



**HAL**  
open science

## Progedo, une infrastructure pour ouvrir les données

Sébastien Oliveau

### ► To cite this version:

Sébastien Oliveau. Progedo, une infrastructure pour ouvrir les données. 8ème conférence document numérique et société, Jun 2022, Liège, Belgique. <hal-03705497>

**HAL Id: hal-03705497**

**<https://hal.science/hal-03705497v1>**

Submitted on 27 Jun 2022

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 CC BY-NC 4.0 - Attribution - Non-commercial use - International License



DATA  
INFRASTRUCTURE

# Une infrastructure pour ouvrir les données

Sébastien Oliveau,  
Directeur de l'infrastructure de  
recherche PROGEDO,  
Maître de conférences à l'Université  
d'Aix-Marseille



8<sup>ème</sup> conférence document numérique et société  
Communications scientifique et science ouverte  
Opportunités, tension et paradoxe

Liège, 23 juin 2022



# OUVRIR

# LA SCIENCE !

La science ouverte est la diffusion sans entrave des publications et des données de la recherche.

Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et -autant que possible- aux données de la recherche.

<https://www.ouvrirlascien>



Huma-Num est une très grande infrastructure de recherche (TGIR) visant à faciliter le **tournant numérique de la recherche en sciences humaines et sociales**. Huma-Num met en œuvre un dispositif humain (concertation collective) et technologique en s'appuyant sur un réseau de partenaires et d'opérateurs.



Acteur central de la **politique nationale sur les données en sciences humaines et sociales**, PROGEDO soutient la réalisation de grandes enquêtes européennes, la mise à disposition des données françaises, le développement de la culture des données et la constitution de briques de compétences de niveau européen.



OpenEdition est une infrastructure complète d'**édition électronique au service de la communication scientifique en sciences humaines et sociales**. Elle rassemble quatre plateformes complémentaires dédiées aux livres, aux revues, aux carnets de recherche et aux événements scientifiques

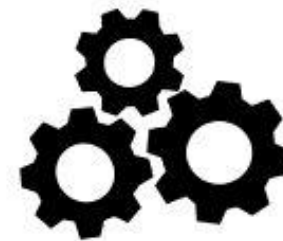


Le Réseau national des MSH fédère 23 Maisons des Sciences de l'Homme réparties sur toute la France, et regroupées autour de cinq principes fondamentaux : **l'Interdisciplinarité**, **l'Internationalisation**, une **Dynamique Interinstitutionnelle**, **l'Implantation territoriale** et une **Identité scientifique**.

# OUVRIR

# LA SCIENCE !

Rendre les données **F**indiquées **A**ccessible **I**nteroperable **R**eusable



# OUVRIR

# LA SCIENCE !

Rendre les données aussi ouvertes que possible...

...les garder aussi fermées que nécessaires

# Qu'est-ce qu'une donnée pour PROGEDO?

Matériaux: ce que l'on récolte sur le terrain

Un matériau peut être ouvert ou non

Un matériau peut-il être FAIR ?

Donnée : matériau mis en forme (et forme spécifique)

Aujourd'hui, la donnée est implicitement numérique

Donnée secondaire ? Donnée issue du traitement d'une donnée existante (données agrégées par exemple)

# Qu'est-ce qu'une donnée pour PROGEDO?

Les données dans le cadre de PROGEDO:

issues d'enquêtes (ou de registres administratifs)

quantitatives

représentatives

(il peut y avoir des exceptions, c'est le principe des règles)

!! Enjeux particuliers autour des  
« microdonnées » !!

# Qu'est-ce qu'une donnée pour PROGEDO?

De nouveaux enjeux:

les données liées aux publications

Besoin exprimé par les chercheurs

Enjeu au niveau de l'édition

les données secondaires

Concept à préciser

Quel sauvegarde souhaitée/nécessaire

# Les données en SHS

## Cadre légal

Loi pour une République  
Numérique

Règlement Général pour la Protection des  
Données

## Implications

Diffusion obligatoire des données  
publiques

Protection des données  
individuelles

## Applicati on

Développement de l'Open  
Data

Développement de procédures d'accès  
restreints

## Accès aux données

opendata.gouv.fr; zenodo; nakala;  
etc.

PROGEDO; CASD; Health Data Hub;  
etc.

# Protéger les données personnelles

## Méthodes

protéger l'accès

pseudonymiser

Rendre anonyme

## Implications

Environnements  
adaptés

Potentiellement  
réidentifiant

Perte  
d'information

# Protéger l'accès

**Ordinateurs non connectés**  
dans les locaux du producteur

Le chercheur se déplace physiquement dans les locaux du producteur et accède à une machine où il peut travailler et imprimer/exporter ses résultats

**Exécution à distance**  
(*remote execution*)

Le chercheur envoie une ligne de commande sur un serveur distant qui opère sans donner accès aux données et renvoie les résultats

**Accès sécurisé à distance**  
 (« bulles » informatiques)

Le chercheur se connecte sur un terminal qui lui donne accès aux données et à un environnement logiciel. Il peut exporter ses résultats.



