

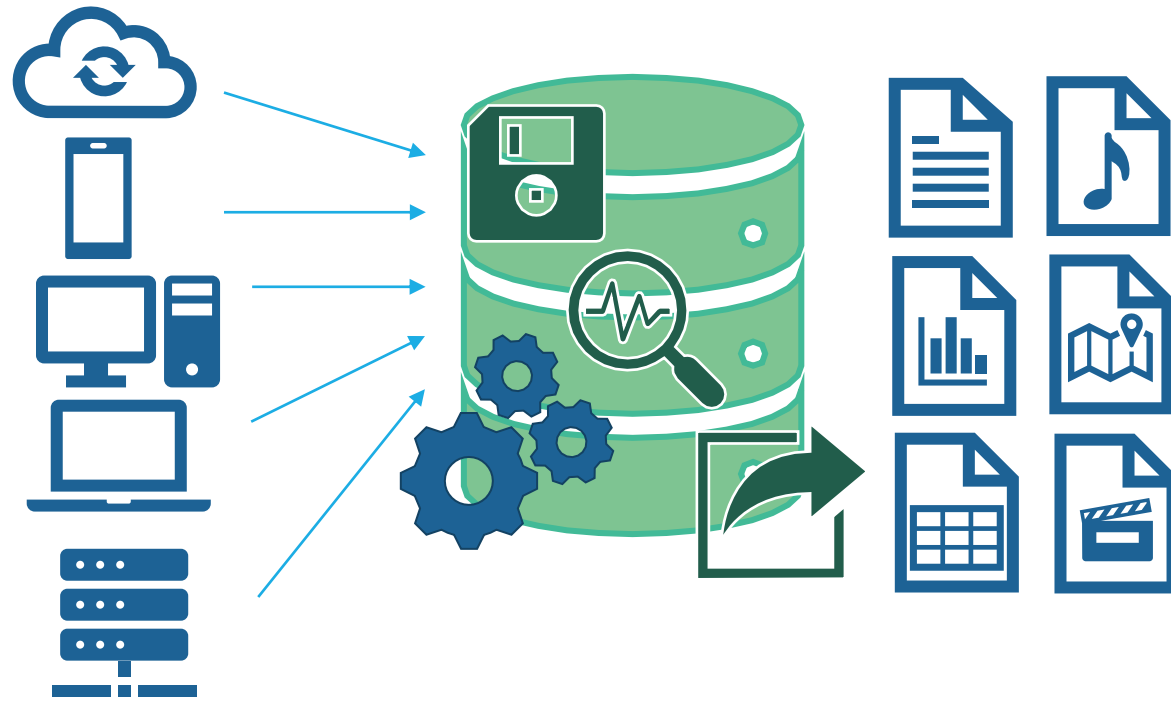
Mettre en place son entrepôt de données avec *Dataverse*

UN OUTIL AU SERVICE DE LA SCIENCE OUVERTE AU SUD



Pourquoi créer son entrepôt ?

Pourquoi créer son entrepôt de données ? Qu'est-ce qu'un entrepôt ?



Un **entrepôt de données de recherche** (Research Data Repository ou Data Repository) est une base de données destinée à **accueillir, conserver, rendre visibles** et **accessibles** des données de recherche.

Son rôle est de permettre le **dépôt** ou la collecte de données, leur **description**, leur **accès**, et leur **partage** en vue de leur **réutilisation**.

Pourquoi créer son entrepôt de données ?

Côté utilisateur: pourquoi utiliser un entrepôt ?

Des outils dédiés aux partage des données

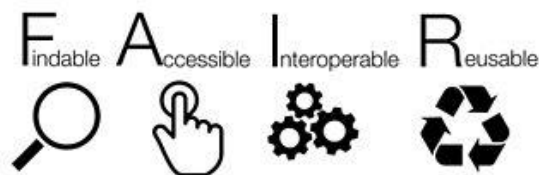
Pérennité



Visibilité



Respect des principes FAIR



Conditions de partage



Citation (DOI)



Pourquoi créer son entrepôt de données ?

Élément de base d'un entrepôt :

le jeu de données (dataset)

Regroupement *cohérent* de « données », contenant :

Fiche descriptive
(métadonnées)

Titre
Résumé
Auteur
Mots-clés
Etc.

Identifiants unique
/ référence

Fichiers de
données

(parfois aucun ou en
dehors de la plateforme)

Documentations
associées

Read me
Dictionnaire des variables
Figures
Schémas
Notes
Guides de l'enquêteurs
Formulaires
etc.

Conditions
d'utilisation

Historique des
versions

Pourquoi créer son entrepôt de données ?

Élément de base d'un entrepôt :

le jeu de données (dataset)

Quel critère pour déterminer les contours d'un jeu de données ?

Données associées à un article

Données d'un projet de recherche

Données d'un composant d'un projet

Données d'une station de mesure

Regroupements géographiques, temporels, par types de données etc.

Pourquoi créer son entrepôt de données ?

Les différents types d'entrepôts ?

- **Généralistes** (ex : Dryad, Figshare, Zenodo)  **DRYAD**  figshare 
- **Thématiques** (ex : GenBank, UniProt, IntAct, GBIF, Movebank)  GenBank  UniProt  IntAct
 GBIF  MOVEBANK
- **Disciplinaires** (ex : Pangaea, WormBase, MycoBank)  PANGAEA.  WormBase 
- **Institutionnels** (ex : Dataverses du Cirad et de l'IRD)  The Dataverse[®] Project

Pourquoi créer son entrepôt de données ?

Trouver et choisir son entrepôt

The screenshot shows the re3data.org website interface. On the left is a 'Filter' sidebar with categories like Subjects, Content Types, Countries, etc. The main area has a search bar and a list of results. The first result is for 'GAWSIS' (GAW Station Information System), with details on subject(s), content type(s), and country (Switzerland). A green callout box at the bottom right of the screenshot reads 'Re3data : 2965 entrepôts'.

Re3data : 2965 entrepôts

The screenshot shows the FairSharing.org website. It features a navigation menu with 'STANDARDS', 'DATABASES', 'POLICIES', 'COLLECTIONS', 'ORGANISATIONS', 'ADD CONTENT', and 'STATS'. The main content area is titled 'Databases' and describes it as 'A registry of knowledgebases and repositories of data and other digital assets.' Below this is a search bar and a list of results, with a green callout box at the bottom right reading 'FairSharing : 1931 entrepôts'.

