



RÉPUBLIQUE  
FRANÇAISE

*Liberté  
Égalité  
Fraternité*

recherche.data.gouv.fr

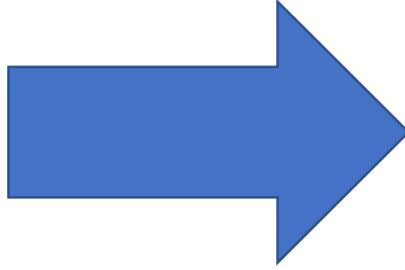
# Atelier d'initiation



UNIVERSITÉ  
DE LORRAINE



- Des injonctions croissantes à l'ouverture des données
- Des besoins croissants de citabilité des données



## Un « entrepôt » ?

- Un espace de publication par défaut
- Une citabilité des jeux de données par l'attribution de DOI et la gestion de métadonnées



### Entrepôt

= traduction française de “repository” (dépôt, répertoire)

≠ “warehouse”

→ Ce n'est **pas** un **espace de stockage** : il n'est en particulier que peu adapté aux données très volumineuses ;

→ Ce n'est **pas** un **espace d'archivage** : Recherche Data Gouv garantit une conservation des données à 5 ans (préservation au sens du CINES = 20 ans)

# DOREL est mort, vive DOREL

Avant nov. 2023

"DOREL" désignait  
l'instance  
institutionnelle de  
Dataverse maintenue à  
l'UL.



Après nov. 2023

"DOREL" désigne  
l'espace institutionnel  
de l'Université de  
Lorraine sur Recherche  
Data Gouv.



# DOREL est mort, vive DOREL

The screenshot shows the old DOREL website interface. At the top, there is a navigation bar with 'RECHERCHE', 'GUIDE D'UTILISATION', 'SUPPORT', 'FRANÇAIS', and 'SE CONNECTER'. The DOREL logo is on the left, and the Université de Lorraine logo is on the right. Below the navigation bar, there is a search bar and a 'Recherche avancée' button. The main content area features a large orange arrow pointing from the old interface to the new one. The left sidebar contains filters for 'Collections (67)', 'Jeux de données (52)', 'Fichiers (15 707)', 'Catégorie de la collection', 'Année de publication', 'Auteur Nom', and 'Sujet'. The main content area displays a list of research results, including 'SpectroLive : optical spectra (autofluorescence and diffuse reflectance) acquired on human skin carcinomas' and 'Physicochemical properties and NMR characterization of propylene - precious metals systems'.

The screenshot shows the new DOREL website interface. At the top, there is a navigation bar with 'RECHERCHE', 'À propos', 'Guide d'utilisation', 'Support', 'Français', 'S'inscrire', and 'Se connecter'. The DOREL logo is on the left, and the Université de Lorraine logo is on the right. Below the navigation bar, there is a search bar and a 'Recherche avancée' button. The main content area features a large orange arrow pointing from the old interface to the new one. The left sidebar contains filters for 'Collections (50)', 'Jeux de données (56)', 'Fichiers (874)', 'Catégorie de collection', 'Source', 'Année de publication', 'Origine des données', and 'Type de données'. The main content area displays a list of research results, including 'Data associated with article - Real-time spectroscopic ellipsometry of plasmonic nanoparticle growth in polyvinyl alcohol thin films' and '1D conduction radiation coupled heat transfer in an absorbing medium : Monte Carlo fonctionnal modeling'.

# Points essentiels sur cette évolution

Ouverture des dépôts dans l'espace institutionnel depuis le 01/11/2023

- Légère accélération du rythme des dépôts depuis cette date
- Dépôt facilité

Fermeture des dépôts dans l'entrepôt institutionnel depuis le 30/01/2024

- Les modifications des jeux de données présents (mises à jour des métadonnées, ajouts de fichiers, nouvelles versions...) restent possibles. Ces jeux sont moissonnés et apparaissent dans Recherche Data Gouv.