# Protéger l'accès

Ordinateurs non  
connectés  
dans les locaux du  
producteur

Exécution à  
distance  
(*remote execution*)

Accès sécurisé à  
distance  
 (« bulles »  
informatiques)

Nécessité de se  
déplacer  
Inégalités  
d'accès

**Not FAIR**

Moindre  
convivialité  
Moindre souplesse  
(difficile de  
proposer des choix de  
logiciels en dehors  
de ce qui est  
proposé, idem pour  
les données)

**FAIR ?**

Environnements  
adaptés  
Coûts

**Aussi FAIR que  
possible**

# Pseudonymiser les données

## Les fichiers « Production et recherche »

En termes de détails, les fichiers "FPR" offrent un niveau intermédiaire d'information entre les données standards anonymisées (plus agrégées) et les données confidentielles plus détaillées.

Les données pseudonymisées sont considérées comme potentiellement ré-identifiantes. A ce titre, elle nécessite une autorisation administrative auprès du **Comité du secret statistique**.

La procédure d'habilitation auprès du Comité du Secret Statistique est gérée par l'intermédiaire des équipes de Quetelet-Progedo-Diffusion. L'accès aux données se fait ensuite depuis l'ordinateur du chercheur.

# Pseudonymiser les données



*PROGEDO assure le financement de Quetelet-Progedo-Diffusion et la coordination scientifique des équipes membres*



*Huma-Num assure l'accueil sur ses serveurs de l'infrastructure informatique de PROGEDO*

Standard de métadonnées

Dublin Core

DDI

Protocoles standards échange

OAI-PMH

Documentation des Process

Certification Core

Trust Seal

## Quelques limites à évoquer

Rapidité des changements => formation

Instabilité législative

Durabilité des infrastructures  
informatiques

Pérennité des financements

# Produire, Partager, Promouvoir

Produire des données



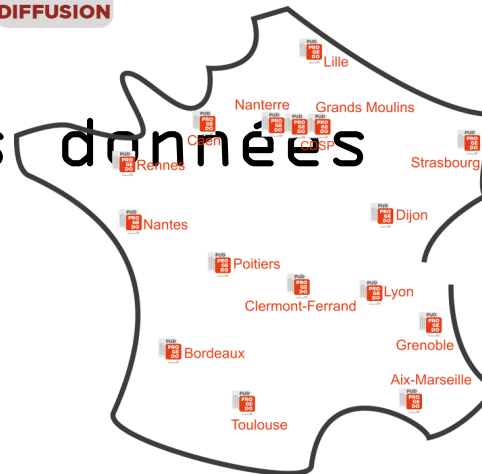
Donner acc



Gérer la diffusion des données



Promouvoir la culture des données





Méthodes & Statistiques Publiques



cessda

Opinion et enquêtes sociopolitiques



Santé



Données historiques économiques



Population et parcours





DATA INFRASTRUCTURE

# Politique Française des données en SHS

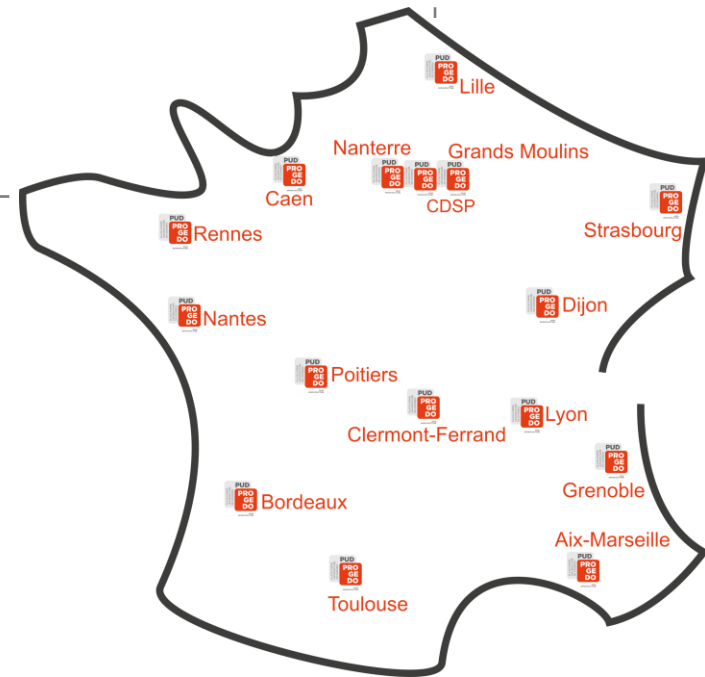
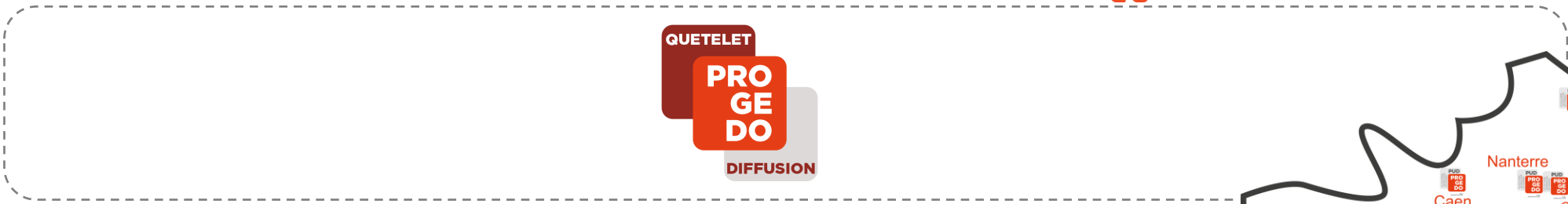
Méthodes & Statistiques Publiques

