



Déposer des données dans l'entrepôt *Recherche Data Gouv*

Cellule Data Grenoble Alpes

atelier de la donnée *Recherche Data Gouv* de l'Université Grenoble Alpes

■ Sommaire

- 1) L'écosystème *Recherche Data Gouv*
- 2) Le dépôt de données
- 3) La modération et la publication
- 4) Lien publications, données, codes
- 5) Quelques ressources

1. L'écosystème *Recherche Data Gouv*

Un écosystème au service du partage et de l'ouverture des données de recherche

FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER

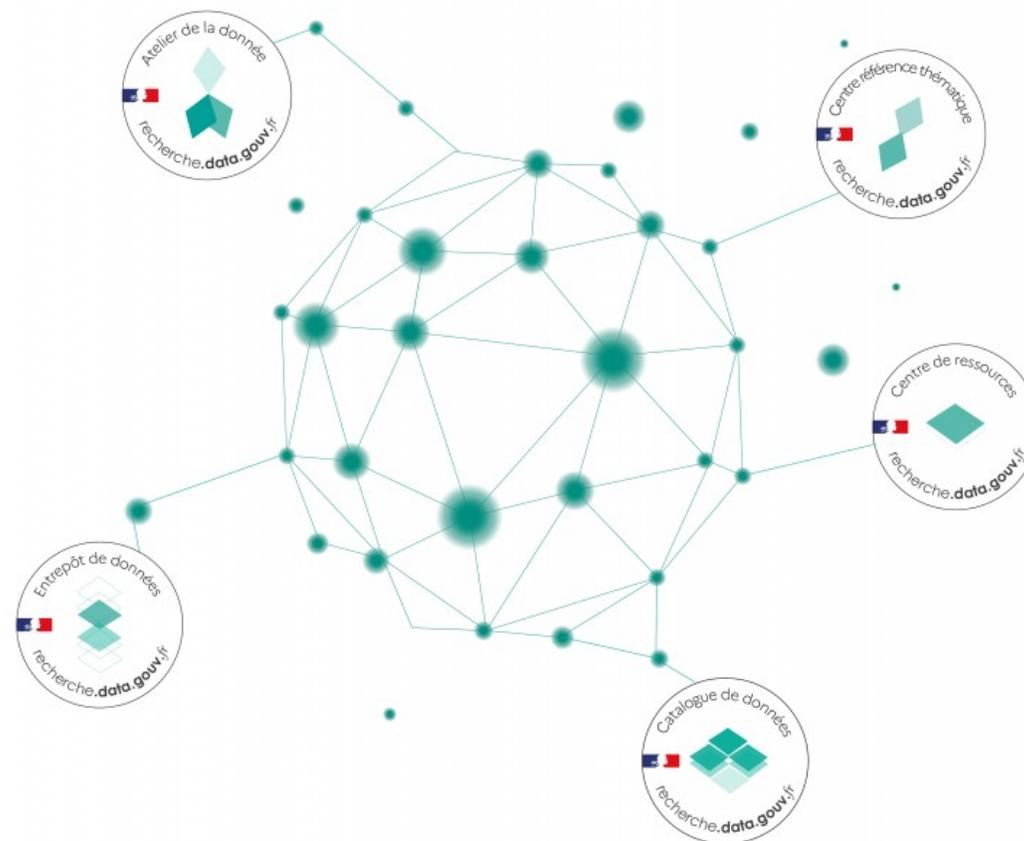


Trois modules pour accompagner les équipes de recherche sur toute question relative aux données :

- › Des **ateliers de la donnée**.
- › Des **centres de référence thématiques**.
- › Des **centres de ressources**.

Deux modules pour déposer, publier et signaler des données :

- › Un **entrepôt** pour déposer et utiliser des données.
- › Un **catalogue** pour rechercher les données publiées sur l'entrepôt ou sur des entrepôts externes.