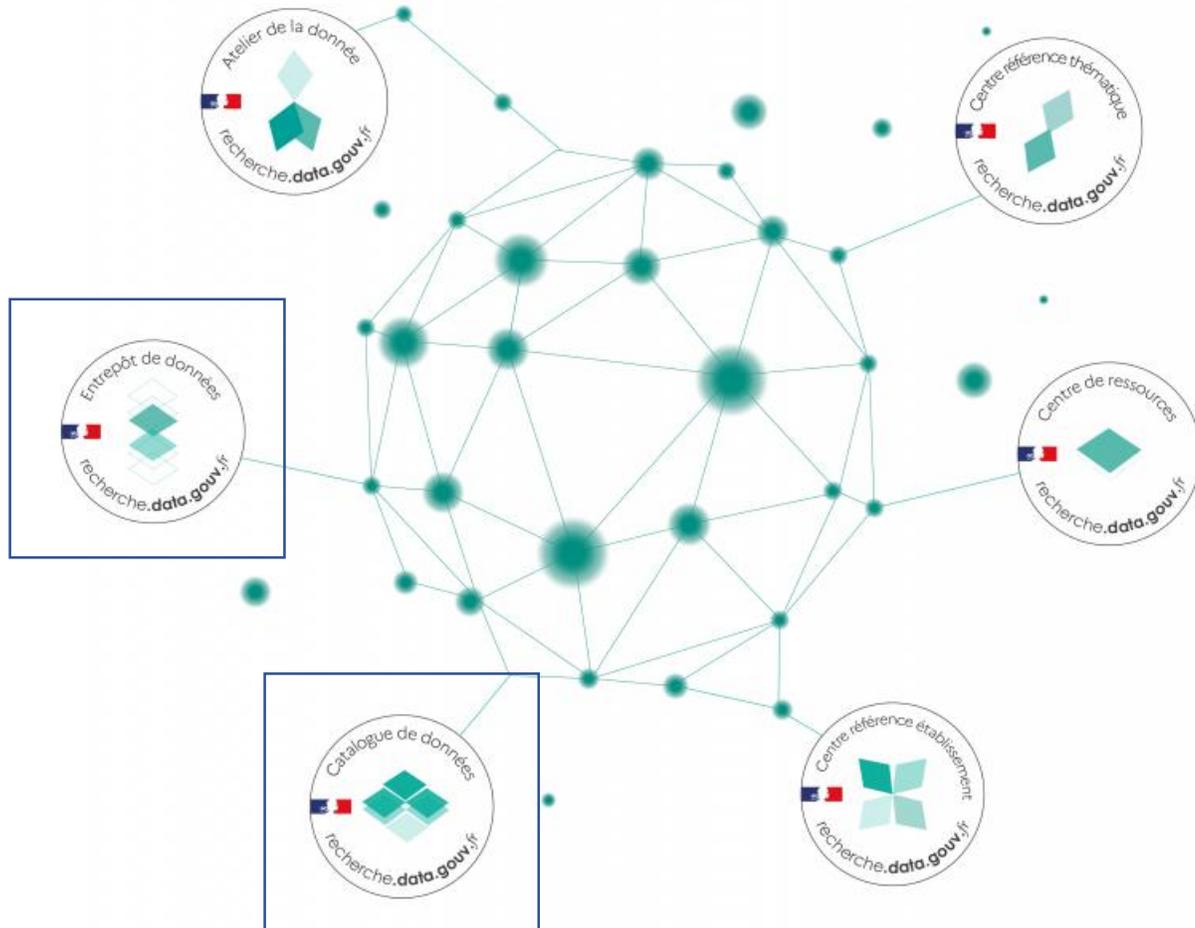
**En cours :**

Migration des jeux de l'entrepôt institutionnel vers l'espace institutionnel

**En cours :**

Alignement de la curation sur les exigences et les modèles de l'espace dit « générique »

