



DOREL

GUIDE DU DÉPOSANT

Table des matières

Qu'est-ce que DOREL ? Qu'est-ce que Dataverse ?	3
DOREL, l'entrepôt de données de l'Université de Lorraine	3
Dataverse, la solution logicielle.....	3
DOREL et le bac à sable.....	4
Pour démarrer : premiers pas, principales notions	5
Se créer un compte.....	5
Demander les droits	Erreur ! Signet non défini.
L'interface de DOREL	6
Premiers pas	7
La gestion des jeux de données	7
Déposer et décrire un jeu de données.....	8
Choisir un modèle (template) si certains ont été définis	9
Choisir la langue de saisie des métadonnées	9
Renseigner les champs essentiels à la description du jeu de données.....	9
Section « Métadonnées de la référence bibliographique / Citation metadata »	10
Pour remplir les mots-clés	12
Utilisation de Loterre (https://www.loterre.fr).....	12
Utilisation de vocabulaires disciplinaires spécifiques.....	12
Section « Fichiers »	13
Téléversement	13
« Ingestion » de fichiers	14
Améliorer la description du jeu de données	15
Blocs de métadonnées complémentaires	16
Métadonnées géospatiales	16
Métadonnées sur les sciences sociales et les sciences humaines	16
Métadonnées liées aux sciences de la vie	16
Métadonnées liées à la revue	16
Choisir la licence de son jeu de données.....	16
Publier le jeu de données.....	17
Le flux de travail de curation.....	17
Le jeu de données est publié !	19
Pour aller plus loin	20
Générer une URL privée.....	20
Créer une URL privée non anonymisée	20
Créer une URL privée anonymisée	21
Mettre des embargos ou restrictions sur des fichiers	21
Mettre un embargo	22
Mettre des restrictions d'accès.....	23
Évolution du jeu de données et numéros de version	24
Créer une nouvelle version d'un jeu de données.....	24
Lier un jeu de données ou une collection dataverse dans une autre collection dataverse	25
Ajouter la référence d'un jeu de données à son profil ORCID.....	27
Optionnel : Libellés (« tags ») de fichiers	28

Qu'est-ce que DOREL ? Qu'est-ce que Dataverse ?

DOREL, l'entrepôt de données de l'Université de Lorraine

DOREL signifie « DONnées de la REcherche Lorraines ».

Il permet :

- la mise en ligne de jeux de données¹ achevés selon le principe "*aussi ouvert que possible, aussi fermé que nécessaire*" en leur attribuant des identifiants numériques pérennes et en permettant notamment leur citation (lien avec des publications, bilans, bibliographies...)
- le référencement et l'identification des données produites et déposées dans d'autres entrepôts.

En lien avec l'exigence nouvelle d'élaborer des plans de gestion de données pour les projets de recherche financés sur fonds publics, il offre une réponse directe à la cinquième partie du [modèle de plan de gestion de données de l'ANR](#) consacrée à l'ouverture et à la sauvegarde des données, en permettant le partage maîtrisé des données et leur conservation à long terme ; et ce en complément des autres entrepôts, comme les entrepôts disciplinaires et ceux des autres acteurs de la recherche.

Il est pensé, dans le système d'information de l'Université, comme étant articulé aux outils existants, que sont l'archive ouverte [HAL](#) pour les publications scientifiques, le mésocentre de calcul [Explor](#), le service de cloud [B'UL](#) et le service de stockage des données de recherche de forte volumétrie [PETA](#).

Il est l'une des briques constitutives des circuits de production et de publication scientifiques sur le site lorrain.

Dataverse, la solution logicielle

Dataverse est un logiciel libre de gestion d'entrepôt de données de la recherche, développé à l'Institute for Quantitative Social Science (IQSS) de Harvard et actuellement utilisé par près de 80 institutions ou services nationaux dans le monde (dont 4 en France).

Il a, également, été choisi comme solution logicielle pour le futur entrepôt national des données de la recherche, Recherche Data Gouv, que DOREL rejoindra à terme.

¹ Jeu de données : « Tout ensemble organisé de données dans un format informatique, défini par un thème ou une catégorie qui reflète ce qui est mesuré/observé/surveillé. La présentation des données dans l'application est rendue possible grâce aux métadonnées. ». <https://codata.org/rdm-glossary/dataset/> .

« Un jeu de données est un ensemble de données ayant une cohérence intellectuelle, qu'elle soit dans l'intention ou dans la forme. Par exemple, une photographie d'un champ, une analyse de sa terre et un relevé hygrométrique de ce champ peuvent constituer un jeu de données cohérent intellectuellement, même s'il est composé d'objets hétérogènes. Un ensemble de tests de matériaux peut aussi constituer un jeu de données, car la forme des données est homogène même si elle porte sur des matériaux très différents les uns des autres. » <https://scienceouverte.couperin.org/donnees-recherche-definitions/>

DOREL et le bac à sable

DOREL

Il se trouve à l'adresse <https://dorel.univ-lorraine.fr>. Les jeux de données publiés sur cette instance le sont de manière définitive (on ne retire pas, sauf cas extrêmement rare, un jeu publié de la circulation) ; les DOI, de même, sont pérennes.

Le bac à sable

Il se trouve à l'adresse <https://bac-dataverse.univ-lorraine.fr>. Comme son nom l'indique, cette instance permet de faire ses premiers pas avec l'outil. Il est recommandé de l'essayer avant de commencer à déposer sur DOREL.

Des jeux de données peuvent y être déposés, mais les DOI qu'ils recevront ne seront pas définitifs, puisque générés via un compte test de Datacite (l'agence gérant les DOI pour les jeux de données). A retenir : les dépôts effectués ne sont **ni pérennes, ni récupérables dans l'instance de production**. Ils seront régulièrement nettoyés.

Pour démarrer : premiers pas, principales notions

Se créer un compte

Se connecter à l'adresse : <http://bac-dataverse.univ-lorraine.fr> pour accéder au bac à sable ou à l'adresse **Error! Hyperlink reference not valid.** pour accéder à DOREL.

Cliquer sur le bouton « Explorer, déposer » puis « Se connecter ».

RECHERCHE ▾ GUIDE D'UTILISATION SUPPORT FRANÇAIS ▾ **SE CONNECTER**

Sur la page suivante, cliquer sur « Votre établissement ». Dans la liste déroulante, sélectionner « Université de Lorraine »



Se connecter sur le service d'authentification locale (UL).



Votre compte est créé.

Attention à bien déposer vos données dans la collection de votre laboratoire !

L'interface de DOREL

The screenshot shows the DOREL web interface. At the top, there is a navigation bar with 'RECHERCHE', 'GUIDE D'UTILISATION', 'SUPPORT', and 'FRANÇAIS'. A user profile for 'THOMAS JOUINEAU' is visible, highlighted with a green box and the number 1. Below the navigation bar, there are statistics (4 021 téléchargements) and a 'Support' button. A warning message states: 'Attention, ceci est un entrepôt de TEST. Aucun dépôt ne sera transféré sur l'entrepôt de production.' A search bar is present with a 'Chercher' button and a link to 'Recherche avancée'. A green box with the number 2 highlights the '+ Ajouter des données -' button. On the left side, there is a sidebar with filters, highlighted with a green box and the number 3. The filters include 'Collections (107)', 'Jeux de données (160)', and 'Fichiers (5 312)'. Below this, there are sections for 'Catégorie du dataverse', 'Source de métadonnées', 'Année de publication', 'Statut de publication', and 'Auteur Nom'. A green box with the number 4 highlights the 'Statut de publication' section. The main content area shows search results, with the first result highlighted by a green box and the number 5. The result is for 'van der Pauw and Hall effect characterization of semiconductor oxides', dated 21 mai 2021. Below it, another result for 'UR AFPA' is highlighted with a green box and the number 6. The interface also shows a 'Tri' button and a 'Version provisoire' label for the first result.

