

Avoir des données conformes à la science ouverte grâce aux principes FAIR

Session ½ : Des données "Faciles à trouver" et "Accessibles"

01/10/2025

Déroulé

Présentation de l'ADCVL

Introduction générale

Des données "Faciles à trouver"

Des données "Accessibles"

Les entrepôts de données

L'Atelier de la Donnée Centre - Val de Loire



Accompagnement proposé



Accompagnement au dépôt des jeux de données



Conseils éthiques et juridiques



FAIRisation des données



Plan de gestion de données



Préservation et sécurité des données



Nous contacter

Guichet unique



Nous suivre

Actus & agenda de la science ouverte

Linkedin

Introduction générale

Qu'est-ce qu'une donnée de recherche ?

Définition de l'OCDE :

« des **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 les résultats de la recherche** »

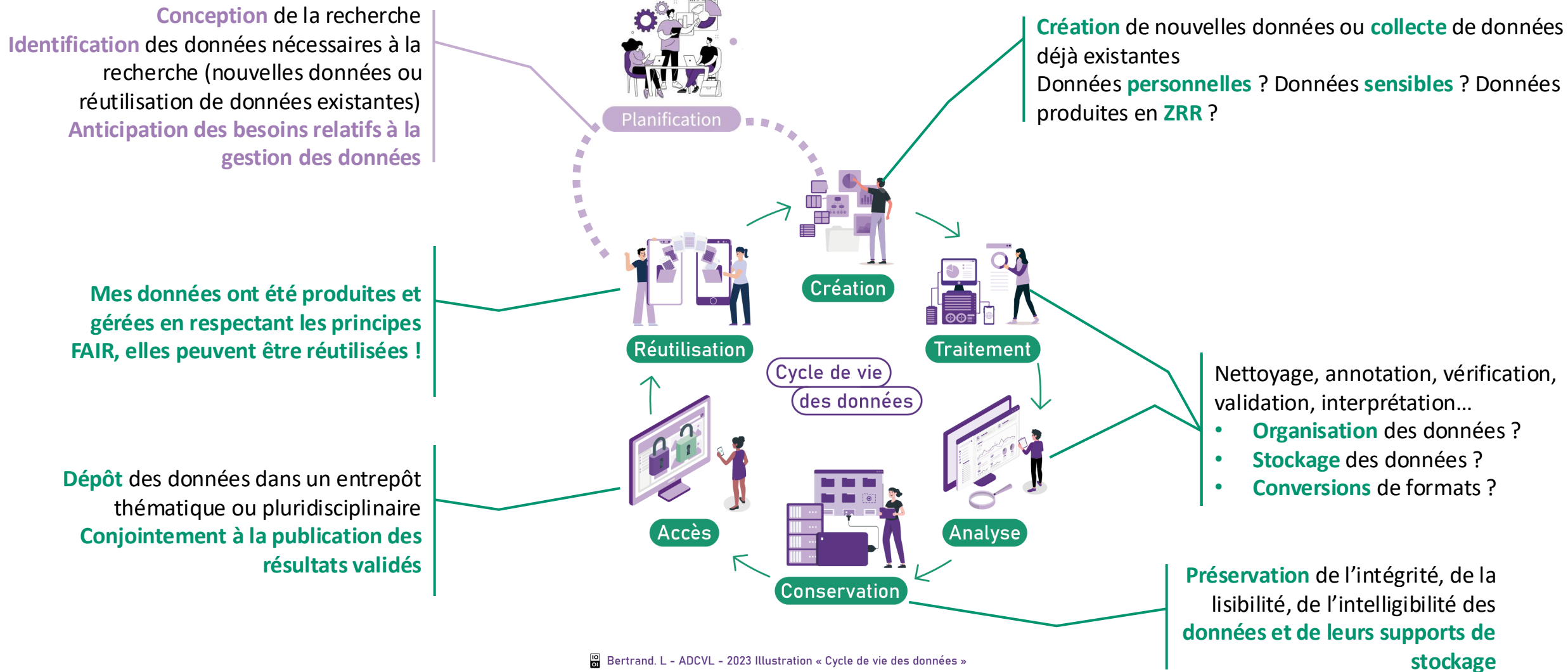
- Donnée audiovisuelle, Collection, Jeu de données, Image, Ressource interactive, Modèle, Objet physique, Service, Logiciel, Son, Texte, workflow
- Brute ou dérivée
- Qualitative ou quantitative
- Produite, recueillie ou préexistante



Ce qui n'est pas une donnée de recherche :

- Carnet de labo
- Publication
- Projet de document scientifique
- Programme de travaux de recherche futurs
- Examen par les pairs
- Communication personnelle avec des collègues

Cycle de vie des données



Les bases d'une métadonnée

Métadonnées
embarquées
produites par les
appareils ou
logiciels

Métadonnées
enrichies
ajoutées par
l'auteur

Schémas standards de métadonnées

- Faits par les communautés scientifiques
- Ensemble d'éléments définis (titre, auteur, date...)
- Contraintes (éléments obligatoires, syntaxe...)

Objectifs

- Retrouver ses données,
- les protéger (propriété),
- les diffuser largement.

Quel standard ?

- Selon le type de données
- Selon le domaine scientifique
- Selon les besoins spécifiques
- Selon le choix de l'entrepôt de partage
- Selon l'étape du cycle de vie des données



Exemples

Généralistes

- **Dublin Core (DC)**
- **DataCite Metadata Schema**

Disciplinaires

- **Data Documentation Initiative (DDI)**
pour décrire les données issues
d'enquêtes et d'autres méthodes
d'observation en sciences sociales,
comportementales, économiques et de
la santé
- **Ecological Metadata Language (EML)**
pour décrire les données dans le
domaine de l'écologie

« Les **métadonnées** sont un ensemble d'informations qui servent à documenter et décrire des données, et à faciliter leur réutilisation. »

Les principes FAIR

Les principes FAIR ont été créés en 2016 par le collectif de chercheurs FORCE11. Ces principes reposent sur une volonté de guider les producteurs de données de la recherche pour rendre leurs données ouvertes.

Les principes FAIR sont donc pensés comme une solution pour l'Open Data, qui avait déjà été formalisée d'une autre façon dans le 5-star Open Data de Tim Berners-Lee en 2010. Ces deux directives sont complémentaires mais doivent être distinguées, l'une explicitant clairement l'usage de technologies identifiées à l'ensemble du web (5-star Open Data), l'autre comme une suite de recommandations sans contrainte d'intégration et spécifiques aux données de la recherche (FAIR).

L'application de l'Open Data dans la recherche vise à permettre une meilleure reproduction des expériences, une meilleure accessibilité des résultats de la recherche et un meilleur croisement des résultats.

