



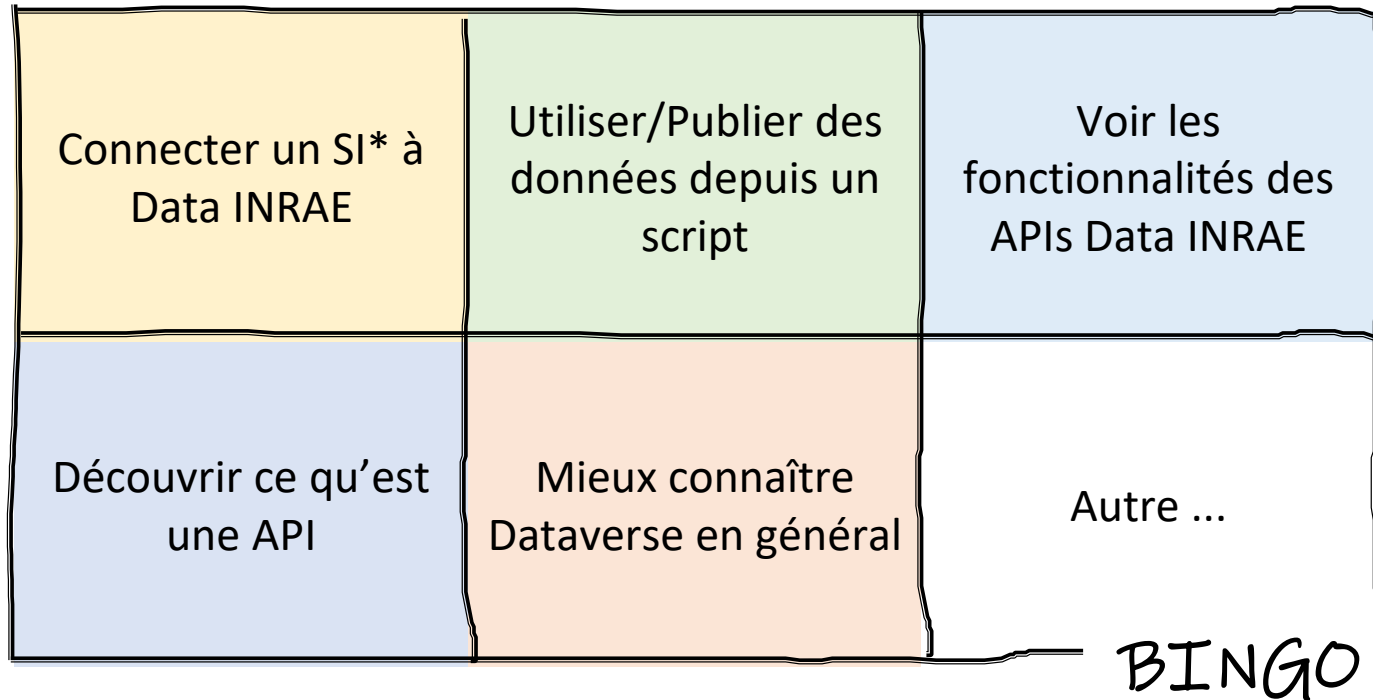
Data INRAE

Introduction aux APIs

DipSO – Pôle Numérique pour la Science
2021

Bingo des besoins

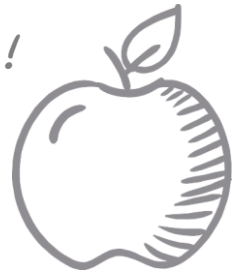
- Je suis...
- Je cherche à...



* Système d'Information

Qu'est-ce qu'une API ?

Facile !



