

Love Data Week 2023

Université de Lorraine

Déposer dans l'entrepôt de données DOREL

Thomas Jouneau

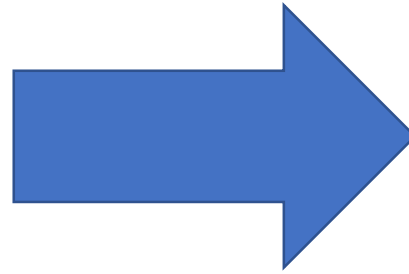
Un événement proposé par l'atelier de la donnée **ADOC Lorraine**
Accompagner aux données les chercheurs et chercheuses en Lorraine

Retrouvez notre offre de services sur notre [site Science Ouverte](#)

Contact : donnees-recherche@univ-lorraine.fr

Pourquoi DOREL?

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



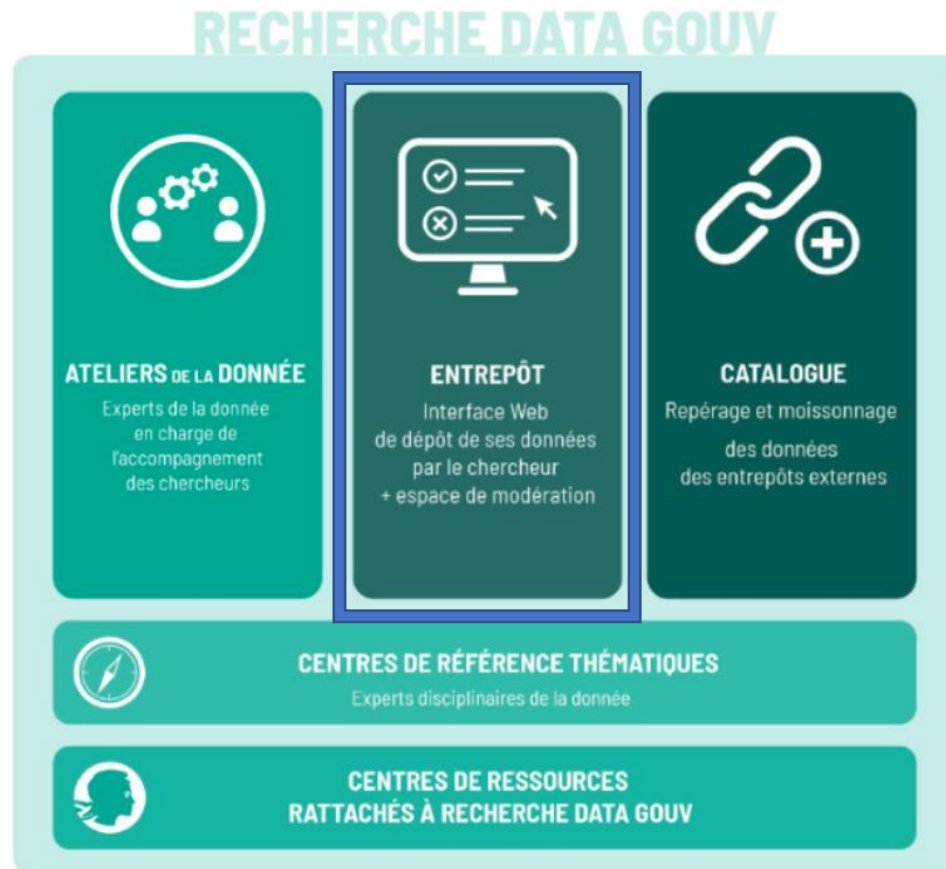
Publication, citabilité

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

Référencement (non atteint)

- *Des possibilités de référencement de jeux de données déposés ailleurs*
- *Un espace de repérage / signalement de l'ensemble des jeux de données d'une unité, d'une équipe...*

DOREL et Recherche Data Gouv



Partie entrepôt
ouverte le 8 juillet 2022

L'Université dispose de son propre espace institutionnel dans lequel les données de DOREL seront signalées par moissonnage

Migration vers RDG
à l'horizon fin 2023-2024

DOREL et RDG : bac à sable vs production

DOREL

Bac à sable

<https://bac-dataverse.univ-lorraine.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

Production

<https://dorel.univ-lorraine.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

Mêmes principes

RDG



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

 RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

recherche.data.gouv.fr

<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).

