoRANum :
[Le RGPD appliqué aux SHS](#)

TRAITER et ANALYSER les données



🎯 Nommage des fichiers

- Les **bonnes pratiques** :
- **25 caractères maxi**
 - **Pas de caractères spéciaux** ou accentués
 - **Pas d'espaces** : utiliser des – et des _
 - **Dates au format ISO** : AAAA-MM-JJ
 - **N° de version** si besoin

🎯 Arborescence

Réfléchir à l'arborescence des fichiers **en amont** et adopter une **arborescence commune** pour le classement des fichiers

🎯 Formats ouverts

Privilégier si possible des formats de fichier **non propriétaires** (= ouverts) et **pérennes** (durables).



Fiche DoRANum : [Liste indicative de formats ouverts et fermés](#)

🎯 Plateformes technologiques MSH Dijon et MSHE Besançon

- [ADN](#) (Archives, Documentation, Numérisation)
- [GeoBFC](#) (Plateforme Géomatique Bourgogne Franche-Comté)
- [NuAnCES](#) (Numérisation et Analyse de Corpus pour la rEcherche Scientifique)
- [PUDD](#) (Plateforme Universitaire de Données de Dijon)



Décrire ses données à l'aide de **métadonnées** dans des portails de référencement



recherche.data.gouv.fr

Déposer ses données dans un **entrepôt de confiance**



Utiliser des **identifiants pérennes**



Attribuer une **licence** à ses données

PUBLIER et PARTAGER les données



Données intermédiaires

→ données produites au cours du processus de recherche, mais dont la conservation pérenne ne s'impose pas

Données finalisées

→ données qui ne seront plus modifiées et dont une partie doit être conservée de manière pérenne



Le Natural Environment Research Council ([NERC](#)) a mis au point une [checklist](#) de 29 critères pour déterminer si une conservation à long terme se justifie.



La conservation pérenne des données a un **coût** qu'il faut anticiper. Dans la communauté de l'ESR, c'est au [CINES](#) qu'il faut s'adresser.

CONSERVER les données



La réutilisation des données nous ramène au **début** de leur cycle de vie : les données sont une **matière première** de la recherche, potentiellement utile à d'autres.

Dans votre processus de recherche, si vous avez vous-même utilisé des données collectées et mise à disposition par d'autres chercheurs.ses, il convient de **créditer** dans votre travail ces données- sources par une **citation**.



Les **portails d'exposition de données** intègrent généralement cette fonction, matérialisée par la présence du pictogramme 

L'outil [DOI Citation Formatter](#) permet d'obtenir une citation complète à partir d'un DOI.



Importance de la **licence**, qui permet la **réutilisation** des données (voir étape « *Publier et partager les données* »)

RÉUTILISER les données





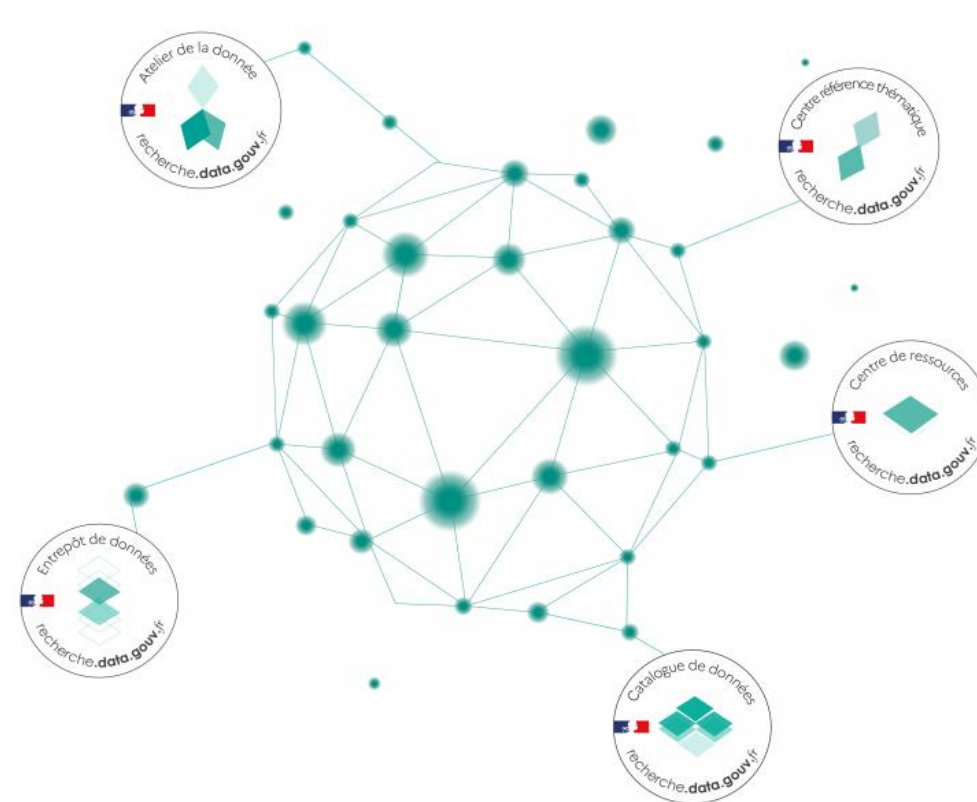
Deux modules pour **DÉPOSER, PUBLIER** et **SIGNALER** des données :