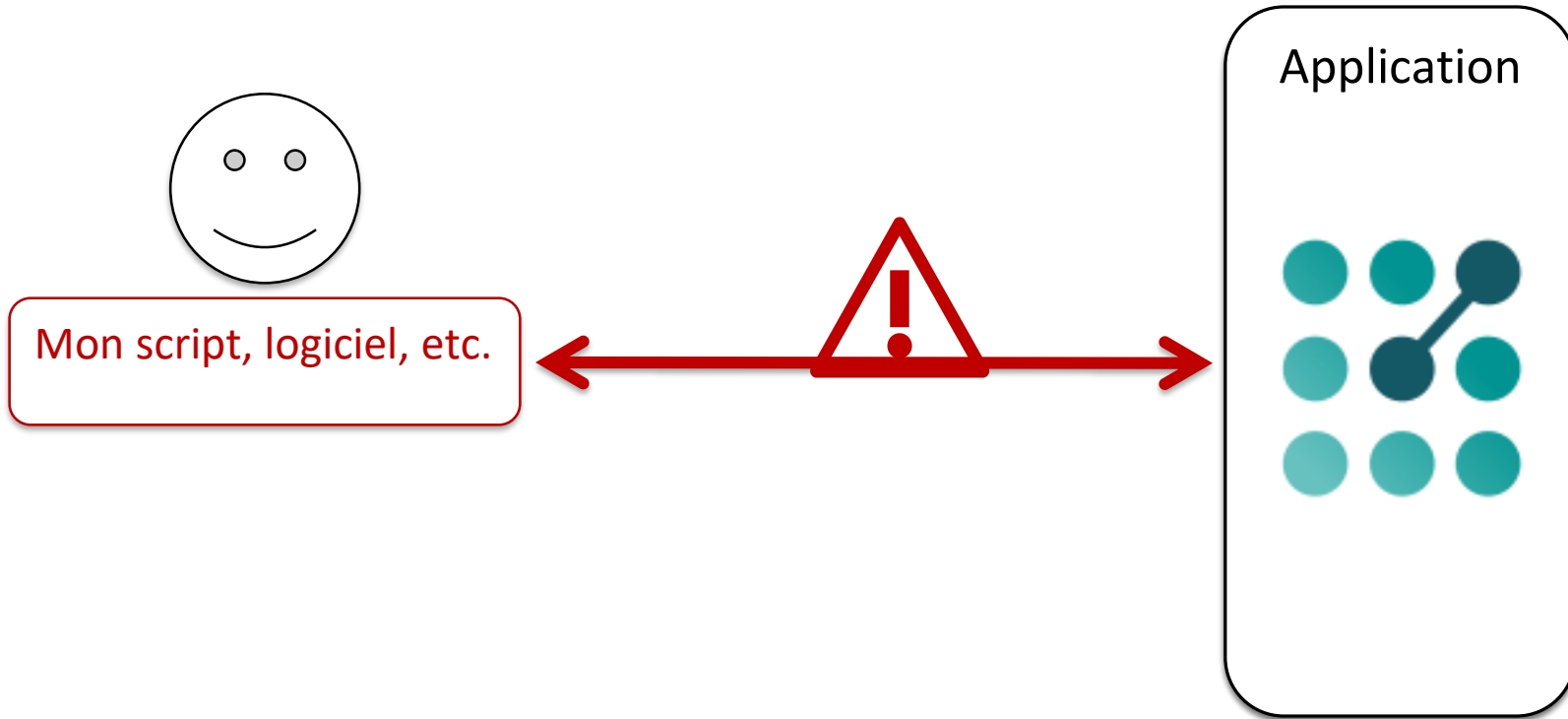
API = **Application** Programming Interface



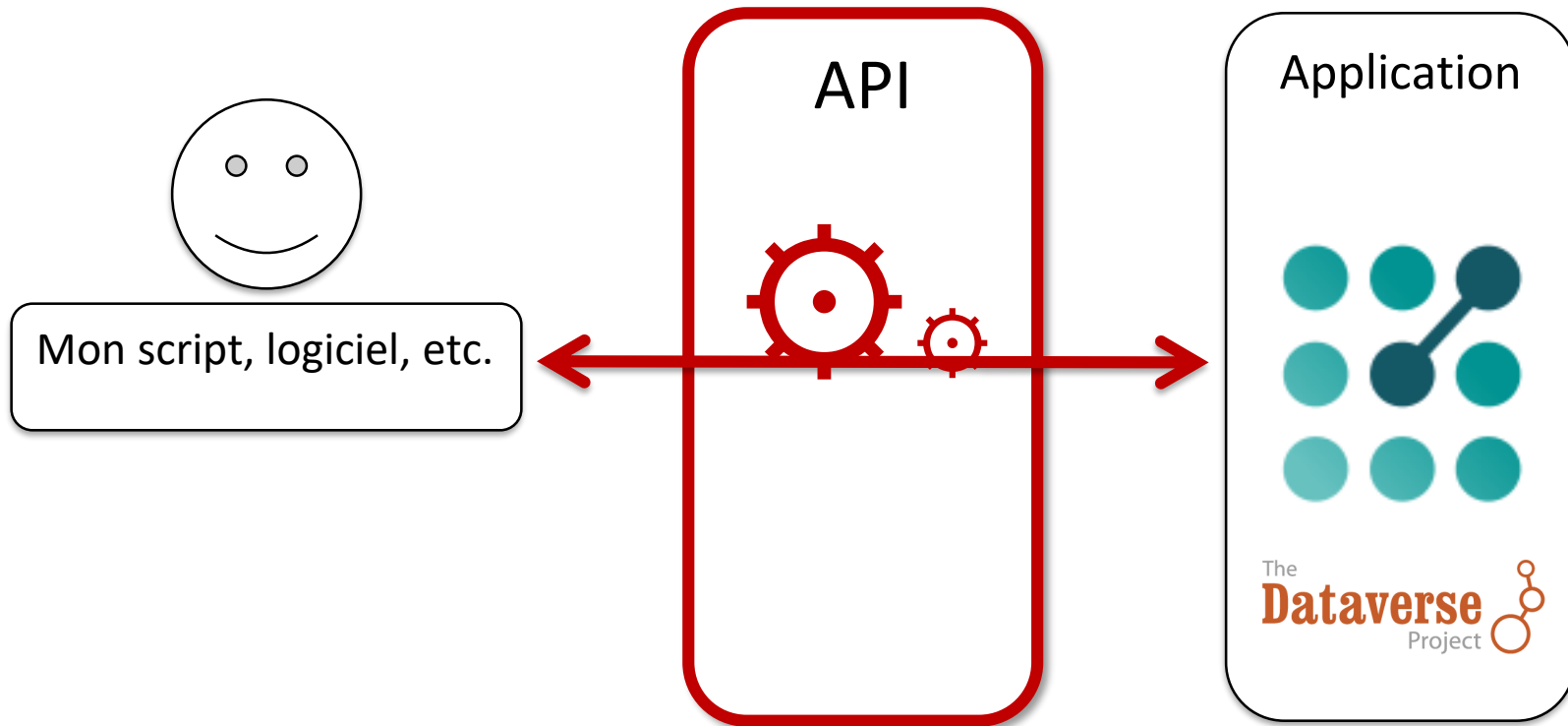
Mon script, logiciel, etc.



