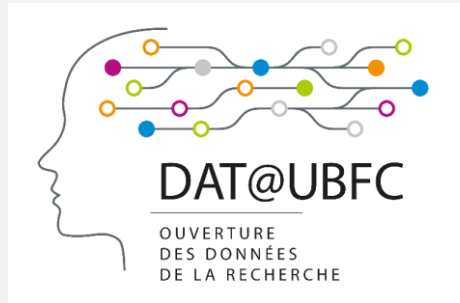


Et si on parlait un peu
OUVERTURE des
DONNÉES de la
RECHERCHE ?

PRINTEMPS
de la DONNÉE 2021 :
Faites fleurir la
science ouverte !



Qui sommes-nous ?



dat@UBFC, c'est :

le service de gestion et de valorisation des données de la recherche d'UBFC

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

Et s'adresse aux 7 établissements de la COMUE et aux 61 laboratoires répartis sur toute la région Bourgogne Franche-Comté

dat@UBFC, c'est :

Sylvie Damy, laboratoire Chrono environnement

Bernard Debray, institut UTINAM

Raphaël Melior et **Nadir Tabbou**, informaticiens OSU THETA

Albane Rossi, chargée d'animation UBFC

Un réseau de **référent.e.s** et de membres du **CoPIL** répartis dans tous les établissements d'UBFC

... et **Hélène Tisserand** et **Rachel Prost**, documentalistes OSU THETA, vos interlocutrices du jour ;-)

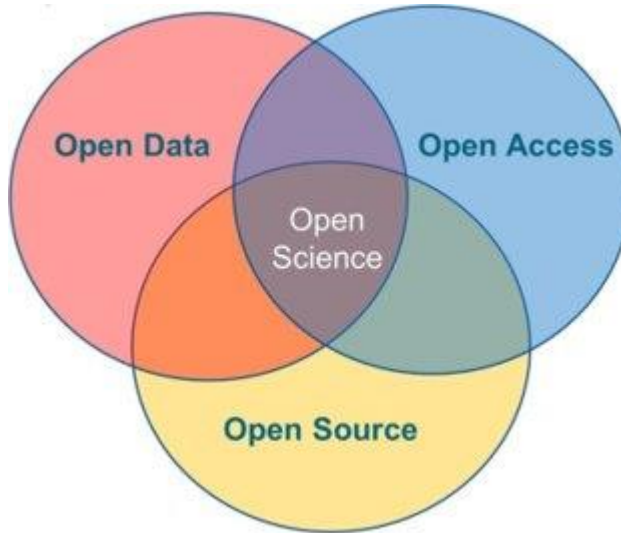
Open Science, Open Data... Quésako ?



Science ouverte (*Open Science* ou *Open Research*) :

Mouvement visant à rendre la recherche scientifique et les données qu'elle produit accessibles à tous.

Selon la Royal Society :
« *combinaison entre données ouvertes, publications scientifiques en libre accès et communication efficace de leurs contenus* ».



« **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

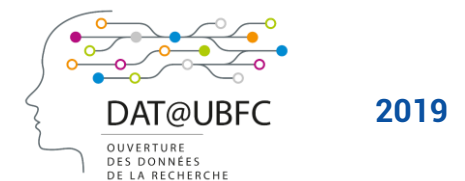
☑ Contexte international



☑ Contexte national



☑ Contexte local



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

1

Définition de l'OCDE :

« Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont 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

