



Outils Constances : Plateforme SNDS, Environnement et Data App

Equipe Constances : Betül Demir – Emeline Lequy – Marylène Baylet

26/05/2026

9^e Journée Scientifique Constances & Gazel



anr

MINISTÈRE
DE L'ENSEIGNEMENT
SUPERIEUR
ET DE LA RECHERCHE

Inserm

 **Université
Paris Cité**

**université
PARIS-SACLAY**

UVSQ

 **l'Assurance
Maladie**
Agir ensemble, protéger chacun

Cnav Retraite
& Action
sociale
Sécurité sociale

01.

Application de validation des données SNDS

Contexte et besoin

- Recettage et intégration SNDS très chronophage
- Vérifications nécessitant plusieurs milliers de lignes SQL
 - plus de 2300 lignes pour les contrôles de clés de jointure,
 - plus de 500 lignes pour les vérifications de dates
- Le recettage et l'intégration d'une extraction pouvaient prendre jusqu'à 2 mois
- Avec plus de 10 ans d'expérience sur le SNDS, nous avons souhaité industrialiser, automatiser et sécuriser ces étapes



Une application Pyspark en 2 parties



**Recettage et intégration
des données SNDS
livrées par la Cnam**



**Pré-traitement adapté à
l'épidémiologie**



Une application en 2 parties



**Recettage et intégration
des données SNDS
livrées par la Cnam**



**Pré-traitement adapté à
l'épidémiologie**



Objectifs de la 1ère partie



Recettage et intégration des données SNDS livrées par la Cnam

- ➔ **Valider les données SNDS** : Tables, variables, formats et périmètres livrés (population, période...)
- ➔ **Contrôles automatiques** des extractions : Unicité des clés de jointure dans les tables « mères » / Nombre de lignes / Distribution des jours, mois, années / Nombre moyen et maximum de lignes par identifiant,
- ➔ **Exporter les données** (dans une base de données ou au format parquet)



Interface de l'application

C:\DataValidator

File Edit Configuration Validation Tools Help

Available studies

Select a study: Refresh

Select an extraction: Refresh

Extraction has a descriptor: None

Recette Rappports

Extraction status:

Total extraction validation task number: NONE Success extraction validation task Number: NONE Extraction Validation Progress: NONE/NONE

task_id	task_name	validation_rule	arguments	status
---------	-----------	-----------------	-----------	--------

Run

Refresh Extraction status Run extraction validation task show extraction validation task run history

Manage tasks

Generate validation tasks from descriptor Remove all obsolete tasks

Dataset validation task status:

Total Dataset number: 0	Total registered descriptor number: 0	Total dataset validation task number: 0	Not-run dataset validation task number: 0
Dataset with descriptor number: 0	Matched descriptor number: 0	Success dataset validation task number: 0	Out-dated dataset validation task number: 0
Dataset without descriptor number: 0	Orphan descriptor number: 0	Fail dataset validation task number: 0	Obsolete dataset validation task number: 0
		Dataset validation task progress: 0/0	

Clear Selection Selection all tasks

Run

Refresh Dataset status Generate validation task from descriptor Run dataset validation task show dataset validation task run history

