



SEMAINE DU LIBRE ACCÈS 2021

# *L'importance de notre façon d'ouvrir la connaissance:*

construisons l'équité structurelle

DU 25 AU 31 OCTOBRE

Le thème de cette année s'aligne avec la récente [Recommandation de l'UNESCO sur une science ouverte](#) dont le **libre accès** est un élément crucial.

# Semaine d'animation organisée par des structures UBFC qui participent à la science ouverte :

L'équipe dat@UBFC

Les services communs de documentation de l'uB, l'UFC et AgroSup

La MSHE

## Ont participé à la préparation de cette semaine :

Emmanuelle ASHTA  
Françoise CHAMBEFORT  
Sylvie DAMY  
Murielle FAUDOT

Marie GILLET  
Céline MAICHER  
Florent MARGUIER  
Kevin OUDARD

Rachel PROST  
Albane ROSSI  
Hélène TISSERAND



# Le programme de la semaine

**LUNDI 25/10** de 12h à 14h

- Et si on parlait un peu ouverture des données de la recherche ?

*Rachel Prost et Hélène Tisserand (dat@UBFC)*

- Voie verte ou dorée, freemium, revue hybride ou full OA... : se repérer dans les modes de publication

*Emmanuelle Ashta (SCD uB)*

**JEUDI 28/10** de 12h à 14h

- Les entrepôts de données

*Albane Rossi (dat@UBFC) et Yuji Kato (MSHE)*

- Ai-je le droit de déposer mon article sur HAL ?

*Emmanuelle Ashta (SCD uB)*

**MARDI 26/10** de 12h à 14h

- Se repérer dans la jungle des identifiants chercheurs

*Céline Maicher (SCD UFC)*

- Valoriser ses publications avec un CV HAL : pourquoi et comment ?

*Florent Marguier et Kevin Oudard (SCD AgroSup)*

**VENDREDI 29/10** de 12h30 à 13h30

- Bonnes pratiques : êtes-vous FAIR ?

*Rachel Prost et Hélène Tisserand (dat@UBFC)*

**MERCREDI 27/10** de 12h à 14h

- Augmenter sa visibilité avec Dissem.in  
*Murielle Faudot (SCD UFC)*

- Les licences Creative Commons  
*Marie Gillet (MSHE)*



SEMAINE du LIBRE ACCES 2021 en Bourgogne-Franche-Comté

# Et si on parlait un peu OUVERTURE des DONNÉES de la RECHERCHE ?

Lundi 25 octobre 2021

#OAWeek



Rachel Prost et Hélène Tisserand  
Documentalistes dat@UBFC



# Open Science, Open Data... Quésako ?

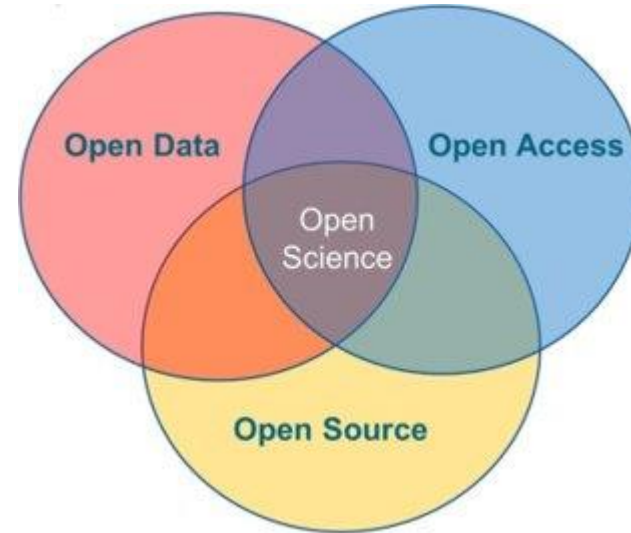


## Science ouverte :

“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 permet à la recherche financée sur fonds publics de conserver la **maîtrise** des résultats qu’elle produit.

Source : Introduction du PNSO 2

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**.”



« Public money, public data »

« Aussi ouvert que possible,  
aussi fermé que nécessaire »



## Données ouvertes (Open Data) :

Données numériques dont l’accès et l’usage sont laissés libres aux usagers.

