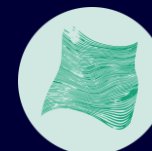


23/05/2023

Les principes F.A.I.R., pourquoi et comment faire ?



Le Printemps
de la Donnée

BU et Learning center – SCD
Fabrique de la science ouverte

la Fabrique de la science ouverte

23 mai - 21 juin 2023



Le Printemps
de la Donnée

Programme

Les bibliothèques universitaires et Learning center de l'Université de Lille vous proposent, dans le cadre du printemps de la donnée, du 23 mai au 21 juin 2023 :

en distanciel

23 mai | Les principes F.A.I.R., pourquoi et comment faire ?

26 mai | Atelier Plan de Gestion des Données express

30 mai | Découverte de Recherche Data Gouv

en présentiel à LILLIAD Learning center Innovation (salle Y)

+ en distanciel

06 juin | Les forges universitaires au coeur de l'ouverture des logiciels de la recherche

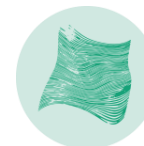
21 juin | LORD : Les acteurs de la donnée ULille

Informations et inscriptions : fabso.univ-lille.fr

À noter dans son agenda
en automne : l'Open Access Week
(plus d'infos à venir).

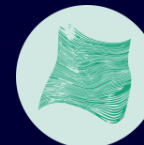
Déroulé de la séance

- 1. Présentation générale**
Les principes FAIR, qu'est-ce que c'est ?
- 2. Contextualisation**
FAIR, pour quoi faire ?
- 3. Être accompagné au Service Commun de Documentation**
Services et outils mis à votre disposition



Les principes FAIR

Qu'est-ce que c'est ?



Quatre principes directeurs

F

Facile à trouver

A

Accessible

I

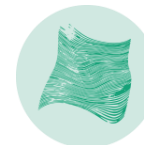
Interopérable

R

Réutilisable

Ces principes servent à couvrir l'ensemble des pratiques visant à la bonne gestion des données, principalement des **données de recherche**, dans un contexte d'ouverture sur Internet.

Cela concerne les différentes façon de **construire, décrire, stocker, présenter, et partager** ses données.



Quatre principes directeurs

Facile à trouver

F1

Un identifiant pérenne et unique est assigné aux (méta)données

F2

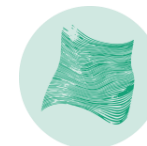
Des métadonnées riches sont associées aux données