Les principes FAIR

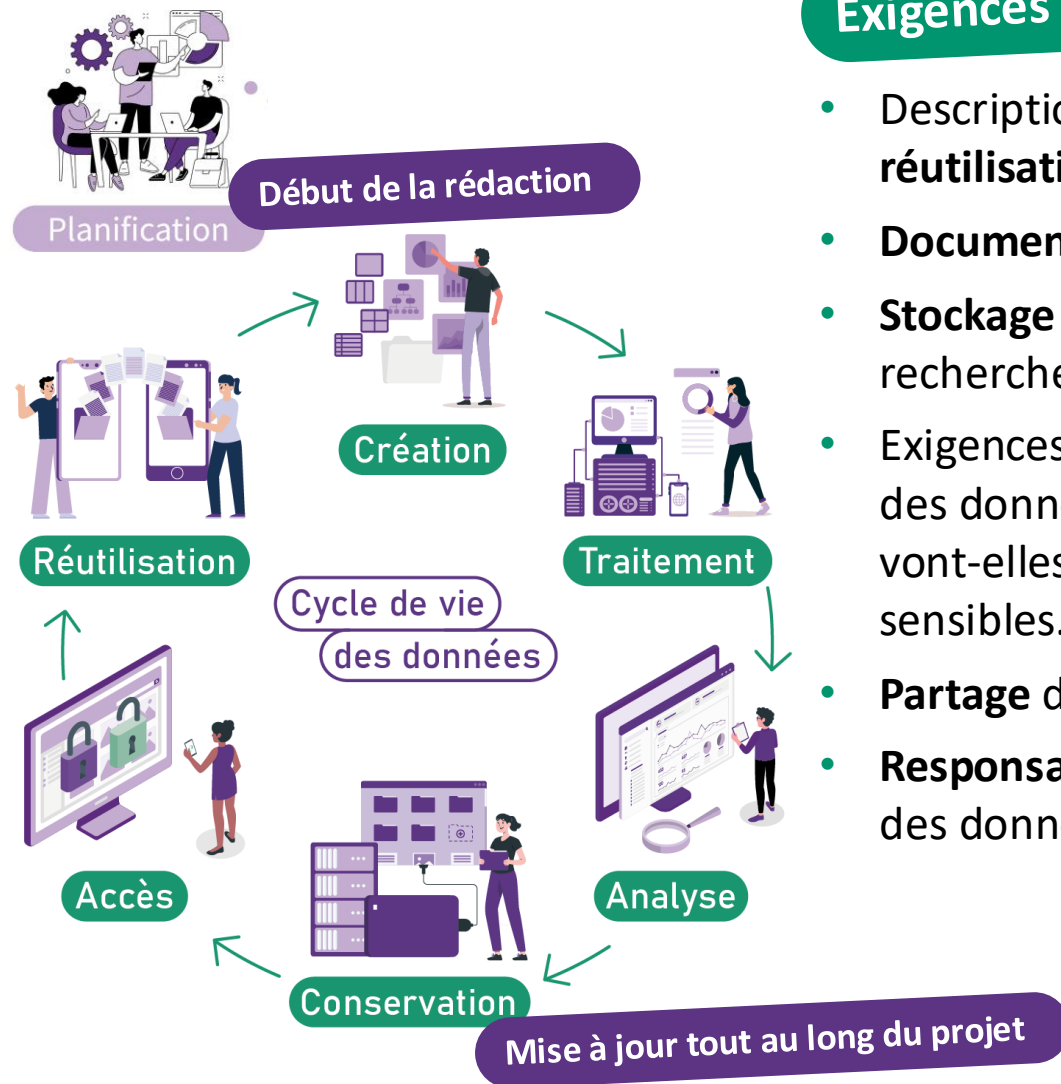
Comme son nom l'indique, les principes FAIR se découpent en quatre grands axes :

- **F – Facile à trouver** : Donner des identifiants persistants ; décrire avec des métadonnées (elles-mêmes identifiables avec des identifiants persistants) ; déposer dans un entrepôt
- **A – Accessible** : Utiliser un protocole standard, libre et ouvert ; équilibrer les besoins d'authentications ; préserver les métadonnées
- **I - Interopérable** : Utiliser un vocabulaire lui-même respectant les principes FAIR pour décrire et indexer ses données ; contextualiser ses données en les liant à d'autres données/ressources ; utiliser
- **R - Réutilisable** : Décrire de façon étendue sa donnée (plusieurs attributs sur une même métadonnée), attribuer une licence de réutilisation, donner la provenance de la donnée dans ses métadonnées, utiliser des standards de la communauté.

Chaque axe est découpé en plusieurs principes qui seront décrits tout au long de la formation.

Le Plan de Gestion de Données (PGD)

- ✓ Outil de **gestion de projet**, évolutif et collaboratif
- ✓ Permet de **définir et mettre en place des stratégies** de gestion, de partage et de conservation des données
- ✓ Participe à la **compréhension des données** pour les personnes extérieures au projet (s'il est partagé)
- ✓ Plusieurs **modèles** proposés par les financeurs et institutions (Ex. ANR, Commission européenne, CNRS...) disponibles sur [DMP OPIDoR](#)



Exigences minimales

- Description des **données**, de leur **collecte** ou **réutilisation**
- **Documentation** et **qualité** des données
- **Stockage** et **sauvegarde** pendant le processus de recherche
- Exigences **légales** et **éthiques**, **codes de conduite** : si des données personnelles sont présentes comment vont-elles être gérées ? Idem pour les données sensibles.
- **Partage** des données et **conservation à long terme**
- **Responsabilités** et **ressources** en matière de gestion des données : les rôles de chaque membre du projet



PGD entité pour décrire la gestion des données à l'intérieur d'un laboratoire, pour un même équipement....

Toutes les données sont-elles diffusables?

Aussi ouvert que possible, aussi fermé que nécessaire

La loi ne concerne que les données publiques.

La règle est la diffusion sauf si des contraintes légales s'appliquent :

- secret des affaires,
- secret statistique,
- secret défense,
- **RGPD**
- Etc.



Le RGPD, qu'est-ce que c'est ?



RGPD = règlement général sur la protection des données

Règlement européen qui s'applique dans tous les Etats Membres de l'Union européenne

A permis d'harmoniser la réglementation au sein de l'Union européenne

Est entré en vigueur le 25 mai 2018

Il encadre les traitements des **données à caractère personnel**

S'applique à tous les établissements publics ou privés européens mais aussi à tous les établissements hors Union européenne dès lors qu'ils traitent des données à caractère personnel de résidents européens.

Qu'est-ce qu'une donnée à caractère personnel

Information qui identifie **directement ou indirectement** une personne

- Directement : nom, prénom
- Indirectement :
 - Par recoupement d'informations : revenus, santé, commune
 - Par l'existence d'un numéro qui permet de retrouver la personne : numéro de sécurité sociale, numéro de téléphone, adresse IP, ORCID, etc.



Données à caractère personnel sensibles

Les données sensibles dans le cadre des données à caractère personnel sont :

- Origine raciale ou ethnique
- Opinions politiques
- Convictions philosophiques ou religieuses
- Appartenance syndicale
- Données génétiques
- Données biométriques
- Données concernant la vie sexuelle
- Données de santé

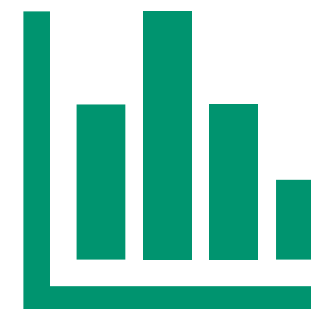


Ces données ne sont pas collectables sauf dans le cas de la recherche

Qu'est-ce qu'un traitement de DCP ?