FairSharing : 1931 entrepôts

Critères à prendre en compte :

- Facilité d'utilisation
- Coût
- Pérennité
- Visibilité (internationale, dans votre communauté scientifique etc.)
- Modalités d'accès
- Formats acceptés
- Certification (CTS)

Pourquoi créer son entrepôt de données ?

Questions préalables

A qui l'entrepôt rend-il service ?

Quels acteurs (internes, instituts partenaires, communauté national, internationale) interagissent avec lui ?

Quels enjeux (notamment stratégiques) ?

Quels types de données sont concernés ?

Quelles contraintes (matérielles, financières, humaines, politiques) doit-on prendre en compte ?

Pourquoi créer son entrepôt de données ?

Points d'attention

1) Quel périmètre pour le futur entrepôt ?

en lien avec les critères d'acceptation des données :

- ▶ Qui peut déposer ? Pour quel public ?
- ▶ Domaines scientifiques couverts
- ▶ Nature des « données » : types, formats, volumes
- ▶ Niveau de sensibilité des données (si applicable)
- ▶ Intérêt des données
- ▶ Choix de licences attribuables aux données

Pourquoi créer son entrepôt de données ?

Points d'attention

2) Anticiper la disponibilité du service sur le long terme

Penser l'entrepôt en tant que service pérenne, à différent niveaux :

- Infrastructures pour l'hébergement des données
- Sécurité : sauvegardes, mises à jour logicielles
- Ressources humaines et compétences
- Continuité du support aux utilisateurs
- Financements en adéquation avec ces besoins
- Engagement de la gouvernance vis-à-vis des déposants
- Engagement vis-à-vis de l'attribution de DOI

Pourquoi créer son entrepôt de données ?

Points d'attention

3) Développer et entretenir la réputation de l'entrepôt

dans une approche guidée par la qualité des dépôts de données

- ✓ Respect des principes FAIR
- ✓ Des bonnes pratiques et standards communautaires
- ✓ Qualité des métadonnées
- ✓ Formaliser les règles applicables par les déposants
- ✓ Aide et conseils pratiques destinés aux déposants
- ✓ Curation systématique avant publication ?
- ✓ Revue et suivi des dépôts (*monitoring*, indicateurs, ...)

Pourquoi créer son entrepôt de données ?

Bénéfices, opportunités et risques d'un entrepôt institutionnel

FORCES

OPPORTUNITÉS

FAIBLESSES

RISQUES

Pourquoi créer son entrepôt de données ?

Bénéfices, opportunités et risques d'un entrepôt institutionnel

FORCES

- Préserver et valoriser le patrimoine des données de son organisation, contrôle des données diffusées
- Montrer son adhésion aux principes de la Science Ouverte
- Intégrer l'outil dans une offre de service (plus large) sur la gestion des données scientifiques
- Favoriser l'adoption de bonnes pratiques dans son organisation

OPPORTUNITÉS

- Se positionner comme une plateforme de référence sur les données d'un domaine
- Développer des collaborations, rejoindre des réseaux ou groupes de travail
- Développer les compétence de ses personnels
- Participer aux évolutions techniques en tant qu'exploitant d'entrepôt

FAIBLESSES

- Evaluation du degré d'implication des différents intervenants du projet (informatique, documentaire, juridique, chercheurs etc.)

Les autres faiblesses dépendront de l'institution (manque de personnel, de réseau relai en interne, périmètre de l'entrepôt etc.)

➡ Evaluer les chances de résoudre les problèmes identifiés

RISQUES

- Ressources pour assurer la continuité du service sur le long terme (pérennité des données, engagement vis-à-vis de la communauté, de Datacite)
- Convaincre les utilisateurs (producteurs de données), sous-utilisation / moyens investis
- Qualité des publications, réputation
- Respect des réglementations applicables



Le choix de *Dataverse*

Le choix de Dataverse

L'application Dataverse

- Logiciel open source, gratuit, développé depuis 2006 à l'Université de Harvard (Etats-Unis)
- Installé sur un serveur, permet d'exploiter un entrepôt généraliste
- Actuellement utilisé par plus de 70 entrepôts publics (7 en France)
- Communauté d'utilisateurs et de développeurs actives (mises à jour fréquentes, développements communautaires)
- Aide à suivre les principes FAIR, à plusieurs niveaux : interopérabilité, identifiants pérennes, licences

72 Installations



<https://dataverse.org>

Le choix de Dataverse

Principes et fonctionnalités

Attribution de DOI

Gestion des dépôts

Métadonnées
adaptées à la
discipline

Autorisation
Restriction d'accès
si nécessaire

Organisation en
arborescence

Lien provisoire
sécurisé pour les
relecteurs

Fédération
d'identité
API

Moissonnage

Gestion des droits
utilisateur

Le choix de Dataverse

Exemples d'entrepôts francophones

Recherche Data Gouv (Recherche Data Gouv) Génération datapaper

376 015 téléchargements

L'entrepôt pluridisciplinaire *Recherche Data Gouv* est une solution souveraine pour le partage d'un entrepôt disciplinaire reconnu. Il est basé sur le logiciel libre *Dataverse*. Le dépôt de données des contributeurs. Un espace générique est dédié aux données produites par les établissements. Retrouvez les actualités et événements de la plateforme *Recherche Data Gouv*.

The *Recherche Data Gouv* multidisciplinary repository is a sovereign solution for sharing data in a repository. It is based on the *Dataverse* software. Data should be deposited in the space as for data produced by institutions which do not such a space yet themselves. Tests should be done. The latest news and events on the *Recherche Data Gouv* platform.

Localisation des jeux de données

DataSuds (IRD) En savoir plus...

Localisation des jeux de données

Par Sujet

- Earth and Environmental Sciences 132
- Agricultural Sciences 58
- Medicine, Health and Life Sciences 40
- Social Sciences 22
- Computer and Information Science 15
- Other 8
- Engineering 3

Portail Cirad du libre accès

Retrouvez l'espace des données de recherche du Cirad dans la Collection Cirad sur l'entrepôt national fédéré des données de la Recherche → recherche.data.gouv

25,687 Downloads

CIRAD Dataverse (L'entrepôt institutionnel des données de la recherche)

Cirad Dataverse permet aux scientifiques du Cirad et à leurs partenaires de déposer, partager et/ou rendre publiques les données produites ou coproduites dans le cadre des travaux de recherche.