1. Bouton « Se connecter » / Accès aux paramètres du compte
2. Ajouter un jeu de données ou une sous-collection dataverse dans une collection ou sous-collection pour lesquels vos droits ont été activés
3. Filtre sur les trois niveaux d'éléments affichables : collection dataverse, jeux de données, fichiers.
4. Facettes sur les principaux champs
5. Jeu de données : ensemble de fichiers organisé de façon cohérente et décrit par des métadonnées
6. Collection (anciennement « dataverse ») : espace dédié à un laboratoire, une équipe, un projet, une plate-forme.

Premiers pas

Lorsqu'on se connecte à DOREL, il est possible de voir un certain nombre d'entrées, de trois types :

- Des **collections (anciennement « dataverses »)** qui servent d'espaces partagés (voir 6).
- Des **jeux de données** (ou ensembles de données, ou datasets)² (voir 4)
- Des **fichiers**

Ces éléments sont organisés hiérarchiquement : les fichiers sont contenus à l'intérieur des jeux de données, qui sont eux-mêmes contenus à l'intérieur des collections dataverse.

Les permissions (déposer, modifier...) sont gérées au niveau de la collection dataverse et de manière fine : on distingue les déposants, les curateurs (qui vont valider la publication d'un jeu de données)...

Ainsi la gestion des collections dataverses « LMOPS », « HISCANT », « BETA »,... peut appartenir complètement à leurs laboratoires respectifs, avec assistance de l'équipe données de recherche.

Une collection Dataverse peut en contenir d'autres, ce qui rend possible la création d'une collection pour un projet, par ex. :

- Une collection dataverse placée à un endroit donné de l'arborescence (par exemple à la racine, donc au niveau 0) peut être liée dans une collection dataverse de niveau 1 (par exemple une collection dataverse de laboratoire)
- Une collection dataverse pour un projet liant plusieurs laboratoires de l'UL peut ensuite apparaître à l'intérieur des collections dataverse des laboratoires.

Le schéma page 26 permet de visualiser un peu mieux ces concepts.

Un « jeu de données » peut contenir, outre les données, du code ou de la documentation permettant de les exécuter, reproduire ou exploiter.

La gestion des jeux de données

La vie d'un jeu de données comprend plusieurs étapes : la création et l'enrichissement ; la publication ; puis, éventuellement, une évolution au gré de versions supplémentaires.

La création et l'enrichissement consistent à mettre en ligne les fichiers de données en les accompagnant de métadonnées descriptives. Dès cette phase, un identifiant pérenne (DOI) est réservé auprès de l'agence DataCite. Rien n'oblige à finaliser un jeu de données en une seule session : il est tout à fait possible de démarrer par les métadonnées et les fichiers les plus simples, puis de mettre à jour, compléter... éventuellement collaborativement.

Bien que cette partie puisse paraître la plus gourmande en termes de ressources et de temps, de très nombreux outils (notamment les modèles ou l'utilisation de l'API) permettent de réduire considérablement le temps à y consacrer.

Une fois le jeu de données prêt pour publication, il peut être soumis au **curateur**. Tant qu'il n'est ni soumis, ni publié, il peut également être **effacé**. Dans le cas de données publiées avant la parution du ou des articles qui lui seront liés, le jeu de données peut être partagé dès à présent avec des relecteurs ou des collègues via la fonction d'URL privée (voir plus bas « [Générer une URL privée](#) »).

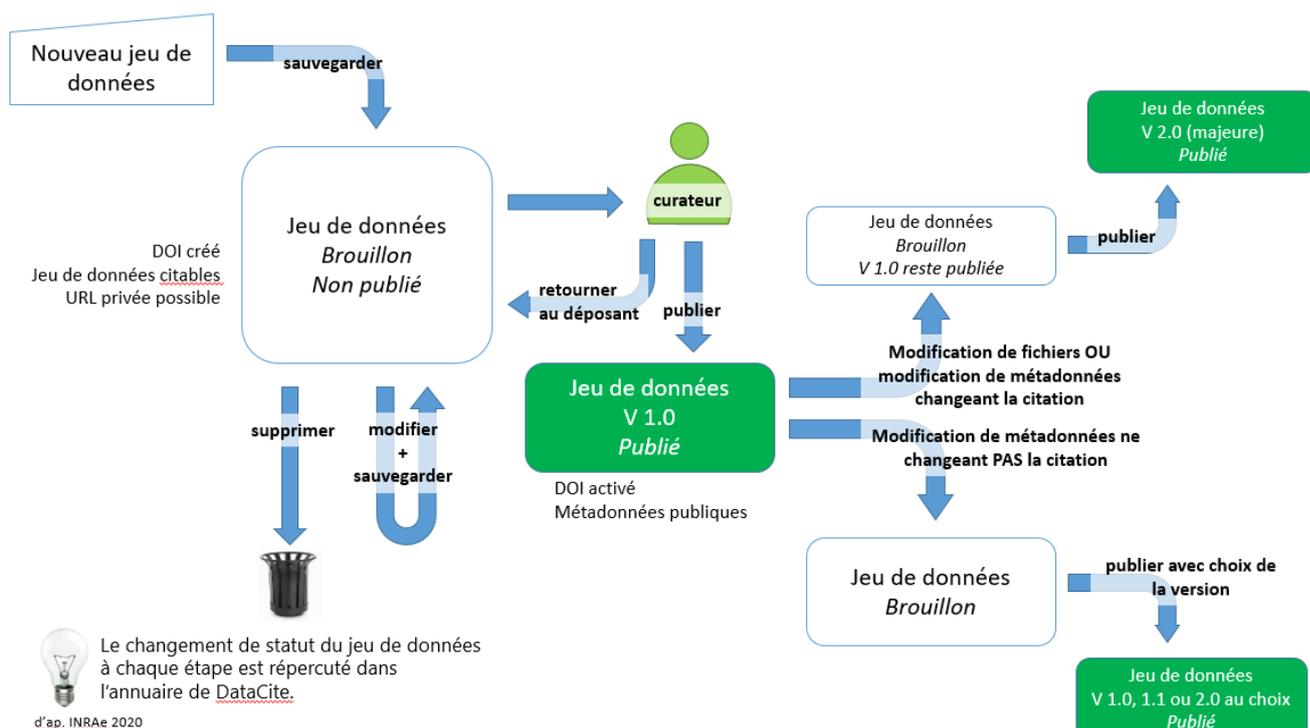
² L'usage commun est « jeu de données » en France et « ensemble de données » au Québec.

La **publication** constitue en quelque sorte la validation du jeu de données créé préalablement.

Un **curateur** ayant cette fonction peut se voir attribuer les droits de publication au sein de votre équipe ou de votre unité, par exemple.

Une fois publié, un jeu de données se voit attribuer un n° de version 1.0 et reste stable dans cet état précis. Le DOI attribué à la création devient définitif.

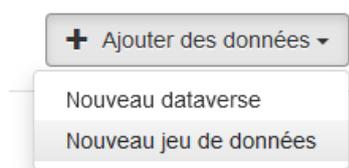
Le jeu publié peut connaître des versions subséquentes, qui peuvent être mineures (modifications ou correction des métadonnées) ou majeures (nouvelles données ajoutées ou remplacées). Le schéma ci-dessous résume ce fonctionnement ; voir également plus bas la section « [Évolution du jeu de données et numéros de version](#) ».