Toute opération réalisée sur des données à caractère personnel :

- Collecte
- Accès,
- Enregistrement,
- Stockage,
- Partage,
- Utilisation,
- Archivage,
- Effacement



Obligation de déclarer les traitements à son délégué à la protection des données (DPO) qui doit tenir un registre des traitements de DCP pour son organisme (sauf exemptions)

Diffusion des données à caractère personnel

En règle générale, la diffusion des données à caractère personnel n'est pas possible sauf exception.



Pour diffuser les données, il est nécessaire de les rendre anonymes.



Supprimer le nom de la personne peut ne pas être suffisant à cause du caractère indirectement identifiable d'autres données.

Réutilisation de données à caractère personnel

La réutilisation de données à caractère personnel est un traitement de données à caractère personnel non collectées directement auprès de la personne concernée.

Possible pour les chercheurs car ce n'est pas considéré comme un changement de finalité mais une information des personnes est nécessaire.



En dehors de la recherche, c'est un changement de finalité donc le consentement des personnes est à obtenir !

L'Exemple de QueCJ (en linguistique)

<https://corli.huma-num.fr/wp-content/outil-site-quecj/index.html>

Des données "Facile à trouver"

Définition

La première étape pour respecter les principes FAIR, est de rendre ses données **trouvables** (soit « **Facile à trouver** »).

-> L'inverse de ce principe serait de **garder ses données sur le disque dur de son ordinateur**.

On pourrait considérer que la publication de données sur le web (dans un entrepôt ou sur un site web spécifique) est la première étape pour respecter ce principe. La donnée devient **trouvable par les humains et les machines** sur le web grâce à sa publication.

Toutefois, des recommandations supplémentaires sont à suivre pour respecter pleinement le F de FAIR.

Définition

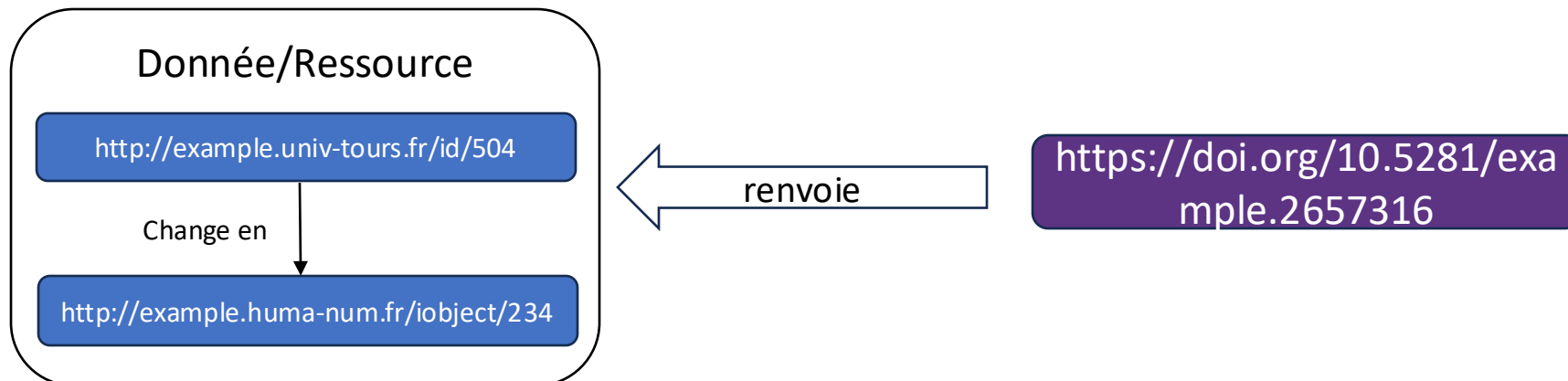
F1 – Les données ont un Persistent Identifier (PID)

Un **PID** ou **Identifiant Pérenne** en français est un moyen d'identifier une donnée comme une ressource stable et unique sur le web.

Un PID permettra de toujours retrouver une donnée peu importe les changements d'hébergement ou d'identifiant dans la base de données d'origine. -> **On évite les erreurs 404**

Plusieurs systèmes d'identifiants pérennes existent : **DOI (Digital Object Identifier)** , **ARK (Archival Resource Key)**, **Handle**

Vous pouvez acquérir un identifiant pérenne pour votre donnée via des entrepôts de données ou des fournisseurs académiques.



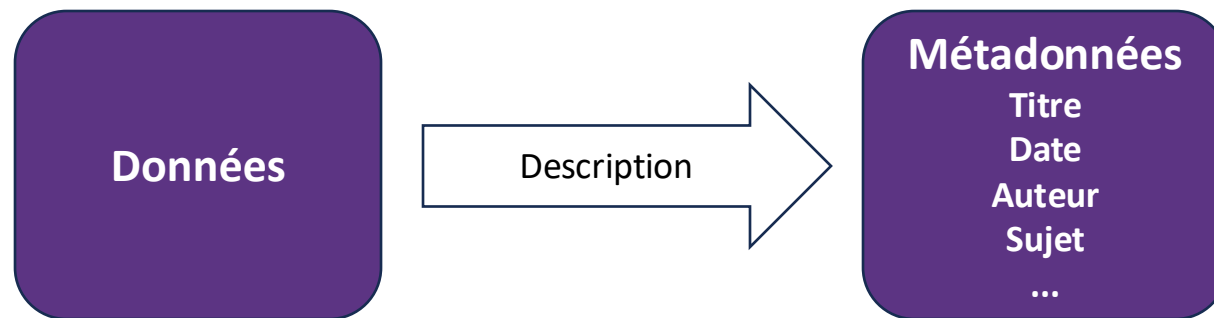
Définition

F2 – Les données sont décrites avec des métadonnées riches

Une métadonnée est un **élément qui décrit une donnée**. C'est ici par exemple, que l'on donne l'auteur, le titre ou la date pour un texte à l'aide du *Dublin Core Schema* par exemple.

Les métadonnées sont les éléments qui permettent de retrouver une donnée via des critères de recherche (les données les plus récentes, produites par X, avec sujet Y, etc.), à la fois pour un utilisateur humain que pour une interface machine.

Pour les métadonnées, **trop n'est jamais assez**, plus une donnée est décrite mieux les principes FAIR seront respectés.