■ Volet entrepôt

- **Recherche Data Gouv : solution souveraine** pour permettre aux chercheurs n'ayant pas d'entrepôt thématique de déposer leur données de recherche.
- **Piloté par le MESR et porté par l'INRAE** sur la partie entrepôt / catalogue. Financé sur le **FNSO**.
- Des universités et organismes impliqués (Grenoble-Alpes, Lorraine, Lille, Paris, Paris-Nanterre, Strasbourg, CNRS, INRAE), membres du **comité de pilotage Recherche Data Gouv**.
- Basé sur une solution logicielle open source **Dataverse**.
- Organisé en **espaces institutionnels** de publication et de signalement des données des établissements partenaires.

■ Dépôt de données en pratique

- **Tous les types de données** de recherche sont acceptés.
- **Taille maximale** des jeux de données : 50 Go par fichier.
- **Accès restreint et embargo** possible pour certains ou tous les fichiers d'un jeu de données.
- Possibilité de créer **une url privée ou en partie anonyme** pour un jeu de données non publié, par exemple pour de la relecture par les pairs.
- Possibilité de faire **un lien depuis un jeu de données vers une autre collection** : par exemple lors d'un projet entre plusieurs laboratoires.
- Chaque administrateur de collection choisit **les métadonnées à remplir** et peut définir **un modèle de dépôt**.

2. Le dépôt de données

■ Prérequis

Vérifications préalables

- Les données n'ont pas déjà été publiées.
- Il n'existe pas d'entrepôt thématique plus pertinent que *Recherche Data Gouv*.
- Vous avez légalement le droit de diffuser les données :
 - Accord des co-auteurs.
 - Respect des contrats et réglementations.
 - Vérification de la présence de données sensibles.

Authentification

- Avoir un compte sur *Recherche Data Gouv* (utiliser l'authentification institutionnelle).
- Avoir les droits de dépôts dans la collection la plus pertinente au dépôt :
Au besoin, demander les droits aux administrateurs de la collection via le bouton contact de la collection.

■ Prérequis

Préparer les données

- Suivre les [recommandations du guide de Recherche Data Gouv](#) sur la documentation, nommage, structure, fichiers, etc.

Métadonnées de description

- Lister les auteurs et contributeurs avec : affiliation, identifiant (ex. orcid), courriel.
- Identifier le nom du producteur de données (généralement un organisme), et du financeur avec les numéros de financement associés.
- Identifier un thésaurus pour les mots clés.
- Publications, données et codes en lien avec le jeu de données à déposer et qui ont déjà été publiés avec un identifiant unique (par ex. DOI).

■ Prérequis

Fichiers

- Rédiger un fichier de type README ou LISEZMOI expliquant le contenu du jeu de données et d'éventuelles consignes d'utilisation.
- Prévoir si nécessaire un dictionnaire de données décrivant les variables mesurées.
- Vérifier si les formats sont ouverts ou largement utilisés. Au besoin, chercher des alternatives ouvertes.

Licence et conditions de réutilisation

- Avoir défini des conditions d'accès et les licences d'usage.
- Déterminer la nécessité de restreindre l'accès à certains fichiers.



Démonstration.

3. La modération et la publication

Modération à l'UGA

Coordonné par la Cellule Data Grenoble Alpes (CDGA)

Personnes impliquées :

- 7 personnes pour la modération au sein de la CDGA.
- 9 modérateurs pour les collections déjà créées à la demande de certains laboratoires.

Consignes de modération : voir le guide du curateur de *Recherche Data Gouv.*

Modération préalable à la publication des jeux déposés et avec la rédaction d'un rapport de curation à destination du déposant.

Validation et publication du dépôt.

