



HAL
open science

Découverte des APIs de l'entrepôt Recherche Data Gouv

Camille Desiles, Virgile Jarrige, Damien Sans, Dimitri Szabo

► **To cite this version:**

Camille Desiles, Virgile Jarrige, Damien Sans, Dimitri Szabo. Découverte des APIs de l'entrepôt Recherche Data Gouv. Printemps de la Donnée 2024, 2024, Paris, France. hal-04702528

HAL Id: hal-04702528

<https://hal.inrae.fr/hal-04702528v1>

Submitted on 19 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Bonjour et bienvenue !

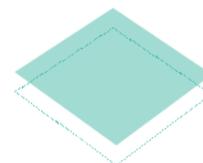
Cette classe virtuelle démarrera
à XX h

Le support est téléchargeable ici :



Découverte des APIs de l'entrepôt Recherche Data Gouv

Sessions 2024



Centre de ressources
entrepôt-catalogue
 Recherche Data Gouv



DANS CETTE PRÉSENTATION

- PRÉSENTATION GÉNÉRALE
- LES CONCEPTS CLÉS
- OUTILS COMPLÉMENTAIRES
- LES APIS



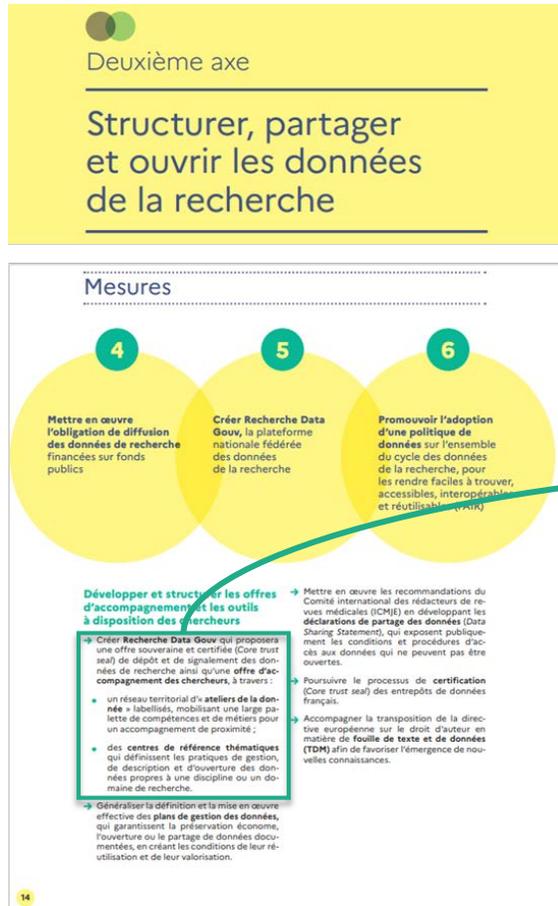
RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