Déposer et décrire un jeu de données

Cherchez la collection dans laquelle le jeu de données doit être déposé, le plus souvent le nom de votre laboratoire.

Cliquer sur « Ajouter des données », puis choisir « Nouveau jeu de données ».



Choisir un modèle (template) si certains ont été définis

Modèle du jeu de données ?

Modifier le modèle effacera tous les champs dans lesquels vous auriez entré des données.

Aucun

Les modèles permettent de préremplir des champs pour accélérer le dépôt. On peut par conséquent créer des modèles par projet, équipe de recherche... où seules les informations variant pour le jeu de données seront à remplir. Si le menu déroulant des modèles n'apparaît pas, il se peut qu'aucun n'ait encore été défini dans la collection Dataverse. N'hésitez pas à contacter l'équipe support ou le curateur de votre laboratoire pour un accompagnement et une mise en œuvre de cette fonctionnalité.

Choisir la langue de saisie des métadonnées

La langue de saisie des métadonnées est la langue de travail du jeu de données. Elle doit correspondre à l'audience attendue du jeu de données.

Langue des métadonnées du jeu de données ?

Sélectionnez la langue que vous souhaitez utiliser pour saisir les métadonnées.

Anglais (English)

Les astérisques indiquent les champs obligatoires

Anglais (English)

Français (French)

Métadonnées de la référence bibliographique

Attention

Une fois la langue choisie, il n'est plus possible de la modifier dans les étapes suivantes.

Il est vivement conseillé de basculer la langue de l'interface dans la langue choisie pour le remplissage des métadonnées.

Renseigner les champs essentiels à la description du jeu de données

La liste des champs pouvant être utilisés pour décrire le jeu de données apparaît. **Tous les champs n'ont pas à être complétés.** Pour vous aider, référez-vous au tableau ci-dessous.

Section « Métadonnées de la référence bibliographique / Citation metadata »

Champ	Sous-champ	Consignes	Exemple	Statut
Titre	Titre	Choisir un titre suffisamment descriptif (type de données, contexte, période de collecte et/ou localisation géographique...). Pour les données de thèse ou liées à un article, il est possible mais non obligatoire d'indiquer le titre de la publication puis de cliquer sur « Ajouter 'données de réplification pour' au titre ».	Consumer perceptions of infant food safety in France	Obligatoire
Auteur	Auteur	Nom d'un auteur de ce jeu de données sous la forme Nom, Prénom (ex. Hugo, Victor).	Rigobert, Elisa	Obligatoire
	Affiliation	L'organisme avec lequel l'auteur est affilié. Pour les laboratoires de l'UL, préciser uniquement le sigle sans espaces ni ponctuations (CRULH, IJL...).	CREM	Obligatoire
	Schéma identifiant	Type (ou « schéma ») de l'identifiant auteur	ORCID	Recommandé
	Identifiant		1111-2222-1234-5678	Recommandé
Personne-ressource (Contact)	Nom	Nom de la personne devant être contactée concernant les données, l'équivalent du champ « corresponding author » pour les publications. Dans la plupart des cas, c'est l'auteur lui-même ; il peut également d'agir d'un simple mail de contact pour certains cas, selon l'organisation pratiquée.	Durand, Jacqueline	Recommandé
	Affiliation	L'organisme avec lequel la personne-contact est affiliée. Pour les laboratoires de l'UL, préciser uniquement le sigle sans espaces ni ponctuations (CRULH, IJL...).	CREM	Recommandé
	Courriel	Courriel de la personne contact pour les données OU courriel générique (alias)	jacqueline.durand@univ-lorraine.fr ; contact-labo@univ-lorraine.fr	Obligatoire
Description	Texte	Court passage descriptif dans la langue de dépôt. L'objectif est globalement le même que pour un résumé d'article : ni trop court ni trop long, il doit inclure l'essentiel de la spécificité et de la caractérisation propre du jeu de données (objet, nature, portée, approche utilisée...).		Obligatoire

	Date	Date de la description au format AAAA-MM-JJ. N'employer QUE si le jeu de données comporte plusieurs descriptions (par exemple, une peut être fournie par le producteur de données et une autre peut être préparée par le dépôt de données où les données sont déposées).		Optionnel
Sujet		Choisir un ou plusieurs sujets appropriés dans la liste. Un seul suffit dans de très nombreux cas.	Sciences sociales	Obligatoire
Mot-clé	Terme	Choisir des termes décrivant au plus près la thématique, l'objet et le contenu du jeu de données. Ces termes peuvent être libres ou, de préférence, être issus de thésaurus ou de vocabulaires existants. Dans ce cas, il est recommandé de préciser le vocabulaire employé ainsi que l'adresse à laquelle il est accessible en ligne. Des exemples de ces vocabulaires sont fournis dans la section « Pour remplir les mots-clés ».	Milk yield Rendement laitier	Obligatoire
	Vocabulaire	Indiquer si possible un vocabulaire existant, et l'URL où l'on peut trouver ce vocabulaire en ligne. Il n'est pas nécessaire de donner l'URL précise du mot employé.	Agrovoc	Recommandé
	Adresse URL du vocabulaire		http://www.fao.org/agrovoc/fr	Recommandé
Publication connexe	Référence	Citation bibliographique de la publication utilisant le jeu de données.		Recommandé
	Type d'identifiant	Renseigner le type d'identifiant concerné. - DOI (le plus fréquent) - Pour les dépôts HAL, laisser vide .	DOI	Recommandé
	Numéro d'identification	Copier-coller l'identifiant choisi (le DOI par exemple). - si DOI : copier tel quel (ex. 10.0000/journal/12345) - si HAL : copier l'identifiant	10.0000/journal/12345 hal-00083055	Recommandé
	Adresse URL	Copier-coller l'URL exacte de la publication.		Recommandé
Notes				Optionnel
Déposant	Remplis automatiquement			Obligatoire
Date de dépôt				Obligatoire

Pour remplir les mots-clés

Les mots-clés sont essentiels, car ils améliorent l'indexation de votre jeu de données par les différents moteurs de recherche. La démarche est strictement identique à celle effectuée pour une publication : choisir des mots clés libres ou issus d'un référentiel pour décrire au mieux le sujet des données déposées.

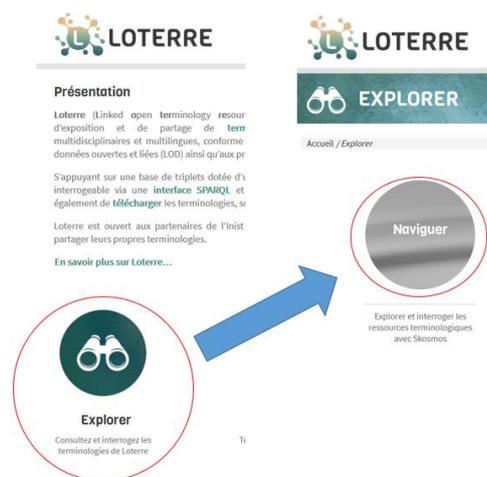
Adopter des termes qui sont extraits d'un thésaurus ou d'un vocabulaire existant, est recommandé lorsque cela est possible. Deux outils peuvent être utilisés, Loterre (développé à l'Inist-CNRS) ; ou directement des vocabulaires spécifiques.

Utilisation de Loterre (<https://www.loterre.fr>)

Loterre (Linked open terminology resources) « est une plateforme d'exposition et de partage de terminologies scientifiques multidisciplinaires et multilingues, conforme aux standards du web des données ouvertes et liées (LOD) ainsi qu'aux principes FAIR. »

Loterre vous permet de trouver des mots-clés standardisés pour décrire vos données, issus de plusieurs vocabulaires de référence. Son utilisation est recommandée.

