

Et si on parlait un peu
BONNES
PRATIQUES :
êtes-vous **FAIR ?**

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 Data, le contexte

☑ Contexte international



☑ Contexte national



☑ Contexte local



Ouverture et partage des données impliquent d'adopter de « bonnes pratiques »

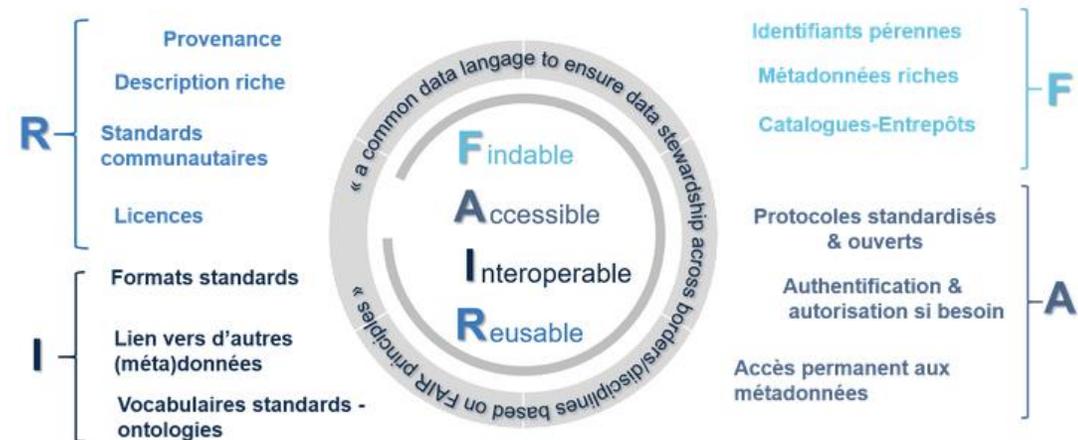
Les principes FAIR



Findable **A**ccessible **I**nteroperable **R**eusable

2014 : Naissance des principes FAIR

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



Source:

Implementation Roadmap for the European Science Cloud (Staff Working Document SWD(2018) 83), 14 March 2018

Traduction Inra <https://www6.inra.fr/datapartage/Produire-des-donnees-FAIR>

Principes FAIR et cycle de vie des données



« Aussi ouvert que possible, aussi fermé que nécessaire »

Le cycle de vie des données
D'après Research data Lifecycle – UK Data Service

Données FAIR : 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



Formation à DMP OPIDoR à UBFC

- École doctorale
- Formation permanente
- En partenariat avec la MSHE



Données FAIR : 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 (données sensibles ?)
- **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

Données FAIR : traiter et analyser les données



🎯 Nommage des fichiers

Les **bonnes pratiques** :

- 30 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).



Voir l'outil [FACILE](#) du [CINES](#)

Données FAIR : publier et partager les données



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



Attribuer une **licence** à ses données



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



Utiliser des **identifiants pérennes**



Données FAIR : conserver les données

Données intermédiaires

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

Données pérennes

→ données dont la valeur justifie une conservation à long terme



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.



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.

Données FAIR : réutiliser 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.se.s, il convient de **créditer** dans votre travail ces données- sources par une **citation**.