Manage tasks

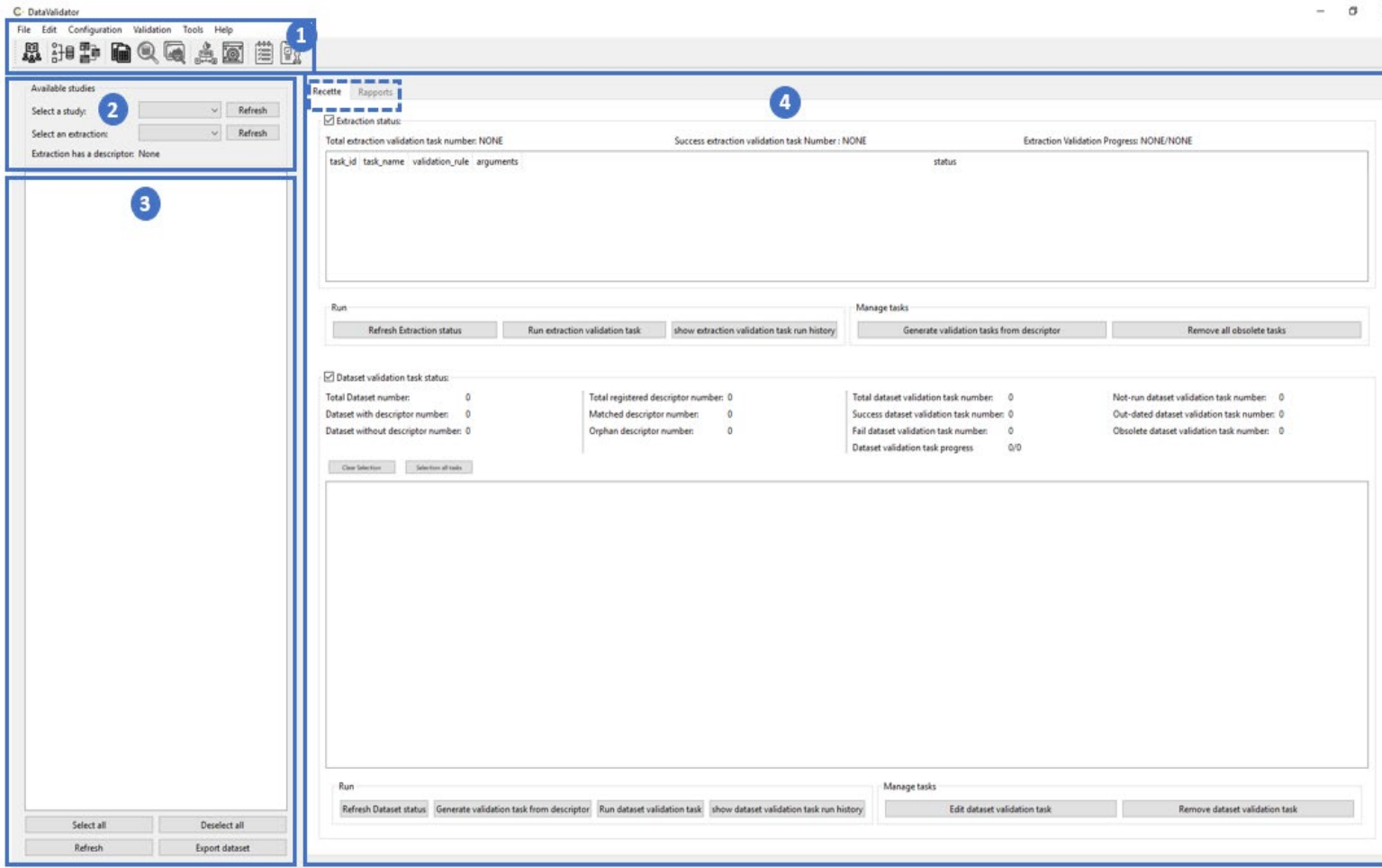
Edit dataset validation task Remove dataset validation task

Select all Deselect all Refresh Export dataset

Avanage_data_validator_v2.0.0.exe



Interface de l'application



- > **1** accès aux fonctionnalités de l'application via les menus et raccourcis.
- > **2** sélection de l'étude et de l'extraction à analyser.
- > **3** visualisation de la liste des tables de l'extraction
- > **4** affichage des résultats de recette et des contrôles.



Conclusion

- Outil utilisable sur le SNDS de la France entière
- Pensé pour être **partagé sur Git, maintenu et accompagné au CASD**
- **Automatisation et standardisation** du recettage SNDS
- **Réduction importante du temps** de validation et d'intégration
- **Sécurisation des contrôles** et amélioration de la **reproductibilité**



Guide utilisateur

L'application de validation des données SNDS

Guide utilisateur

CONSTANCES - CASD

2026-02-05



- Guide utilisateur complet rédigé pour accompagner les utilisateurs.
- Documentation de l'installation, du recettage et des contrôles.
- Description détaillée des fonctionnalités de l'application.
- Support pour le déploiement dans les bulles CASD.



État d'avancement et perspectives

- **Première partie terminée, en phase de tests utilisateurs**
 - Cohorte en population générale : **Constances**
 - Cohorte clinique de l'APHP : **Frenchie**
 - Projet **Graph4Health** (Centre de recherche en économie et statistique - CREST) : **11 ans de SNDS France entière**
- Deuxième partie dédiée au pré-traitement épidémiologique (modèle relationnel, dates de soins...) en **cours de développement**
- L'application **partagée sur Git**, maintenu et accompagné au CASD



02.

Plateforme Environnement
résidentiel et professionnel

Traitements en santé environnement (professionnel et résidentiel)

Données des volontaires :

Episodes professionnels

Adresses résidentielles



Expositions à appairer :

Matrices emploi exposition
françaises et internationales

Cartes, bases de données françaises
et internationales



Méthodes d'appariement :

Par année/période
Peut dépendre d'âge/sexe/...



Par période d'intérêt
Différentes modalités (valeur brute
ou retravaillée, distance ...)