- ⇒ NB : A la date de rédaction de ce guide (automne 2021), Loterre souffre toujours d'un problème sur le champ de recherche en page d'accueil. Il est conseillé de cliquer sur « Explorer » puis « Naviguer » ou se rendre directement à la page <https://www.loterre.fr/skosmos/fr/>.



Utilisation de vocabulaires disciplinaires spécifiques

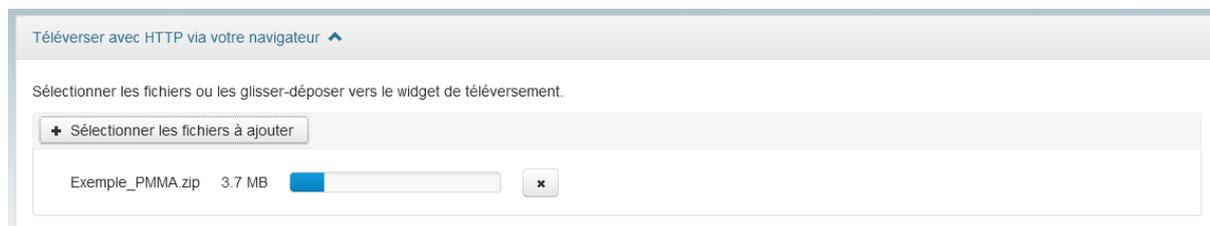
Il existe des milliers de vocabulaires contrôlés spécialisés, et il est impossible de les passer tous en revue. Le tableau ci-dessous en liste quelques-uns à titre purement indicatif.

Discipline	Vocabulaire suggéré (champ « Vocabulary »)	URL
Multidisciplinaire	UNESCO	http://vocabularies.unesco.org/browser/thesaurus/en/
Sciences physiques	PhySH	https://physh.aps.org/browse?facetIds=Research%2520Areas
Archéologie	PACTOLS	https://masa.hypotheses.org/pactols
Sciences économiques	JEL	https://www.aeaweb.org/jel/guide/jel.php
Environnement	GEMET	https://www.lexicool.com/gemet-multilingual-environment-thesaurus.asp?IL=1
Agriculture	AGROVOC	https://agrovoc.fao.org/browse/agrovoc/en/
Sciences de la vie et santé	MeSH	https://www.loterre.fr/skosmos/JVR/fr/
Psychologie	Thésaurus APA	

Section « Fichiers »

Téléversement

C'est ici que les fichiers de données peuvent être téléversés. Plusieurs options existent : téléverser les fichiers successivement et individuellement ; ou les téléverser sous la forme d'une archive compressée au format .zip³.



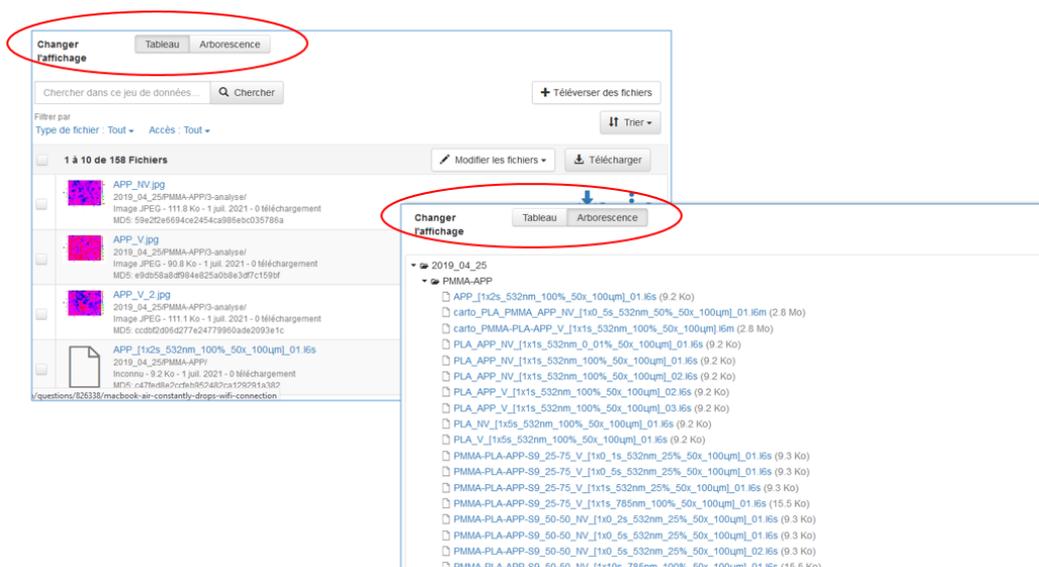
Si la seconde option est choisie, l'arborescence interne de l'archive (si elle existe) sera préservée. Attention : les formats de compression autres que .zip (.rar, 7z...) ne sont pas décompressés et dans ce cas l'archive apparaîtra comme un fichier unique.

Les fichiers une fois chargés, il est possible d'en ajouter ou modifier certaines propriétés : nom, description chemin d'accès (auxquels s'ajoutent les libellés, voir ci-dessous). L'arborescence interne au jeu de données est en effet renseignée via ce champ et peut par conséquent être ajoutée ou modifiée tant que le jeu n'est pas publié.



Dans la page du jeu de données, il sera alors possible de choisir entre la vue à plat des fichiers ou la représentation d'une arborescence.

³ Un utilitaire, DVUploader, permet également de téléverser une arborescence complète depuis un poste : merci de vous rapprocher de l'équipe support pour plus de détails.



« Ingestion » de fichiers

Certains formats de données sont automatiquement convertis en fichiers texte tabulés (.tab). Le terme retenu est celui d'« ingestion ». Une fois cette opération effectuée, il est possible de télécharger depuis le jeu de données, soit le fichier original, soit le fichier converti. Les fichiers .tab créés par Dataverse comportent des guillemets dans l'encadrement des cellules.

Les formats suivants font l'objet d'une tentative automatique de conversion⁴ :

- SPSS
- Stata
- données R
- Excel (.xlsx seulement)
- .csv

Dans l'exemple ci-dessous, le fichier MaitriseSDSC.csv a été automatiquement converti au format .tab. Les deux fichiers (l'original, le .tab) sont tous deux accessibles au téléchargement.



MaitriseSDSC.tab

semis_direct_sous_couvert/data/
Données tabulaires - 92.2 Ko - 22 juil. 2021 - 0 téléchargement

96 Variables, 143 Observations - UNF:6:IXCT2cmHHdQ3Xh8VY+m6Aw==



Options de téléchargement 

Format original du fichier (Valeurs séparées par des virgules)

Valeurs séparées par tabulations

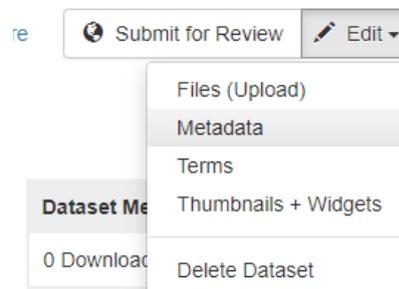
⁴ La liste à jour se trouve sur la page suivante : <https://guides.dataverse.org/en/latest/user/tabulardataingest/supportedformats.html>

Améliorer la description du jeu de données

Le jeu apparaît désormais dans la collection dataverse du laboratoire. Vous ne verrez pas les jeux de données des autres, et c'est normal, car pour le moment, les jeux ne sont pas publiés, et restent avec les statuts « Version provisoire » (en anglais : Draft) et « Non publié » (en anglais : « Unpublished »).