# Une part de l'écosystème Recherche Data Gouv



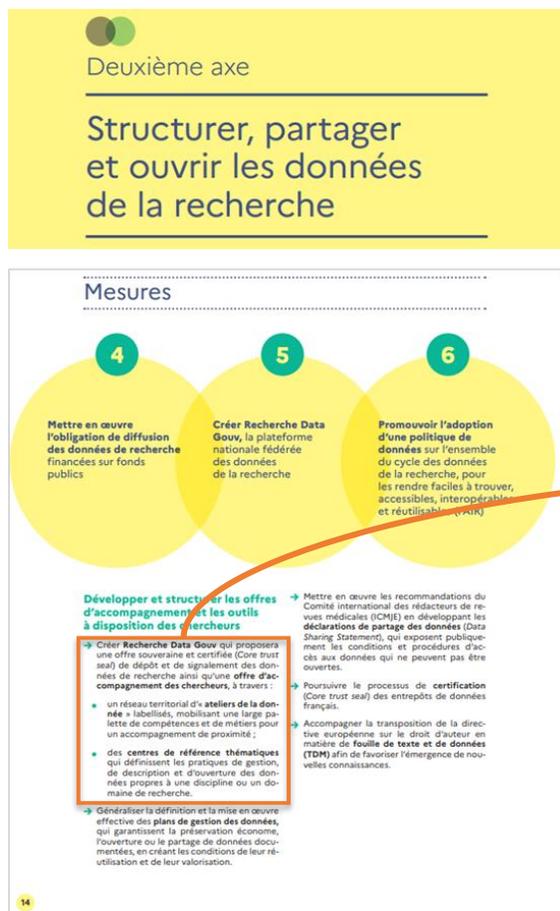
## 4 Modules d'accompagnement :

- Ateliers de la donnée
- Centres de référence thématiques
- Centres de référence établissement
- Centres de ressources

## 2 Modules de partage :

- Entrepôt
- Catalogue (à venir)

# L'entrepôt Recherche Data Gouv



Créer **Recherche Data Gouv** qui proposera une offre souveraine et certifiée ([\*Core trust seal\*](#)) de dépôt et de signalement des données de recherche ainsi qu'une **offre d'accompagnement des chercheurs**, à travers :

un réseau territorial d'« **ateliers de la donnée** » labellisés, mobilisant une large palette de compétences et de métiers pour un accompagnement de proximité ;

des **centres de référence thématiques** qui définissent les pratiques de gestion, de description et d'ouverture des données propres à une discipline ou un domaine de recherche.