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**.

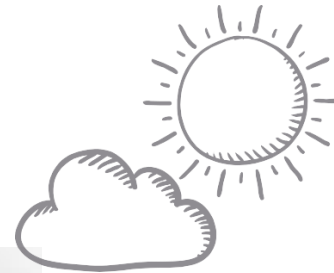
—

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

Dans votre navigateur, tapez :

<https://data.inrae.fr/api/metadatablocks>

et voilà !

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

{json}

Un peu de vocabulaire

`https://data.inrae.fr/api/v1/search?q=* &fq=publicationStatus:Published`



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

Les concepts de Data INRAE

Dataverse ou dataverse ?



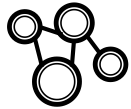
Dataverse, moteur de Data INRAE



Projet Open Source



Conçu par Harvard



Utilisé au CIRAD, IRD, CNRS et internationalement

The
Dataverse
Project 



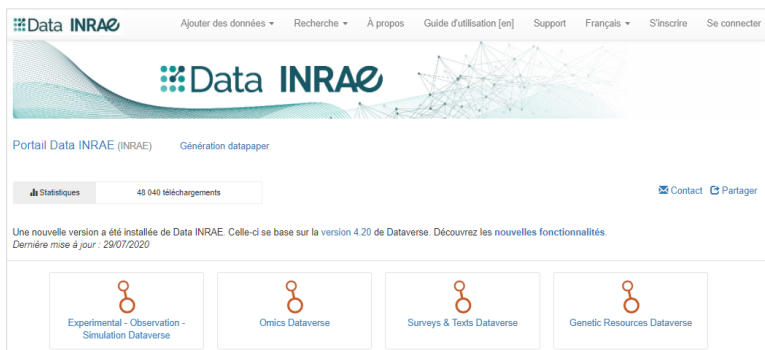
Data INRAE : Rappels

- Facilite la recherche, la découverte, l'accès et la réutilisation des données de l'établissement dans le respect des bonnes pratiques et des règles éthiques et réglementaires en vigueur,
- Périmètre : toute personne ou système d'information INRAE, partenaires dans les UMR et projets.

Deux environnements :

Instance de production

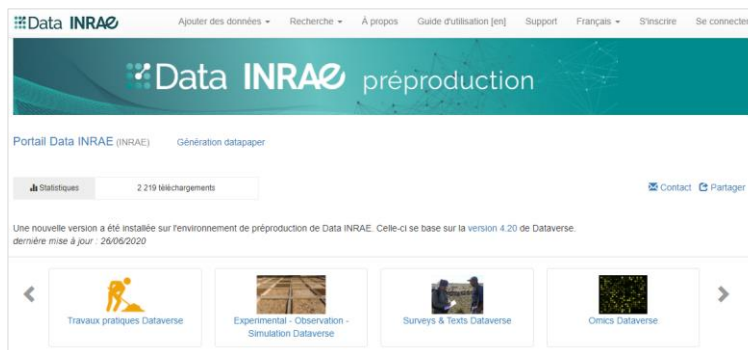
<https://data.inrae.fr>



The screenshot shows the production environment homepage. At the top, there is a navigation bar with the Data INRAE logo and links for 'Ajouter des données', 'Recherche', 'À propos', 'Guide d'utilisation [en]', 'Support', 'Français', 'S'inscrire', and 'Se connecter'. Below the navigation bar is a large banner with the Data INRAE logo and a network diagram. Underneath the banner, there are two tabs: 'Portail Data INRAE (INRAE)' and 'Génération datapaper'. A statistics bar shows '48 040 téléchargements'. There are also links for 'Contact' and 'Partager'. A message at the bottom states: 'Une nouvelle version a été installée de Data INRAE. Celle-ci se base sur la version 4.20 de Dataverse. Découvrez les nouvelles fonctionnalités. Dernière mise à jour : 29/07/2020'. At the bottom, there are four data categories: 'Experimental - Observation - Simulation Dataverse', 'Omics Dataverse', 'Surveys & Texts Dataverse', and 'Genetic Resources Dataverse'.