Il faut à présent compléter les descriptions. Entrer sur la fiche du jeu puis cliquer sur Edit > Metadata / Modifier l'ensemble de données > métadonnées.

Tout un nouvel ensemble de champs supplémentaires apparaît. Nul besoin de tout remplir mais la sélection ci-dessous est toutefois très utile et il est suggéré de la renseigner. Les modèles, décrits plus hauts, permettront de pré-remplir ces champs, ce qui accélère grandement le dépôt de jeux de données similaires.



Champ	Sous-champ	Consignes	
Contributeur	Type	Choisir un rôle dans la liste déroulante.	Recommandé
	Nom	Collaborateurs, différents des auteurs, ayant contribué au jeu de données. Nom sous la forme Nom, Prénom OU renseigner l'institution contributrice.	Recommandé
Langue		Choisir une ou plusieurs langues dans la liste déroulante. Ce champ doit servir à qualifier la ou les langue(s) utilisée(s) dans les fichiers de données (la langue des métadonnées ayant été précisée à la création).	Recommandé
Date de production		Date à laquelle les données ont été produites.	Recommandé
Endroit de production		Lieu où les données ont été produites.	Recommandé
Renseignements sur la subvention	Organisme subventionnaire	<ul style="list-style-type: none"> Financement ANR : Agence nationale de la recherche Financement Horizon 2020 ou Europe : Union européenne / European Union (si le jeu est renseigné en anglais) Autre financement (DRAC, collectivité,...) : indiquer le plus clairement possible le nom du financeur, en développant. Ex. : Caisse des dépôts et consignations. 	Recommandé si applicable
	N° de la subvention	Le numéro de subvention ou de contrat lié au financement.	Recommandé si applicable
Type de données		Choisir dans la liste déroulante, le type de données concerné: survey data, census/enumeration data, aggregate data, clinical data, event/transaction data, program source code, machine-readable text, administrative records data, experimental data, psychological test, textual data, coded textual, coded documents, time budget diaries, observation data/ratings, process-produced data, or other. Le champ peut être répété.	Recommandé
Logiciel	Nom Version	Nom et version du logiciel utilisé. Ex. : OriginLab 2021b.	Recommandé

Blocs de métadonnées complémentaires

Plusieurs blocs de métadonnées complémentaires peuvent être activés au niveau de la collection dataverse. Ces schémas de métadonnées complémentaires apparaissent sous la forme de « blocs » de métadonnées qui ne sont pas présents lors de la création initiale mais lorsque l'on retourne sur la fiche du jeu de données pour le compléter. Les métadonnées de ces blocs complémentaires définies comme obligatoires apparaissent dès la création initiale.

Les blocs de métadonnées inclus dans Dataverse sont les suivants.

Métadonnées géospatiales

Ce bloc est intéressant si le jeu de données concerne un lieu ou une aire géographique précis. Ces derniers peuvent être définis par le renseignement d'un toponyme, ou par une aire définie en longitude et latitude. La définition d'une aire permet, par exemple, in fine, l'affichage d'un jeu de données sur une carte (telle que Google Maps ou OpenStreetView).

Champ	Sous-champ	Consignes, exemple	
Couverture géographique	Pays/Nation	« France »	Recommandé si applicable
	État/province	« Grand Est » ; « Lorraine ».	
	Ville	« Nancy »	
	Autre	« Massif du Fossart »	
Unité géographique	Niveau de regroupement géographique le plus bas porté par le jeu de données. Exemple : village, pays, région.		
Zone de délimitation géographique	Longitude est et ouest, Latitude nord et sud.	Pour un secteur géographique ne correspondant pas à une commune ou une subdivision administrative (ex. : Massif du Fossart), il est recommandé d'encadrer la zone par un rectangle « au plus près » du secteur effectivement concerné.	

Métadonnées sur les sciences sociales et les sciences humaines

Elles sont basées sur le [modèle de métadonnées DDI](#). Ce bloc est particulièrement intéressant pour les données d'enquête ou les données sociologiques. Merci d'écrire à donnees-recherche@univ-lorraine.fr pour toute demande d'accompagnement sur l'utilisation de ce bloc.

Métadonnées liées aux sciences de la vie

Elles sont basées sur le modèle de métadonnées *ISA-Tab*. Merci d'écrire à donnees-recherche@univ-lorraine.fr pour toute demande d'accompagnement sur l'utilisation de ce bloc.

Métadonnées liées à la revue

Ce bloc permet de compléter le cas échéant les données d'article lié du bloc de métadonnées principales.

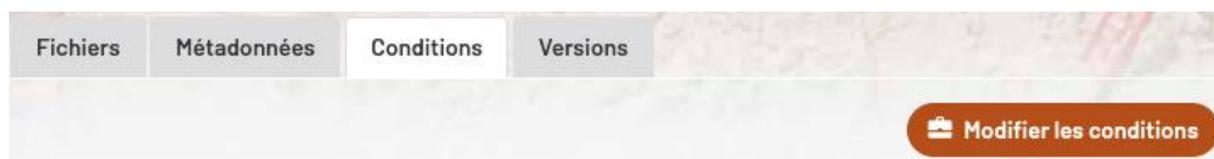
Choisir la licence de son jeu de données

Depuis la mise à jour d'avril 2022, DOREL permet une plus grande liberté de choix quant à la licence attribuable à un jeu de données.

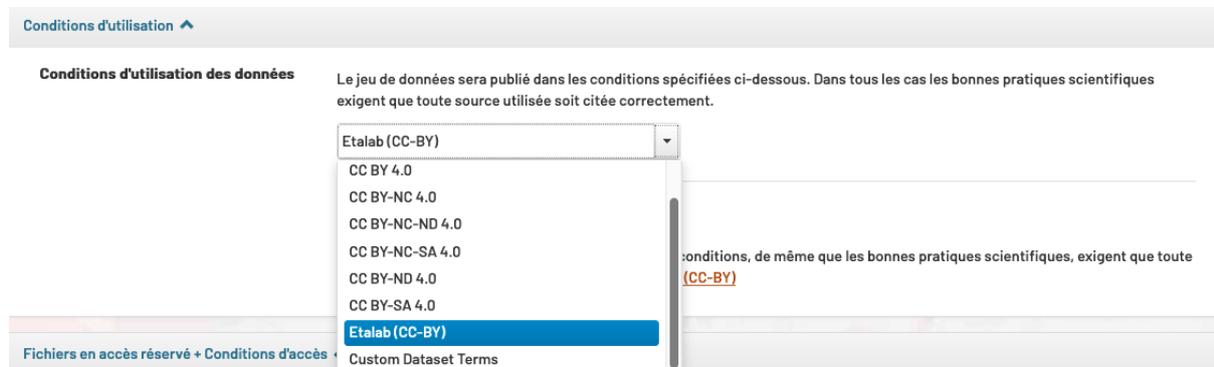
Par défaut, un jeu de données créé dans DOREL se voit attribuer la licence Etalab⁵. Cette licence créée dans le cadre de la politique française d'ouverture des données publiques permet le partage et la réutilisation des données, tout en rendant obligatoire la mention de paternité. Elle est, de fait, compatible avec la licence Creative Commons CC-BY.

Dans certains cas, l'on peut souhaiter attribuer une autre licence à un jeu de données.

Pour ce faire, depuis la fiche du jeu de données, cliquer sur « Conditions » puis sur « Modifier les conditions ».



Le menu déroulant permet de choisir entre Etalab et l'ensemble des licences Creative Commons⁶.