Extraction et diffusion vers les équipes de recherche : tables de données et de documentation



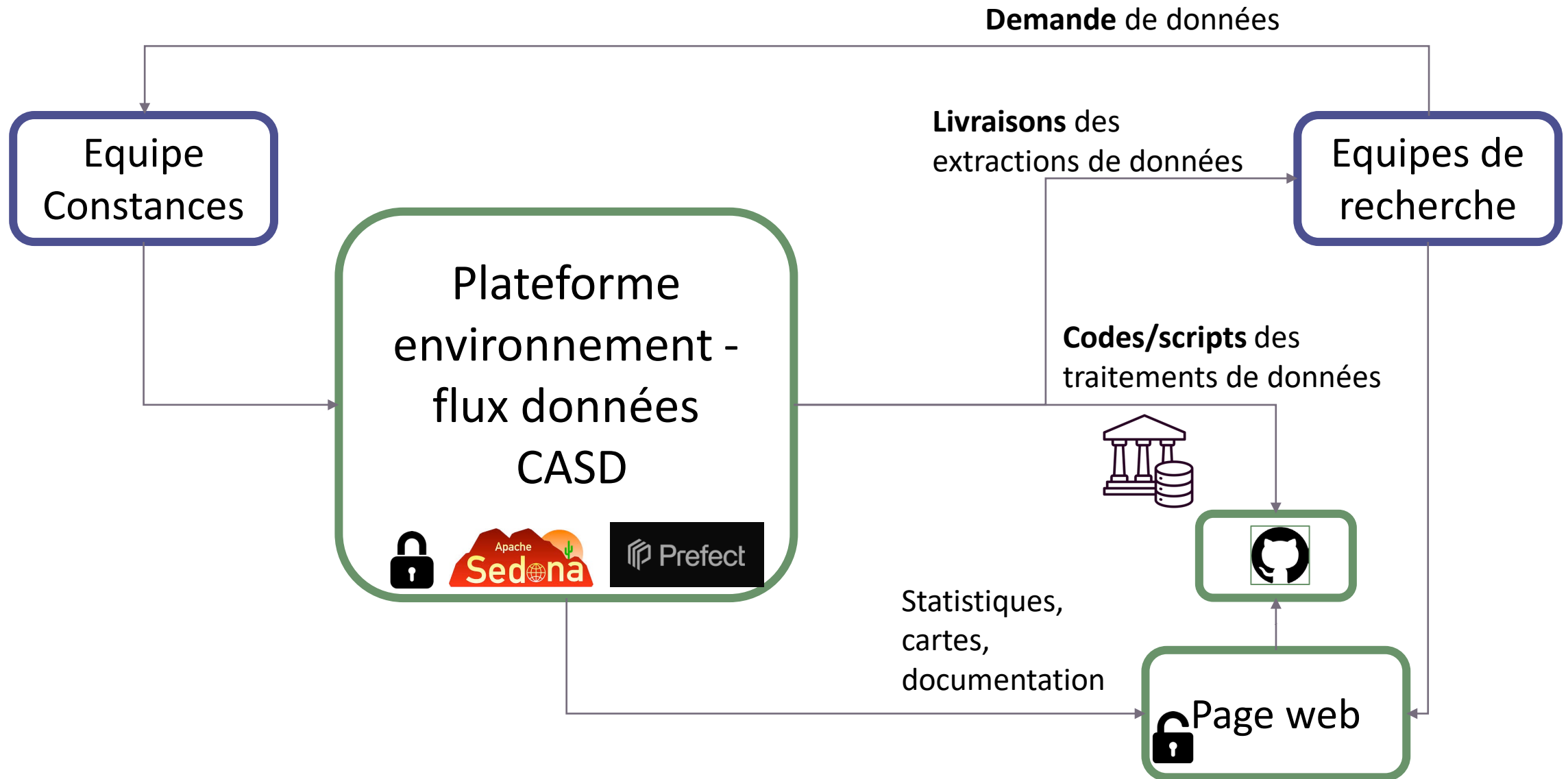
Constats sur ce processus

Ensemble des traitements réalisés avec des scripts R au fil de l'eau par projet :

- Risque d'erreurs de manipulation
 - Reproductibilité limitée (modification d'un script, déplacement d'un jeu de données)
- Besoin de formaliser la gouvernance des données et de davantage de documentation (processus formels, description des données)



Vers la plateforme environnement



Nouveau processus

Ensemble des traitements réalisés avec des scripts automatisés, déclenchables à la demande

Beaucoup moins d'erreurs de manipulation

- Reproductibilité optimisée (modification d'un script, déplacement d'un jeu de données)

→ gouvernance des données et documentation (processus formels, description des données)

harmonisées avec les autres procédures de la cohorte Constances



03.