- Un **ENTREPÔT** pour déposer et utiliser des données
- Un **CATALOGUE** pour rechercher les données publiées sur l'entrepôt ou sur des entrepôts externes



Trois modules pour **ACCOMPAGNER** les équipes de recherche sur toute question relative aux données :

- Des **ATELIERS de la DONNÉE**
- Des centres de référence thématiques
- Des centres de ressources



Les SERVICES ACTIFS :

Service 1 : **Portail d'exposition des données**

- Description et exposition de jeux de données
- **Interopérabilité** avec des portails nationaux et internationaux

A venir : interopérabilité avec RDG

Service 2 : **Actions de formation**

- **3 formations** proposées à l'École doctorale et à la formation permanente uB et UFC
- **Formations ponctuelles** sur demande (ex : Institut Agro Dijon, Bibliest)

Service 3 : **Actions de sensibilisation**

- **Événements thématiques** (Printemps de la donnée, *Open Access Week*)
- **Interventions** dans les laboratoires sur demande

Service 4 : **Accompagnement des chercheur.se.s**

- Description des données sur le portail
- Rédaction / relecture des PGD
- Réflexion sur la **gestion des données** au sein des laboratoires
- **Accompagnement des porteur.se.s de projet** ANR / Europe

Service 5 : **Diffusion de l'information** sur l'ouverture des données de la recherche

- **Site Web** <https://data.ubfc.fr/>
- Canaux de communication **UBFC et Science ouverte**
- **Partage** au sein du réseau des Ateliers de la donnée
- Twitter @dataUBFC

Les SERVICES à VENIR :



Service 6 : **Développement d'un réseau d'expert.e.s** pour la gestion des données spécifiques



Service 7 : **Service de stockage** en lien avec Recherche Data Gouv



MERCI
pour votre écoute !



Nous **CONTACTER**

dataubfc-doc@ubfc.fr



Nous **SUIVRE**

@dataUBFC



Point d'entrée **UNIQUE** sur les **DONNÉES** de la recherche à UBFC

<https://data.ubfc.fr/>



CATALOGUE des **DONNÉES** de la **RECHERCHE** en Bourgogne Franche-Comté

<https://search-data.ubfc.fr/>



Vendredi **09 décembre 2022**
10h00

CIMEOS (EA 4177)
Présentation de l'atelier de la donnée dat@UBFC



Source des **PICTOGRAMMES** utilisés :

<https://iconmonstr.com/>



Nos **FORMATIONS** :

INTRODUCTION à la **GESTION** et à l'**OUVERTURE** des **DONNÉES** de la **RECHERCHE**
(GODoRe)

RÉDIGER un **DMP**
(Data Management plan ou Plan de Gestion de Données)

INTRODUCTION aux **BASES** de **DONNÉES** **RELATIONNELLES**



3h00



[BELFORT 27/01/23](#)



[BESANÇON 01/02/23](#)



[DIJON 26/01/23](#)



3h00



[BELFORT 29/03/23](#)



[BESANÇON 24/03/23](#)



[DIJON 23/03/23](#)



5 x 3h00



[BESANÇON + visio](#)



Avril et mai 2023