Dans certains cas, d'autres licences ou conditions spécifiques doivent être précisées. Dans ce cas, sélectionner la dernière entrée du menu déroulant (Conditions personnalisées), puis utiliser les champs qui s'afficheront automatiquement au-dessous. Il est recommandé de renvoyer vers des termes de licence existants et d'en préciser simplement le lien dans le champ « Conditions d'accès ».

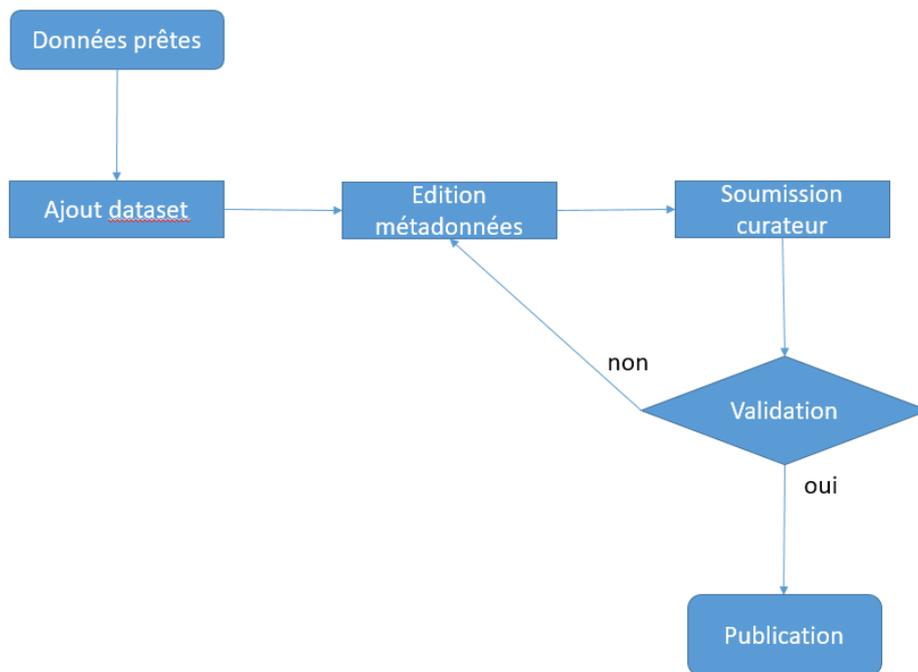
Publier le jeu de données

Le flux de travail de curation

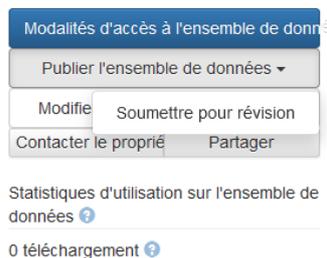
Une fois le jeu de données prêt, il doit être soumis au curateur pour publication. Il est proposé à chaque laboratoire de définir une ou deux personnes en charge de la curation des données de la collection dataverse attribué à l'unité de recherche. Dans ce cas, l'équipe Données de la recherche ne sera plus en charge de la curation et la procédure de création et de publication se modifie légèrement, pour intégrer la phase de validation dans le flux de travail du laboratoire.

⁵ <https://www.etalab.gouv.fr/licence-ouverte-open-licence/>

⁶ Pour plus de renseignements sur les licences Creative Commons et leurs spécificités, se référer au site officiel : <https://www.creativecommons.org>.



Après création du jeu de données, cliquer sur « Publier » fait apparaître pour le déposant l’unique option « Soumettre pour révision ».



Un avertissement s’affiche :

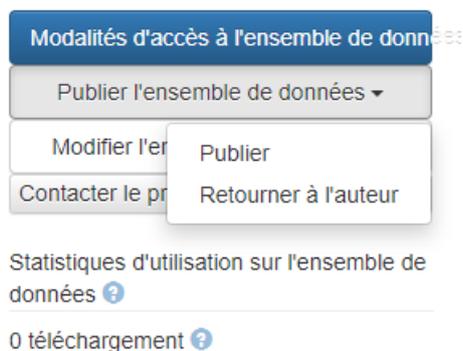


Une fois cette opération réalisée, le jeu de données acquiert le statut « En révision » :



Le curateur reçoit une notification par mail et peut, après examen, valider la publication ou renvoyer à l’auteur pour révision. A noter que l’interface ne permet pas de préciser les raisons de ce « refus » provisoire, un mail séparé est donc nécessaire.

Les critères à vérifier ainsi que l'ensemble de la procédure sont détaillés dans le **guide du curateur**.



Le jeu de données est publié !

La publication du jeu de données entraîne plusieurs conséquences immédiates :

- Le jeu de données est visible par tous, y compris les utilisateurs non connectés.
- Il passe du statut « Draft » (Brouillon) au statut « Published » dans Datacite et est automatiquement passé en version 1.0.
- Un DOI pérenne est attribué au jeu de données, ce qui permet notamment de lier le jeu de données à des publications scientifiques.

Pour aller plus loin

Générer une URL privée

Tant que le jeu de données n'est pas publié, il est possible de générer une URL privée permettant de donner accès aux fichiers à un tiers, sans que ce dernier ait besoin de disposer d'un compte sur DOREL.

Cette fonctionnalité peut être utile dans le cadre des procédures de revue par les pairs, ou pour tout autre besoin de communication des fichiers avant publication.

Deux types d'URL privée, anonymisée et non anonymisée, sont proposées.

Pour créer une URL privée, il est nécessaire soit de vous adresser au curateur de votre laboratoire, soit d'écrire à donnees-recherche@univ-lorraine.fr. Cette obligation est malheureusement liée aux limitations du logiciel Dataverse.

Créer une URL privée non anonymisée

Dans le cas d'une URL privée non anonymisée, l'ensemble des éléments du jeu de données (métadonnées et fichiers) sont accessibles à toute personne disposant de l'URL.

En voici un exemple :

Citation Metadata ^	
Dataset Persistent ID ?	doi:10.80427/FK2/OHVQEI
Title ?	Social media use among teens
Author ?	Doe, Jane (CREM) - ORCID: 0000-0000-1234-6789
Contact ?	Use email button above to contact. Doe, Jane (CREM)
Description ?	The Social Media Use Among Teens survey was conducted by the Youth Communication Development Project to understand social media communication behaviours among youth in Canada. The survey collected responses from Canadian youth using an online questionnaire that asks about social media use including, platform type, frequency of use, activity type, and location of use. This information is supplemented with the respondent's demographic and household characteristics.
Subject ?	Social Sciences
Keyword ?	social media teens communication
Notes ?	This survey was organized online. Mode of interview has been found to impact results, therefore it is not recommended that these results are compared with other survey results where the interview mode was telephone based
Language ?	English
Depositor ?	Doe, Jane
Deposit Date ?	2022-03-31

Créer une URL privée anonymisée

Dans le cas d'une URL privée anonymisée, un certain nombre de métadonnées sont masquées, permettant notamment la revue par les pairs en double aveugle.

Les personnes destinataires ne verront que les métadonnées ne comportant pas d'information d'authentification. En voici un exemple :

Citation Metadata ^	
Dataset Persistent ID ?	doi:10.80427/FK2/OHVQEI
Title ?	Social media use among teens
Author ?	withheld
Contact ?	withheld
Description ?	The Social Media Use Among Teens survey was conducted by the Youth Communication Development Project to understand social media communication behaviours among youth in Canada. The survey collected responses from Canadian youth using an online questionnaire that asks about social media use including, platform type, frequency of use, activity type, and location of use. This information is supplemented with the respondent's demographic and household characteristics.
Subject ?	Social Sciences
Keyword ?	social media teens communication
Notes ?	This survey was organized online. Mode of interview has been found to impact results, therefore it is not recommended that these results are compared with other survey results where the interview mode was telephone based
Language ?	English
Depositor ?	withheld
Deposit Date ?	2022-03-31

Mettre des embargos ou restrictions sur des fichiers

Dans certains cas (données à caractère personnel non anonymisées, données de santé, données tierces...), certaines données ne doivent ou ne peuvent être diffusées librement ou immédiatement.