# La plateforme Recherche Data Gouv

 Portail

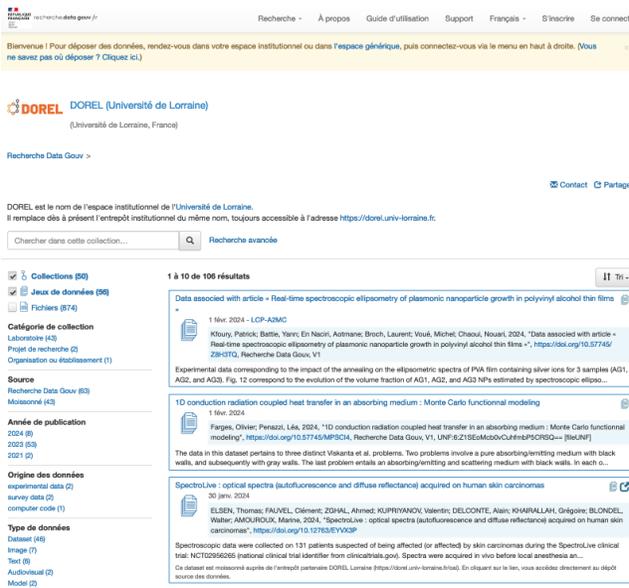


## Actualités

 <b>Lien entre données et projet ANR</b>  Lier ses données à son projet ANR ? Si, c'est possible avec Recherche Data Gouv !  Il est à présent possible de relier automatiquement ses jeux de données déposés dans...  Publiée le lundi 11 mars 2024	 <b>Nouvelle Classe Virtuelle</b>  Vous avez découvert l'API de Recherche Data Gouv ? Devenez pro avec cette nouvelle formation !  La formation « Découverte des APIs de l'entrepôt Recherche Data Gouv », proposée depuis...  Publiée le jeudi 29 février 2024	 <b>Partage d'expériences Data Univ Eiffel</b>  Retour sur la construction de l'Atelier Data Univ Eiffel  Le GTSD Données de Couperin est parti à la rencontre de Céline Rousselet, responsable...  Publiée le mardi 23 janvier 2024	 <b>[Vidéo] L'entrepôt de données</b>  À quoi sert l'entrepôt Recherche Data Gouv ?  Découvrez (ou redécouvrez) l'entrepôt de données de Recherche Data Gouv, un...  Publiée le lundi 18 décembre 2023	 <b>Le 12 mars 2024 à 10:00</b> <b>Classe virtuelle : déposer un jeu de données</b>  De 10h à 12h	 <b>Le 21 mars 2024 à 10:00</b> <b>Classe virtuelle : administrer une collection</b>  De 15h à 17h	 <b>Le 26 mars 2024 à 10:00</b> <b>Classe virtuelle : découverte des APIs</b>
--	--	---	---	---	---	---

S'informer

 Entrepôt



Publier

 Catalogue



Rechercher, signaler

# La plateforme Recherche Data Gouv

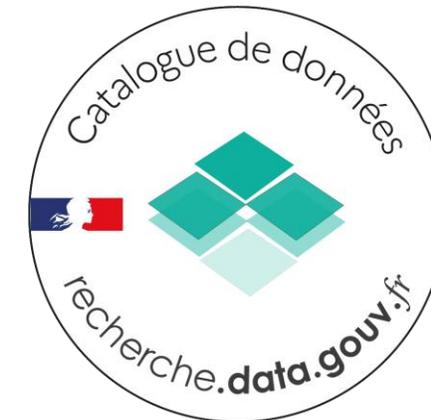
## Datacenter national

- Ile de France, Toulouse + sauvegarde à l'UniStra
- Dispositif cible : partage sur des data centers de certaines co-tutelles de l'UMS