Volonté politique et citoyenne d’ouvrir l’accès aux données publiques pour promouvoir la transparence, la démocratie, le progrès de la connaissance.

# Open Data, le contexte

## Contexte international



2002  
Budapest Open  
Access Initiative



2003  
Déclaration de Berlin sur  
le libre accès à la  
connaissance



2016  
Amsterdam Call for  
Action on Open Science

## Contexte national



1978



2015



2016



2018 et  
2021



2020 et  
2021

## Contexte local



datacenter régional

Labellisation  
2018



2014



2019



# Les données de la recherche... c'est quoi au juste ?

1

## Définition de l'OCDE :

« Enregistrements factuels (valeurs numériques, textes, images et sons [...] utilisés comme **sources principales pour la recherche scientifique** et généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats de recherche.** »

3

## Définition de l'Université de Bristol :

« Research data is **information** that is involved directly in **funded or unfunded research activities**. Research data is often arranged or formatted in a such a way as to make it **suitable for communication, interpretation and processing**. Put more simply, research data is all of the information that you use as an **integral part of your research.** »

2

## Définition du CIRAD :

« L'ensemble des informations collectées, observées ou créées sous une forme numérique dans le cadre d'un projet de recherche. »

4

## Définition d'un *jeu de données* ou *dataset* :

« Agrégation [...] de données brutes ou dérivées présentant une certaine "unité", rassemblées pour former un ensemble cohérent »

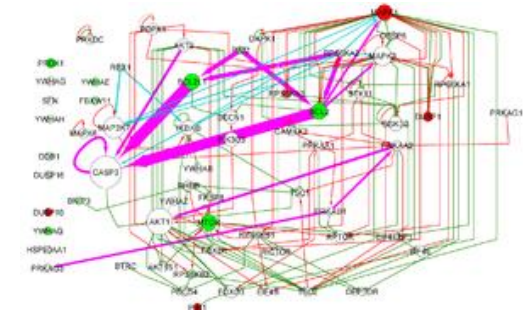
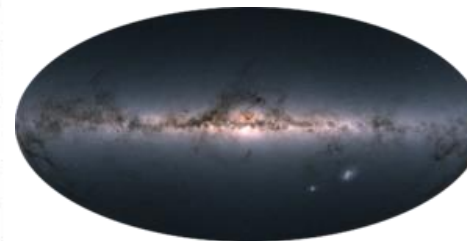
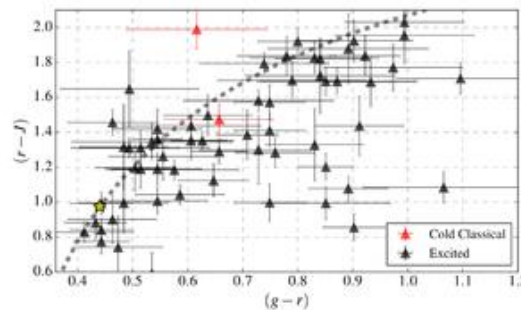
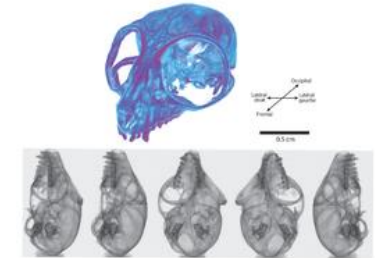
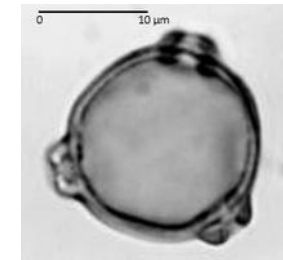
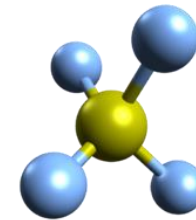
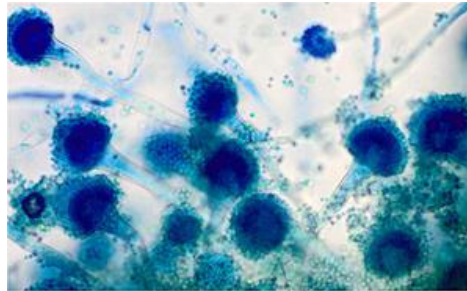
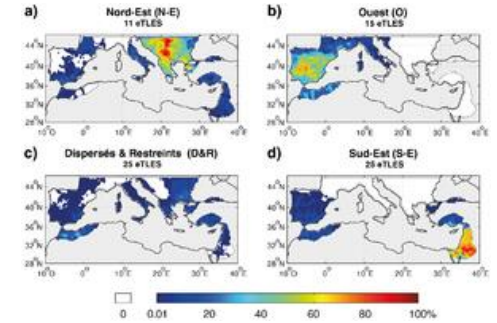
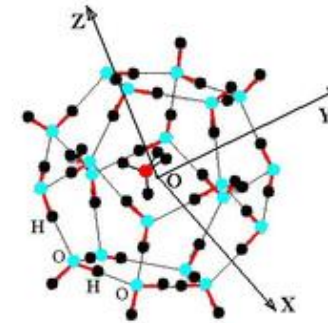
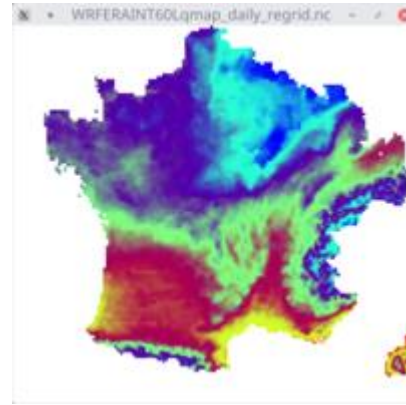
Rémi Gaillard