PRÉSENTATION DU PROJET NATIONAL RECHERCHE DATA GOUV



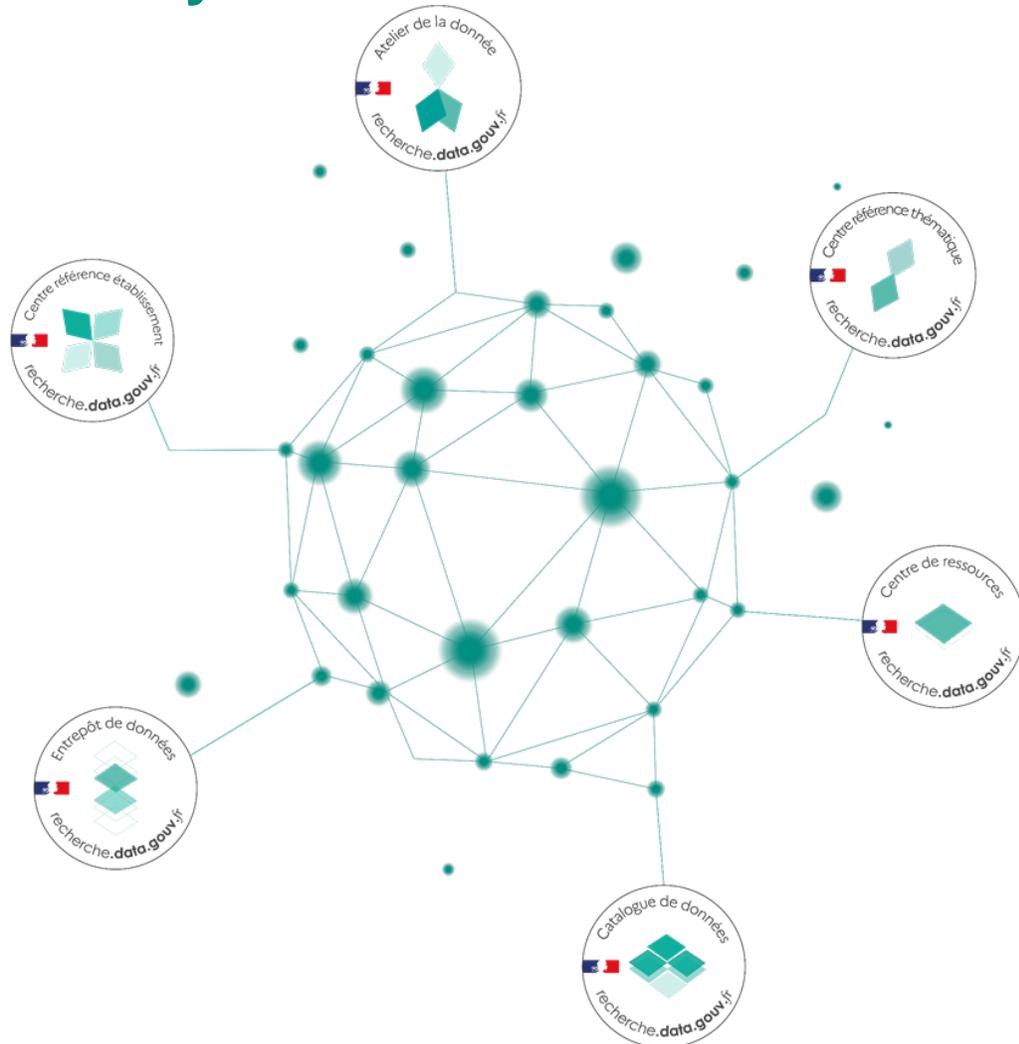
Un projet national



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.

L'écosystème Recherche Data Gouv



4 Modules d'accompagnement :

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

2 Modules de partage :

- Entrepôt
- Catalogue (à venir)



L'entrepôt



Environnement de production

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

Environnement bac-à-sable

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

- Documenter et partager tout type de données de recherche ;
- Rechercher, découvrir, accéder et réutiliser ;
- Toutes fonctionnalités accessibles **via interface web** ou **API**.



RÉPUBLIQUE
FRANÇAISE recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

LES CONCEPTS CLÉS DE L'APPLICATION DATAVERSE



@RechercheDataGv

Dataverse, moteur de l'entrepôt et du catalogue

116 Installations



Source : <https://dataverse.org/> - 11/04/2024

The
Dataverse
Project



Projet Open Source



Conçu par Harvard

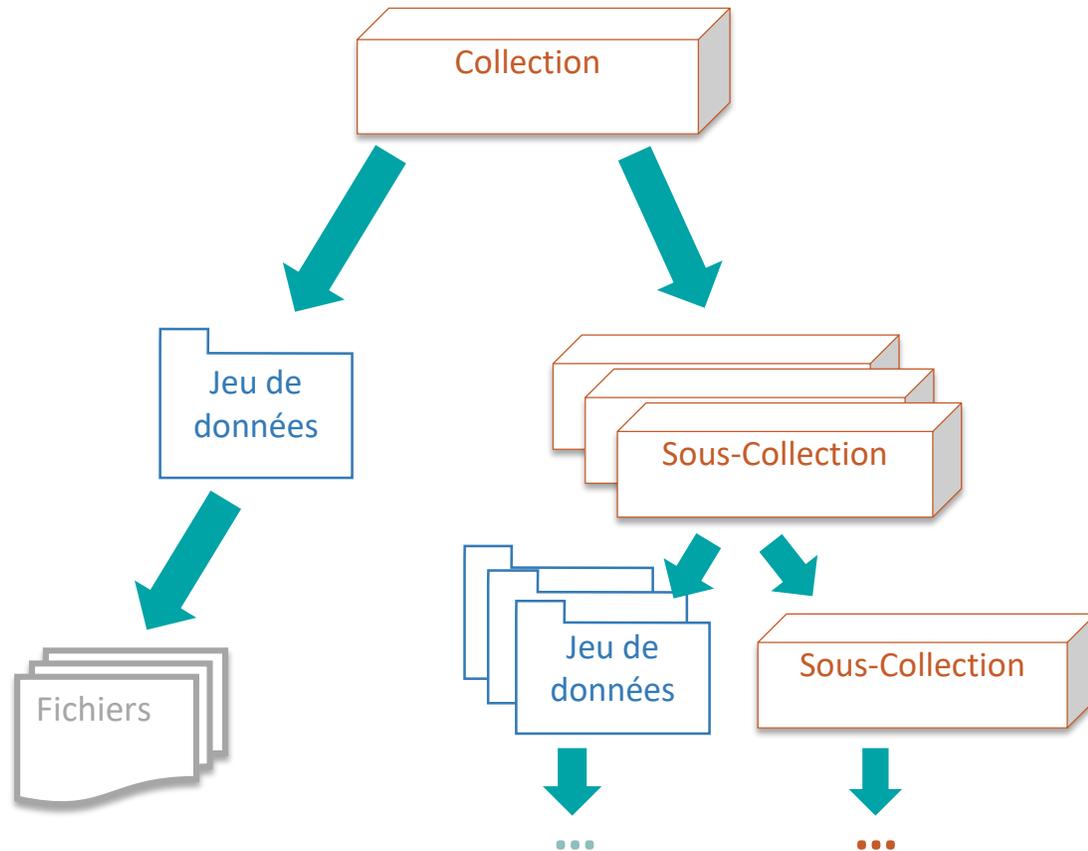


34 Institutions (18 pays)

dans la gouvernance

(dont INRAE et le CNRS)