Rapport de curation



	OK	!A
<p>Title : Titre complet le plus informatif possible Commentaire :</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Alternative title : Présence d'un titre en anglais si le titre principal est en français Commentaire :</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Link to data : Adresse URL où les données sont accessibles si aucun fichier de données n'est associé au jeu de données. Commentaire :</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Contact : Courriel conforme Commentaire :</p>	<input type="checkbox"/>	<input type="checkbox"/>
Author		
<ul style="list-style-type: none"> Name : Nom, Prénom ou Nom de l'organisme Commentaire : 	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> Affiliation : Acronyme du laboratoire; Tutelle 1, Tutelle 2...; Pays ou Tutelle(s) pour un organisme, si pertinent Commentaire : 	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> Identifiant : L'usage de l'identifiant <u>ORCID</u> est recommandé. Si l'identifiant est renseigné, vérifier que le type d'identifiant est également renseigné Commentaire : 	<input type="checkbox"/>	<input type="checkbox"/>
Contributor		
<ul style="list-style-type: none"> Name : Nom, Prénom ou Nom de l'organisme Commentaire : 	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> Affiliation : Acronyme du laboratoire; Tutelle 1, Tutelle 2...; Pays ou Tutelle(s) pour un organisme, si pertinent Commentaire : 	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> Identifiant : L'usage de l'identifiant <u>ORCID</u> est recommandé. Si l'identifiant est renseigné, vérifier que le type d'identifiant est également renseigné Commentaire : 	<input type="checkbox"/>	<input type="checkbox"/>
Producer		
<ul style="list-style-type: none"> Name : Nom de l'organisme, ou Nom, Prénom de la personne si à titre privé, portant la responsabilité administrative ou financière 	<input type="checkbox"/>	<input type="checkbox"/>

Rapport de curation



Science ouverte
Grenoble Alpes

FICHIERS		
	OK	!A
• Vérifier que les noms de fichiers respectent les bonnes pratiques mentionnées dans le guide de saisie.	<input type="checkbox"/>	<input type="checkbox"/>
• Vérifier si une convention de nommage des fichiers est documentée, par exemple dans un <u>README</u> , et si elle est respectée. Si non, le suggérer au déposant.	<input type="checkbox"/>	<input type="checkbox"/>
• Vérifier que les formats de fichiers sont ouverts ou largement utilisés.	<input type="checkbox"/>	<input type="checkbox"/>
• Vérifier la bonne ingestion des fichiers tabulés (pas de format <u>csv</u> avec points-virgules, erreurs repérables lors de la prévisualisation du fichier, cohérence du nombre de variables et d'observations).	<input type="checkbox"/>	<input type="checkbox"/>
• Vérifier la présence d'un dictionnaire de données décrivant les variables mesurées (par exemple dans un <u>README</u>) si nécessaire.	<input type="checkbox"/>	<input type="checkbox"/>
• Vérifier la présence d'une documentation (par exemple dans un <u>README</u>).	<input type="checkbox"/>	<input type="checkbox"/>
• Indiquer ci-dessous les fichiers pour lesquels des corrections sont à apporter.	<input type="checkbox"/>	<input type="checkbox"/>

- o Nom de fichier :
- o Commentaire :

CONDITIONS		
	OK	!A
<p>Licence : L'usage de la Licence Ouverte <u>Etalab</u> 2.0 est recommandée par défaut, se reporter à la rubrique du guide. Préciser les Conditions d'utilisation du jeu de données. Cependant, si des données sont réutilisées, vérifier si cette licence est compatible avec les licences de réutilisation attribuées aux données sources. Voir les recommandations DoRANum.</p> <p>Commentaire :</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Si le jeu de données comprend à la fois des fichiers de données et du code, une licence doit être attribuée au code, décrite dans un fichier, <u>Licence.txt</u> ou <u>Licence.md</u> par exemple, déposé dans le même répertoire que le code.</p> <p>Commentaire :</p>	<input type="checkbox"/>	<input type="checkbox"/>

4. Liens publications, données, codes

Liens publications et codes

- Déposer les publications dans HAL.
- Déposer les données dans un entrepôt de confiance, comme *Recherche Data Gouv.*
- Archiver les codes et les logiciels sur Software Heritage et les décrire dans HAL.



5. Ressources

Ressources

- **Bac à sable** de *Recherche Data Gouv* :
<https://demo.recherche.data.gouv.fr>
- **Espace institutionnel UGA** (Data Repository Grenoble Alpes) :
<https://entrepot.recherche.data.gouv.fr/dataverse/univ-grenoble-alpes>
- Adresse support de la **Cellule Data Grenoble Alpes** :
sos-data@univ-grenoble-alpes.fr
- **Site de la Science Ouverte de l'UGA** :
<https://scienceouverte.univ-grenoble-alpes.fr>



Merci pour votre écoute !
-
Des questions ?

Cellule Data Grenoble Alpes :
sos-data@univ-grenoble-alpes.fr