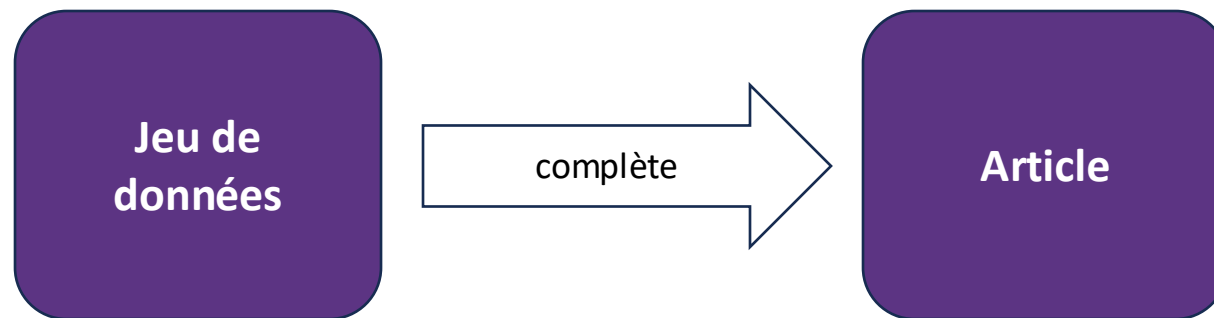
Définition

F3 – Utiliser des Identifiants Pérennes dans les métadonnées

Quoi de mieux pour décrire une donnée que de la **lier avec une autre donnée** ?

Dans un monde parfait, chacune de vos données possède un identifiant pérenne, que vous pouvez citer dans les métadonnées d'une autre donnée.

Trop n'est jamais assez, il ne faut pas hésiter à citer ces identifiants pérennes pour commencer à créer un **nuage de données**.



Définition

F4 – Entreposer vos données pour les indexer

Le but de réaliser les principes de F1 à F3 est d'indexer vos données et de les interroger dans un moteur de recherche, afin de les rendre **Facile à trouver**.

Plusieurs systèmes permettent de réaliser cela, notamment les **entrepôts de données** comme *Nakala*, *Dataterra*, *Zenodo*, etc. Ces systèmes sont indexés par de plus grands portails, qui eux-mêmes sont mieux mis en avant par les moteurs de recherche de nos navigateurs web.

On a donc un passage à l'échelle **progressif** qui permet à vos données d'être **visibles** dans la masse du web.

Vous pouvez réaliser votre propre système qui permettra **d'être indexé et de passer à l'échelle**, ou utiliser les **entrepôts de données** qui vous guideront dans le respect des **principes FAIR** (et notamment F1-F3 discuté) en complément.

Ces données sont-elles "faciles à trouver" ou non ?

<https://www.wooclap.com/fr/>

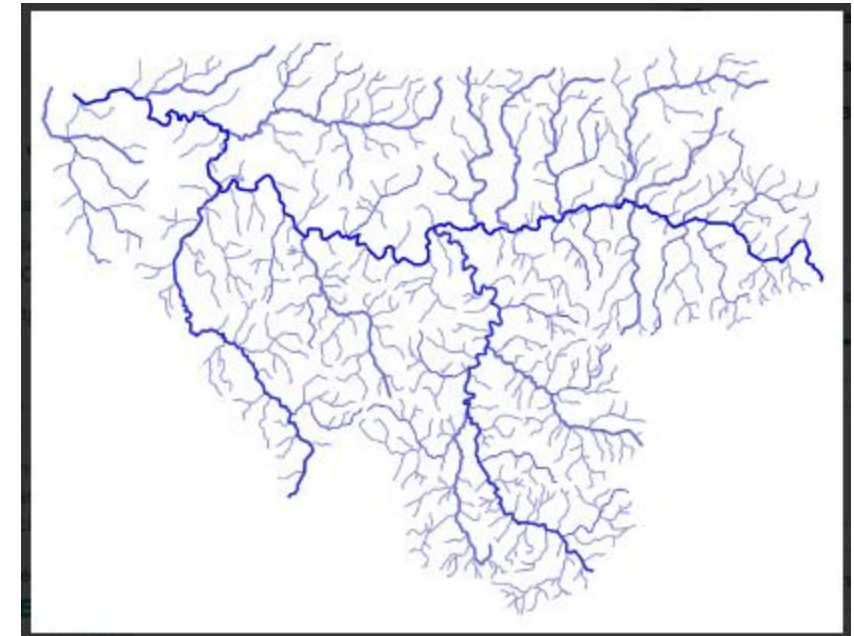
Code Wooclap de l'évènement : **SEMLUZ**

Exemple d'une donnée "Facile à trouver"

- Deux **données de recherche** portant sur des **informations géographiques** au **format image**
- Une issue de l'archéologie et l'autre de l'hydrologie
- Toutes les deux **mises en ligne mais de manière très différentes**



<https://lat.gitpages.huma-num.fr/coenotur/cormery.html>



<https://geodata.inrae.fr/datahub/dataset/1c7945a5-0e93-4807-b839-0a070d9af639>

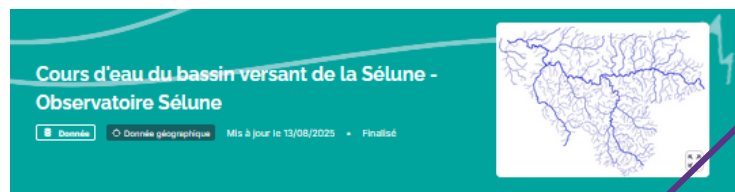
D'importants réaménagements furent entrepris au cours des XIII^e-XIV^e siècles, durant lesquels le transept et le chevet furent passablement modifiés. Dans l'emprise du bras nord du transept des sondages archéologiques menés en 2017 ont permis de mettre au jour de nombreux fragments de verres peints à la grisaille décorés de motifs végétalisant qui ornent les vitraux du transept de l'église gothique. Bien que la partie orientale de l'église abbatiale ait disparu, l'abondante documentation de l'époque moderne, notamment les dessins des tombeaux abbatiaux réalisés pour Galignières et conservés à la BnF, permettent de documenter le chœur de l'édifice de culte qui était marqué par un important usage funéraire. Au XV^e siècle le bras sud du transept de l'église, qui menaçait de s'effondrer, fut épaulé par la construction d'une tour massive en forme de fer à cheval qui était dotée de trois archères-canonnières. Aux XVII^e et XVIII^e siècles, assez peu d'aménagement furent entrepris, à l'exception de la chapelle Saint-Symphorien qui fut agrandie à deux reprises.



Figure 3. Plan des vestiges en élévation de l'espace claustral de Cormery (© Th. Pouyet).

Les bâtiments monastiques de Cormery sont parmi les mieux conservés de Touraine avec trois grands ensembles préservés : le réfectoire à l'ouest, les celliers au nord et la salle capitulaire à l'est. Dans l'aille septentrionale, le bâtiment des celliers, qui remonte au XII^e siècle, était partiellement enterré et mesurait 28 m de long sur 12 m de large. Divisé en quatre travées voûtées avec des ogives, l'intérieur était éclairé par des baies aux ouvertures très étroites en façade, mais profondément ébrasées à l'intérieur. La disposition des celliers au nord, plutôt inhabituelle dans un monastère bénédictin, pourrait s'expliquer par la proximité de la rivière Indre au nord, qui est supposée navigable jusqu'au XII^e siècle avant l'implantation de nombreux moulins sur son cours.

- Donnée utilisée pour illustrer un développement scientifique
- La donnée est insérée dans une page web mais pas dans un espace dédié
- Pas d'identifiant unique
- Découvrabilité inexistante : Introuvable par des moteurs de recherche, seul accès par un lien direct



L'effacement des barrages de la Sélune fait l'objet d'un programme scientifique initié en 2012 dans lequel des paramètres environnementaux sont suivis afin de fournir un réel retour d'expérience sur cette opération de restauration.