Les objets



- **collection** : contient des jeux de données et/ou des sous-collections
- **jeu de données** : rassemble les fichiers et la documentation
- **fichier** : Un **fichier** de données ou de documentation (e.g. README)



Les métadonnées en bref

Identifiant pérenne ?	doi:10.57745/PM55G5
Date de publication ?	2023-01-26
Titre ?	One-step synthesis of spin crossover nanoparticles using flow chemistry and supercritical CO2
Point de contact ?	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact. TOULIN, Stéphane (CNRS)
Auteur ?	Daro, Nathalie (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0002-3707-9686 Chastanet, Guillaume (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0001-6829-4066 Marre, Samuel (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0001-8889-187X Aymonier, Cyril (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0003-1775-0716 Vaudel, Tony (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) Afindouli, Luc (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France)
Contributeur ?	Data Collector : TOULIN, Stéphane (Univ. Bordeaux, CNRS, Bordeaux INP, ICMCB, UMR 5026, F-33600 Pessac, France) - ORCID: 0000-0002-2087-7403
Producteur ?	Centre National de la Recherche Scientifique (CNRS) https://www.cnrs.fr/  Université de Bordeaux (UBX) https://www.u-bordeaux.fr/  Bordeaux INP https://www.bordeaux-inp.fr/ 
Localisation de la production ?	Pessac
Distributeur ?	Entrepôt-Catalogue Recherche Data Gouv
Description ?	Data set and images from - Publication : Daro N., Vaudel T., Afindouli L., Marre S., Aymonier C. and Chastanet G. One-step synthesis of spin crossover nanoparticles using flow chemistry and supercritical CO2. Chemistry - European Journal 2020 vol.26, n°69, p.16286-16290. DOI : 10.1002/chem.202002322 / HAL : https://hal.science/hal-02945200

- Données décrivant les données
- Facilite la découverte
- Pour **comprendre** les données et les **conditions de production**

Blocs de métadonnées

- Métadonnées générales :
 - 6 sont **obligatoires** : **Titre, Auteur Nom, Point de contact Courriel, Description Texte, Sujet, Type de données**
 - liens avec d'autres objets (**Lien vers les données, Autre identifiant, Publication associée, Jeu de données associé**)
 - etc.
- Métadonnées spécifiques :
 - *Métadonnées géospatiales*
 - *Métadonnées de sciences humaines et sociales*
 - *Métadonnées de science de la vie*
 - *Métadonnées de revue*
 - *Métadonnées de ressources sémantiques*
 - *Métadonnées d'astronomie et astrophysique*
 - *Métadonnées de workflow de calcul*

Conformes aux standards
DDI , Dublin Core,
DataCite, ISA-TAB



RÉPUBLIQUE
FRANÇAISE

recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

LES OUTILS COMPLÉMENTAIRES



@RechercheDataGv

Pour les site webs : les widgets

Widget : Window gadget

- Permet d'afficher des éléments de Recherche Data Gouv dans un site web
 - **Contenu d'une collection**
 - **Recherche dans une collection**
 - **Contenu d'un jeu de données**
 - **Citation d'un jeu de données**
- Code à ajouter simplement dans une page web
- Récupérables sur l'interface dans le jeu de données ou la collection

Les widgets

La collection doit être publiée pour accéder aux widgets.

Thème Widgets

ⓘ Copier et coller ce code dans le code HTML de votre site web. Pour apprendre davantage sur les widgets, consulter la section [Thème et widgets](#) du guide d'utilisation.

Boîte de recherche dans l'entrepôt-catalogue Recherche Data Gouv
Permet aux visiteurs de votre site Web d'effectuer une recherche dans la collection.

```
<script src="https://entrepot.recherche.data.gouv.fr/resources/js/alias=espacegenerique&dvUrl=https://entrepot.recherche.data.gouv.fr/entrepot-catalogue"></script>
```

Search my dataverse... Find

Liste des collections
Permet aux visiteurs de votre site Web de voir vos collections et jeux de données, de les trier ou encore de les parcourir.

```
<script src="https://entrepot.recherche.data.gouv.fr/resources/js/alias=espacegenerique&dvUrl=https://entrepot.recherche.data.gouv.fr/entrepot-catalogue"></script>
```

Terminé

Experimental - Observation - Simulation Dataverse Omics Dataverse Surveys & Texts Dataverse Genetic Resources Dataverse

Search this dataverse... Find Advanced Search

1 to 10 of 83,719 Results

Publication Status
Published (78,183)
Draft (5,521)
Unpublished (5,499)
Deaccessioned (3)

Potential environmental impacts of water deprivation and land use for food consumption in France and Tunisia
Amiot-Carlin, Marie Josephe, 2019, "Potential environmental impacts of water deprivation and land use for food consumption in France and Tunisia", <https://doi.org/10.15454/9VWITQ>, Portal Data Inra, DRAFT VERSION