Instance de pré-production (tests)

<https://data-preproduction.inrae.fr>



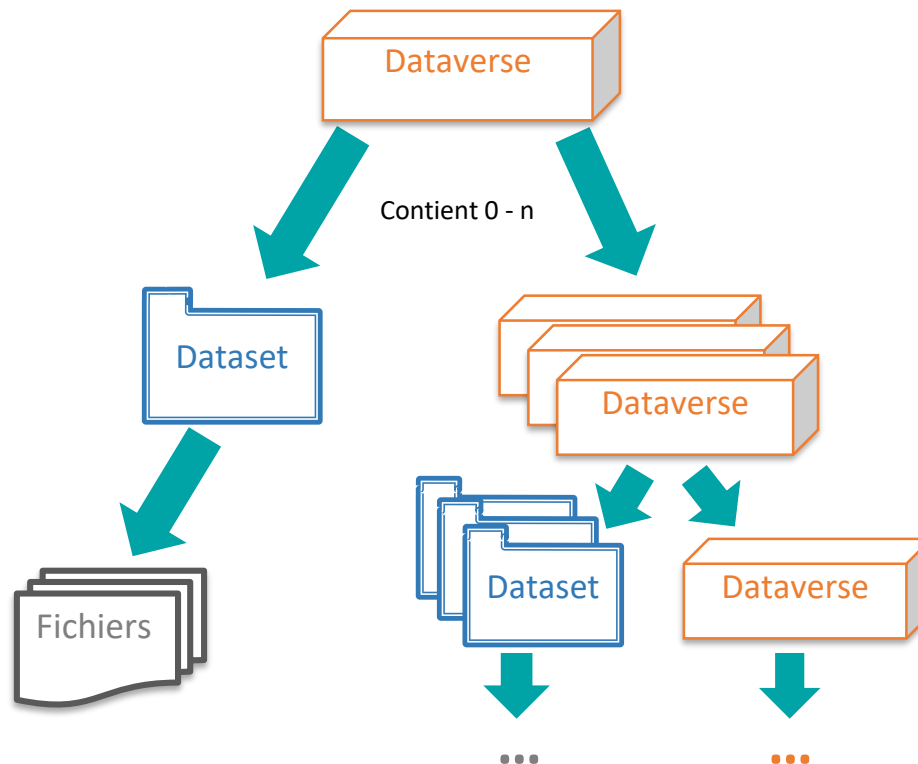
The screenshot shows the pre-production environment homepage. It has a similar layout to the production environment but with a dark teal header and the text 'Data INRAE préproduction'. The statistics bar shows '2 219 téléchargements'. The message at the bottom states: 'Une nouvelle version a été installée sur l'environnement de préproduction de Data INRAE. Celle-ci se base sur la version 4.20 de Dataverse. dernière mise à jour : 26/05/2020'. At the bottom, there are four data categories: 'Travaux pratiques Dataverse', 'Experimental - Observation - Simulation Dataverse', 'Surveys & Texts Dataverse', and 'Omics Dataverse'.

⚠ en intranet :
accès depuis un site ou via le VPN

Les objets dans Dataverse :

dataverse, dataset et fichier

- **fichier** : Un **fichier** de données ou de documentation (ex. README)
- **dataset** : **Jeu de données**, rassemble les fichiers et les métadonnées.
- **dataverse** : **Collection** de datasets et/ou de sous-dataverses



Les métadonnées

- **Common metadata (citation)**
 - génériques, incluent en particulier les métadonnées nécessaires à la citation et les liens avec d'autres objets (Link to data, Other ID, Related Publication, Related Datasets)
 - 5 sont obligatoires : title, nom d'auteur, email du contact, texte de description, subject
 - compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, Dublin Core's DCMI Metadata Terms
- **Geospatial Metadata (geospatial)**
 - compliant with DDI Lite, DDI 2.5 Codebook, DataCite, Dublin Core
- **Social Science and Humanities Metadata (socialscience)**
 - compliant with DDI Lite, DDI 2.5 Codebook, Dublin Core.
- **Life Sciences Metadata (biomedical)**
 - based on ISA-Tab Specification
- **Journal Metadata (journal)**
- **Semantic resource* (semantics)**

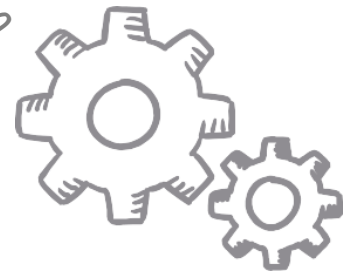
Conformes aux **standards**
DDI , Dublin Core,
DataCite, ISA-TAB

Schéma json : <https://data.inrae.fr/api/metadatablocks/{bloc}>

Valeurs contrôlées : <https://ist.blogs.inrae.fr/datainrae-guide/metadonnees-controlees-par-une-liste-de-valeurs>

Pour quoi faire ?

Hé Manu, tu m'appelles ?