F3

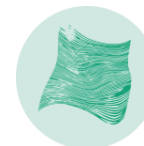
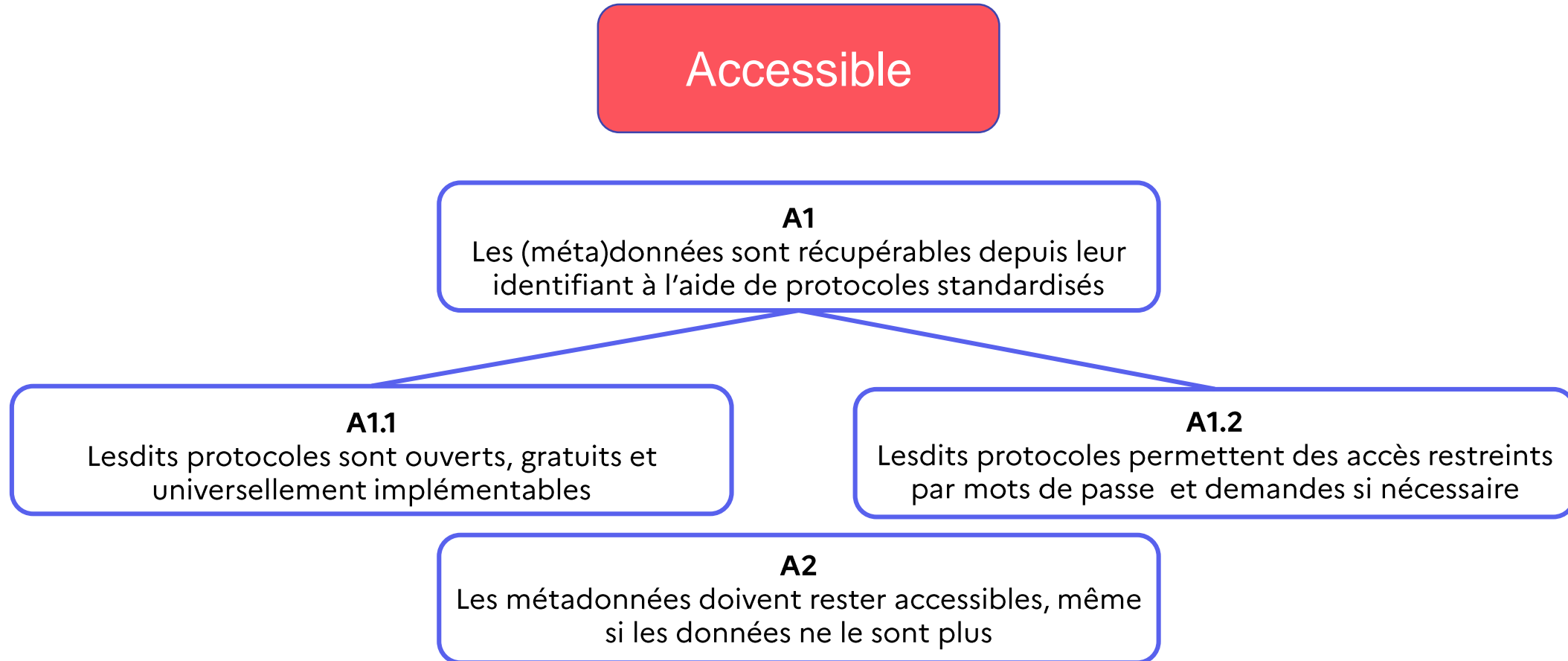
Les (méta)données sont enregistrées dans une ressource consultables

F4

Les métadonnées indiquent l'identifiant pérenne associé aux données



Quatre principes directeurs



Quatre principes directeurs

Interopérable

I1

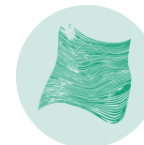
Les (méta)données sont décrites à l'aide d'un langage formel, accessible, partagé et largement applicables

I2

Les (méta)données utilisent un vocabulaire qui suit les principes FAIR

I3

Les métadonnées incluent des références pertinentes à d'autres données



Quatre principes directeurs

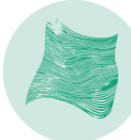
Réutilisable

R1
Les (méta)donnée ont une pluralité d'attributs précis et pertinents

R1.1
Les (méta)données sont publiées avec une licence de réutilisation claire et accessible

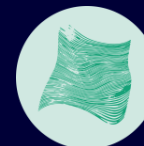
R1.2
Les (méta)données sont associées à leur provenance

R1.3
Les (méta)données atteignent les prérequis des divers standards communautaires



Contextualisation

FAIR, pour quoi faire ?



Origine des cinq principes

Chronologie

2014

Lancement en 2014 de l'initiative « Data FAIRport », basée sur l'ouverture des données, par un groupe de travail au sein communauté FORCE 11

2016

Publication d'un article dans la revue « Nature », officialisant les principes FAIR dans leur forme actuelle.

Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

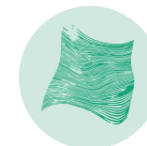
Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, ... Barend Mons  + Show authors

Scientific Data 3, Article number: 160018 (2016) | [Cite this article](#)

495k Accesses | 4627 Citations | 2049 Altmetric | [Metrics](#)

2020

Incitation pour l'application dans les grandes institutions de la recherche dans le monde (Horizon 2020, ANR, etc.).



Mise en application de 2020 à 2022

ANR

Une seule mention de l'anagramme « FAIR » au sein du modèle de Plan de Gestion de Données de l'ANR.

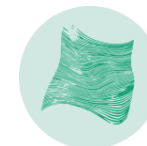
6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Mais les principes se retrouvent de façon plus insinués dans diverses sections du Plan de Gestion de Données.