CIRAD Dataverse allows CIRAD's researchers and partners to store data produced or co-produced in the framework of collective research work or projects.

Par Sujet

- Earth and Environmental Sciences 132
- Agricultural Sciences 58
- Medicine, Health and Life Sciences 40
- Social Sciences 22
- Computer and Information Science 15
- Other 8
- Engineering 3

[Recherche.data.gouv](https://recherche.data.gouv.fr) (2022)
construit sur la base de DataInrae (2016)

[Dataverse.ird.fr \(Datasuds\)](https://dataverse.ird.fr) (2019)

[Dataverse.cirad.fr](https://dataverse.cirad.fr) (2018)



25/10/2022

METTRE EN PLACE SON ENTREPÔT DE DONNÉES DATAVERSE



Le choix de Dataverse

Exemples d'entrepôts disciplinaires

Dataverse

Search User Guide Support Sign Up Log In

MELDATA



Metrics

The Dataverse port Reusable (FAIR) on LIVESTOCK, the data as long as permission from the nonconformities, de The user should ak

Citation Requirement

In accordance with citation. Such citati and as shown on t includes position p whenever possible.

Disclaimer

While utmost care i event shall the data

Search this datav

Metrics 165

CIMMYT institutional network of sc



Cereal Systems Initiative Asia (CSISA) Research

Search this dataverse...

Dataverses (8)

Dataverse



International Potato Center (CIP)

CIP is a mem research for a

Metrics 13,142 Downloads

CIP institutional network of scientific datasets and softw

Search this dataverse... Advanced Search

Dataverses (8)

Datasets (554)

Files (4,457)

Dataverse Category

Research Project (4)

Organization or Institution (2)

Researcher (2)

1 to 10 o

Dataset highlight

Dataverse

Dataverse

Looking for the numbers behind our science? Browse through our datasets, or search by key terms.



Center for International Forestry Research (CIFOR)

Metrics 47,714 Downloads

The root dataverse.



Database of Vegetation map



Global Comparative Study on REDD+



The Sustainable Wetlands Adaptation and Mitigation Program (SWAMP)



Forest Spatial Information Catalog (FSIC)

Search this dataverse... Find Advanced Search

Dataverses (14)

Datasets (281)

Files (2,116)

Dataverse Category

Research Project (7)

Laboratory (2)

Organization or Institution (2)

Research Group (1)

Publication Year

2019 (101)

1 to 10 of 296 Results

A global dataset on the information underlying the barriers and pathways of the United Nations Decade on Ecosystem Restoration Strategy.

Jun 28, 2022

Meli, P.; Schweizer, D.; Gaarigata, M.R., 2022. "A global dataset on the information underlying the barriers and pathways of the United Nations Decade on Ecosystem Restoration Strategy". <https://doi.org/10.17528/CIFOR/DATA.00286>, Center for International Forestry Research (CIFOR), V1

Bibliographic database of peer-reviewed and gray literature publications over 1990 and 2021, presenting information related to ecosystem restoration, and that may be associated as evidence for the UN Decade on Ecosystem Restoration. The database constitutes the full list of an in...

Tropical and Subtropical Wetlands Distribution

Entrepôts des **CGIAR** (à partir de 2018)



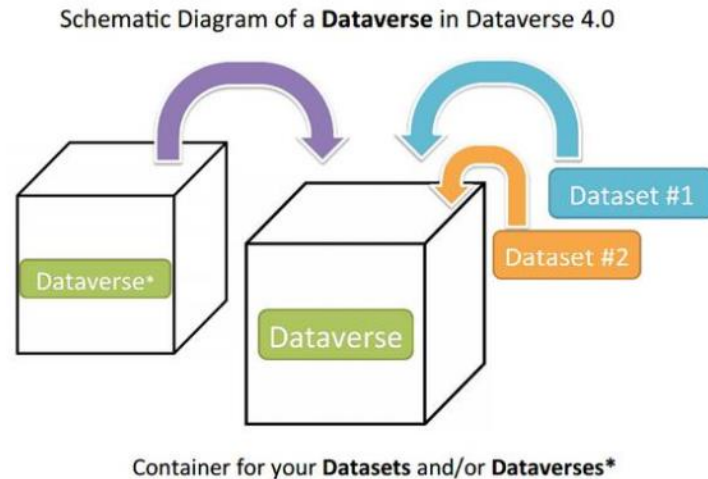
25/10/2022

METTRE EN PLACE SON ENTREPÔT DE DONNÉES DATAVERSE

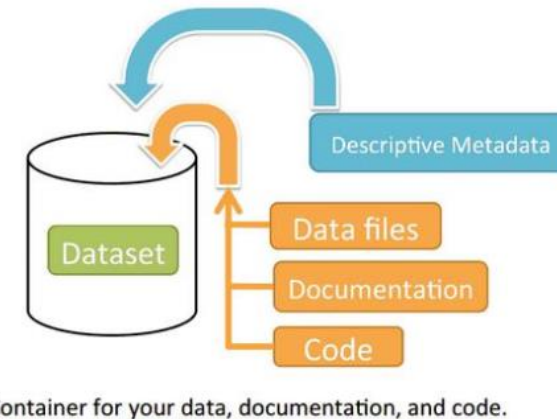


Le choix de Dataverse

Structuration d'un entrepôt de données Dataverse



Schematic Diagram of a **Dataset** in Dataverse 4.0

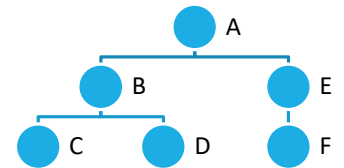


Collections
= dossiers
= répertoires de l'entrepôt
= « *dataverses* »

Chaque collection peut contenir :