Les différentes APIs

- Search API
- Native API
- Data Access API
- Metrics API
- SWORD API
- MyData API ⚠ non documentée

Rechercher des jeux de données

- API = Search & MyData

Cas d'usages :

- 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 15](#)
- Peu de métadonnées renvoyées : si besoin, utiliser l'API Native ensuite
- Attention à la pagination : tout les résultats ne sont pas renvoyés

Gérer un dataverse

- API = Native

Cas d'usages :

- Créer, modifier et publier un dataverse
- Gérer les informations, droits et autres possibilités du portail
- Supprimer le dataverse
- (prochainement) : récupérer les données des Guestbooks

Point d'attention à l'usage :

- Authentification généralement nécessaire
- CORS (voir [diapo 34](#))

Gérer un dataset

- API = Native & SWORD

Cas d'usages

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

Point d'attention à l'usage :

- Authentification généralement nécessaire
- CORS (voir [diapo 34](#))

Gestion des fichiers

- API = Native & Data Access

Cas d'usages

- 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 (voir [diapo 34](#))

Métriques

- API = Metrics

Cas d'usages

- Récupérer des métriques sur le portail (créations, téléchargements...)
- (préproduction) Récupérer des métriques pour un dataverse
- (préproduction) Récupérer des métriques d'utilisation pour un dataset

Point d'attention à l'usage :

- Données de préproduction pas forcément renseignées

Comment faire
en pratique ?

Élémentaire mon cher Watson !



Outils



Good ol' console (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

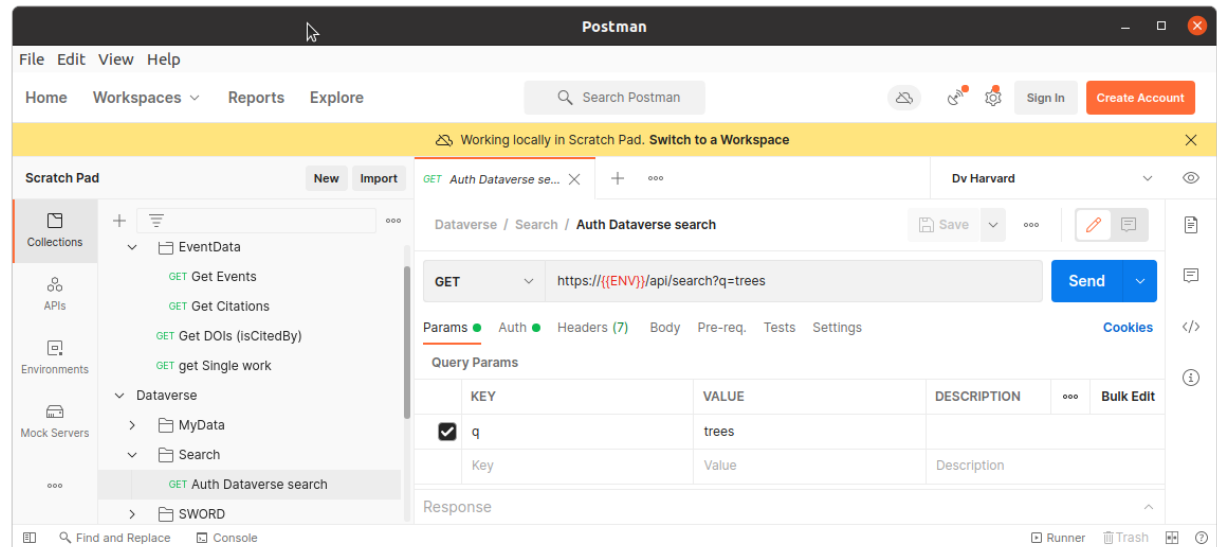
✓ Correspond à la documentation Dataverse



Le Moins : praticité

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

Postman



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

- Outil à installer depuis <https://www.postman.com/>



Le Plus : User Friendly

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



Le Moins : automatisation

- Outil séparé d'une potentielle chaîne de traitement

Clients



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 comme « curl » !



Le Plus : intégration

✓ Correspond le plus souvent à l'usage final

(DVUploader)



Pas un outil API en soit.

Outil de dépôt de fichiers en ligne de commande.

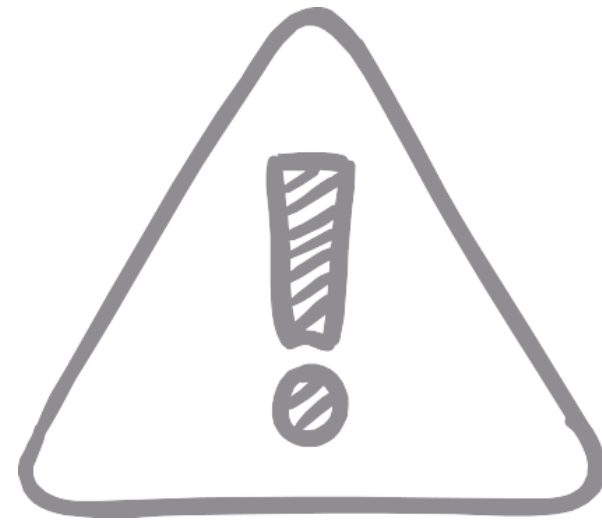
Permet de téléverser :

- un grand nombre de fichiers (>1000)...
- ...sur tout un répertoire/dossier...
- ...de façon récursive : avec les sous dossiers.