## Calendrier à 3 ans, mise en place de Recherche Data Gouv :

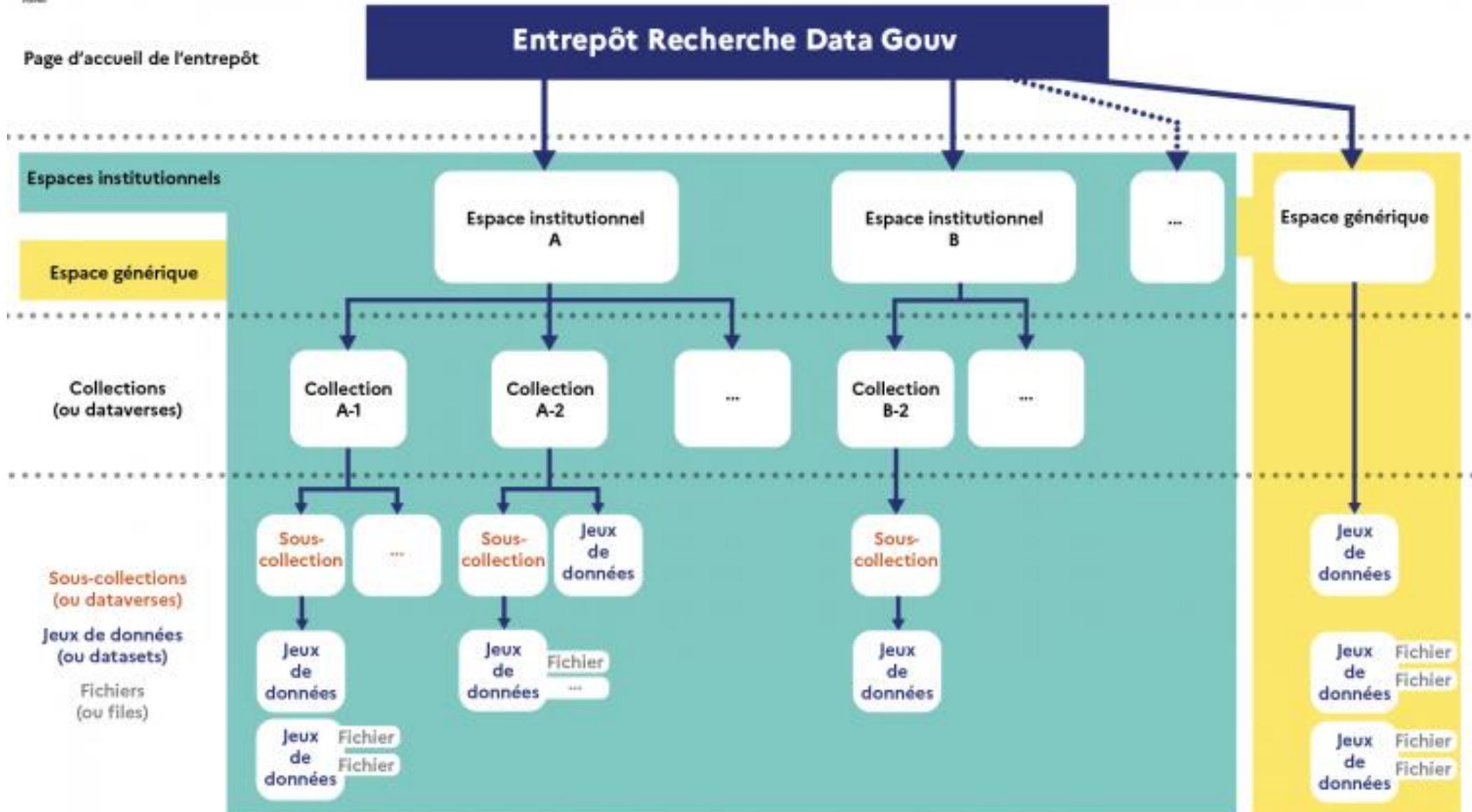
- Début du projet : mai 2021
- Lancement de l'entrepôt : 8 juillet 2022
- 2022-2023 : campagne d'alimentation
- 2023-2024 : démarrage du projet catalogue de données externes

**2024 : Création d'une UMS** portant l'entrepôt, le catalogue et le portail





## Organisation de l'entrepôt Recherche Data Gouv



# DOREL : choix concernant l'organisation

- Une collection par laboratoire (à la racine de l'espace institutionnel)
- Des collections à la racine ou des sous-collections sont possibles pour:
  - Des équipes de travail
  - Des projets
  - Des plates-formes ou instrumentations
  - (d'autres cas sont possibles, n'hésitez pas à nous consulter)

# Viser une conformité FAIR maximale



■ **Entrepôt** : DOI ; métadonnées indiquant l'identifiant ; entrepôt permettant la recherche  
■ **Déposant** : métadonnées riches et utilisant des vocabulaires



■ **Entrepôt** : données accessibles avec le DOI ; métadonnées toujours disponibles  
■ **Déposant** : données ouvertes (éventuellement sous conditions)



■ **Entrepôt** : métadonnées standardisées et *machine readable*  
■ **Déposant** : format de fichiers ouverts et *machine-readable* ; vocabulaires avec URI



■ **Entrepôt** : licences disponibles  
■ **Déposant** : provenance des données ; standards correspondant à la communauté

# A l'Université de Lorraine

## Une équipe de travail DN/DirDoc pour la curation en central

- 1 administrateur fonctionnel et curateur (DirDoc)
- 1 administrateur technique (DN)
- 1 technicien (DN)
- 3 curatrices (DirDoc, MSH)