- des **jeux de données** (*datasets*)
- des **sous-collections**

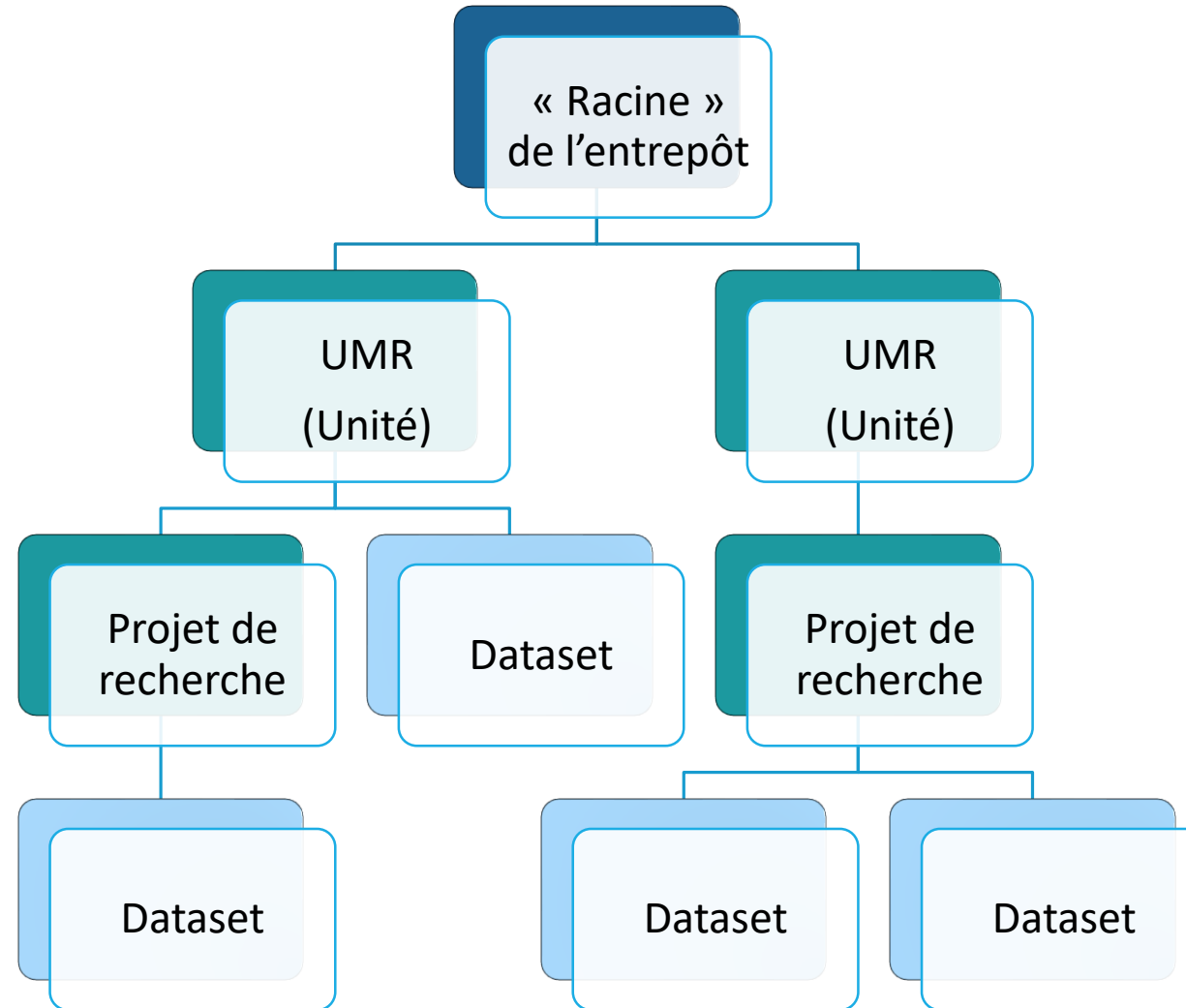
⇒ arborescence



Fonctionnement d'un entrepôt Dataverse au quotidien

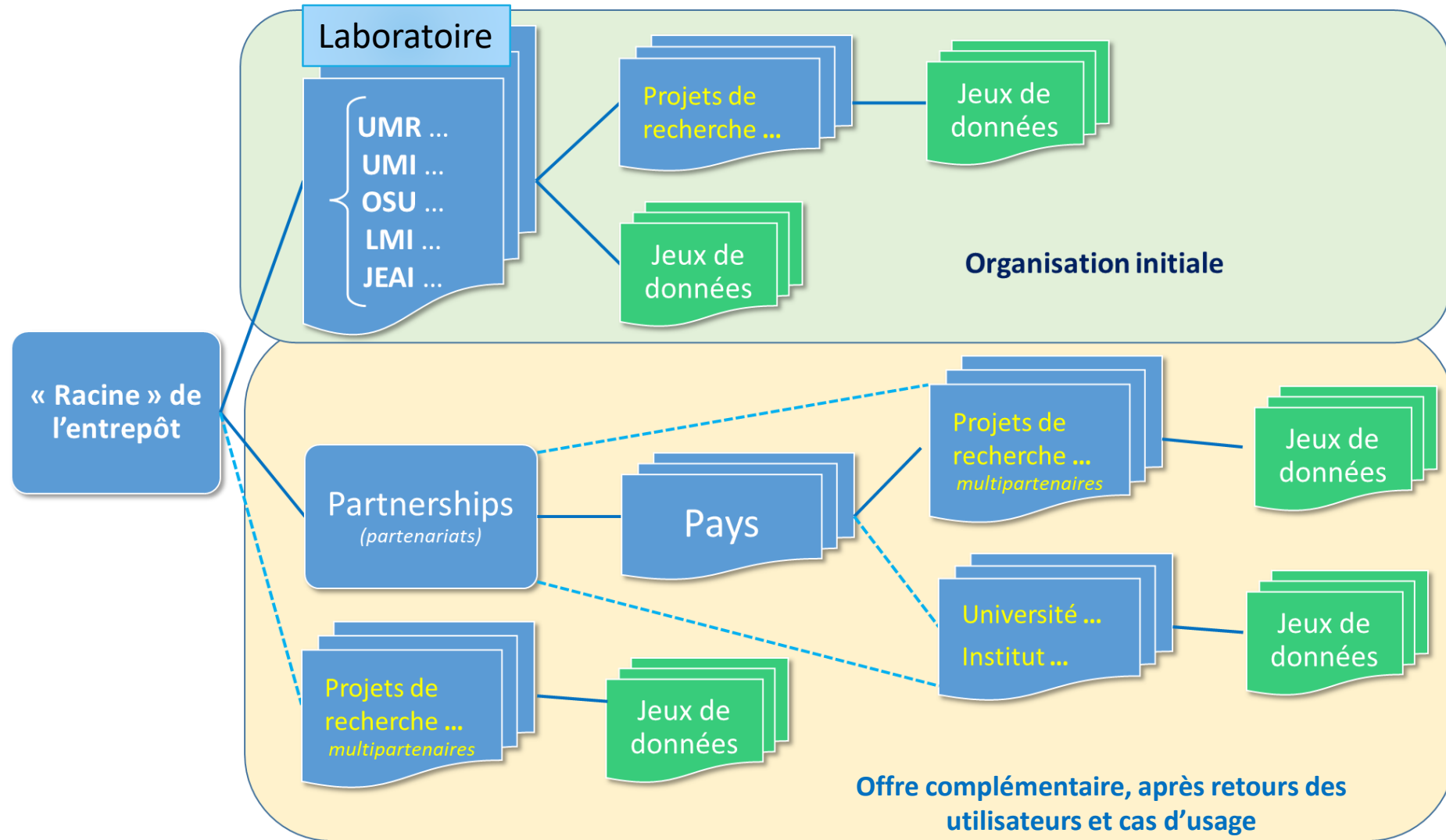
Exemple du Cirad Dataverse : organisation structurelle