Data App

La Data App : l'outil de pilotage multisource de la cohorte

 Mettre a disposition les **indicateurs** de la cohorte en **temps réels** pour faciliter le suivi de l'activité et la réalisation des rapports

 Assurer la **transparence** et la **traçabilité** du partage des données



La Data App : l'outil de pilotage multisource de la cohorte

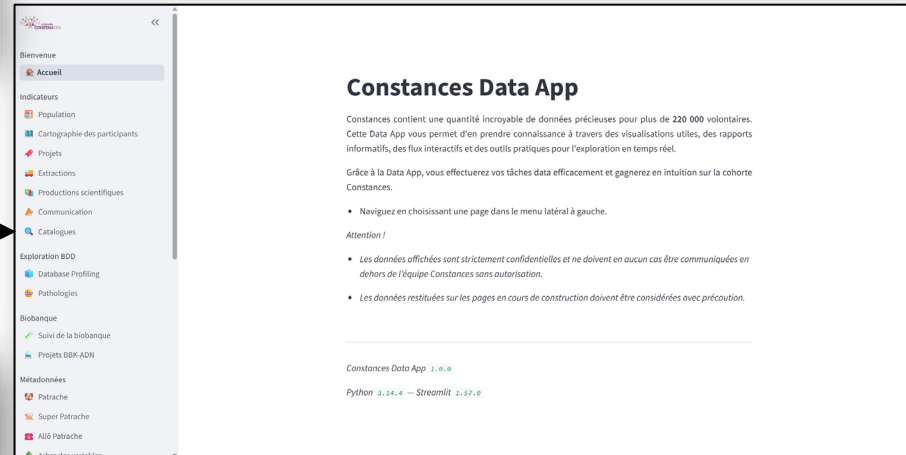