The dataset concerns the environmental impacts of water consumption and land use for the production of food and beverage ingredients used for human consumption in France and in Tunisia. The origin of ingredients are considered to

Data is stored at Portal Data Inra. Powered by **The Dataverse Project**

Contact Partager Lier Modifier

Espace Générique
espacegenerique

Renseignements généraux
Thème + widgets
Autorisations
Groupes
Modèles de jeux de données
Registre des visiteurs du jeu de données
Collections en vedette

Les widgets 1/4 : Contenu d'une collection

Recherche Data Gouv >

Chercher dans cette collection...  Recherche avancée

 Collections (2)

 Jeux de données (66)

 Fichiers (4 229)

Catégorie de collection
Organisation ou établissement (1)

Année de publication
2023 (22)
2022 (15)

Statut de publication
Version provisoire (36)
Publié (31)
Non publié (31)
En révision (5)

1 à 10 de 68 résultats ↑↓ Tri ▾

Public research data from ebi-umr 7267 (CRNS, Université de Poitiers) 
18 juil. 2023

Responses of the hyper-diverse community of canopy-dwelling Hymenoptera to oak decline Version provisoire Non publiée 
10 juil. 2023

 Le Souchu, Elodie; Cours, Jérémy; Sallé, Aurélien, 2023, "Responses of the hyper-diverse community of canopy-dwelling Hymenoptera to oak decline", <https://doi.org/10.57745/H8CBG1>, Recherche Data Gouv, VERSION PROVISOIRE

This dataset contains the R scripts and both the list of Hymenoptera species traits and the list of plot and stand characteristics used for the paper

Les données sont archivées par : Recherche Data Gouv.

Fourni par  The Dataverse Project  Hymenoptera

Les widgets 2/4 : Recherche dans une collection

Les widgets 3/4 : Contenu d'un jeu de données

Recherche Data Gouv > Espace Générique >

Données de réplication pour : Trait-environment associations diverge between native and alien breeding bird assemblages on the world's oceanic islands.

Version 1.0



RAULT, Charlotte; LEPRIEUR, Fabien; BARBARO, Luc; KREFT, Holger; MOUQUET, Nicolas; PAPAIX, Julien; SEKERCIOGLU, CAGAN; VIOLLE, Cyrille; BARNAGAUD, Jean-Yves, 2023, "Données de réplication pour : Trait-environment associations diverge between native and alien breeding bird assemblages on the world's oceanic islands.", <https://doi.org/10.57745/1BRHFC>, Recherche Data Gouv, V1, UNF:6:kTrGqusY0zmC5asi9vB0Eg== [fileUNF]

Citer le jeu de données ▾

Pour en apprendre davantage sur le sujet, consulter le document Data Citation Standards [en].

Les données sont archivées par : Recherche Data Gouv.

Fourni par  The Dataverse®
Project



@RechercheDataGv

Les widgets 4/4 : Citation d'un jeu de données



RAULT, Charlotte; LEPRIEUR, Fabien; BARBARO, Luc; KREFT, Holger; MOUQUET, Nicolas; PAPAIX, Julien; SEKERCIOGLU, CAGAN; VIOLLE, Cyrille; BARNAGAUD, Jean-Yves, 2023, "Données de réplication pour : Trait-environment associations diverge between native and alien breeding bird assemblages on the world's oceanic islands.", <https://doi.org/10.57745/1B>
RHFC: Recherche Data Gouv V1 UNF:6:kTrGousY0zmC5asl9vB0Fq== [file] INF

Les données sont archivées par : Recherche Data Gouv.

Fourni par  The
Dataverse
Project

Pour déposer des fichiers : DVUploader

Outil de dépôt de fichiers en ligne de commande.

Permet de téléverser :

- un grand nombre de fichiers (>1000)
- à partir d'un répertoire de dossier sous-dossiers
- de façon récursive, en reproduisant l'arborescence

[Plus d'informations](#)

Pour explorer les données : prévisualisation & outils externes

- Prévisualisations existantes

Tableaux (.csv, .tab*, .tsv) Texte brut (.fasta, .fastq, .txt,) Audio (.mp3, .mpeg, .ogg, .wav) Markdown (.md)	Image (.gif, .jpeg, .png) Video (.mp4, .ogg, .quicktime) Page web (.html)	PDF (.pdf) Archive (.zip)
---	---	------------------------------

- Condition d'ajout

- **Accessible via une webapp**
- **Récupération des informations via une URL de Dataverse**
- **Plus d'informations**

- Les outils externes (Data Explorer, JupyterHub, Binder)

- **Seul Data Explorer (tableaux) installé**
- **Documentation**



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

LES APIS : POUR QUOI FAIRE ?