- Collection / « dataverse »
- Jeu de données



Fonctionnement d'un entrepôt Dataverse au quotidien

Exemple de Datasuds (IRD) : organisation structurelle



Le choix de Dataverse

Démonstration de Dataverse Interface d'exploration et de recherche

Recherche générale

Chercher dans ce dataverse... **Recherche avancée**

Recherches ciblées

- Dataverses (55)
- Jeux de données (77)
- Fichiers (836)

Catégorie de Dataverse

- Laboratoire (27)
- Projet de recherche (18)
- Organisation ou établissement (5)
- Groupe de recherche (4)
- Département (1)

Date de publication

- 2020 (90)
- 2019 (10)

Sujet

- Sciences de la terre et de l'environnement (73)
- Sciences de l'agriculture (36)
- Médecine, santé et sciences de la vie (25)
- Informatique et science de l'information (9)
- Sciences mathématiques (4)

Thématique scientifique

- Environmental sciences (44)
- Soil science (26)

1 à 10 de 132 Résultats

Une collection

UMR LEMAR (Univ. Bretagne Occidentale, CNRS, IRD, Ifremer)
12 oct. 2020

LABORATORY OF ENVIRONMENTAL MARINE SCIENCES LEMAR is a strongly interdisciplinary laboratory that brings together ecologists, biologists, biogeochemists, chemists, physicists and marine environment lawyers. Its objectives are (1) to conduct research on topics of fundamental resea...

Un jeu de données

Design of the PI2Puino casing system: STL and SCAD 3d print models
2 oct. 2020 - Project ANR PI2P

Rebaudo, Francois; Soulard, Titouan, 2020, "Design of the PI2Puino casing system: STL and SCAD 3d print models", <https://doi.org/10.23708/XBJPK>, DataSuds, V1

The PI2Puino is an integrated board that includes a LoRa module (with an embedded antenna), an Arduino-compatible microcontroller and a battery holder. It allows gathering on-site measurements and sending them to a gateway. It was designed specifically for the PI2P project. These...

Design of the PI2Puino integrated board: Gerber files
2 oct. 2020 - Project ANR PI2P

Soulard, Titouan; Rebaudo, Francois, 2020, "Design of the PI2Puino integrated board: Gerber files", <https://doi.org/10.23708/13OVHV>, DataSuds, V1

The PI2Puino is an Arduino-like integrated board which allows to gather on-site measurements and send them to a gateway. It was designed specifically for the PI2P project. The PI2Puino includes a LoRa module (with an embedded antenna), an Arduino-compatible microcontroller and a...

Project ANR PI2P
30 sept. 2020 UMR EGCE

Predicting Insect Pest Phenology: calibrating laboratory-based model parameter estimates from field-based monitoring of

(exemple : entrepôt DataSuds)

Le choix de Dataverse

Démonstration de Dataverse

Présentation d'un jeu de données (dataset)

The screenshot shows the Dataverse interface for the dataset 'Base de données insulaires mondiale (BIM)'. Annotations include:

- Identifiant:** Points to the DOI logo and the dataset title.
- Métadonnées:** Points to the description, subject, and related publications sections.
- Licence:** Points to the Creative Commons BY license icon.
- Export citation:** Points to the 'Export citation' button.
- Onglets : 4 sections principales:** Points to the 'Fichiers', 'Métadonnées', 'Conditions', and 'Versions' tabs.
- Accès fichiers:** Points to the 'Télécharger' buttons for the data files.