Plus d'infos sur :

<https://github.com/GlobalDataverseCommunityConsortium/dataverse-uploader>

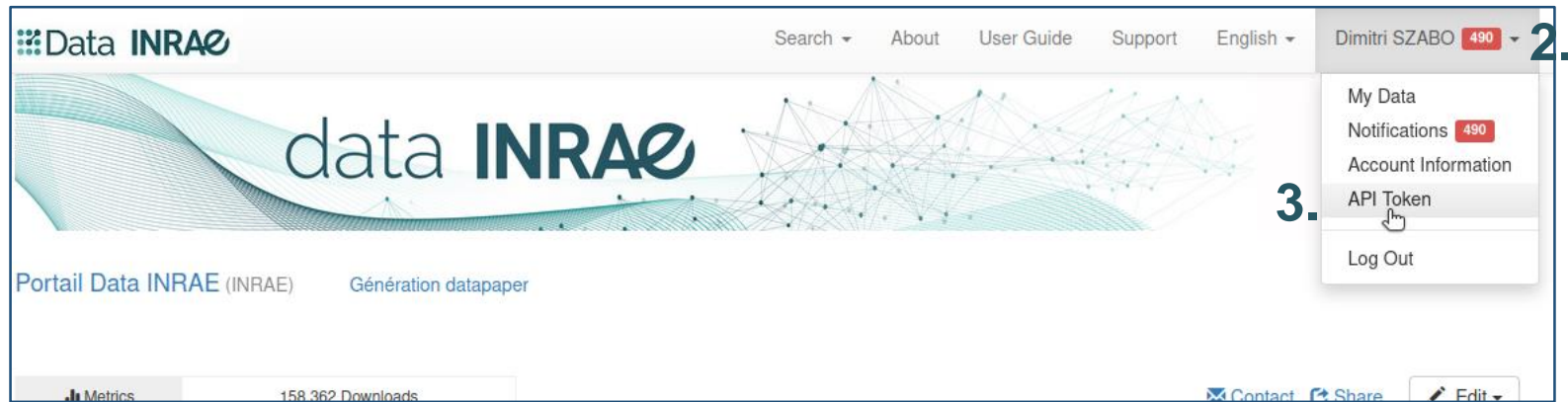
Points d'attentions



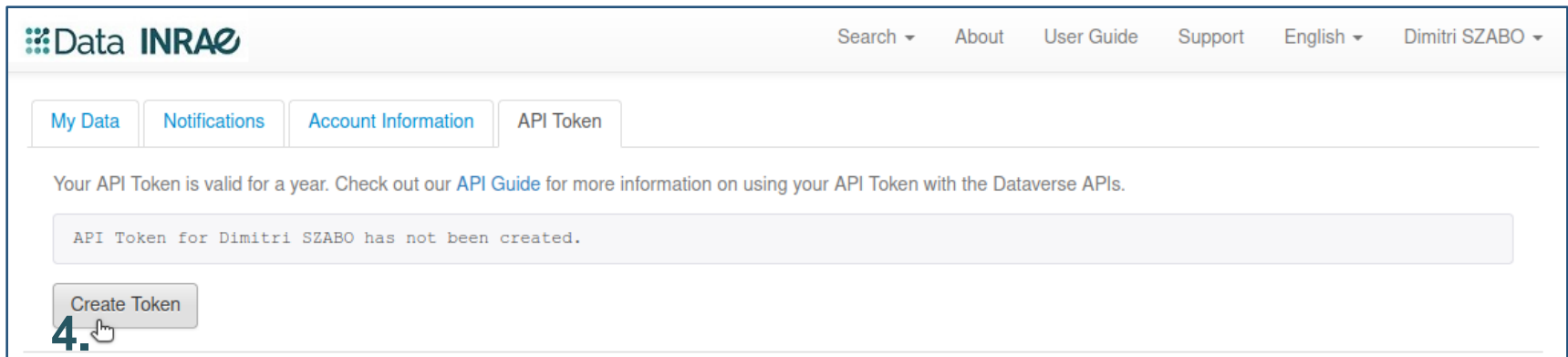
S'authentifier : le jeton API

1/3 L'obtenir

Sur Data INRAE, 1. se connecter puis :



The screenshot shows the top navigation bar of the Data INRAE website. The user profile 'Dimitri SZABO' is logged in, with a notification badge showing '490'. A dropdown menu is open, listing options: 'My Data', 'Notifications 490', 'Account Information', 'API Token', and 'Log Out'. A hand cursor is pointing at the 'API Token' option, which is highlighted. A large number '3.' is overlaid on the right side of the menu.



The screenshot shows the 'API Token' page in the user profile. The page has tabs for 'My Data', 'Notifications', 'Account Information', and 'API Token'. Below the tabs, there is a message: 'Your API Token is valid for a year. Check out our [API Guide](#) for more information on using your API Token with the Dataverse APIs.' Below this message is a text box containing the text: 'API Token for Dimitri SZABO has not been created.' At the bottom left, there is a 'Create Token' button with a hand cursor pointing at it. A large number '4.' is overlaid on the bottom left corner.