2. DOCUMENTATION ET QUALITÉ DES DONNÉES

5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME

Les principes « FAIR » sont donc présents dans les recommandations de l'ANR même si assez peu mises en avant.



Mise en application de 2020 à 2022

Horizon Europe / 2020

La partie principale du Plan de Gestion de Données est centrée autour des principes « FAIR » dans le modèle européen.

Chaque principe y est détaillé ouvertement.

2. FAIR DATA

F

2.1. Making data findable, including provisions for metadata

A

2.2.1. Making data accessible : Repository 2.2.2. Making data accessible : Data
2.2.3. Making data accessible : Metadata

I

2.3. Making data interoperable

R

2.4. Increase data re-use

Conséquences d'une gestion vertueuse

Pas qu'une question d'obligation venant des financeurs, aussi des effets vertueux et tangibles directement pour l'équipe de recherche

Cela permet de mettre en avant vos recherches de façon plus poussée qu'un simple dépôt des post-prints dans une archive ouverte.

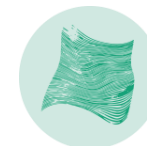
Les données mises à disposition et correctement renseignées peuvent être réutilisées, menant les ré-utilisateurs à se pencher sur vos publications pour comprendre le contexte de production.

Cela favorise la gestion des données en interne, assurant une organisation saine et pérenne pour les membres de votre propre structure.

Ces réutilisations augmentent aussi les taux de citations de vos travaux, et permettent de mettre en valeur tous les aspects de vos recherches.

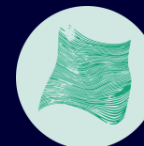
L'aspect le plus fondamental des principes FAIR peut parfaitement être atteint sans surcoût financier, en investissant temps pour échanger avec nos équipes.

Cela est surtout valable pour votre premier projet respectant les principes FAIR. Une fois les principes en tête, il peut être aisé d'agir rapidement en autonomie.



Être accompagné au Service Commun de Documentation

Services et outils à votre disposition



Les services mis à votre disposition

Formations

Formations en petits comme en grands groupes dans le cadre de formations programmées ou sur demande.

Relecture

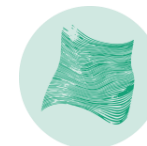
Relectures de vos documents demandés par les financeurs autour de la bonne gestion des données de recherche (PGD, relecture du projet avant dépôt, etc.).

Accompagnement

Accompagnement plus personnel pour guider vers les bonnes pratiques, pour le choix d'un entrepôt, pour le dépôt de vos données, pour la FAIRisation de vos bases de données, etc.

Orientation

En cas de demande sortant de nos compétences, nous sommes en mesure de vous indiquer à qui vous adresser directement pour répondre au reste de vos questions



Exemples d'outils mis à votre disposition

Guide des bonnes pratiques

Site des bibliothèques: <https://bu.univ-lille.fr/chercheurs-doctorants/science-ouverte/donnees-de-recherche>

Template de « readme »

Site des bibliothèques

FAIR Checker

<https://fair-checker.france-bioinformatique.fr/check>

Modèle de plan de gestion de données

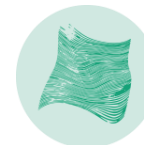
(Bientôt) Sur le site des bibliothèques et sur DMP OPIDoR