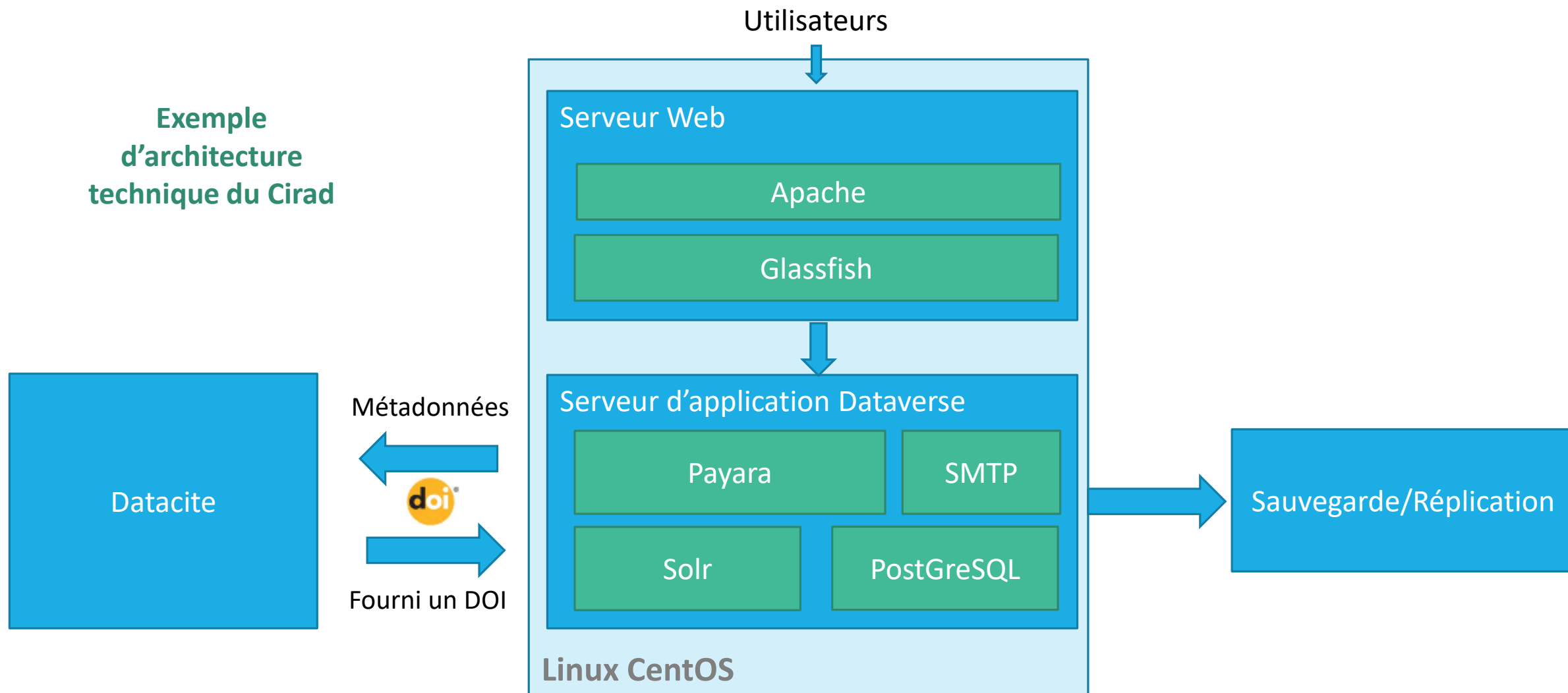


Etudes de cas : le Cirad et l'IRD

1. ASPECTS TECHNIQUES (INSTALLATION, MAINTENANCE)

Aspects techniques

Installation : aspects techniques de la plateforme, etc.



Aspects techniques

Installation : infrastructures, hébergement, etc.



2 plateformes opérationnelles:

- Production, préproduction
- Projet de plateforme test/formations

Hébergement

- ❖ Serveur virtuel, Linux centOS, CPU 8 cœur, 16Go de ram

Stockage :

- ❖ Disque de 250Go (point de montage SAN, extensible si besoin)

Aspects techniques

Installation : infrastructures, hébergement, etc.



3 plateformes opérationnelles:

- Production, pré-production, tests (bac-à-sable)

Hébergement

- ❖ 4 serveurs virtuels CentOS 7 CPU 8 cœur, 16go de ram (production)
- ❖ 1 serveur virtuel (test)

Stockage :

- ❖ Stockage réseau NFS 2 To (extensible si nécessaire)

Aspects techniques

Installation : infrastructures, hébergement, etc.

Différents types d'installation et d'automatisations sont documentés et partagés par la communauté :

- Ansible
- Docker
- Puppet



<https://guides.dataverse.org/en/latest/installation/>

Aspects techniques

Réservation de DOIs : fournisseurs de DOIs



Service payant :

- En France, géré au niveau national (INIST-CNRS) pour le CIRAD et l'IRD
- Frais annuel d'abonnements : 500€ à 2.000€ suivant le type d'organisations
- Frais de services : plusieurs tiers suivant la consommation annuelle de DOIs

Annual DOI range	Tier	Organization Fee	DOI Fee	Annual Service Fee
0 - 1.999	Tier 1	500€	0,80€ per DOI	500€ + 0,80€ per DOI
2.000 - 10.000	Tier 2	500€	1600€	2.100€
10.001 - 100.000	Tier 3	500€	2.500€	3.000€
100.001 - 250.000	Tier 4	500€	3.500€	4.000€
250.001 - 1.000.000	Tier 5	500€	8.500€	9.000€
1.000.001 - 2.000.000	Tier 6	500€	13.000€	13.500€
2.000.001 - 5.000.000	Tier 7	500€	18.500€	19.000€
5.000.001 - 10.000.000	Tier 8	500€	25.500€	26.000€

Aspects techniques

Réservation de DOIs : les DOIs dans Dataverse

- Création de DOI automatique depuis Dataverse en statut draft
- Activation du DOI à la publication du dataset
- Une fois publié, pas de possibilité de le supprimer
- Obligation contractuelle de maintien des liens

iSDAsoil soil texture class (USDA system) 0-20cm for Africa, preprocessed for use in SARRA-O crop model

Draft **In Review** **Unpublished**

Jérémy LAVARENNE, 2022, "iSDAsoil soil texture class (USDA system) 0-20cm for Africa, preprocessed for use in SARRA-O crop model", <https://doi.org/10.18167/DVN1YSVTS2>, CIRAD Dataverse, DRAFT VERSION ⓘ

Cite Dataset ▾ Learn about Data Citation Standards.

Access Dataset ▾
Publish Dataset ▾
Edit Dataset ▾
Contact Owner Share

Dataset Metrics ⓘ
0 Downloads ⓘ

Description ⓘ

This raster map is adapted from "iSDAsoil: soil texture class (USDA system) for Africa predicted at 30 m resolution at 0-20 and 20-50 cm depths" v0.13 (doi:10.5281/zenodo.4094616) for use in SARRA-O crop model. It was obtained by first performing raster alignment with the TAMSAT v3.1



iSDAsoil soil texture class (USDA system) 0-20cm for Africa, preprocessed for use in SARRA-O crop model Dataset

Jérémy LAVARENNE,

Dataset published 2022 via CIRAD Dataverse

This raster map is adapted from "iSDAsoil: soil texture class (USDA system) for Africa predicted at 30 m resolution at 0-20 and 20-50 cm depths" v0.13 (doi:10.5281/zenodo.4094616). It was obtained by first performing alignment with the TAMSAT v3.1 precipitation product (<https://www.tamsat.org.uk>), through upscaling of the original 30m resolution to 0.0375° (approximately 4 km). Mode was used as aggregation function, so that pixels in the obtained raster represent the main soil class in the extent of the newly computed pixels. Then, no-class encoding was updated from 255 to 0. Finally, class identifiers were updated by applying a factor 1E6 to obtain 8-digit encoding needed by SARRA-O crop model.

Created October 14, 2022 at 10:01:45 UTC. **Draft**

10.18167/dvn1/ysvts2

Aspects techniques

Sécurité de l'installation



- Service pour la **diffusion de données**, mais pas spécifiquement pour la conservation sécurisée (≠ un coffre fort numérique)

- 3 possibilités de configuration pour l'authentification
 - Avec des **comptes locaux**
 - Authentification par **fédération** (Shibboleth/SAML)
 - Avec **OAuth** via les services proposés par ORCID, Github, Google

Your Institution

CIRAD

Please select...