@RechercheDataGv

Les différentes APIs

- Search API : Recherche
- Native API : Fonctionnalités principales
- ⚠ Direct Upload/Replace : Téléversement des fichiers
- Data Access API : Téléchargement de fichiers
- Metrics API : Statistiques d'utilisation
- SWORD API : Dépôt de fichiers et jeux de données
- MyData API ⚠ non documentée

Rechercher des jeux de données

API = Search ou MyData

Cas d'usage :

- Recherche simple ou avancée, avec ou sans authentification
- Récupérer les données d'un utilisateur (MyData)

Points importants à l'usage :

- Pour rechercher sur des métadonnées spécifiques, voir le schéma [diapo 13](#)
- Peu de métadonnées renvoyées : si besoin, utiliser l'API Native ensuite
- Attention à la pagination : tous les résultats ne sont pas renvoyés

Gérer une collection

API = Native

Cas d'usage :

- Créer, modifier et publier une collection
- Gérer les informations, droits et autres possibilités du portail
- Supprimer la collection
- Récupérer les données des Registres des visiteurs

Point d'attention à l'usage :

- Authentification généralement nécessaire
- CORS (certaines commandes non utilisables en frontend web)

Gérer un jeu de données

API = Native ou SWORD

Cas d'usage

- Créer, modifier et publier un jeu de données
- Gérer les métadonnées, droits et autres possibilités du portail
- Supprimer le jeu de données

Point d'attention à l'usage :

- Authentification généralement nécessaire
- CORS (certaines commandes non utilisables en frontend web)

Gestion des fichiers

API = Native ou Data Access

Cas d'usage

- Créer, remplacer, modifier et supprimer des fichiers
- Administrer (et demander) l'accès
- Télécharger des fichiers ou certaines de leurs variables

Point d'attention à l'usage :

- Authentification généralement nécessaire
- CORS (certaines commandes non utilisables en frontend web)

Métriques

API = Metrics

Cas d'usage

- Récupérer des métriques sur le portail (créations, téléchargements...)
- Récupérer des métriques pour une collection
- Récupérer des métriques d'utilisation pour un jeu de données

Point d'attention à l'usage :

- Données de bac à sable pas forcément renseignées

API MyData : Documentation

- Endpoint : {serveur}/api/mydata/retrieve
- Paramètres :
 - **Key = Jeton de l'utilisateur; pour cette API il ne doit pas être passé dans le header**
 - **role_ids = Rôles de l'utilisateur, plusieurs valeurs possibles parmi :**
 - 1 = Admin
 - 2 = File Downloader
 - 3 = Dataverse + Dataset Creator
 - 4 = Dataverse Creator
 - 5 = Dataset Creator
 - 6 = Contributor
 - 7 = Curator
 - 8 = Member
 - **dvobject_types = Type d'objet, plusieurs valeurs possibles parmi :**
 - DataFile
 - Dataset
 - Dataverse
 - **published_states = Etat de l'objet, plusieurs valeurs possibles parmi :**
 - Published
 - Unpublished
 - Draft
 - Deaccessioned
 - In+Review
 - **per_page = Nombre de résultats renvoyés par page**



RÉPUBLIQUE
FRANÇAISE

recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

LES APIs : QU'EST-CE QUE C'EST ?