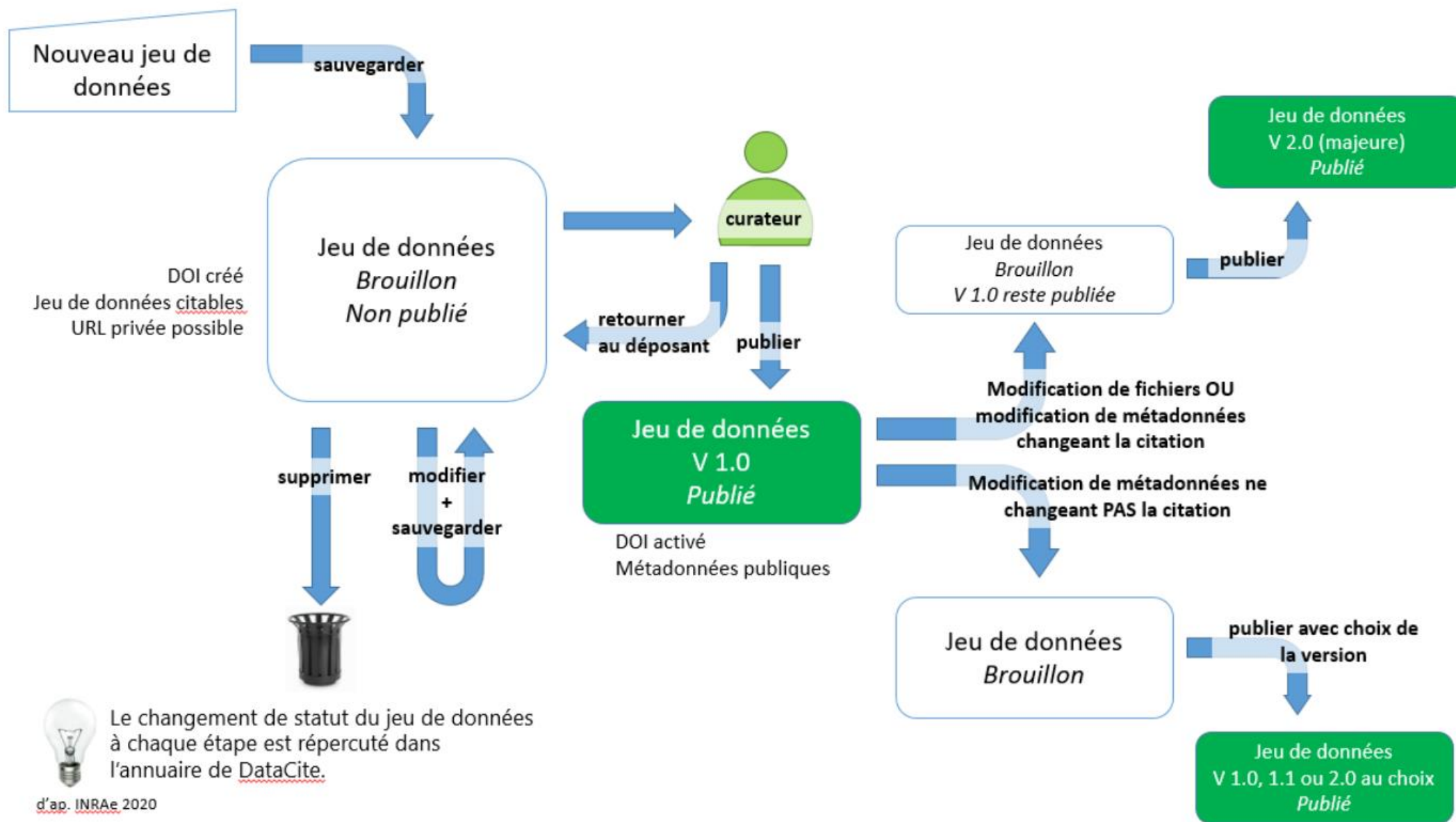
!! Eviter les fichiers Excel à plusieurs feuilles!

>> 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.