Please select...

ABES - French Bibliographic Agency for Higher Education

Académie militaire de Saint-Cyr Coëtquidan

ADIUT - Assemblée des Directeurs d'IUT

Agence Universitaire de la Francophonie (AUF)

Other options

ORCID is an open, non-profit, community-based effort to provide a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. ORCID is unique in its ability to reach across disciplines, research sectors, and national boundaries and its cooperation with other identifier systems. Find out more at orcid.org/about.

This repository uses your ORCID for authentication (so you don't need another username/password combination). Having your ORCID associated with your datasets also makes it easier for people to find the datasets you have published.

Other options

Aspects techniques

Sécurité de l'installation

La configuration dépend de la politique de la plateforme :

- **Au Cirad** : Ouverture au personnel du Cirad et à ses partenaires
 - Authentification Shibboleth avec une adresse @cirad.fr uniquement
 - Compte local uniquement pour l'administration
- **A l'IRD** : Ouverture à la fédération Renater
 - Authentification par la fédération Renater
 - Compte local uniquement pour l'administration



Maintenance : préservation/sauvegarde

- Sauvegarde quotidienne des serveurs virtuels (applications + base + données)
- Conservations des sauvegardes :
 - 8 dernières sauvegardes quotidiennes
 - Les 5 dernières sauvegardes du 1er jour des semaines précédentes
 - Les 4 dernières sauvegardes du 1er jour des mois précédents



Aspects techniques

Montées de version

Dataverse : 5 à 7 versions / an (évolutions techniques, fonctionnelles, correctifs, etc.)

➡ Au Cirad, depuis 2018, 1 montée de version réalisée (avril 2021)

Une montée de version nécessite :

- 1 ETP technique (DSI) + 1 ETP fonctionnel
- 1 phase de configuration et paramétrages des évolutions
- 1 suivi de la conformité des données après migration
- 1 recettage de la montée de version pour valider techniquement et fonctionnellement l'opération