Outil de sélection d'entrepôt de Cornell

Site de Cornell : <https://finder.research.cornell.edu/storage>

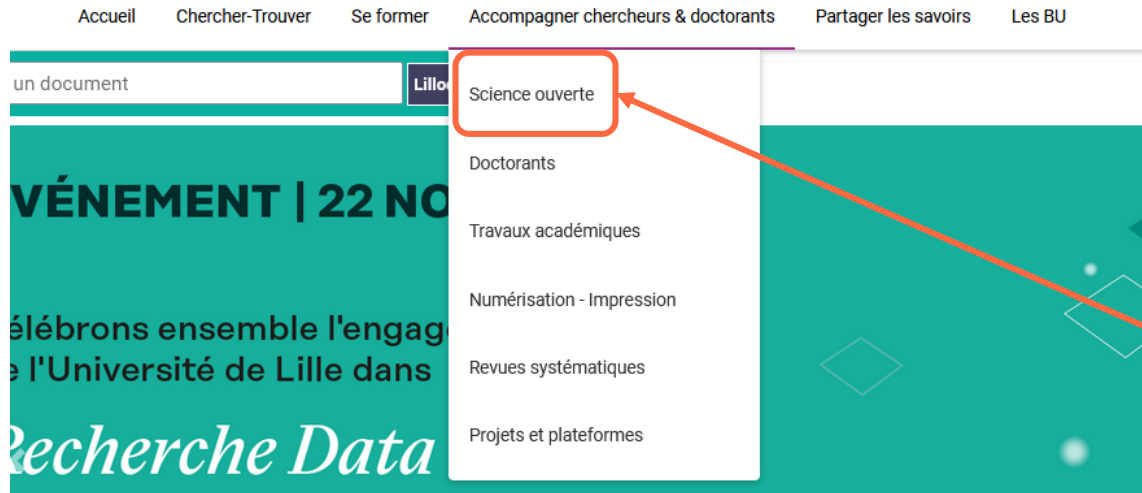
La feuille de route FAIR

Sur demande ou après les formations



Les outils mis à votre disposition

Guides des bonnes pratiques



Tous les guides des bonnes pratiques abordant les thématiques FAIR peuvent être trouvés sur le site des bibliothèques. <https://bu.univ-lille.fr/>

Premièrement, il faut se diriger vers l'onglet « *Science ouverte* » du site, dans la partie « *Accompagner chercheurs et doctorants* ».



Ensuite, il faut se diriger vers l'onglet « *Gérer et diffuser des données de recherche* ».

Les outils mis à votre disposition

Guides des bonnes pratiques

Bonnes pratiques et principes FAIR

La **FAIR**isation des données (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) correspond à un ensemble de bonnes pratiques, lancées par l'organisation Force 11, à appliquer de façon plus ou moins soutenue tout au long des projets de recherche. L'application de ces différents principes doit normalement assurer la bonne **accessibilité**, **compréhension** et **réutilisation** de ces données de recherche.

Une présentation plus poussée de ces principes, de leur composition et de leur mise en place est trouvable sur cette page :

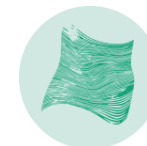
- ⇒ [Préparation des données FAIR pour la réutilisation et la reproductibilité](#)

Vous pouvez aussi consulter un de nos guides plus détaillés sur des points précis de la FAIRisation :

- ⇒ Citer des données
- ⇒ Bonne gestion de fichiers
- ⇒ Formats des fichiers de données
- ⇒ Guide de rédaction de métadonnées de type « readme »
- ⇒ Introduction à la propriété intellectuelle pour la gestion des données
- ⇒ Métadonnées et description des données
- ⇒ Partager et archiver les données pendant le projet de recherche
- ⇒ Préparation des données tabulaires pour la description et l'archivage
- ⇒ Sauvegarde et stockage des données

Enfin, il est possible de dérouler le menu « *Bonnes pratiques et principes FAIR* » pour avoir accès à la liste des guides.

L'onglet « *Préparation des données FAIR pour la réutilisation et la reproductibilité* » permet d'accéder à une page principale résumant les principes et amenant aux autres pages de façon logique.



Les outils mis à votre disposition

Template de « readme »

Sommaire des bonnes pratiques :

- [Préparation des données FAIR pour la réutilisation et la reproductibilité](#)
- [Citer des données](#)
- [Bonne gestion de fichiers](#)
- [Formats des fichiers de données](#)
- [Guide de rédaction de métadonnées de type « readme »](#)
- [Introduction à la propriété intellectuelle pour la gestion des données](#)
- [Métadonnées et description des données](#)
- [Partager et archiver les données](#)
- [Préparation des données tabulaires pour la description et l'archivage](#)
- [Sauvegarde et stockage des données](#)

Modèle de Readme

ReadmeCornell_FR.txt

Un modèle de « *readme* », traduit et adapté depuis le modèle anglais et libre et l'université de Cornell est disponible sur le site.