Bases de données Constances



Outil de gestion bibliographique

zotero

Outil de gestion de projet



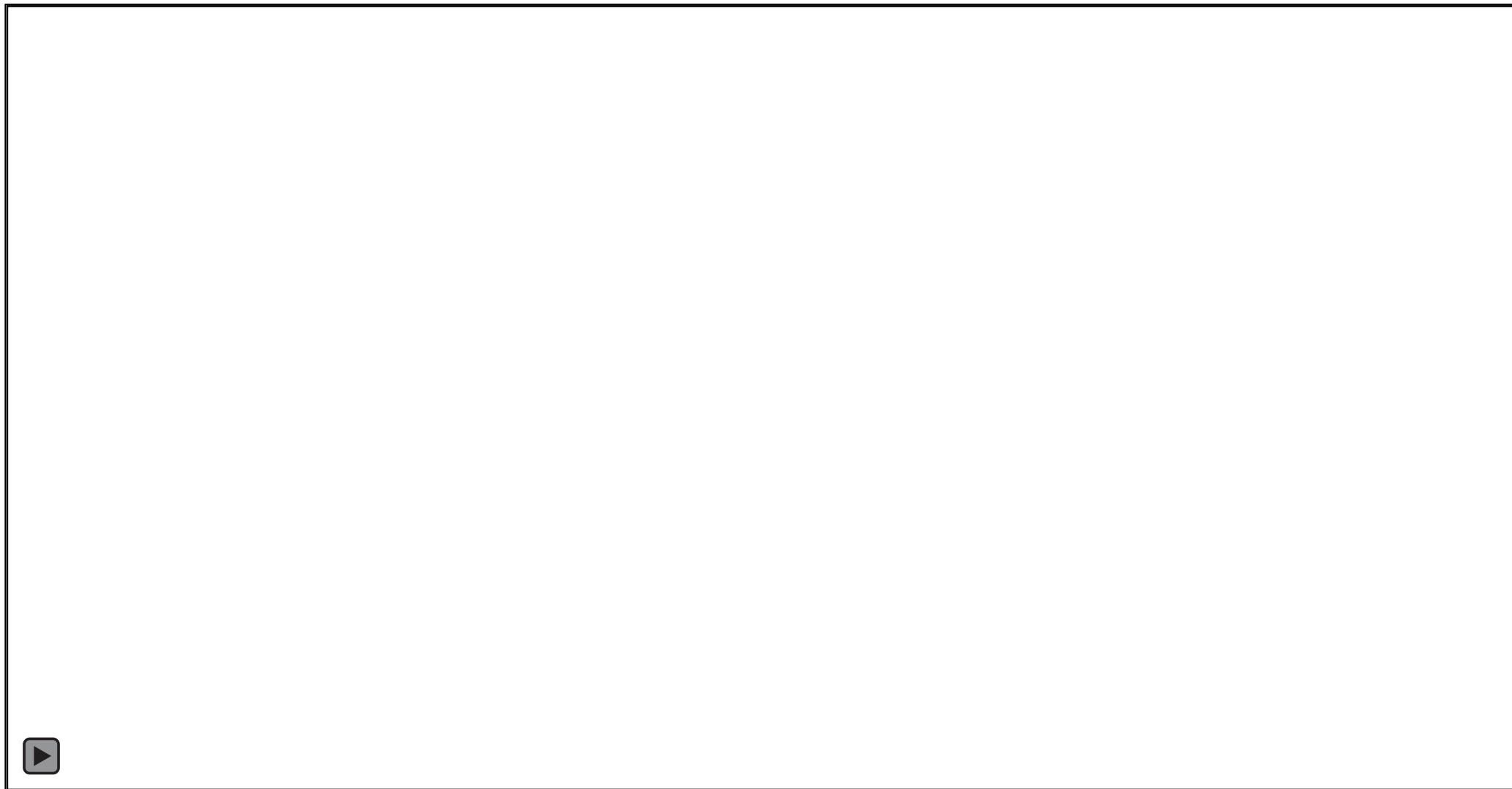
Outil de communication



 **Données agrégées, accessible uniquement depuis le réseau de l'équipe Constances**



La Data App : Démonstration

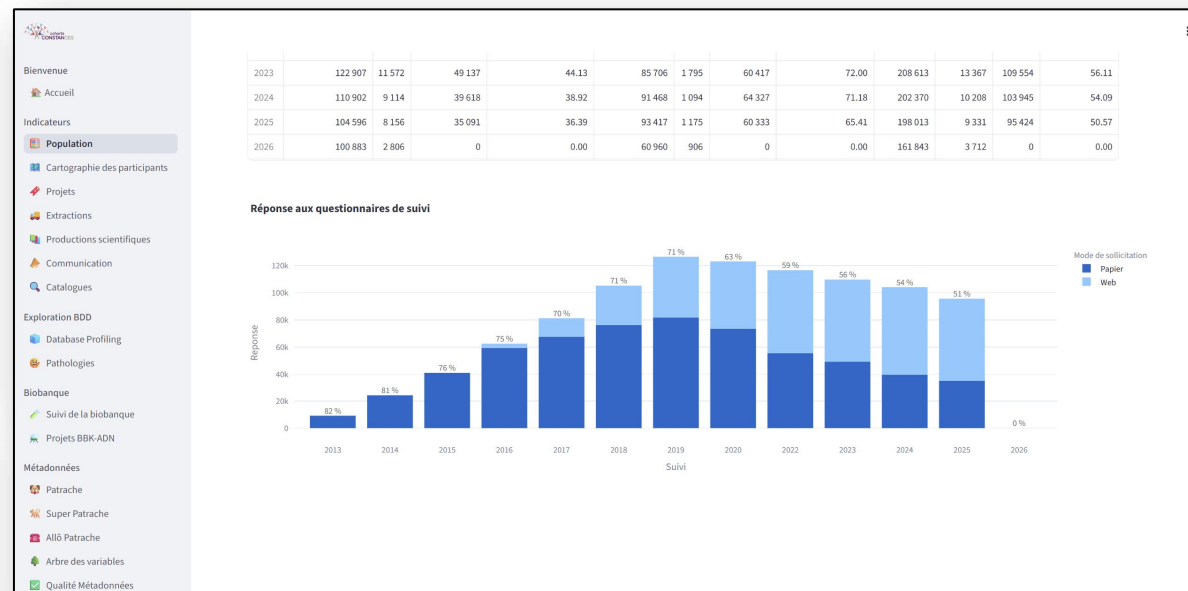
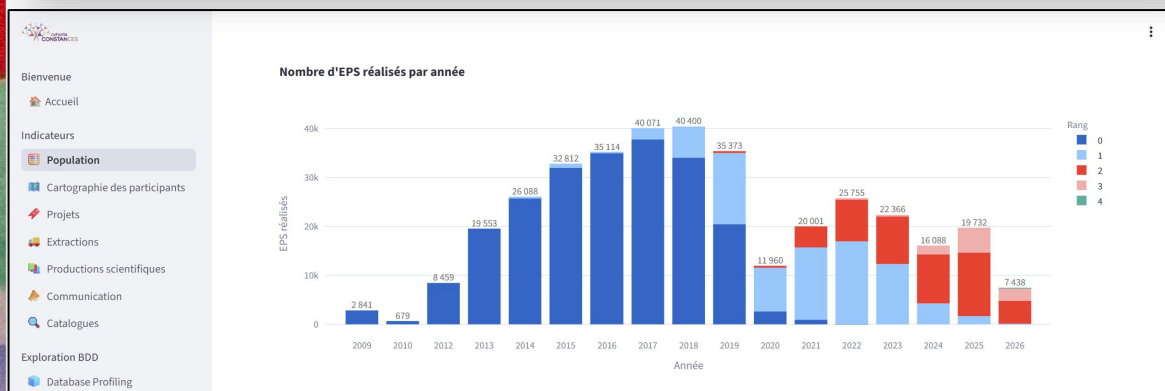
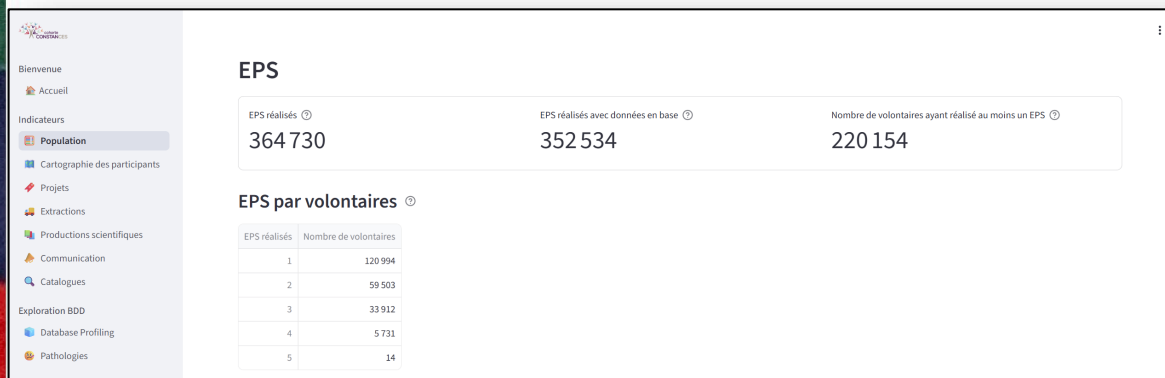