En 2019 s'est mis en place un observatoire de données, chargé d'assurer la

[Lire la suite](#)

Mots-clés

données ouvertes hydrologie limite hydrographie

Licences et conditions d'utilisation

Pas de restriction d'accès public

Usage libre sous réserve des mentions obligatoires sur tout document de diffusion :
"Source : UMR DECOD INRAE - Agrocampus Ouest"

Licence ouverte : <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

À propos de la donnée

Ces cours d'eau sont extraits de la BD Topage sur l'emprise de la zone d'étude. Des codes, utilisables dans le cadre du programme Sélune ont été ajoutés, ainsi que les codes issus de l'ancien référentiel hydrographique (BD Carthage).

Producteur de la donnée

UMR ESE INRAE - Agrocampus Ouest

<https://programme-selune.com/fr/>

Jean-Marc.Roussel@inrae.fr

Date de création
12/02/2025

Fréquence de mise à jour des données
Les données sont mises à jour selon les besoins

Langues
Anglais

Autres informations

Modifié le
13/08/2025 14:03:16

Fiche de métadonnées d'origine

<https://geodata.inrae.fr/geonetwork/srv/fr/catalog.search#/metadata/1c7945a5-0e93-4807-b839-0a070d9af639>

Catalogue d'origine

UMR ESE INRAE - Agrocampus Ouest

Identificateur de ressource unique

1c7945a5-0e93-4807-b839-0a070d9af639

eaux intérieures, hydrographie, environnement

Mots-clés

données ouvertes

hydrologie

limite

hydrographie

À propos de la donnée

Ces cours d'eau sont extraits de la BD Topage sur l'emprise de la zone d'étude. Des codes, utilisables dans le cadre du programme Sélune ont été ajoutés, ainsi que les codes issus de l'ancien référentiel hydrographique (BD Carthage).

Producteur de la donnée

UMR ESE INRAE - Agrocampus Ouest

<https://programme-selune.com/fr/>

Jean-Marc.Roussel@inrae.fr

Identificateur de ressource unique

1c7945a5-0e93-4807-b839-0a070d9af639

- La donnée a été déposée sur un **outil dédié aux jeux de données** : Geodata INRAE
- la donnée a une existence propre (on peut renvoyer vers elle)
- la donnée est accompagnée de **métadonnées enrichies** : mots-clés, auteur, description ...
- Présence d'un Identifiant unique**
- Découvrabilité forte** : trouvable par les moteurs de recherche, par l'outil de recherche interne de l'outil de dépôt de données eul accès : URL directe

Des données "Accessible"

Définition

Dans le contexte de la recherche et de la science ouverte, l'accessibilité vise à garantir que les données peuvent être **localisées, consultées et téléchargées** sans obstacle technique ou juridique.

Cela signifie que les utilisateurs, les machines et les systèmes doivent pouvoir accéder aux données et à leurs métadonnées de manière fiable et durable, quel que soit le temps écoulé ou les changements d'infrastructure.

L'enjeu est d'éviter la perte de données, les liens brisés (erreurs 404) et les barrières techniques qui empêchent leur réutilisation.

Le principe A possède plusieurs recommandations à suivre pour le respecter.

Définition

A1 – Les données sont récupérables par leur identifiant

Les données doivent être accessibles via un protocole standard (*HTTP, HTTPS, FTP*) en utilisant leur identifiant pérenne (PID). Cela permet à tout utilisateur ou programme de récupérer la donnée ou ses métadonnées en un clic, sans dépendre de l'URL d'origine. Cette récupération pérenne sert à la fois pour l'affichage de la donnée dans un navigateur ou pour réutiliser les données → **On évite les données introuvables ou bloquées par des restrictions techniques.**

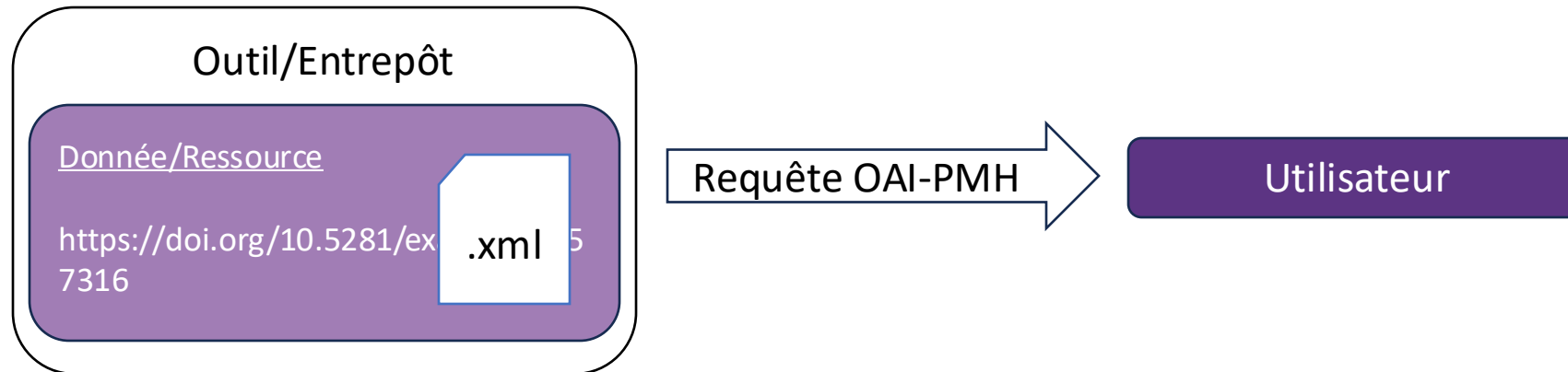
Là encore, la solution reste de déposer sa donnée ou son jeu de donnée dans un entrepôt de données certifié, qui garantira l'attribution d'un PID et l'accès à la donnée via celui-ci.

Définition

A1.1 – Le protocole d'accès est ouvert, libre et universellement implémentable

Le protocole utilisé pour accéder aux données doit être standard, gratuit et ne pas dépendre de technologies propriétaires. Par exemple, privilégier HTTP/HTTPS plutôt qu'un protocole propriétaire ou un format nécessitant un logiciel payant. L'utilisation de formats ouverts (CSV, JSON, XML) et de protocoles standardisés (REST, OAI-PMH) donne un accès équitable à tous.

La première chose à ne pas faire en tant que producteur de données, est d'utiliser un logiciel propriétaire pour publier ses données ou les enregistrer dans un format propriétaire. Si le logiciel disparaît, les données ne pourront pas être réutilisées et seront perdus.