Vous pouvez y accéder dans la « *sidebar* » depuis n'importe quelle page des bonnes pratiques.

Le « *readme* » se présente alors sous forme d'un fichier texte qu'il ne reste qu'à compléter et à joindre aux données lors des dépôts sur l'entrepôt sélectionné.

Si les caractères ne sont pas encodés sur votre navigateur, enregistrer le document avec **CTRL+S**, au lieu de le copier.

Ce JEUEDONNÉESreadme.txt a été généré le YYYY-MM-DD par NOM
<Les textes d'aides sont entre crochets, et sont à supprimer avant la sauvegarde>

INFORMATION GÉNÉRALE

Titre du jeu de données :

Information sur l'auteur

Informations de contact de l'auteur principal

Nom :

Institution :

Adresse :

Email :

Informations de contact des associés et co-auteurs

Nom :

Institution :

Adresse :

Email :

Informations de contact alternative

Nom :

Institution :

Adresse :

Email :

Date de récupération des données (date unique, plage, date approximative) <format su

Lieu de récupération des données <latitude, longitude, ville/région, état, pays...> :

Information sur les fonds ayant permis la récupération des données :

INFORMATION D'ACCÈS ET DE PARTAGE

Licences/restrictions placée sur les données :

Liens vers les publications qui citent ou utilisent les données :

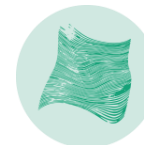
Liens vers les autres localisation publiquement accessible des données :

Liens/rerelations avec des ensembles de données auxiliaires

Les données étaient-elles dérivées d'une autre source ? oui / non

Si oui, liste des sources :

Format de citation recommandé pour ce jeu de données :



Le Printemps
de la Donnée

BU et Learning center – SCD
La Fabrique de la science ouverte

Les outils mis à votre disposition

FAIR-Checker

Improve the FAIRness of your web resources

FAIR checker

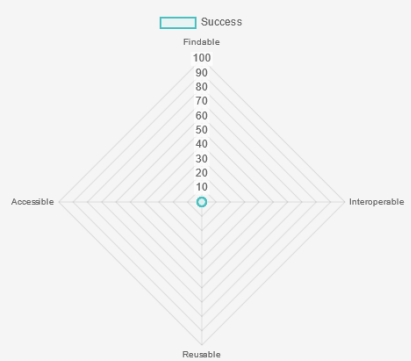
Enter resource identifier (URL/DOI)

[Test all metrics](#)

[Clean results](#)

[Dataset Dataserve](#) [Workflow](#) [Publication Datasite](#) [Dataset](#) [Tool](#)

Radar chart of metrics completion



Success

Findable

100

90

80

70

60

50

40

30

20

10

0

Accessible

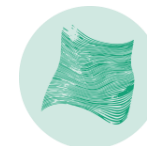
Interoperable

Reusable

FAIR Checker est un outil gratuit permettant en un clic d'avoir une estimation de la qualité de vos données vis-à-vis des principes FAIR, simplement à partir d'une URL ou d'un DOI.

Si possible, l'outil vous indique quels points font défaut à vos données afin de vous mettre sur une piste d'amélioration.

L'outil n'est pas 100% fiable, mais si vous ne savez pas par où commencer, cela peut constituer une piste intéressante.



Les outils mis à votre disposition

Outil de sélection d'entrepôt de Cornell

<https://data.research.cornell.edu/>




Data Storage Finder

Our tool is intended to help you find options for data storage at Cornell.

Browse data storage

Data Storage Finder

Evaluate options for data storage at Cornell

-  All services presented on this finder tool are vetted and supported by Cornell University.
-  To explore data options available to Weill Cornell Medicine Cornellians please visit the [WCMC storage wizard](#).
-  We welcome [feedback](#) on this tool.