La Data App : la population

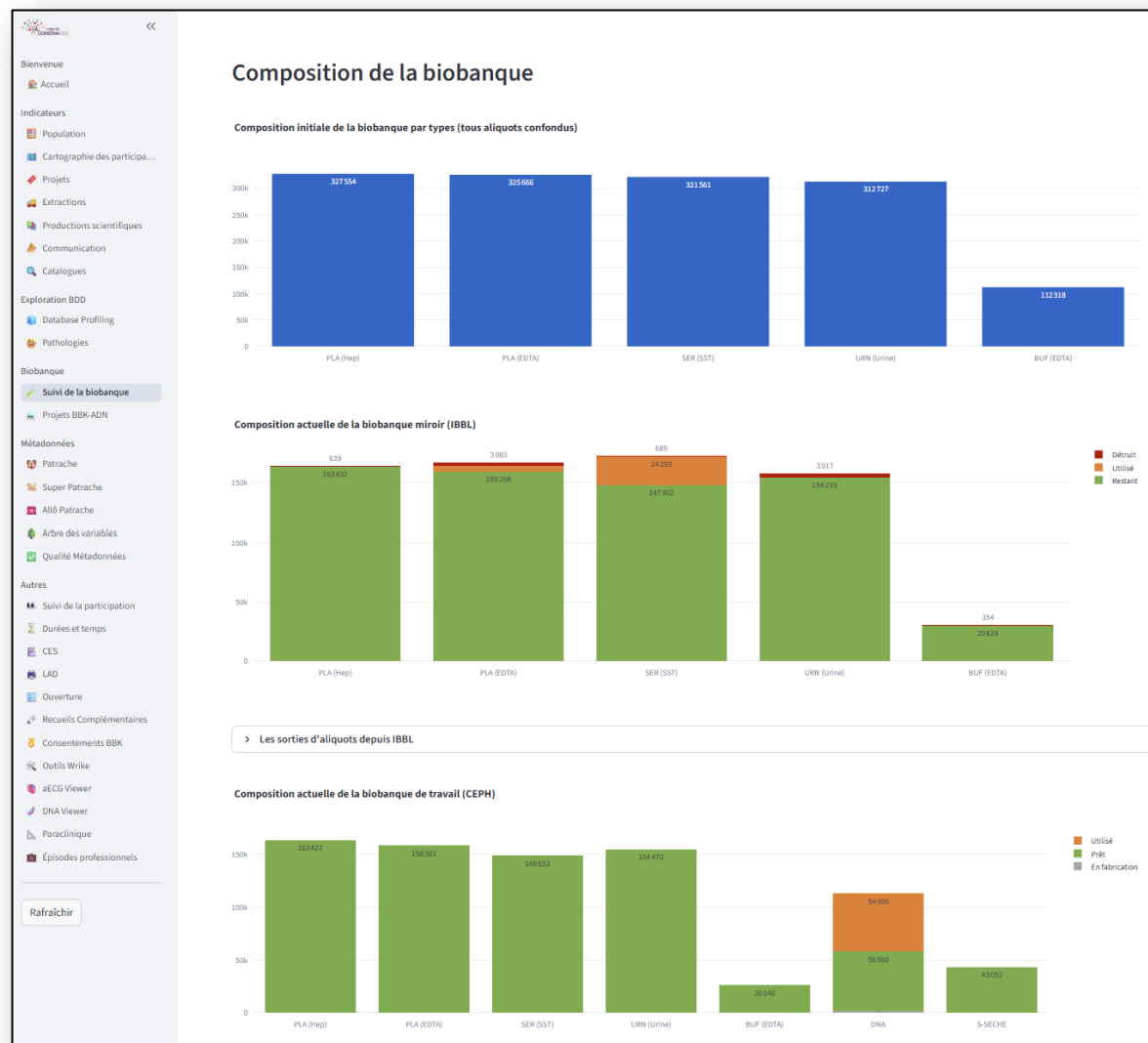
Examens de santé

Questionnaires





La Data App : les stocks de la biobanque