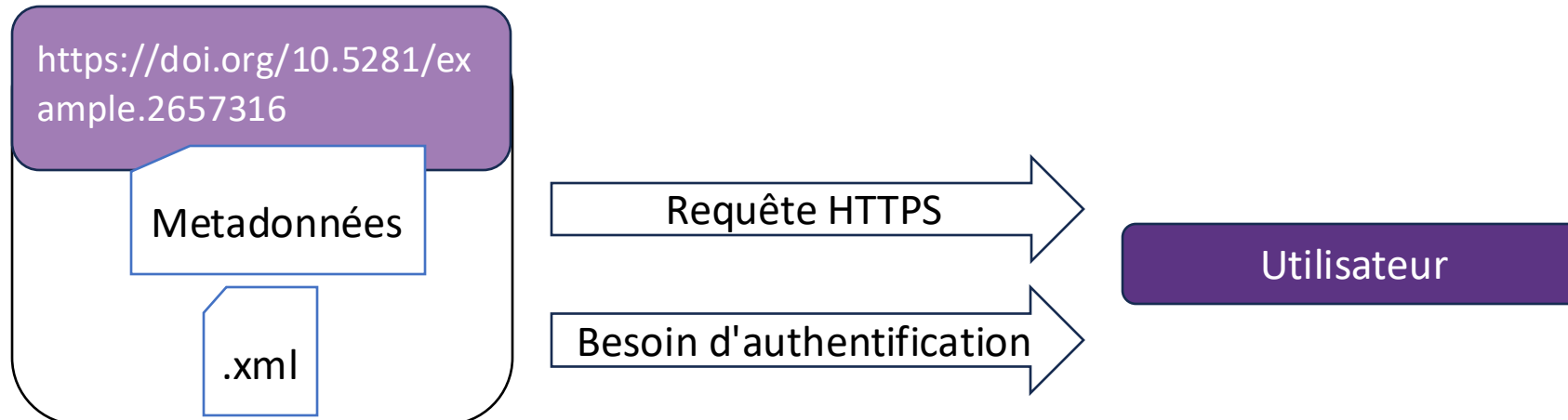
Définition

A1.2 – Les données sont accessibles via un protocole d'authentification standardisé si nécessaire

Si les données sont sensibles (données personnelles, médicales), leur accès doit être sécurisé via des protocoles d'authentification standardisés, comme Shibboleth ou OAuth2.

Par exemple, utiliser EduGAIN pour une authentification unique dans les collaborations internationales. Les entrepôts sécurisés (comme EGA pour les données génomiques) permettent de gérer les niveaux d'accès différenciés.

Le but est d'ouvrir et de rendre accessible les données le plus possible sans dépasser les limites juridiques qui peuvent s'y appliquer.

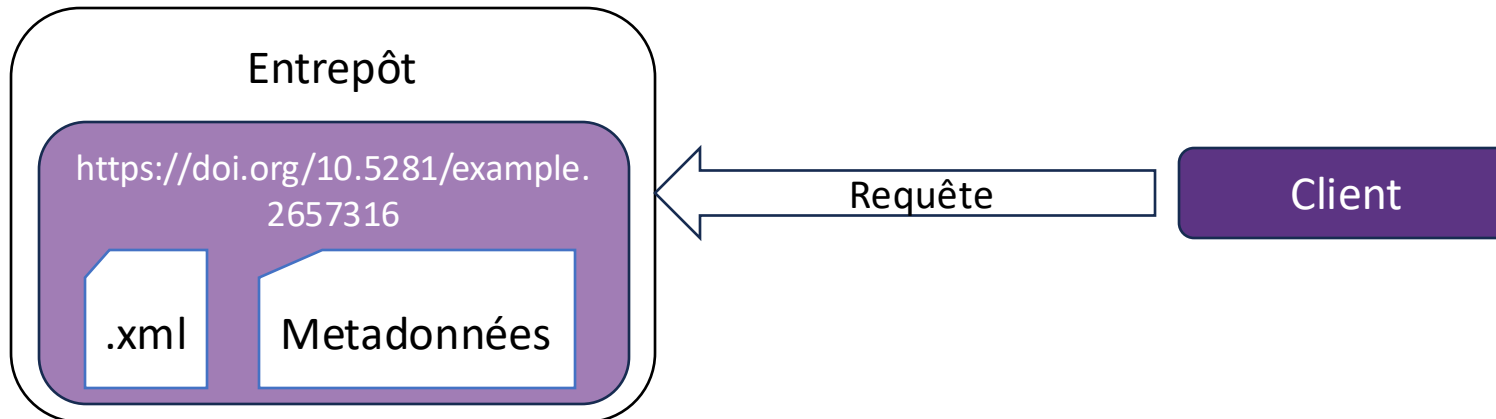


Définition

A2 – Les métadonnées sont accessibles même si les données ne le sont plus

Même si les données brutes sont supprimées ou déplacées, leurs métadonnées doivent rester accessibles pour documenter leur existence et leur contexte.

Un catalogue comme affiche les métadonnées d'un jeu de données supprimé, avec une mention de sa disponibilité passée. L'archivage des métadonnées dans des entrepôts dédiés garantit leur pérennité. Les métadonnées et les données resteront disponibles, et permettront d'y conserver un accès ne dépendant pas d'une situation interne instable.

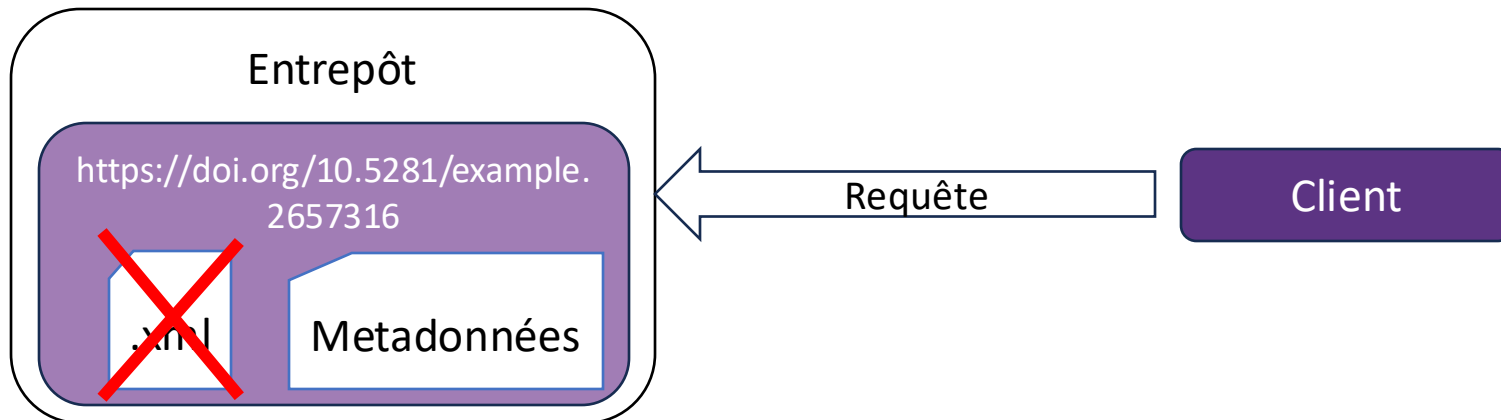


Définition

A2 – Les métadonnées sont accessibles même si les données ne le sont plus

Même si les données brutes sont supprimées ou déplacées, leurs métadonnées doivent rester accessibles pour documenter leur existence et leur contexte.