@RechercheDataGv

API

Application Programming Interface

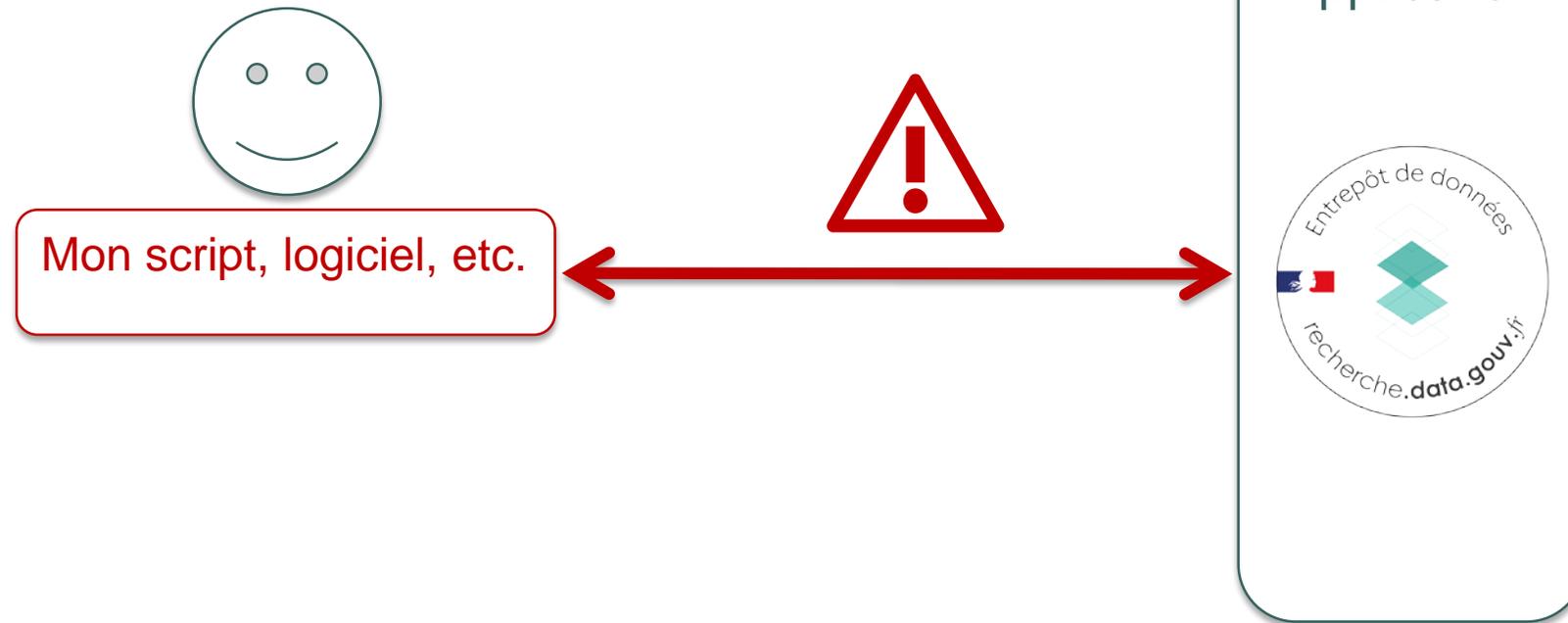


Mon script, logiciel, etc.



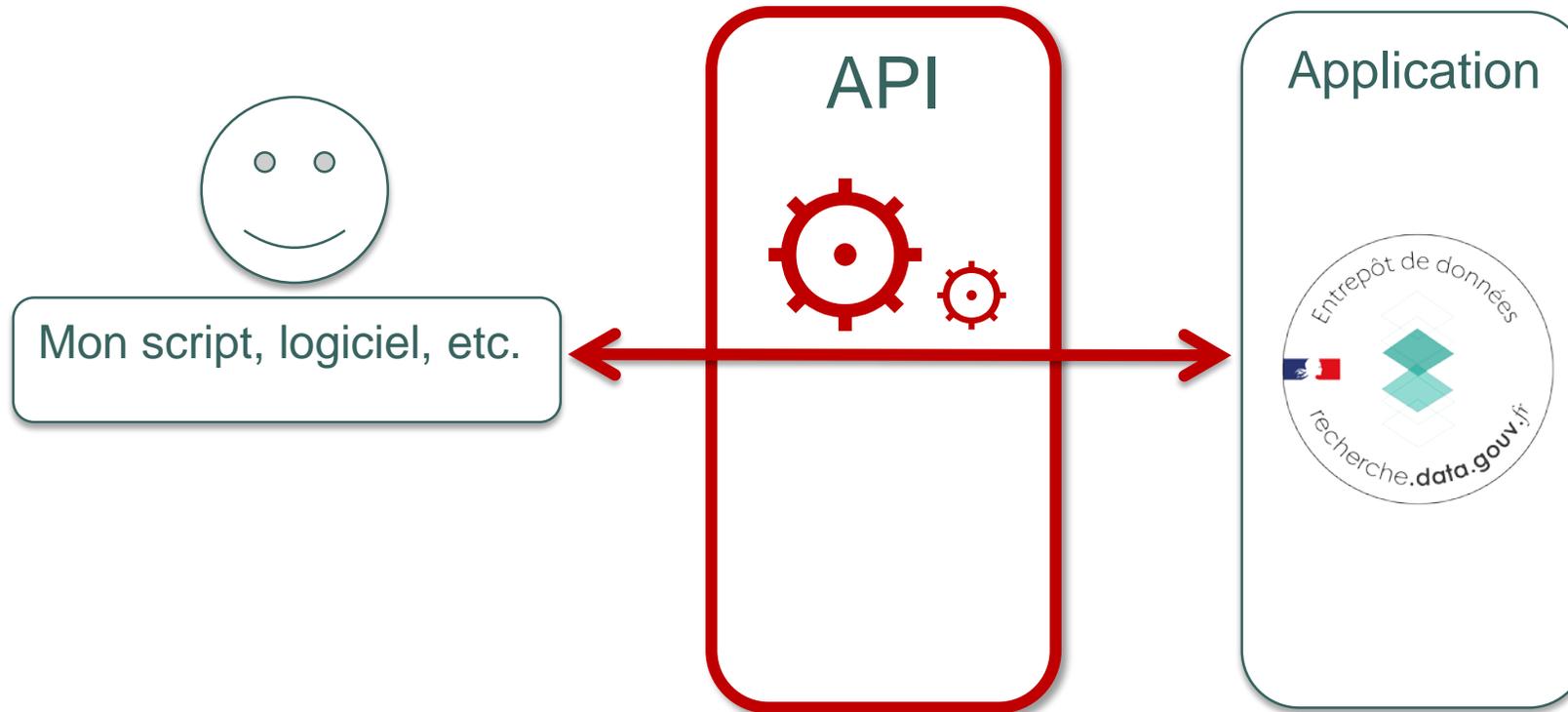
API

Application Programming Interface



API

Application Programming Interface



Analogie

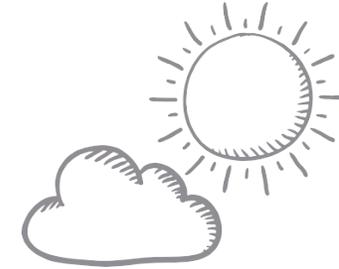
Un **annuaire téléphonique** permet à un **humain**, de demander une information à un autre **humain**, par **téléphone**.