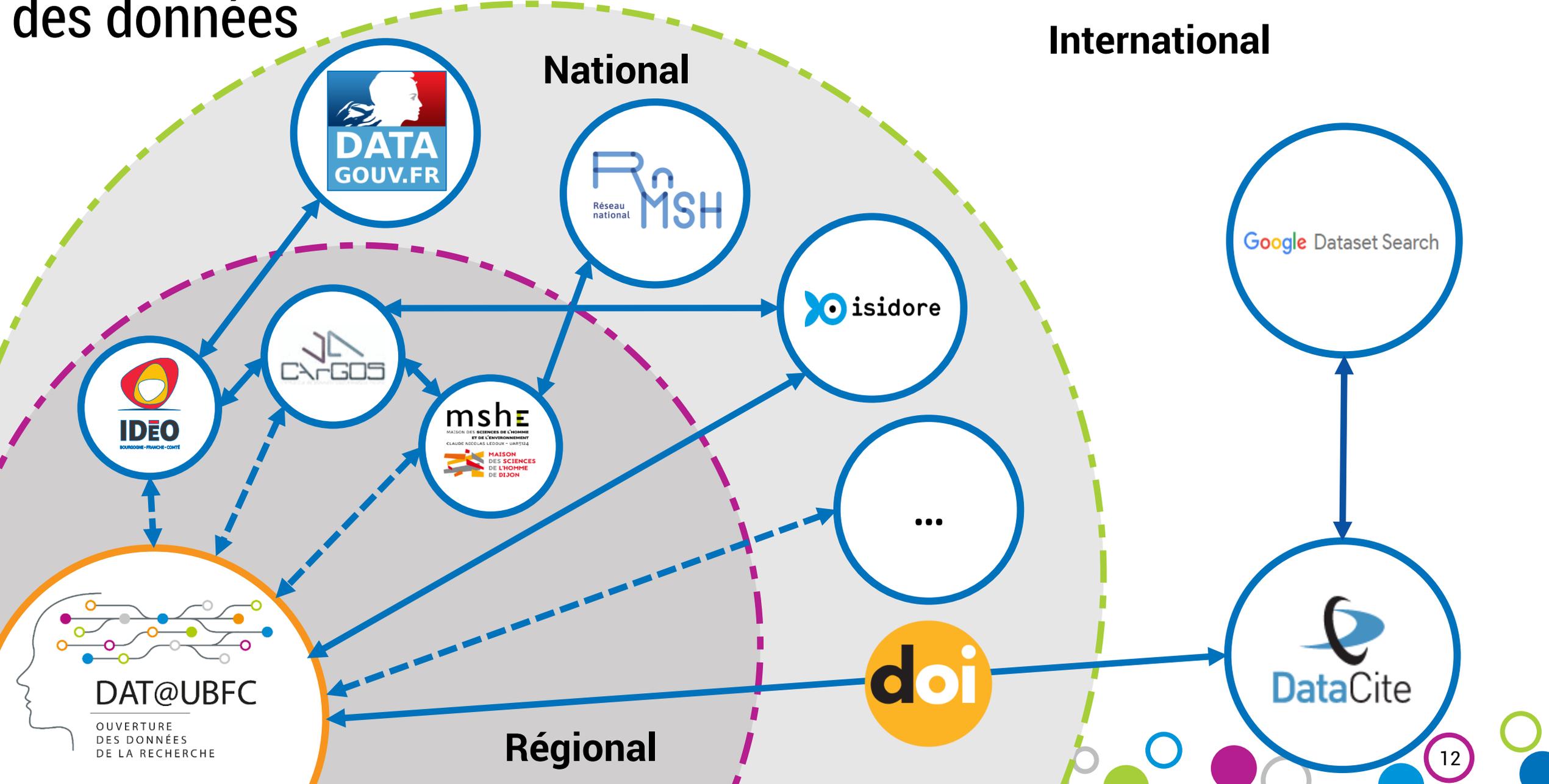


Les **portails d'exposition de données** intègrent généralement cette fonction, matérialisée par la présence du pictogramme 

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



Données FAIR : focus sur l'interopérabilité et la réutilisation des données



L'intérêt des bonnes pratiques

VISIBILITÉ, IMPACT

valorisation du travail de recherche et augmentation du taux de citation grâce au partage

QUALITÉ

les bonnes pratiques produisent des données de qualité et diminuent le risque de perte des données

REPRODUCTIBILITÉ et RÉUTILISATION

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

INTÉGRITÉ de la RECHERCHE

adoption de pratiques transparentes, respect du droit et de l'éthique

COLLABORATIONS

susciter des partenariats avec de nouveaux.elles collaborateur.rice.s

INNOVATION

vers de nouvelles pratiques de recherche : data mining et data driven science

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 :

Mark D. Wilkinson et al. *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 15/03/2016
<https://www.nature.com/articles/sdata201618.pdf>

Rémi Gaillard. *De l'Open data à l'open research data : quelle(s) politique(s) pour les données de recherche*. 2014. Disponible sur :
<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>

DMP OPIDoR : <https://dmp.opidor.fr/>

FAIRsharing : <https://fairsharing.org/>

Ontobee : <http://www.ontobee.org/>

FACILE – CINES : <https://facile.cines.fr/>

CINES – archivage : <https://www.cines.fr/archivage/>

NERC – Natural Environment Research Council : <https://nerc.ukri.org/>

NERC Data value checklist :
<https://nerc.ukri.org/research/sites/environmental-data-service-eds/policy/data-value-checklist/>

« The State of Open Data » – figshare :
https://figshare.com/collections/State_of_Open_Data/4046897

DOI Citation Formatter : <https://citation.crosscite.org/>

« The citation advantage of linking publications to research data » - The Alan Turing Institute : <https://doi.org/10.1371/journal.pone.0230416>

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