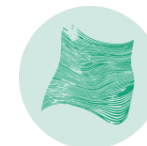
Le portail recherche du site de l'université de Cornell offre un large panel d'outils et de documentation pour l'ouverture et la bonne gestion des données de recherche.

Un outil très intéressant est notamment disponible sur la page d'accueil, le « *Data Storage Finder* ».

Cet outil permet, en remplissant un formulaire, de trouver les entrepôts de données correspondants à nos besoins.

Attention, cela concerne des entrepôts internationaux, et dont la plupart ont pour vocation à ne servir que de stockage temporaire lors de la durée de vie de vos projet.

De plus, le formulaire est adapté à la communauté scientifique de Cornell, ainsi l'outil est surtout utile en tant que base pour découvrir la variété d'offre d'entrepôts.



Les outils mis à votre disposition

Outil de sélection d'entrepôt de Cornell

L'onglet de gauche met à votre disposition un formulaire où vous pouvez sélectionner les réponses correspondants à vos besoins.

Vous pouvez laisser des questions sans réponse, tandis que de multiples cases peuvent parfois être cochées simultanément.

À chaque nouvelle sélection, les résultats se mettent à jour en temps réel sur la droite, grisant les entrepôts ne correspondant pas aux critères.

Describe your data

Answer these questions to help identify data storage services that are suitable for your needs. Checking these boxes will change the list of available services. If you are uncertain how to answer, leave the question blank to maximize your resulting options.

[Clear Answers](#)

1. What is the classification of your data?

- Public / Low Risk
- Sensitive / Moderate Risk
- Confidential or Restricted / High Risk
- HIPAA-Regulated

2. Do you need backups, snapshots or replication of your data?

- I need one or more backup/snapshot copies of the data, and need to be able to restore data from previous points in time (high durability).
- I need to have replicate copies of the data to minimize downtime (high availability).

<p>Amazon Web Services Elastic Block Store</p> <p>Storage for use with Amazon EC2</p>	<p>Amazon Web Services Elastic File System</p> <p>Storage for use with multiple Amazon EC2 instances</p>	<p>Amazon Web Services Glacier</p> <p>Cloud based archival storage</p>	<p>Amazon Web Services S3</p> <p>Flexible, scalable object storage</p>	<p>BioHPC Cloud</p> <p>Storage for BioHPC lab computing services</p>
<p>CAC Archival Storage</p> <p>Single copy, non-mountable storage</p>	<p>CAC Red Cloud Storage</p> <p>Storage for Red Cloud compute instances</p>	<p>CCSS Research Servers</p> <p>Storage for CCSS computing environment</p>	<p>CISER Data & Reproduction Archive</p> <p>Publicly shared and restricted data and code packages repository</p>	<p>CUGIR</p> <p>Publicly shared geospatial data storage</p>
<p>CUL eCommons</p> <p>Publicly shared data repository</p>	<p>Cornell Box</p> <p>Online file sharing and collaboration</p>	<p>Cornell Restricted Access Data Center</p> <p>Storage for CRADC (confidential) computing environment</p>	<p>EZ-Backup Static Storage</p> <p>Archival storage and backup storage</p>	<p>Google Drive</p> <p>Cornell Google Workspace file storage and sharing</p>