S'authentifier : le jeton API

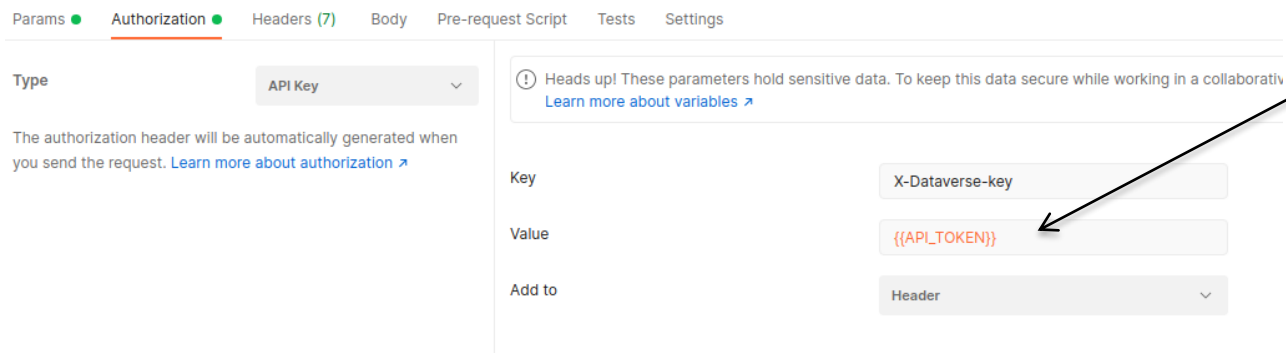
2/3 L'utiliser

Une fois créé, il faut le copier (un bouton sera disponible dans le futur).

⚠ des sauts de lignes sont copiés avec !!

Le mode de renseignement du jeton dépend ensuite de l'outil utilisé.

- Exemple dans cURL :
en-tête /header* « X-Dataverse-key » :
`curl -H X-Dataverse-key :{jeton} -X GET {ma commande}`
- Exemple dans Postman :
Dans la requête, onglet « Authorization »



The screenshot shows the Postman interface for configuring an API Key. The 'Authorization' tab is selected. The 'Type' is set to 'API Key'. A warning message states: 'Heads up! These parameters hold sensitive data. To keep this data secure while working in a collaborative environment, we've automatically generated a unique API key for you. Learn more about variables'. The 'Key' field contains 'X-Dataverse-key', the 'Value' field contains '{{API_TOKEN}}', and the 'Add to' dropdown is set to 'Header'. An arrow points from the word 'jeton' to the value field.

*éléments permettant d'envoyer au serveur des informations supplémentaires.

S'authentifier : le jeton API

3/3 Le renouveler

⚠ Le jeton a une **durée de validité d'un an**,
Au delà le jeton est révoqué et il faut en créer un autre sur le portail (voir [diapo 30](#)).

- Renouvellement sur le portail :

Dans la même fenêtre qu'en [diapo 30](#).

Bouton « Créer de nouveau le jeton » / « Recreate token »

- Renouvellement par API :
`curl -H X-Dataverse-key :{jeton} -X POST {serveur}/api/users/token/recreate`
- Vérifier la date d'expiration par API :
`curl -H X-Dataverse-key :{jeton} -X GET {serveur}/api/users/token`