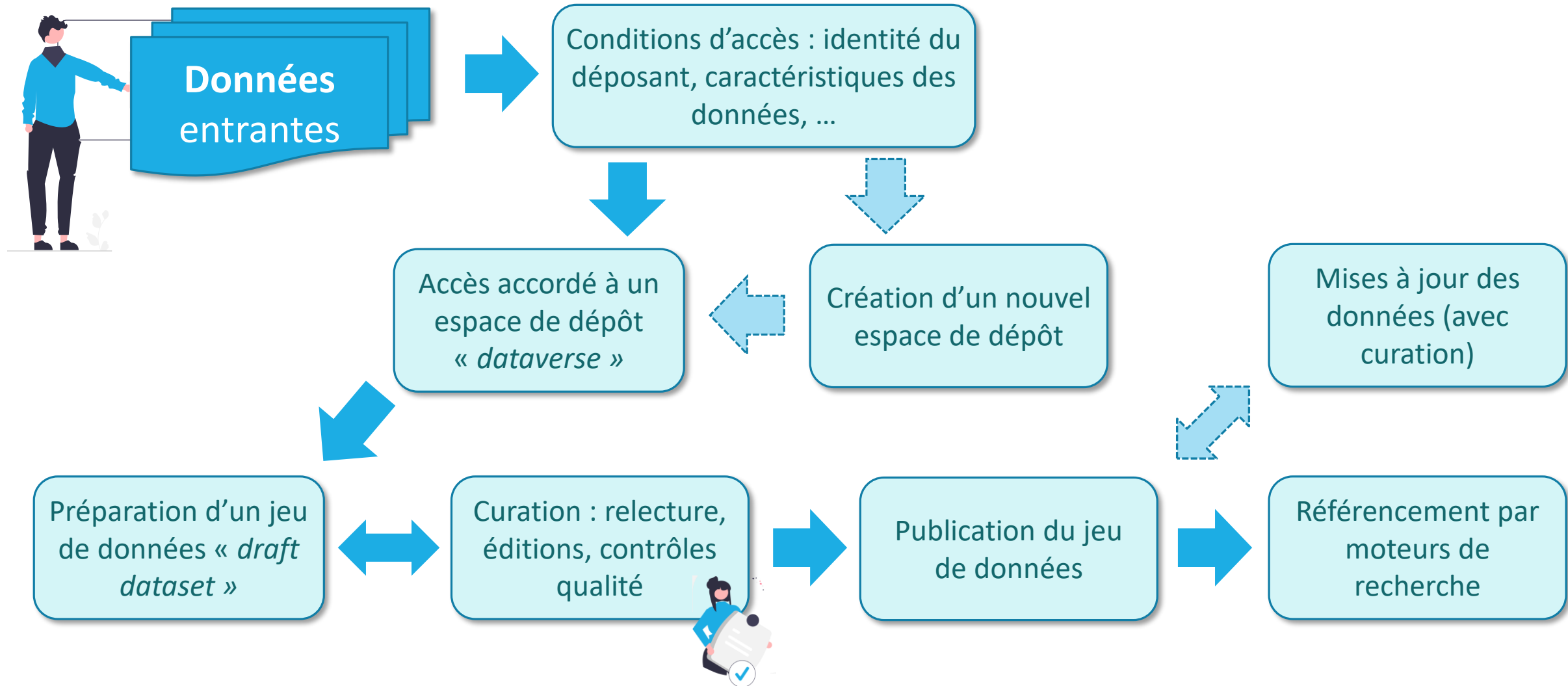


Etudes de cas : le Cirad et l'IRD

2. LE FONCTIONNEMENT D'UN ENTREPÔT DATAVERSE AU QUOTIDIEN

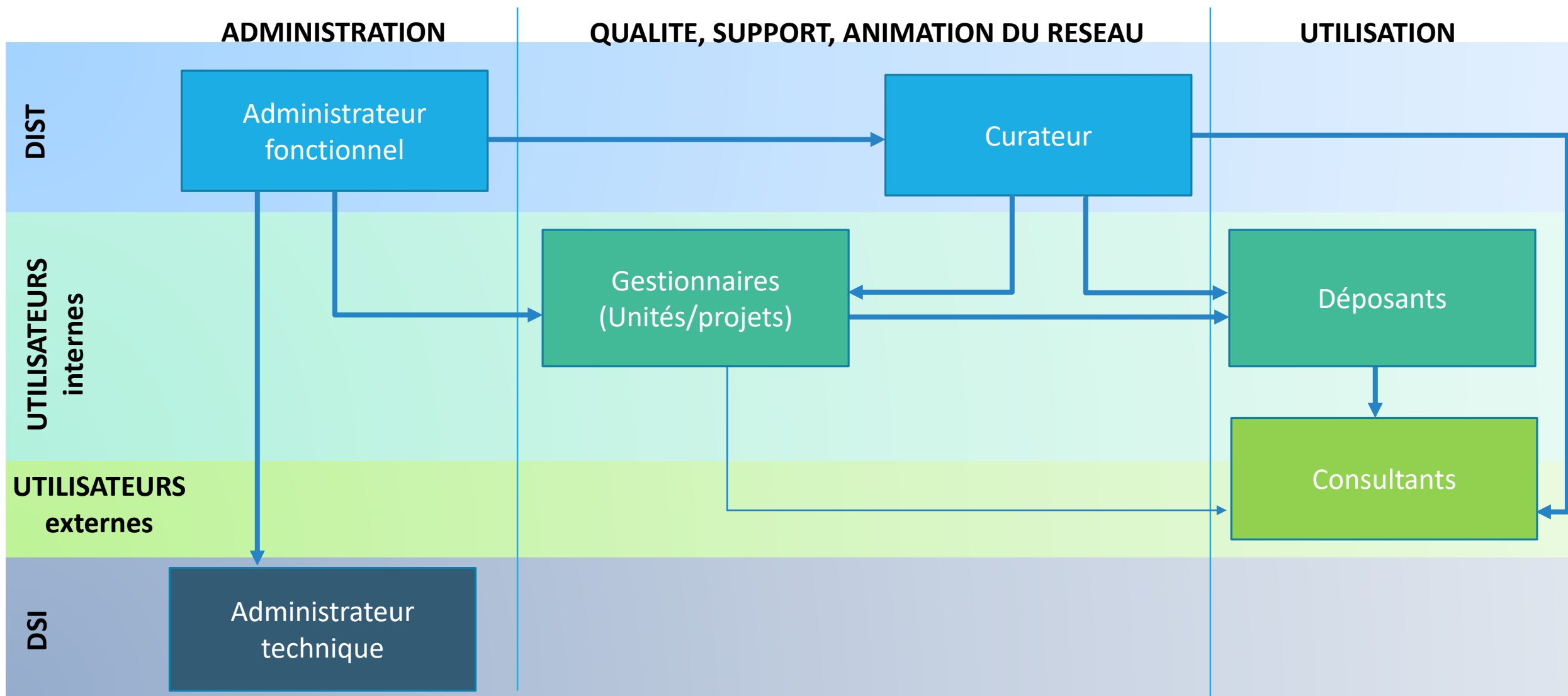
Fonctionnement d'un entrepôt Dataverse au quotidien

Flux des données - *Data workflow*



Fonctionnement d'un entrepôt Dataverse au quotidien

Exemple du Cirad Dataverse : organisation fonctionnelle



Fonctionnement d'un entrepôt Dataverse au quotidien

Importance des personnels d'appui

Compétences dans les équipes

Personnel spécialisé

Documentaire

Juridique

Disciplinaires

Curateurs

Formateurs

Relais dans les
équipes

Fonctionnement d'un entrepôt Dataverse au quotidien

Importance des personnels d'appui

1) Former

Les chercheurs tout au long du cycle de vie des données, pour assurer la qualité des données et métadonnées :

- ▶ Cycle de vie des données
- ▶ Plan de gestion de données
- ▶ Gérer et préparer ses données (bonnes pratiques, outils, standards de métadonnées)
- ▶ Utiliser l'entrepôt institutionnel suivant les différents niveaux de responsabilité (déposant, gestionnaire, curateur etc.)

Fonctionnement d'un entrepôt Dataverse au quotidien

Importance des personnels d'appui

2) Accompagner

Au quotidien à la gestion de l'entrepôt :

- ▶ A la définition des datasets
- ▶ Au choix des métadonnées
- ▶ Au choix des licences de diffusion

Au moyen de :

- ❖ Rédaction de guides et outils
- ❖ Permanences
- ❖ Etc.

Fonctionnement d'un entrepôt Dataverse au quotidien

Exemple de modalités d'accompagnement des déposants

Service : assistance, formation et communication (1 ETP)

- Permanence email et téléphonique
- Rendez-vous en présentiel ou par visioconférence (1h)
 - conseils personnalisés
 - aide à la préparation des métadonnées et des fichiers
- Configuration des collections en lien avec les référents du laboratoire ou chefs de projet
- Organisation d'ateliers « *datathons* » et de formations
- Présentation de l'entrepôt, par exemple au cours d'une session d'un séminaire interne d'un laboratoire
- Communication sur les actualités autour des données de recherche



Fonctionnement d'un entrepôt Dataverse au quotidien

Importance des personnels d'appui

3) Gérer la qualité du contenu

Garante de la réputation et de la visibilité de l'entrepôt :

- ▶ Gérer le workflow des métadonnées
- ▶ Relecture, correction, validation des métadonnées, voire des données
- ▶ Création de templates pour le dépôt des données

Passant par des choix stratégiques :

- ▶ Définition et évolution d'une politique de dépôt
- ▶ Choix de la standardisation des métadonnées

Fonctionnement d'un entrepôt Dataverse au quotidien

Importance des personnels d'appui

4) Suivi de l'utilisation

Afin d'établir des indicateurs utiles à votre institution :

- ▶ Statistiques d'usages (dépôts, téléchargements)
- ▶ Indicateurs spécifiques aux types de données et domaines scientifiques couverts

Sous réserve de :

- ❖ Qualité des métadonnées suffisantes
- ❖ Domaine de la datamétrie encore jeune, beaucoup reste à développer

Et maintenant ?

CONCLUSION

Nous contacter

Pascal Aventurier (IRD) pascal.aventurier@ird.fr

Céline Barthélemy (Cirad) celine.barthelemy@cirad.fr

Luc Decker (IRD) luc.decker@ird.fr

Gwenaël Doux (Cirad) gwenael.doux@cirad.fr