# Hétérogénéité des données de la recherche

Le champ des données de la recherche est vaste et hétérogène :

« La principale difficulté que représente la définition, dans le champ de la recherche scientifique, de la notion de « donnée de recherche », est qu'il faut s'efforcer d'identifier ce qui peut rassembler des éléments aussi divers qu'un cliché de coléoptère prélevé à Madagascar, un spectro héliogramme produit à Meudon, des informations sur les gènes d'une moisissure, les relevés météorologiques d'un vaisseau ayant traversé l'Atlantique au XVIIIe siècle ou l'enregistrement d'un dialecte rare. »

Rémi Gaillard





# Typologie des données de la recherche

Il existe des données **brutes, traitées** ou **analysées**.

Leur typologie dépend de la manière dont elles sont produites et de leur valeur supposée :

## Les données expérimentales :

- Obtenues à partir d'équipements de laboratoire
- Souvent reproductibles mais coûteuses à reproduire
- Exemples : chromatogrammes, séquençage ADN, ...

## Les données dérivées ou compilées :

- Issues du traitement de données brutes
- Souvent reproductibles mais coûteuses à reproduire
- Exemples : fouille de textes, bases de données, ...

## Les données d'observation :

- Capturées en temps réel
- Uniques et impossibles à reproduire
- Exemples : neuro-imagerie, photographies astronomiques, données d'enquêtes, relevés de conditions atmosphériques, ...

## Les données computationnelles ou de simulation :

- Générées par des modèles informatiques ou de simulation
- Souvent reproductibles si le modèle est correctement documenté
- Exemples : modèles météorologiques, simulations sismiques, modèles économiques, ...

## Les données de référence :

- Collection ou accumulation de petits jeux de données qui ont été revus par les pairs, annotés et mis à disposition
- Exemples : GenBank, BD de cristallographie, corpus de textes, archives d'images historiques, ...

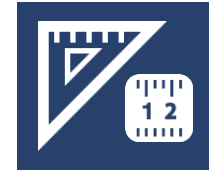
# Dans le sillage des données de la recherche



Publications



Carnets de laboratoire



Métadonnées



Algorithmes



Code source  
Logiciels



Flux de travail

**Algorithmes** : étapes et règles computationnelles devant être suivies dans les calculs ou autres opérations de résolution de problèmes, en particulier par un ordinateur

**Code source** : ensemble d'instructions de programme informatique lisibles par l'homme, qui énonce un algorithme de sorte qu'il puisse être exécuté par un ordinateur

**Logiciel** : désigne à la fois le code et les fichiers exécutables et bibliothèques générées à partir du code source

**Flux de travail** : description précise des étapes d'une méthode utilisée pour générer des résultats de recherche à l'aide de ressources analogiques [...] et numériques [...].