La Data App : le suivi du partage des données

Outil de gestion de projet



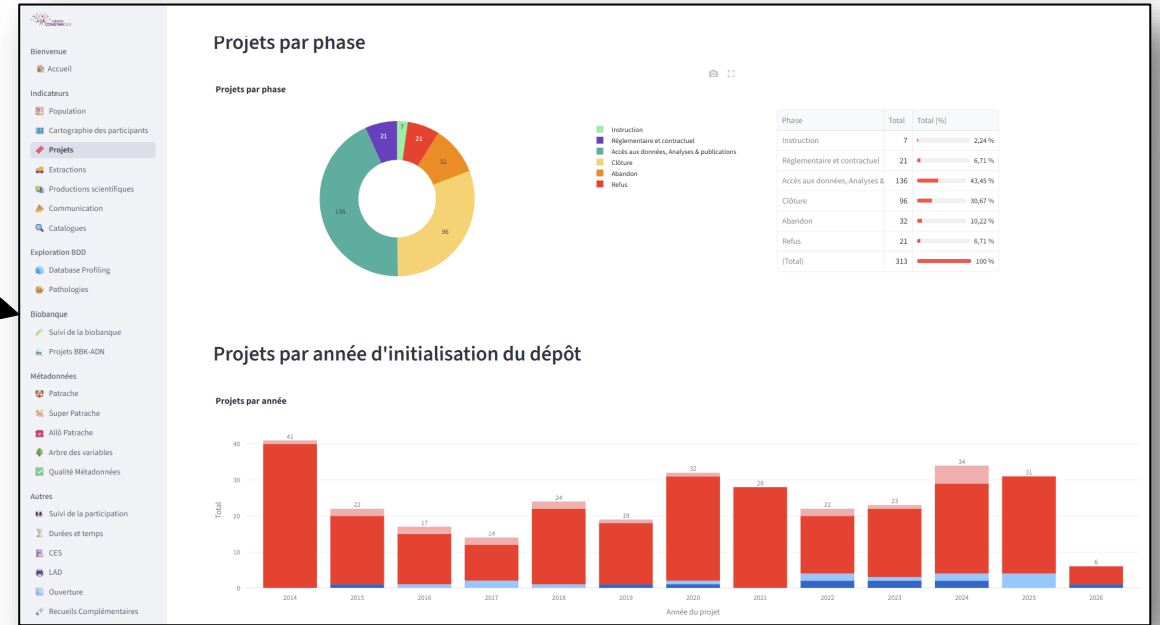
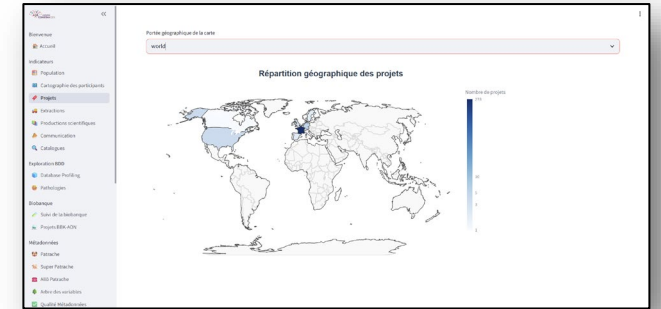
1. Projets