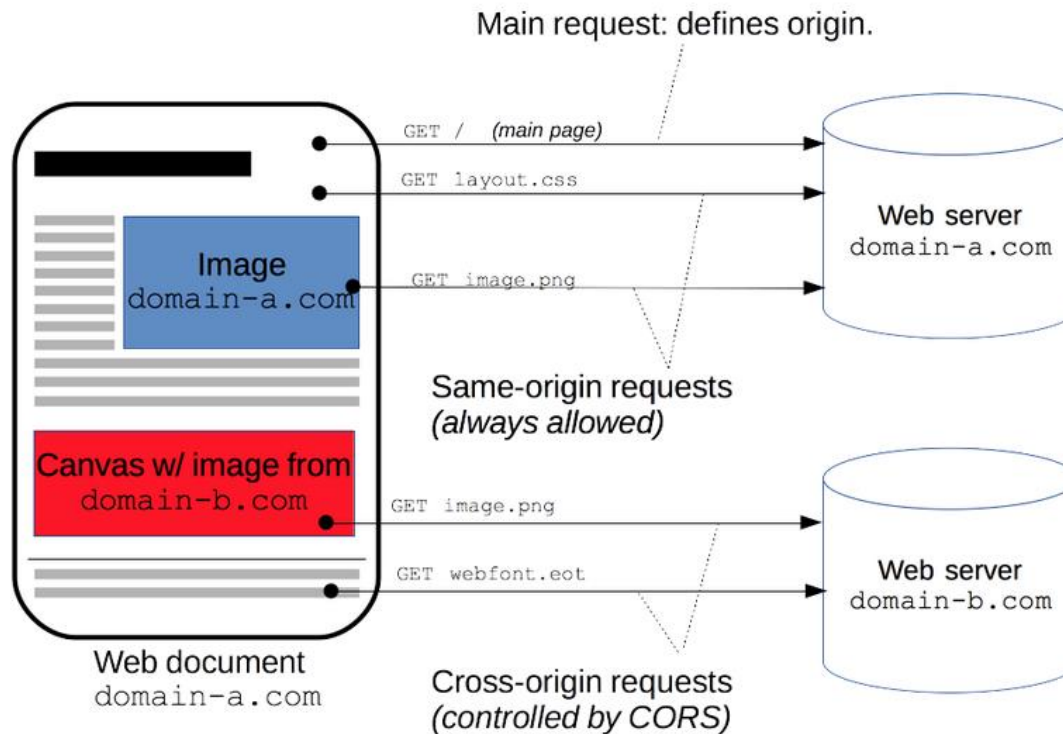
Les métadonnées spécifiques

- Respecter les blocs et valeurs contrôlées (voir [diapo 15](#))
- Appliquer en le schéma pour déposer ou importer des datasets
 - Exemple limité :
<https://nextcloud.inrae.fr/s/ZEX3GcZnrJC2iYe>
 - Exemple complet (toutes métadonnées remplies) :
<https://nextcloud.inrae.fr/s/zo7arWKqqtNs89a>
- Attention aux évolutions !

CORS (Cross Origin Data Sharing)

1/2 Qu'est-ce-que CORS ?

Sans CORS, un site web ne peut pas appeler des ressources d'un autre serveur (par exemple data.inrae.fr).



Source (et plus d'explications) :

<https://developer.mozilla.org/fr/docs/Web/HTTP/CORS>

consultation le 26/10/21

CORS (Cross Origin Data Sharing)

1/2 CORS et APIs Data INRAE

Seules certaines commandes ont CORS activé,
ces commandes sont **toutes des lectures (GET)**,

Il est alors possible de les lancer dans un navigateur ou scripts web.

Dans le cas d'une appli web, pour les autres commandes,
il faut passer par un appel depuis le backend.

Au secours ma commande ne marche pas !

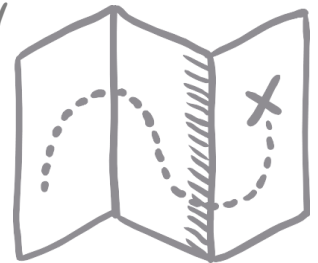
Quelques solutions courantes testables :

- Que disent l'erreur et la documentation ?
- (curl) Mettre entre guillemets les URLs
- Vérifier avec une durée plus importante entre appels APIs

... Quelles sont les vôtres ?

Et maintenant ?

I'm going on an adventure !



En plus des APIs, les widgets

Widget : Window gadget

- Permet d'afficher des éléments de Data INRAE dans un site web
 - Contenu d'un dataverse
 - Recherche dans un dataverse
 - Contenu d'un dataset
 - Citation d'un dataset
- Code à ajouter simplement dans une page web
- Récupérables sur l'interface dans le dataset/verse & utiliser

La Documentation des APIs

Guide Dataverse :

<https://guides.dataverse.org/en/5.3/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 la préproduction !

- Pour tester la publication, utiliser si possible toujours le même dataset
- Si possible, supprimer les tests au fur et à mesure



Support utilisateur et formations



Support utilisateur (Portail et APIs) :
datainrae@inrae.fr



Déposer dans Data INRAE
Administrer un dataverse
<https://ist.inra.fr/produit/datainra-formation/>



Formations présentielle

- Organisées sur/par un centre
- À la demande d'un collectif



Inscription à la liste datainrae-users : <http://tiny.cc/ListeDataINRAE>



Portail Data INRAE : <https://data.inrae.fr/>
Site de préproduction (Intranet) : <https://data-preproduction.inrae.fr/>

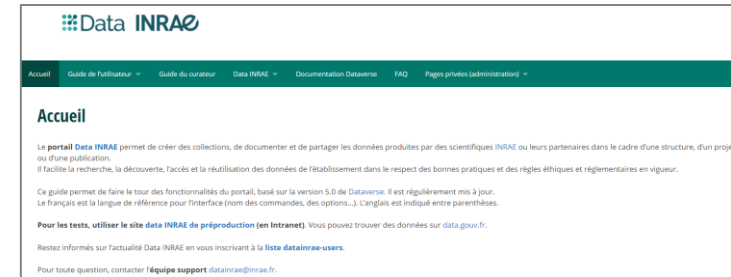


@DataINRAE



<https://datapartage.inrae.fr>

Guide Utilisateur
<https://ist.blogs.inrae.fr/datainrae-guide>



Prochaines évolutions de Data INRAE (impacts APIs)



Mars 2022 RechercheDataGouv

- Portail national
- Bac-à-sable public
- Changement des adresses

Suite 2022 Fonctionnalités

- Récupération des données Guestbook par API
- Mise à jour API Metrics
- Mises à jour dépôt et téléchargements fichiers via API
- Récupération métadonnées JSON-LD API

Data INRAE

- Nouvelle identité
- Dépôt de gros fichiers
- Meilleur référencement



Merci de votre écoute !

DipSO-Formation@inrae.fr