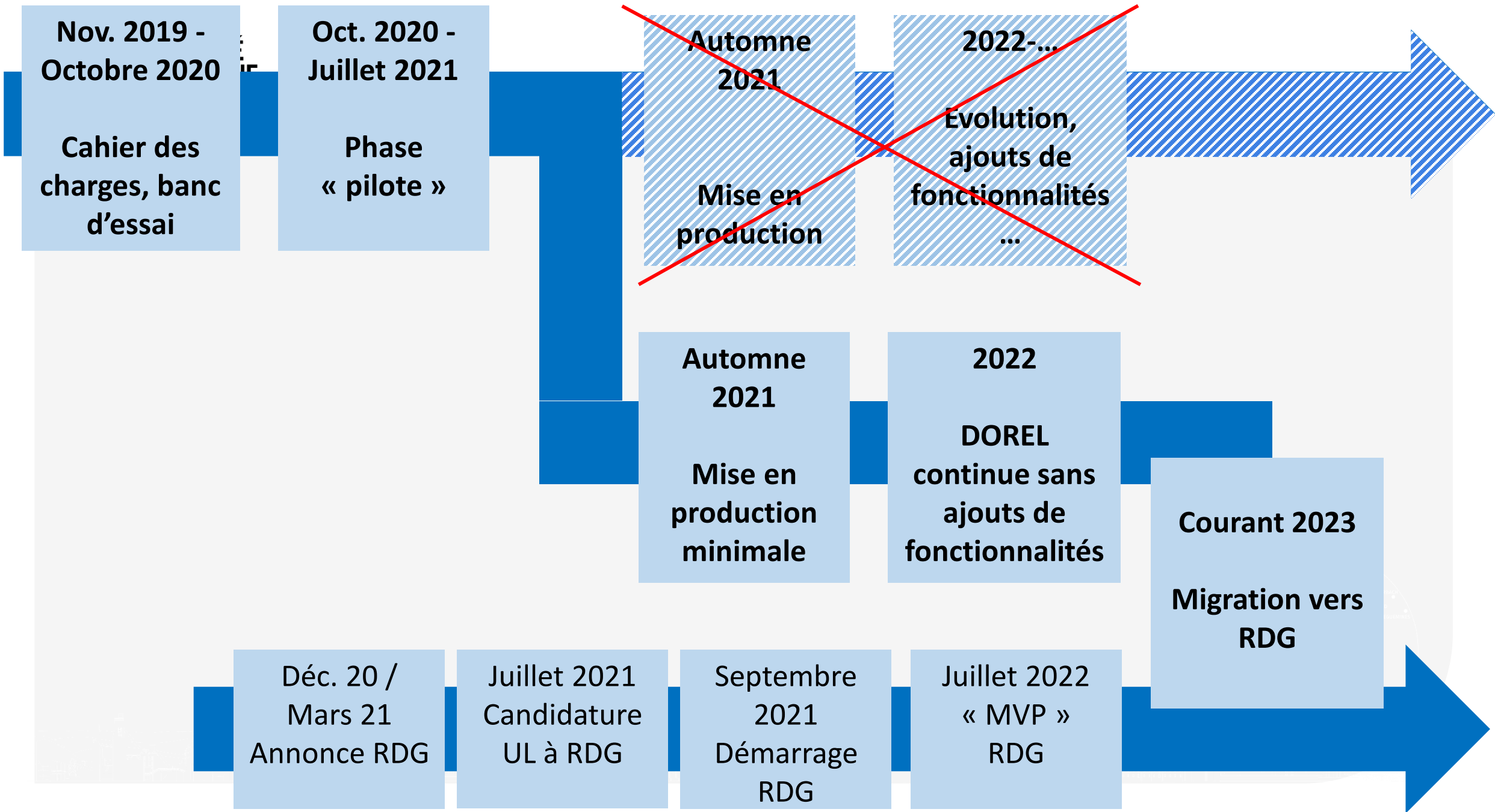
DOREL et RDG : le flux de travail d'un jeu de données



Démonstration

Pour aller un peu plus loin...

- Les guides d'utilisation de DOREL : <http://scienceouverte.univ-lorraine.fr/boite-a-outils/>
- Lier son article dans HAL à un jeu de données dans DOREL : <https://hal.univ-lorraine.fr/page/donnees-de-recherche>
- La documentation officielle de Dataverse (DOREL et RDG) :
 - <https://guides.dataverse.org/en/latest/user/index.html>



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.