Une **API**, permet à un **ordinateur** de demander une information à un **autre ordinateur**, par **internet**.

Source : api.gouv.fr, consultation le 18/10/2021

Votre premier appel API

Vous l'avez fait sans le savoir !



Votre premier appel API vers l'entrepôt

Dans votre navigateur, tapez :

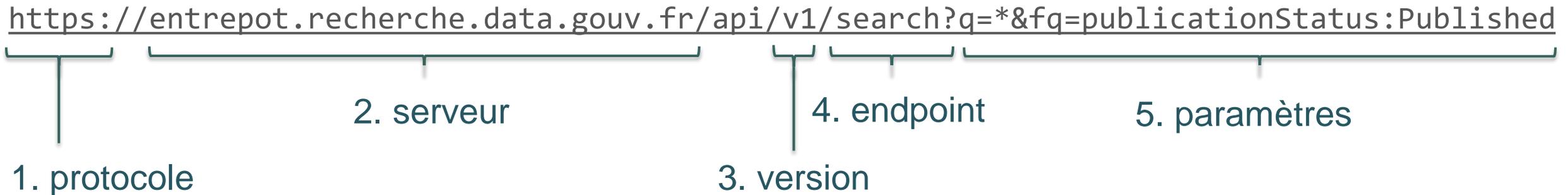
<https://entrepot.recherche.data.gouv.fr/api/metadatablocks>

et voilà !

Bonus : Si votre navigateur affiche plusieurs onglets, ouvrir l'onglet "Raw Data"

{json}

Un peu de vocabulaire



1. Protocole : par quoi on échange
2. Serveur : avec qui/quoi on échange
3. Version : permet d'utiliser l'ancienne version d'une API si elle change
4. Endpoint : point d'entrée, quelle commande on effectue
5. Paramètres : avec quelles spécificités

Quelques commandes supplémentaires

Pour obtenir le schéma du bloc de métadonnées Générales :

<https://entrepot.recherche.data.gouv.fr/api/metadatablocks/citation>

Tester de réaliser l'appel avec un autre bloc :

<https://entrepot.recherche.data.gouv.fr/api/metadatablocks/{bloc}>

Tester une recherche :

<https://entrepot.recherche.data.gouv.fr/api/search?q=banana>



Tester les paramètres pour une recherche

Tester une nouvelle recherche :

<https://entrepot.recherche.data.gouv.fr/api/search?q=tree>

Filtrer sur un champ de métadonnées :

<https://entrepot.recherche.data.gouv.fr/api/search?q=tree&fq=title:tree>

Limiter le nombre de résultats par page :

https://entrepot.recherche.data.gouv.fr/api/search?per_page=1&q=tree&fq=title:tree

Les méthodes :

- Verbes HTTP correspondant à 4 opérations (CRUD)

Opération	Traduction	Verbe
Create	Créer	POST
Read	Afficher	GET
Update	Mettre à jour	PUT
Delete	Supprimer	DELETE

- Les exemples précédents sont en réalité des GET implicites :

GET

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



RÉPUBLIQUE
FRANÇAISE

recherche.data.gouv.fr

*Liberté
Égalité
Fraternité*

LES APIS : EN PRATIQUE



@RechercheDataGv

Les bibliothèques

Des bibliothèques dans différents langages de programmation pour construire des programmes.

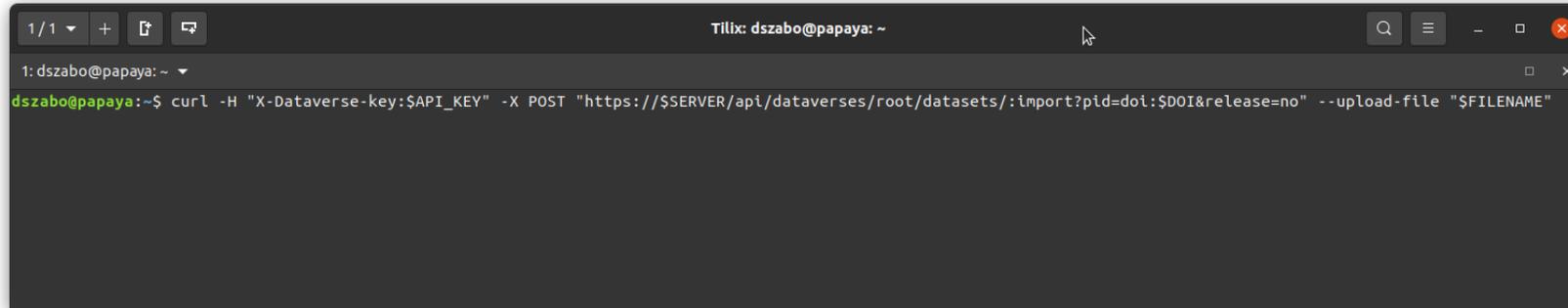
- R : atom4R, dataverse
- Python : pyDataverse
- Java
- Javascript
- Ruby

Si le client n'existe/ne convient pas, utiliser des bibliothèques génériques comme « curl » !