DOREL permet de répondre à ces cas de figure, tout en permettant la conservation, la gestion de l'accès et la citation de ces jeux de données.

Il est possible de restreindre l'accès à des fichiers de manière très fine (fichier par fichier) :

- en précisant une date d'ouverture (**embargo**)
- en posant une **restriction** d'accès sans date de fin.

Ces restrictions valent pour l'ensemble des personnes n'ayant pas de droits sur le jeu de données, qu'elles soient identifiées dans DOREL ou non.

Il n'est possible de mettre des restrictions ou des embargos que tant que le jeu de données n'a pas été publié.

La restriction :

- n'a pas de limite dans le temps et doit être levée par le déposant ou le curateur ;
- permet d'autoriser la demande de communication, pour un utilisateur connecté à DOREL.

L'embargo :

- est limité dans le temps (date à préciser au moment du dépôt) ;
- au bout de la durée spécifiée, le jeu est automatiquement ouvert.

Mettre un embargo

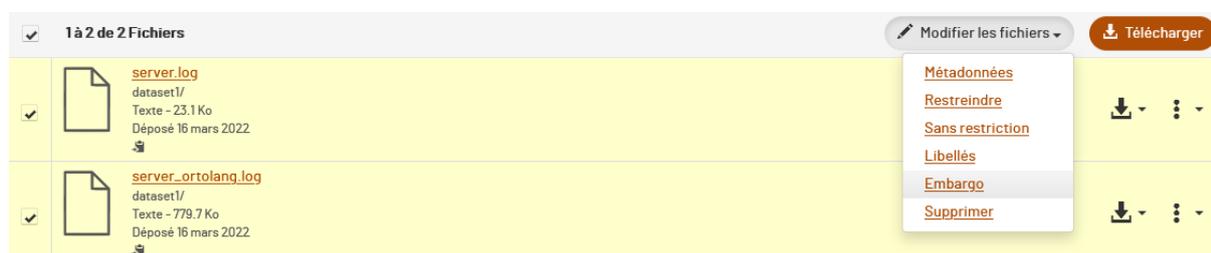
L'embargo est utilisé pour limiter l'accès aux données pendant un temps défini.

Quelques exemples :

- Lorsque les données déposées sont liées à un article en attente de publication.
- Lorsque les données sont liées à des brevets.
- Lorsqu'un contrat ou une convention avec un partenaire impose une confidentialité des données pendant une durée définie (par exemple, 6 mois après la fin du projet).

D'autres cas de figure peuvent exister.

Depuis la page du jeu de données, sélectionner tout ou partie des fichiers déposés, puis cliquer sur « Modifier les fichiers » puis « Embargo » :



Sélectionner une date. La raison de l'embargo DOIT être précisée.

Par défaut, DOREL fixe la durée maximale d'embargo à **X** ans. Il est toutefois possible de définir un embargo plus long : merci en ce cas de contacter la cellule donnees-recherche@univ-lorraine.fr .

Edit Embargo ✕

Edit the planned embargo for the selected file or files. Once this dataset version is published, you will need to contact an administrator to change the embargo date or reason of the file or files.

Add or Change

Select the embargo end-date *

Enter a short reason why this embargo exists

Mettre des restrictions d'accès

La restriction d'accès est utilisée pour limiter l'accès aux données sans limite de temps.

En raison de son caractère strict, la restriction d'accès n'est à utiliser que pour les quelques exceptions à l'ouverture par défaut prévue par la loi⁷.

Quelques exemples :

- Les données correspondant à des informations à caractère personnel, sauf si anonymisation, consentement des personnes ou autorisation légale ou réglementaire.
- Les données relevant de la protection du potentiel scientifique et technique de la nation (notamment certaines données issues des laboratoires situés en zone à régime restrictif – ZRR)⁸.

Dans la fiche du jeu de données, sélectionner tout ou partie des fichiers déposés, puis cliquer sur « Modifier les fichiers » puis « Restreindre » :

The screenshot shows a file management interface with two files selected. The first file is 'server.log' (Texte - 23.1 Ko, Déposé 16 mars 2022, MD5: 969...aff). The second file is 'server_ortolang.log' (Texte - 779.7 Ko, Déposé 16 mars 2022, MD5: 43c...f90). A dropdown menu is open over the first file, showing options: 'Métadonnées', 'Restreindre', 'Sans restriction', 'Libellés', 'Embargo', and 'Supprimer'. The 'Restreindre' option is highlighted. There are also buttons for 'Modifier les fichiers' and 'Télécharger'.

Il est demandé de renseigner les conditions d'accès au fichier, puis de cocher la possibilité de demander un accès au questionnaire du jeu de données.

⁷ Pour une liste complète des cas concernés, voir les documents proposés sur la page Boîte à outils du site Science Ouverte de l'UL, rubrique « Questions juridiques » : <https://scienceouverte.univ-lorraine.fr/boite-a-outils/>

⁸ Toutes les données issues des laboratoires en ZRR ne sont pas exclues du régime d'ouverture par défaut des données. Pour plus d'informations ou un avis concernant l'éligibilité à la restriction d'ouverture des données concernées, consulter le référent ZRR de l'unité de recherche.

Restrictions d'accès au fichier ✕

Conditions d'accès ⓘ

Data under access restrictions due to XXXXX
 For any communication request please use the "Request access" button or write to XXXXX.

Demander l'accès ⓘ Autoriser la demande d'accès

Continuer Annuler

Les restrictions peuvent être ôtées par la procédure inverse (sélection des fichiers concernés et choix de l'option « Sans restriction » dans le menu déroulant).

Évolution du jeu de données et numéros de version

Un jeu de données publié se voit attribuer la version 1.0. Cette publication est définitive : en bonne pratique, les données publiées dans cette première version doivent rester accessibles.

Ceci ne signifie pas que le jeu de données n'évoluera pas, bien au contraire : il est tout à fait possible de créer de nouvelles versions, et de les publier pour les rendre accessibles de la même façon que la première. Cette modification sera soumise au même cycle de curation que décrit précédemment. Après la publication d'un jeu de données modifié, les versions antérieures restent consultables dans l'onglet « Versions » du jeu de données :

Jeu de données	Résumé	Contributeurs	Publié
2.0	Métadonnées bibliographiques additionnelles : (1 Ajouté); Fichiers (Ajouté(s) : 1); Voir les renseignements	Thomas Jouneau	2 sept. 2021
1.0	Il s'agit de la première version publiée.	Thomas Jouneau	2 sept. 2021

Créer une nouvelle version d'un jeu de données

Une modification d'un jeu de données intervient lorsqu'on édite les métadonnées d'un jeu publié (Modifier le jeu de données > Métadonnées) ou lorsqu'on y ajoute, retranche ou remplace des fichiers.



Deux règles régissent la numérotation des versions au-delà de la 1.0.

Ces règles peuvent se ramener à une seule très simple : « si la citation du jeu de données change, le n° de version majeure doit changer aussi ».

1. **Modification des métadonnées ou des conditions d'accès (licence) sans modification de données** : le choix est donné au moment de la publication de mettre à jour le jeu de données :
 - a. avec le même n° de version
 - b. avec un n° de version mineur (1.1, 1.2...)
 - c. avec un n° de version majeur (2.0, 3.0...)