## Un réseau de curateurs dans les laboratoires



**Marine AMOUROUX**

CRAN  
Centre de Recherche en  
Automatique de Nancy



**Lilia AYADI**

IMoPA  
Ingénierie Moléculaire et  
Physiopathologie Articulaires



**Arnaud BIANCHI**

IMoPA  
Ingénierie Moléculaire et  
Physiopathologie Articulaires



**Sidi HAMADY**

LMOPS  
Laboratoire matériaux  
optiques, photonique et  
systèmes



**Thomas KAUFFMANN**

LMOPS  
Laboratoire matériaux  
optiques, photonique et  
systèmes



**Cécile MANGAVEL**

LIBio  
Laboratoire Ingénierie  
des Biomolécules



**Julie MANSUY**

BETA  
Bureau d'Économie  
Théorique et Appliquée



**Pierre PERALDI-MITTELETTE**

Institut François Gény  
Droit



**Romain PIERRONNET**

CEREFIGE  
Centre Européen de  
Recherche en Économie  
Financière et en Gestion  
des Entreprises



**Isabelle PIGNONE**

Archives Henri Poincaré-  
PReST  
Philosophie et Recherches sur  
les Sciences et les  
Technologies



**Bruno ROCK**

CRM2  
Cristallographie, Résonance  
Magnétique et Modélisations



**Julien VINCENT**

LEM3  
Microstructures, Mécanique et  
sciences des Matériaux



**Lucie VOINSON**

HISCANT-MA  
Histoire et Cultures de  
l'Antiquité et du Moyen Âge

# Démo vs production

## Démo

<https://demo.recherche.data.gouv.fr>

Pour l'entraînement, l'expérimentation...

Les jeux de données ne sont PAS conservés

Les DOI sont sur un compte de test et ne sont pas conservés



<https://demo.recherche.data.gouv.fr>

## Production

<https://entrepot.recherche.data.gouv.fr>

Pour la publication de données achevées

Les jeux de données sont conservés  
(et sont définitivement publiés sauf exception rare)

Les DOI sont définitifs et peuvent servir à la citation



<https://entrepot.recherche.data.gouv.fr>

# La création d'un jeu de données : une valse à quatre temps

- Premier temps :  
je **remplis les métadonnées essentielles**
- Second temps :  
je **téléverse mes fichiers**
- Troisième temps :  
je **complète mes métadonnées**
- Quatrième temps :  
je **soumets pour validation**



# Les fichiers

Il n'y a pas de types de fichiers prohibés. Il faut toutefois privilégier les fichiers en format ouvert.

Les fichiers doivent être compréhensibles et réutilisables par un chercheur de la discipline. Dans un grand nombre de cas, un README est recommandé.

Les fichiers de type tableaux, ou relevés d'enregistrement, doivent être formatés avec le plus grand soin (Excel, csv, tsv, SPSS, R). DOREL ingère (indexe) ces données et les convertit dans un format adapté à la conservation (texte UTF-8 tabulé).