Le Plus : intégration

✓ Correspond le plus souvent à l'usage final

Le terminal



```
Tilix: dszabo@papaya: ~  
1: dszabo@papaya: ~  
dszabo@papaya:~$ curl -H "X-Dataverse-key:$API_KEY" -X POST "https://$SERVER/api/dataverses/root/datasets/:import?pid=doi:$DOI&release=no" --upload-file "$FILENAME"
```

- Prérequis : curl
- Comment y accéder (Windows) : chercher “Invite de commande” ou Win + r, taper “cmd” puis entrée

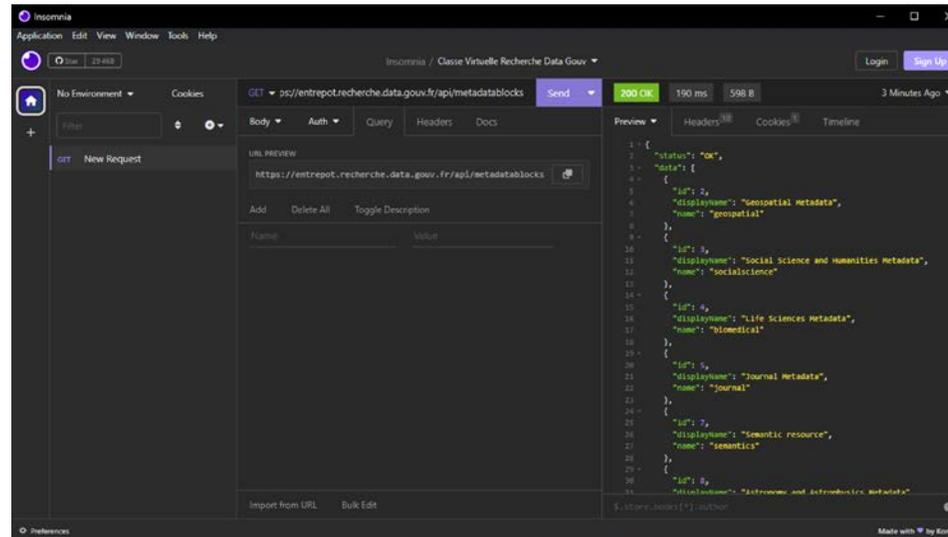
Le Plus : découvrir

- ✓ Correspond à la documentation Dataverse

Le Moins : praticité

- Connaissance console et autres outils (ex. jq) vite nécessaire

Les Clients API



**Nécessaire pour la
classe virtuelle
« Utilisation des APIs »**

Plateforme d'APIs. Permet des requêtes plus modulaires qu'avec la console

- Outil à installer depuis <https://insomnia.rest/download> par exemple

Le Plus : User Friendly

- ✓ Distinction et configuration des variables, environnements, sauvegarde...

Le Moins : automatisation

- Outil séparé d'une potentielle chaîne de traitement

Commandes terminal : pour aller plus loin

Utiliser des variables...

- sous Windows :

```
SET SERVER=https://demo.recherche.data.gouv.fr  
curl %SERVER%/api/metadatablocks
```

- sous Linux ou macOS :

```
SERVER=https://demo.recherche.data.gouv.fr  
curl $SERVER/api/metadatablocks
```

Sauvegarder le résultat :

```
curl https://demo.recherche.data.gouv.fr/api/metadatablocks -o {dir}/resultat.json
```

Récupérer les valeurs contrôlées des métadonnées :

```
curl https://nextcloud.inrae.fr/s/DG4SnAGB2JAgkNw/download/datainraeControlledVocabularies.json  
-o {dir}/controlledValues.json
```

Commandes terminal : au secours, ma commande ne marche pas !

Quelques solutions courantes testables :

- Mettre entre guillemets les URLs
- Que disent l'erreur et la documentation ?
- Si variables, tester le texte de la commande avec : `echo {la_commande}`
- Vérifier avec une durée plus importante entre appels APIs



Les autres formations et guides Recherche Data Gouv

- [Guide DVUploader](#)
- [Formation : Utiliser les APIs Recherche Data Gouv](#)

La documentation des APIs

Guide Dataverse :

<https://guides.dataverse.org/en/5.14/api/>

⚠ bien préciser la version de Dataverse utilisée.

Les différentes APIs y sont documentées séparément
(cf. section “Pour quoi faire”)

Comment tester ?

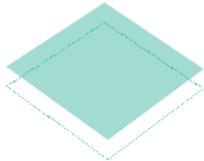
- Pour les tests et d'autant plus en écriture :
Utiliser le bac à sable !
- Pour tester la publication, utiliser si possible toujours le même jeu de données
- Si possible, supprimer les tests au fur et à mesure





**Merci
pour votre écoute**

Classe Virtuelle : Découverte des APIs



Centre de ressources
entrepôt-catalogue
 Recherche Data Gov

support-recherchedatagouv@inrae.fr



Support réalisé par : Camille Desiles (CNRS-INIST) ; Virgile Jarrige (Unistra) ; Damien Sans (INRAE) ; Dimitri Szabo (INRAE)

Les institutions membres


MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
*Liberté
Égalité
Fraternité*

En partenariat avec :