Publier le jeu de données

⚠ Êtes-vous sûr(e) de vouloir publier à nouveau ce jeu de données ?

Indiquer s'il s'agit d'une mise à jour mineure ou majeure de la version.

Version mineure (1.1) Version majeure (2.0) Mettre à jour la version actuelle (ceci écrasera définitivement la dernière version publiée).

Continuer

Annuler

2. **Modification des données avec ou sans modification des métadonnées** : une version majeure (2.0, 3.0...) est imposée.

Publier le jeu de données

⚠ Êtes-vous sûr(e) de vouloir publier à nouveau ce jeu de données ?

En raison de la nature des modifications apportées à la version provisoire actuelle, il s'agira d'une version majeure (2.0).

Continuer

Annuler

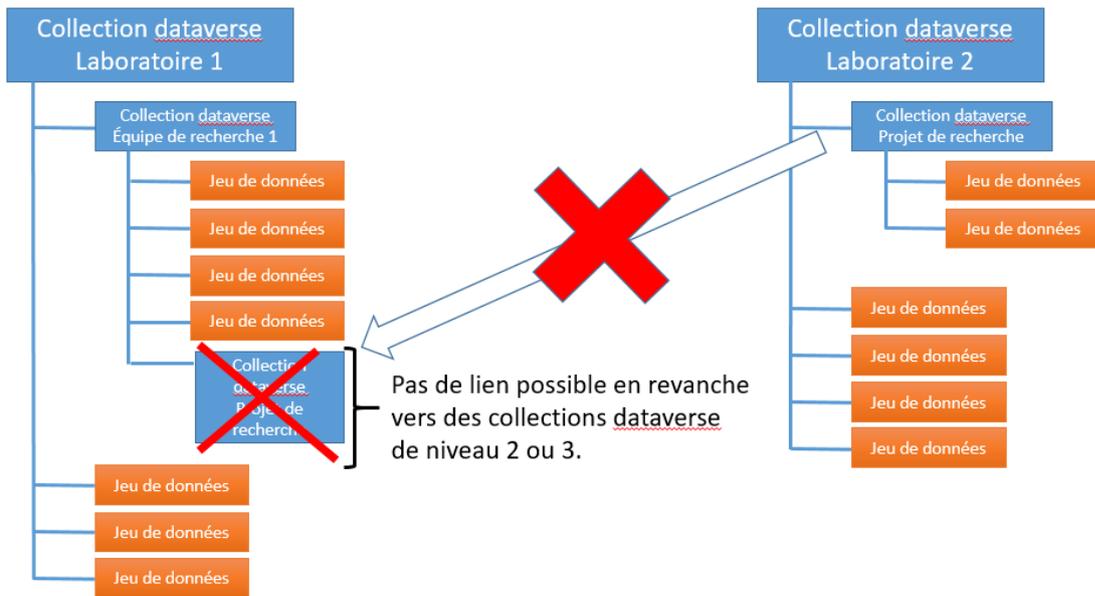
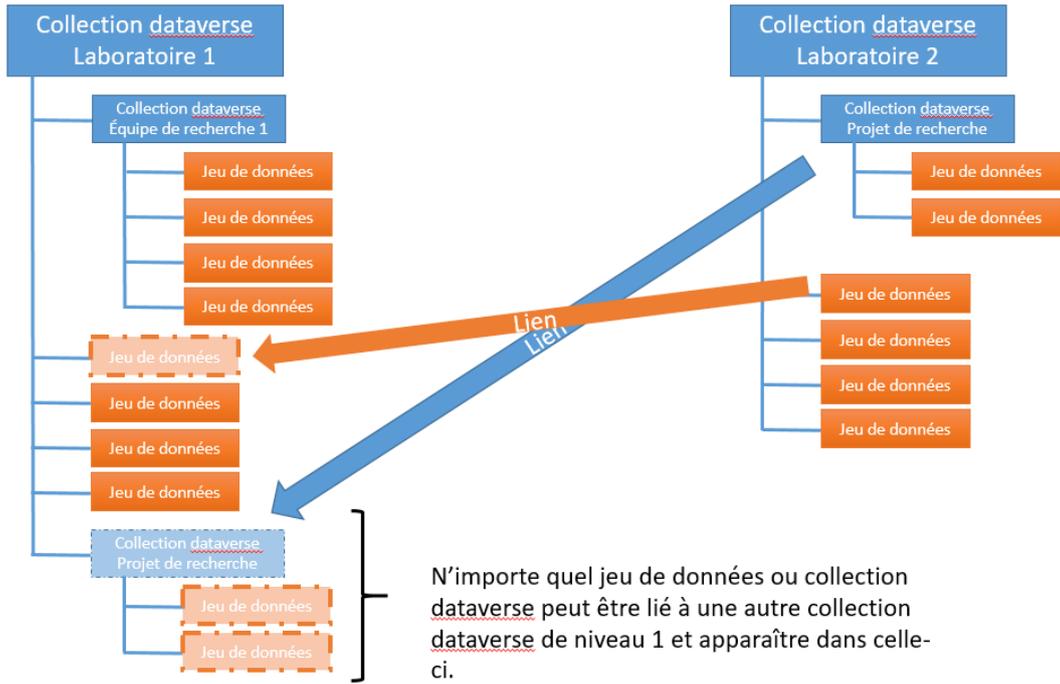
Le schéma de la section « [La gestion des jeux de données](#) » résume ce fonctionnement.

Lier un jeu de données ou une collection dataverse dans une autre collection dataverse

Il est possible de créer des liens de jeux de données ou de collections dataverse, dans une collection dataverse de son choix. Dans l'état actuel de cette fonctionnalité, il n'est possible de créer ces liens que vers des collections dataverse de niveau 1. Des jeux de données ainsi liés seront considérés comme faisant également partie de la collection cible⁹.

Pour réaliser un tel lien, merci de contacter l'équipe support : donnees-recherche@univ-lorraine.fr.

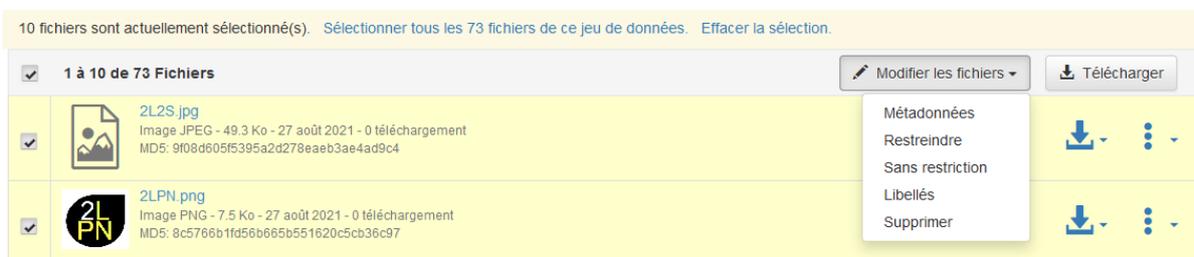
⁹ On peut faire le parallèle du fonctionnement de ces liens avec celui, par exemple, des liens symboliques (« alias ») dans les systèmes Unix.



Optionnel : Libellés (« tags ») de fichiers

Les libellés de fichiers permettent de signaler de façon simple des caractéristiques des fichiers téléversés. Leur renseignement reste optionnel.

Sélectionner les fichiers à libeller. Depuis le haut de la liste des fichiers cliquer sur « Modifier les fichiers » puis « Libellés ».



Dans la boîte de dialogue qui s’ouvre, sélectionner un ou plusieurs des trois libellés existants, ou en créer de nouveaux.