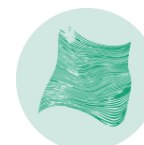
Les outils mis à votre disposition

La feuille de route FAIR

	Facile à trouver	Accessible	Interopérable	Réutilisable
	<ul style="list-style-type: none"> - Assignation d'un identifiant pérenne (PID/DOI). - Dépôt dans un entrepôt de données adapté aux besoins. - Création de métadonnées externes riches, liées aux jeux de données. 	<ul style="list-style-type: none"> - Les données doivent être aussi ouvertes que possible, aussi fermées que nécessaire 	<ul style="list-style-type: none"> - Création de métadonnées externes riches, liées aux jeux de données. - Sélection d'un entrepôt de données connu et reconnu auprès des pairs et des institutions. 	<ul style="list-style-type: none"> - Choix d'une licence pour stipuler clairement les conditions de réutilisation des données. - La provenance des données doit être décrite avec précision. - Ajout d'un fichier "readme" si nécessaire pour la bonne compréhension et réutilisation des données.
Palier 1	<p>Accompagnement du SCD :</p> <p>[Formation] Former sur les bonnes pratiques en termes de rédaction de métadonnées externes pour un projet de recherche avec un guide des bonnes pratiques sur le site de l'université.</p> <p>[Formation] Faire connaître Recherche Data Gouv, recommandé comme solution par défaut pour le stockage de ses données de recherche.</p> <p>[Fiches et outils] Mettre à disposition une checklist permettant de bien cerner les besoins de chaque projet et de sélectionner un entrepôt de données adapté et ne dépendant d'aucun éditeur.</p>	<p>Accompagnement du SCD :</p> <p>[Formation] Former/Inciter/Informer sur l'importance de déposer ses données dans un entrepôt public, surtout lorsqu'elles doivent être ouvertes. Si les données nécessitent une restriction, inciter à déposer dans un entrepôt public gérant les restrictions d'accès, plutôt que sur des serveurs et machines internes aux laboratoires.</p> <p>[Formation / Information] Introduction à l'entrepôt pluridisciplinaire Recherche Data Gouv permettant de gérer efficacement les droit d'accès de ses données, et même d'y appliquer des périodes d'embargo.</p>	<p>Accompagnement du SCD :</p> <p>[Fiches et outils] Mettre à disposition un guide des bonnes pratiques pour la rédaction des métadonnées, afin d'en comprendre les enjeux et la pertinence, sur le site de l'université.</p>	<p>Accompagnement du SCD :</p> <p>[Formation] Former sur les bonnes pratiques de rédaction des métadonnées</p> <p>[Fiches et outils] Mettre à disposition des informations sur les différentes licences conseillées et les plus utilisées sur le site de l'université. Mise en avant des différences de reconnaissance des licences à l'échelle nationale et internationale (Etalab / Creative Commons).</p> <p>[Fiches et outils] Mise à disposition d'un "readme" à l'aide d'un template anglais et français, sur le site de l'université.</p>

La feuille de route FAIR est un document mis à disposition par nos équipes pour servir de guide pour les chercheurs tentant de respecter au mieux les principes FAIR.

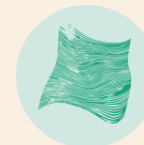
Ce document a l'avantage de décomposer les grands principes en petits jalons, organisés en trois paliers progressifs, plus aisément atteignables. Pour chacun d'entre eux est indiqué la liste des ressources humaines, numériques et documentaires mises à disposition pour vous aider à les atteindre.



Merci de votre écoute

Pour contacter le service science ouverte pour une question, une précision, ou une demande d'assistance : scd-aap@univ-lille.fr

- Lien vers le site de FORCE11 : <https://force11.org/info/the-fair-data-principles/>
- Lien vers les ressources numériques du site des Bus : <https://bu.univ-lille.fr/chercheurs-doctorants/science-ouverte/donnees-de-recherche>
- Lien vers les ressources numériques Cornell: <https://data.research.cornell.edu>
- Lien vers l'outil FAIR checker : <https://fair-checker.france-bioinformatique.fr/check>



la Fabrique de la science ouverte

23 mai - 21 juin 2023



Le Printemps
de la Donnée

Programme

Les bibliothèques universitaires et Learning center de l'Université de Lille vous proposent, dans le cadre du printemps de la donnée, du 23 mai au 21 juin 2023 :

en distanciel

23 mai | Les principes F.A.I.R., pourquoi et comment faire ?

26 mai | Atelier Plan de Gestion des Données express

30 mai | Découverte de Recherche Data Gouv

en présentiel à LILLIAD Learning center Innovation (salle Y)

+ en distanciel

06 juin | Les forges universitaires au coeur de l'ouverture des logiciels de la recherche

21 juin | LORD : Les acteurs de la donnée ULille

Informations et inscriptions : fabso.univ-lille.fr

À noter dans son agenda
en automne : l'Open Access Week
(plus d'infos à venir).