Un catalogue affiche les métadonnées d'un jeu de données supprimé, avec une mention de sa disponibilité passée. L'archivage des métadonnées dans des entrepôts dédiés garantit leur pérennité. Les métadonnées et les données resteront disponibles, et permettront d'y conserver un accès ne dépendant pas d'une situation interne instable.



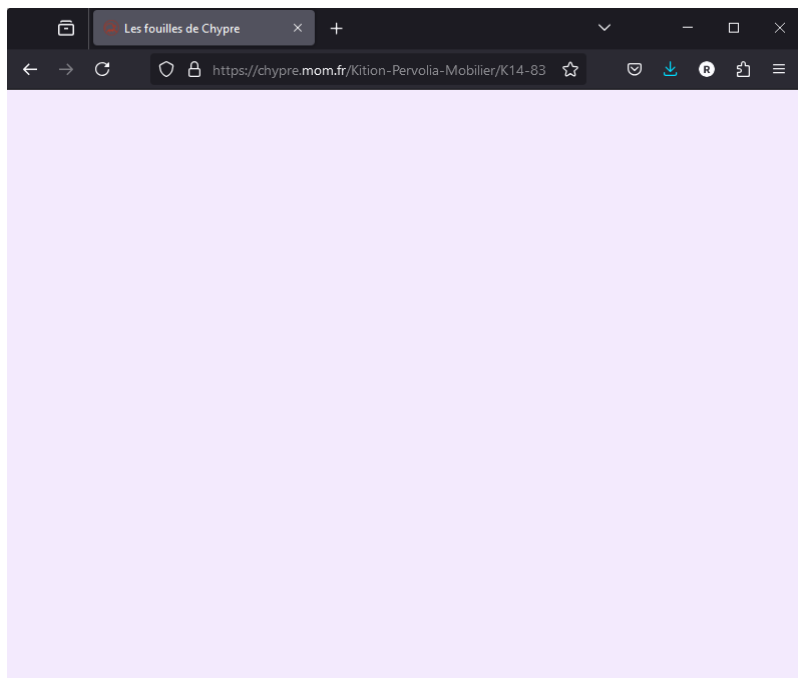
Ces données sont-elles "accessibles" ou non ?

<https://www.wooclap.com/fr/>

Code Wooclap de l'évènement : **YIXOTU**



Exemple d'une donnée "Accessible"



<https://chypre.mom.fr/Kition-Pervolia-Mobilier/K14-83>

Explorateur Sparnatural

Service SPARQL : <http://openarchaeo-graphdb.huma-num.fr/repositories/openarchaeo>

Mobilier $\rightarrow$ avait comme utilisé $\rightarrow$ Type $\rightarrow$ stèle $\rightarrow$ +

Afficher l'éditeur SPARQL / Exporter la requête en JSON / Charger la requête en JSON

Résultats

Table Grid Stats Map Timeline 5 results Page size: 50

HumanMadeObject_1

- 1 Mobilier K13-04
- 2 Mobilier K13-12 : bord de stèle
- 3 Mobilier K02-46
- 4 Mobilier K14-83
- 5 Mobilier K14-03

Showing 1 to 5 of 5 entries

Portail Sparnatural

- Données "inaccessibles" sur le site du projet mais métadonnées "Accessibles" via l'API donc rempli les conditions du "A"

Cat OPIDoR

pour des données FAIR tout au long de leur cycle de vie

Cat OPIDoR

Cat OPIDoR est un catalogue qui recense et décrit les services français qui contribuent à la gestion, au partage et la réutilisation des données scientifiques tout au long de leur cycle de vie.

Le catalogue permet aux producteurs et aux utilisateurs de données d'identifier aisément les services dont ils ont besoin dans le cadre de leurs projets de recherche.

Wiki sémantique avec moteur de recherche, cartographie, accès par domaines et sous-domaines scientifiques

<https://cat.opidor.fr/>

Les autres outils

Les entrepôts de données
pour rendre ses données "Faciles à trouver" et "Accessibles"

Les entrepôts de données

Pourquoi ?

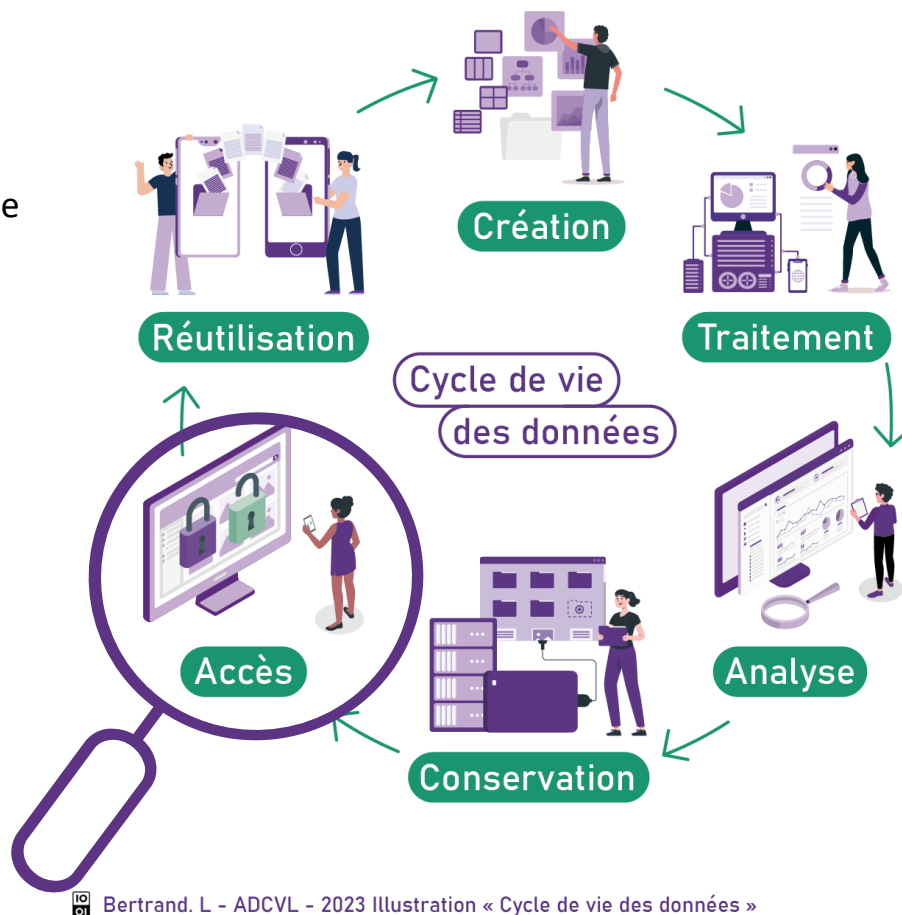
Exposer ses données

Citer ses données dans ses publication grâce aux **identifiants pérennes (PID)**

- Favoriser la réutilisation de ses données
- Augmenter la visibilité de ses travaux
- Intégrité scientifique
- Répondre aux exigences des financeurs

Quelles données ?