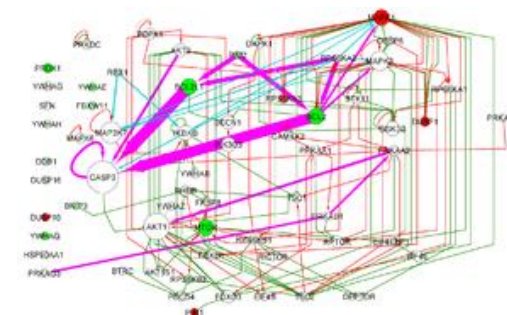
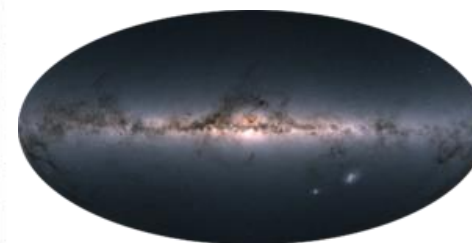
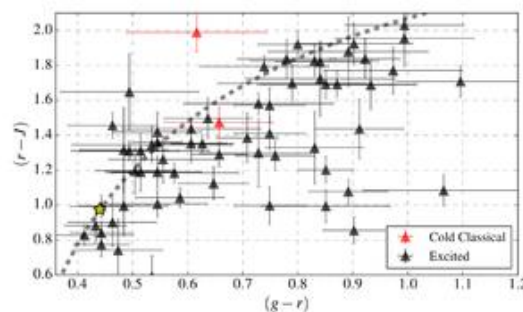
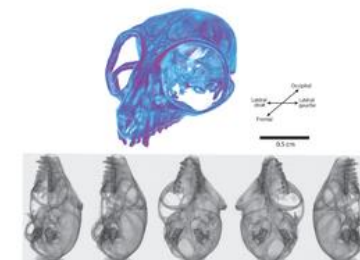
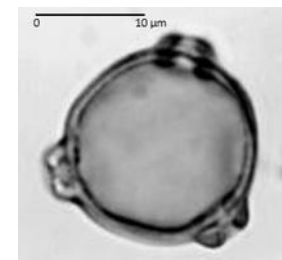
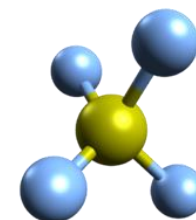
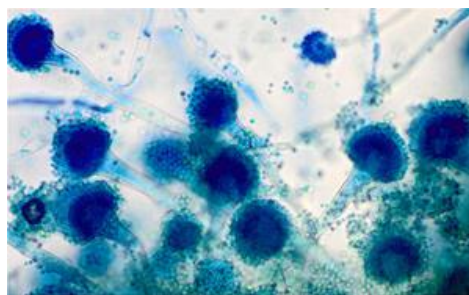
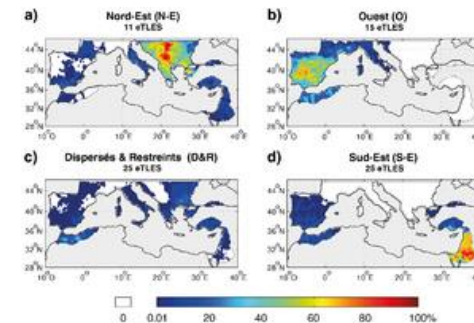
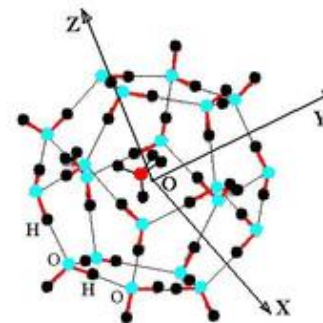
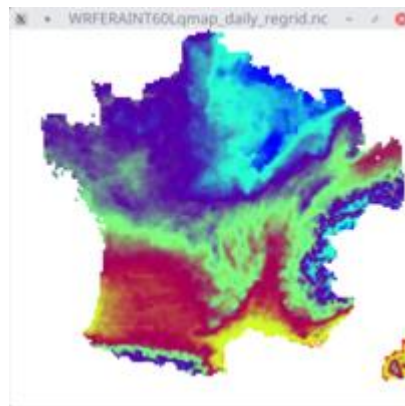
5

Définition des DR : difficulté n°1 - hétérogénéité

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, ...

Définition des DR : difficulté n°2 - juridique

Pas de définition juridique des données de la recherche :

« Les données produites par les établissements de recherche et d'enseignement dans le cadre de leur mission de service public sont considérés comme des **documents administratifs** [...]. »

« Lorsque la loi parle de documents administratifs, cela englobe également les données produites par ces établissements quelles qu'elles soient : données brutes, données élaborées et métadonnées. Cependant la loi précise que ne peuvent être accessibles au public que les documents « achevés ». »

Guide d'analyse du cadre juridique en France

Ne sont pas considéré.e.s comme données de la recherche :



Les données administratives



Les carnets de laboratoire



Les publications



Les objets matériels



Les supports de cours



Les communications professionnelles avec des collègues

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 personnelles

→ 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



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 !



En savoir plus sur dat@UBFC : <https://data.ubfc.fr/>
(site en cours de construction)

Contactez l'équipe : dataubfc-equipe@ubfc.fr

**Et à bientôt pour le prochain RDV
du Printemps de la donnée 2021**

Sources utilisées et éléments cités dans ce diaporama :

Plan National pour la Science Ouverte https://www.ouvrirlascience.fr/wp-content/uploads/2018/08/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf

Plan Données de la Recherche du CNRS 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

Définition des données de la recherche de l'OCDE <http://www.oecd.org/fr/science/inno/38500823.pdf>

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#ylyxVO>

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