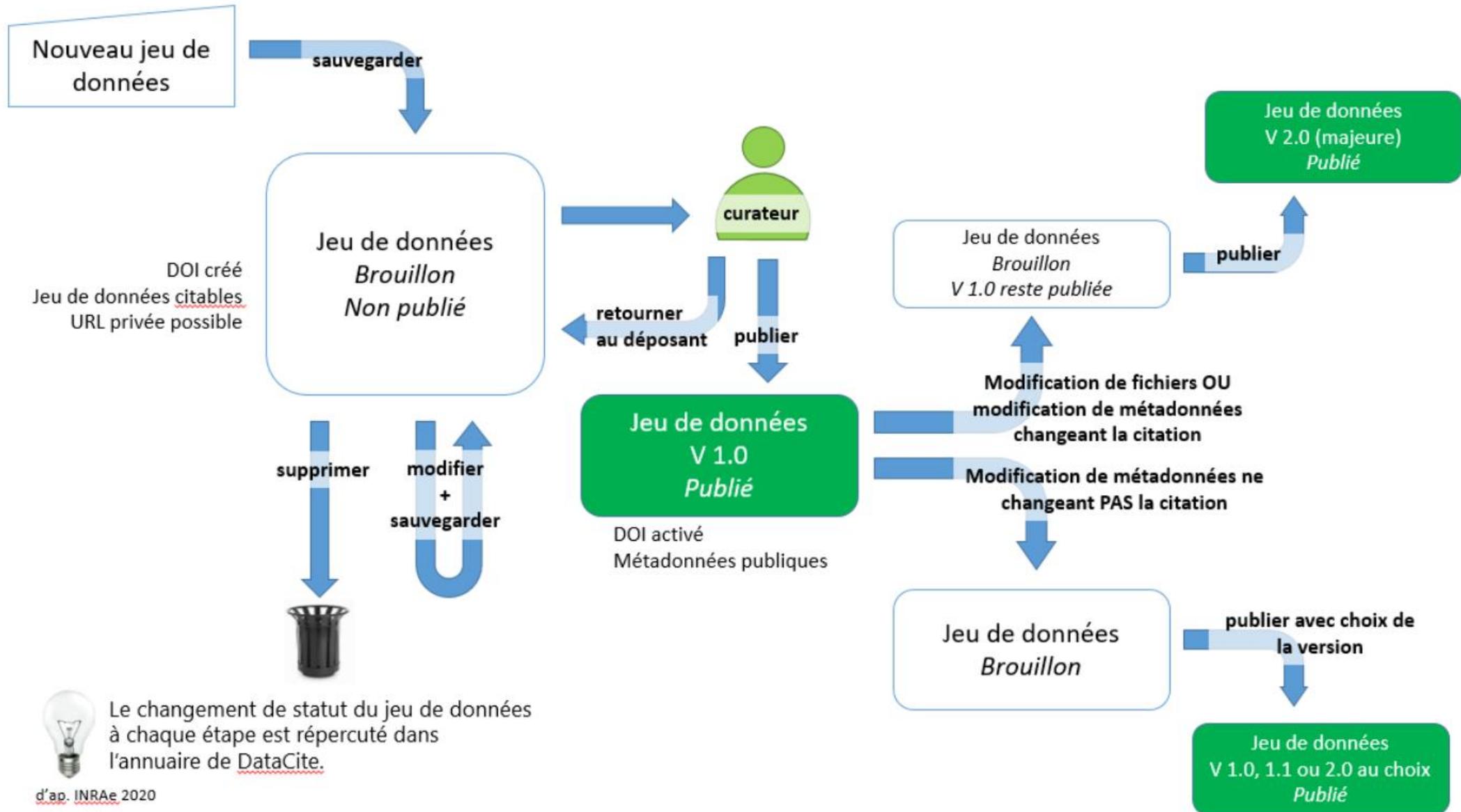
!! Eviter les fichiers Excel à plusieurs feuilles!

>> Des outils supplémentaires (DVUploader, RDM Integration...) permettent de faciliter le téléversement : consultez-nous!

# Les métadonnées

- Elles se remplissent en deux temps ; les métadonnées “supplémentaires” ont leur importance (logiciel utilisé, financements...) : prenez en soin!
- Elles sont adaptées à la description de jeux de données de différentes disciplines, DOREL (et Recherche Data Gouv) étant des entrepôts multidisciplinaires.
- Il est conseillé de renseigner les métadonnées absentes du schéma proposé dans le README d’accompagnement.

# Le flux de travail d'un jeu de données



Le changement de statut du jeu de données à chaque étape est répercuté dans l'annuaire de [DataCite](#).







## Supported File Formats

Tabular Data ingest supports the following file formats:

File format	Versions supported
SPSS (POR and SAV formats)	7 to 22
STATA	4 to 15
R	up to 3
Excel	XLSX only (XLS is NOT supported)
CSV (comma-separated values)	(limited support)

See the subsections in the left sidebar for more information on each of these supported formats.