Private 9 members Add bookmark

PHASE 0 - INES Pour JS Test clôture Suivi des fiches-projet PHASES 5-6-7 All

Name	Op. Phase	↑ 0q. Sous-phase
Empty 6		
1e contact		
2. Projets en cours		
2025-A-273-1	2. Dépôt	
2025-A-276-1	2. Dépôt	
2025-A-300-1	2. Dépôt	
2026-A-315-1	2. Dépôt	
2025-A-274-1	3. Instruction	3.2. Instruction - Attente retour U...
2025-A-290-1	3. Instruction	3.1. Instruction - Attente retour PI
2025-A-297-1	3. Instruction	3.1. Instruction - Attente retour PI
2025-A-299-1	3. Instruction	3.2. Instruction - Attente retour U...

⚡ API temps réel



La Data App : le suivi du partage des données

Free 30-day licenses

3. Suivi des extractions

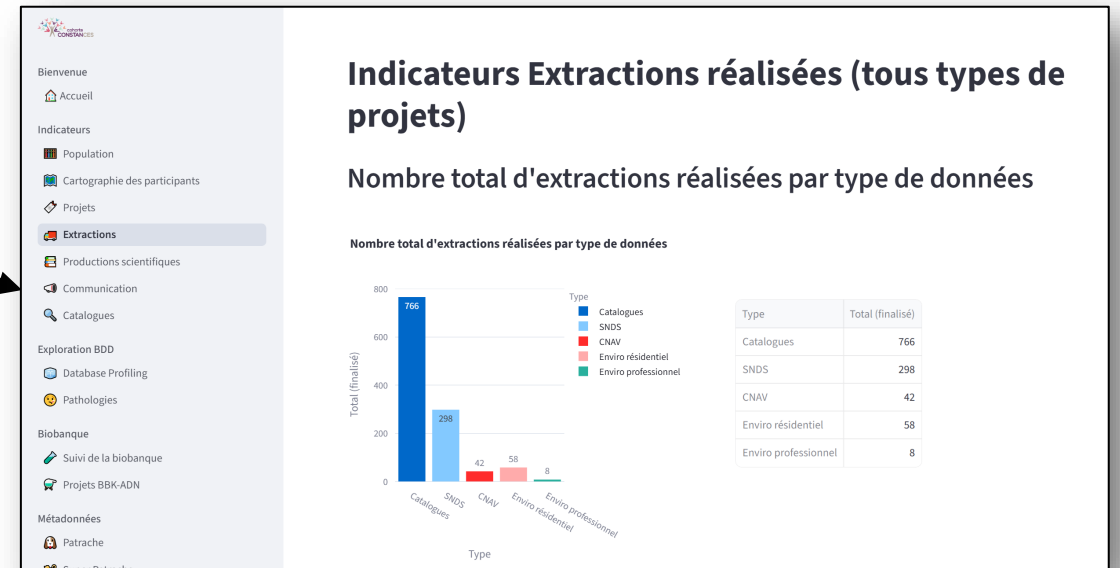
1. Catalogues

Tableau... Table + 2 filters Created date Group Fields Subitems Public link

5m. Historique Ca... : not Oui Created date: after 60d ago Add filter

Name	Created date	Status	Completed date	Due date
2022-A-196-1	01/04/2026	Completed	02/04/2026	02/04/2026
2025-A-269-1	07/04/2026	Completed	15/04/2026	15/04/2026
2025-A-306-1	07/04/2026	Completed	15/04/2026	15/04/2026
2018-I-010-6	14/04/2026	Completed	16/04/2026	16/04/2026
2017-A-079-1	15/04/2026	Completed	15/04/2026	06/03/2018
2017-A-079-1	15/04/2026	Completed	15/04/2026	22/03/2018
2017-A-079-1	15/04/2026	Completed	15/04/2026	08/06/2018
2024-A-237-1	15/04/2026	Completed	15/04/2026	17/06/2024
2024-A-237-1	15/04/2026	Completed	15/04/2026	20/03/2025
2024-A-237-1	15/04/2026	Completed	15/04/2026	04/07/2024
2024-A-237-1	15/04/2026	Completed	15/04/2026	15/05/2025

⚡ API
temps réel





La Data App : la traçabilité en quelques secondes



Outils Constances :

Plateforme SNDS, Environnement et Data App



Merci de votre attention