# Quand faut-il partager ses données de recherche ?



**Diffusion obligatoire**

## Données géographiques

→ Directive européenne INSPIRE

**Données environnementales relatives à des émissions de substances dans l'environnement**

→ Convention d'Aarhus



**Diffusion sous conditions**

## Embargo

→ Durées différentes selon les disciplines

**Données à caractère personnel**

→ RGPD, secret statistique

**Droit d'auteur**



**Diffusion interdite**

## Secret professionnel

→ Secret médical, de l'instruction, bancaire, fiscal, ...

## Secret des affaires

→ Secret des procédés, des informations économiques et financières, des stratégies commerciales ou industrielles)

**Secret défense** et données présentant un risque pour la sécurité publique

## Et en cas de doute :

→ Consultez l'outil [Aide à la décision sur la diffusion des données de la recherche](#) du CIRAD

→ Consultez Emmanuel Paris, le DPO UBFC pour les questions liées au RGPD : [juridique@ubfc.fr](mailto:juridique@ubfc.fr)



# Ouvrir ses données de recherche, ça rapporte quoi ?

## VISIBILITÉ

valorisation du travail de recherche et augmentation du taux de citation

## INTÉGRITÉ

meilleure garantie contre la fraude scientifique et vérification possible des résultats

## PUBLICATIONS

de nombreux éditeurs demandent la publication des données en lien avec un article

## QUALITÉ

l'ouverture nécessite d'adopter de bonnes pratiques de gestion de ses données

## TRANSPARENCE

diffuser aux citoyen.ne.s et aux pouvoirs publics les résultats de la recherche pour redonner confiance en la science

## RÉUTILISATION

gain de temps et d'argent en réutilisant ses propres données ou celles d'autres chercheur.se.s

**Merci de nous avoir écoutées !**



A tout de suite pour le prochain RDV de la Semaine du libre accès 2021 en Bourgogne-Franche-Comté :

**« Voie verte ou dorée, freemium, revue hybride ou full OA...  
se repérer dans les modes de publication »**

Par Emmanuelle Ashta (SCD uB)

## Sources utilisées et éléments cités dans ce diaporama :

Plan National pour la Science Ouverte 1 [https://www.ouvrirlascience.fr/wp-content/uploads/2018/08/PLAN\\_NATIONAL\\_SCIENCE\\_OUVERTE\\_978672.pdf](https://www.ouvrirlascience.fr/wp-content/uploads/2018/08/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf)

Plan National pour la Science Ouverte 2 [https://cache.media.enseignementsup-recherche.gouv.fr/file/science\\_ouverte/20/9/MEN\\_brochure\\_PNSO\\_web\\_1415209.pdf](https://cache.media.enseignementsup-recherche.gouv.fr/file/science_ouverte/20/9/MEN_brochure_PNSO_web_1415209.pdf)

Plan Données de la Recherche du CNRS [https://www.cnrs.fr/sites/default/files/pdf/Plaqueette\\_PlanDDOR\\_Nov20.pdf](https://www.cnrs.fr/sites/default/files/pdf/Plaqueette_PlanDDOR_Nov20.pdf)

Rapport Bothorel « Pour une politique publique de la donnée » [https://www.mission-open-data.fr/uploads/decidim/attachment/file/36/Mission\\_Bothorel\\_Rapport.pdf](https://www.mission-open-data.fr/uploads/decidim/attachment/file/36/Mission_Bothorel_Rapport.pdf)

Définition des données de la recherche de l'OCDE <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0347>

Définition des données de la recherche du CIRAD <https://coop-ist.cirad.fr/gerer-des-donnees/s-initier-en-ligne-aux-donnees-de-la-recherche/1-familiarisez-vous-avec-le-concept-de-donnees-de-la-recherche>

Définition des données de la recherche de l'Université de Bristol <http://www.bristol.ac.uk/staff/researchers/data/>

Définition et typologie des données de la recherche par Rémi Gaillard <https://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf>

Guide d'analyse du cadre juridique des données de la recherche en France <https://hal.inrae.fr/hal-02791224/>

Outil d'aide à la décision sur la diffusion des données de la recherche du CIRAD <https://www.loginos.net/base#ylxxVO>

Source des pictogrammes : <https://iconmonstr.com/>