- Finalisées, validées,
- Conformes au RGPD
- Non confidentielles
- Non protégées par un brevet



10 01 Bertrand. L - ADCVL - 2023 Illustration « Cycle de vie des données »

Où ?

- Entrepôts thématiques, pluridisciplinaire ou institutionnels

Comment choisir ?

- **Pratiques de la communauté scientifique**
- Modèle **économique** (gratuit/payant)
- **Types** de données acceptés
- Attribution d'un **identifiant pérenne**
- Conditions de **réutilisation acceptées** (licences)
- Modes de **visibilité** offerts (totale, embargo)
- **Certification / labellisation**
« [entrepôt de confiance](#) »

Exemples d'entrepôts

NAKALA : l'entrepôt de données des SHS

Caractéristiques de l'entrepôt

- **Entrepôt de données de recherche** français de référence pour les **Sciences Humaines et Sociales**
- Géré par l'IR* **Huma-Num** et fait partie de l'écosystème national "Recherche Data Gouv"
- Métadonnées au format **Dublin Core qualifié**
- **Pas de modération/curation systématique des dépôts** mais possibilité de **demande une modération** pour obtenir un label "jeu de données de qualité"
- **Dépôt manuel** ou **en lot** via outil Mynkl
- Attribution d'un **DOI**

Pour quelles données ?

- Données qualitatives, quantitatives, textuelles, images, sons, vidéos ...
- Limite de 10Go / fichiers sans limite de taille globale
- Ouvertes, sous embargo et secrètes

Liens



- [Accès à l'entrepôt](#)
- [Documentation](#)
- [Formations](#)

EasyData : l'entrepôt des sciences de la Terre

Caractéristiques de l'entrepôt

- **Entrepôt de données de recherche** français pour les données de l'**Environnement et du Système Terre**.
- Porté par l'ensemble des partenaires de l'**IR DATA TERRA** et des pôles de données (CNRS, CNES, IRD, INRAE, Météo-France, Ifremer, IGN, BRGM, ...)
- Métadonnées **compatibles et exportables** vers d'autres formats et schémas de métadonnées (GeoDCAT-AP, DataCite, ...)
- **Modération systématique** des dépôts avant publication
- Attribution d'un **DOI**

Pour quelles données ?

- données dites « orphelines » ou de longue traine de l'**Environnement et du Système Terre**.
- Limite de **5Go** / fichier et de **100Go** pour la totalité des fichiers déposés.
- Ouvertes, sous embargo

Liens

- [Accès à l'entrepôt](#)
- [Documentation](#)



EaSy Data
Earth System Data Repository

Zenodo : l'entrepôt généraliste (déconseillé)

Caractéristiques de l'entrepôt

- **Entrepôt pour le partage des résultats de la recherche (publications et données de recherche)** pour lesquels on ne dispose pas de dépôts institutionnels, disciplinaires ou thématiques
- Proposé par le **CERN** (European Organization for Nuclear Research), issu d'une collaboration avec **OpenAire** (Open Access Infrastructure for Research in Europe) et **financé par la Commission Européenne**
- Métadonnées au format **DataCite Metadata Schema, Dublin Core**
- **Pas de curation par défaut**, curation nécessaire avant publication selon les communautés de dépôt
- Attribution d'un **DOI**



Pour quelles données ?

- Texte, Son, Image, Vidéo, Code informatique, Métadonnées
- Limite de **50Go** / fichier
- Ouvertes, sous embargo, fermées

Liens



- [Accès à l'entrepôt](#)
- [Documentation](#)

Caractéristiques de l'entrepôt

- **Entrepôt pour le partage des résultats de la recherche** (**publications et données de recherche**) pour lesquels on ne dispose pas de dépôts institutionnels, disciplinaires ou thématiques
- dépôt des données dans des **espaces institutionnels**, ou à défaut dans **l'espace générique**
- métadonnées aux formats **Dublin Core, DDI, DataCite Metadata Schema, OAI-ORE**,
- **curation systématique** des données déposées avant publication,
- attribution d'un **DOI**

Pour quelles données ?

- Texte, Son, Image, Vidéo, Code informatique, Métadonnées
- Limite de **50Go** / jeu de données
- Ouvertes, sous embargo, fermées

Liens



- [Accès à l'entrepôt](#)
- [Documentation](#)
- [Formations](#)



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

recherche.data.gouv.fr

Exemple de lien publication / code / données

Publication

Besson, S. (2025). Enquête sur les projets en Humanités numériques. Zenodo.
<https://doi.org/10.5281/zenodo.15726734>

Besson. Enquête sur les projets en Humanités numériques. MSH Lyon - Saint-Étienne; Huma-Num. 2024. [hal-04886620](#)

Données

Besson, S. (2025). Jeux de données issus d'une enquête sur les projets en Humanités numériques (Version 3) [Data set]. NAKALA - <https://nakala.fr> (Huma-Num - CNRS). <https://doi.org/10.34847/NKL.1BEDV5UJ>

Code

Sylvain Besson. Script for analyzing the survey of digital humanities projects on the Lyon Saint-Étienne site. 2025,
([swh:1:dir:92208f0de86c44fb37ae765fb86d3d4dc95a2b3f;origin=https://hal.archives-ouvertes.fr/hal-04935698](https://hal.archives-ouvertes.fr/hal-04935698);visit=[swh:1:snp:1bd223f633c6aa68b441d54059ca312c787ce318;anchor=swh:1:rel:c1d71f12492f8327bdf4a961169647e1dbc513c8;path=/](https://hal.archives-ouvertes.fr/hal-04935698)). ([hal-04935698](#))

Enquête sur les projets en Humanités numériques

en

fr

Sylvain Besson (1)

[Afficher plus de détails](#)



1 MSH LSE - Maison des Sciences de l'Homme Lyon Saint-Etienne

Ressources associées ▾



Cite [10.34847/nkl.1bedv5uj](#) Jeu de données



— Besson, S. (2025). *Jeux de données issus d'une enquête sur les projets en Humanités numériques* (Version 1) [Data set]. NAKALA - <https://nakala.fr> (Huma-Num - CNRS). <https://doi.org/10.34847/NKL.1BEDV5UJ>



Est citée par [hal-04935698](#) Logiciel



— Sylvain Besson. Script for analyzing the survey of digital humanities projects on the Lyon Saint-Étienne site. 2025, ([swh:1:dir:92208f0de86c44fb37ae765fb86d3d4dc95a2b3f;origin=https://hal.archives-ouvertes.fr/hal-04935698](https://hal.archives-ouvertes.fr/hal-04935698);visit=[swh:1:snp:1bd223f633c6aa68b441d54059ca312c787ce318;anchor=swh:1:rel:c1d71f12492f8327bdf4a961169647e1dbc513c8;path=/](https://hal.archives-ouvertes.fr/hal-04935698)). ([hal-04935698](#))



Est identique à [10.5281/zenodo.15726734](#)

— Besson, S. (2025). Enquête sur les projets en Humanités numériques. Zenodo. <https://doi.org/10.5281/ZENODO.15726734>

Merci

Retrouvez-nous sur <https://adcvl.pu-cvl.fr/>



@Atelier de la donnée Centre-Val de Loire