Opinion et enquêtes sociopolitiques

Santé

Données historiques économiques et financières

Population et parcours de vie



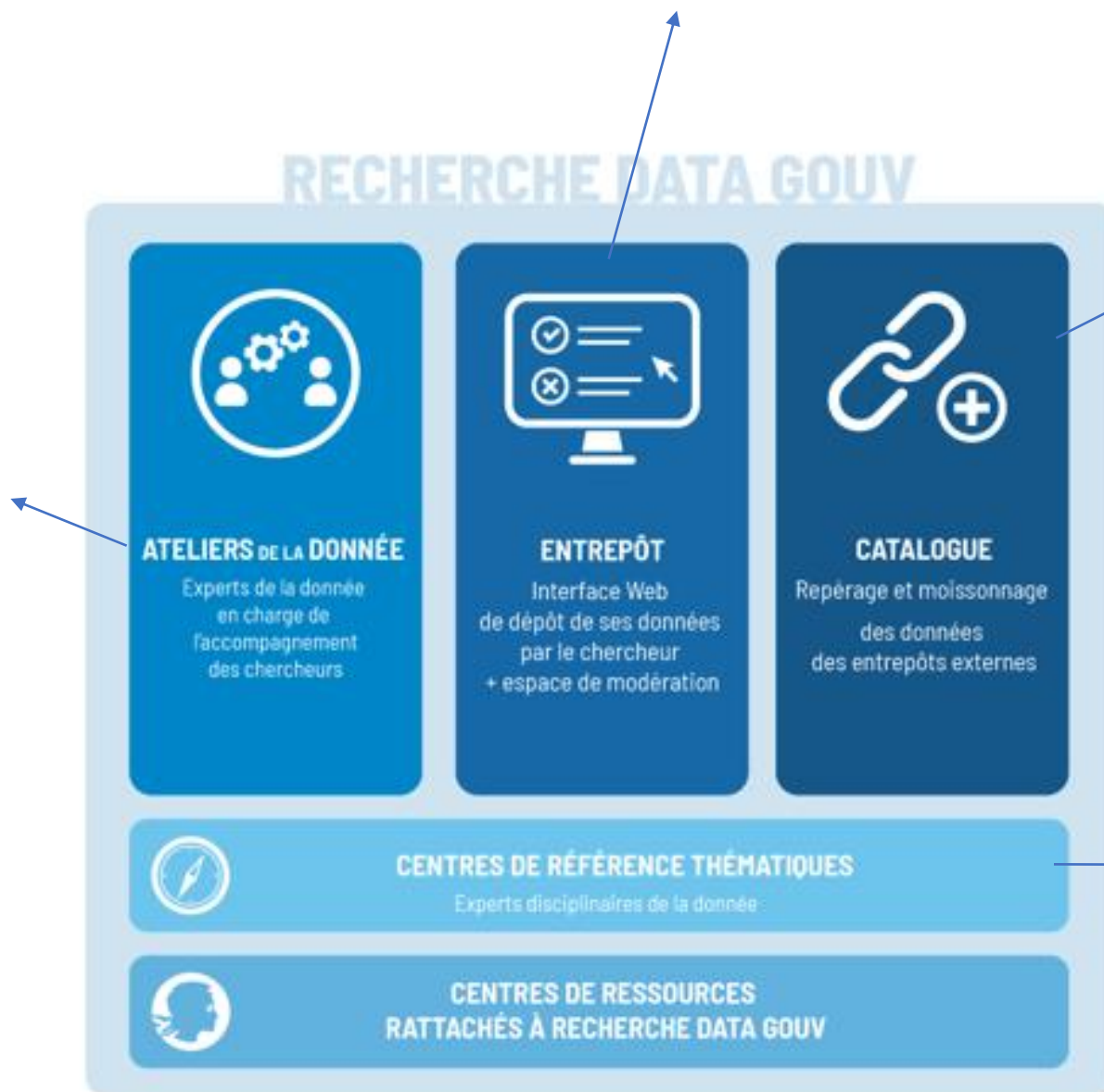
Plateformes Universitaires de Données



et Universités partenaires

Pour les enquêtes quantitatives en SHS: <https://data.progedo.fr>  
Pour les données SHS de manière plus générique: <https://nakala.f>

Projets de site,  
incluant les PUD le  
cas échéant



Pour les enquêtes quantitatives en SHS:  
<https://data.progedo.fr>

De manière plus générique en SHS:  
<https://isidore.science/>



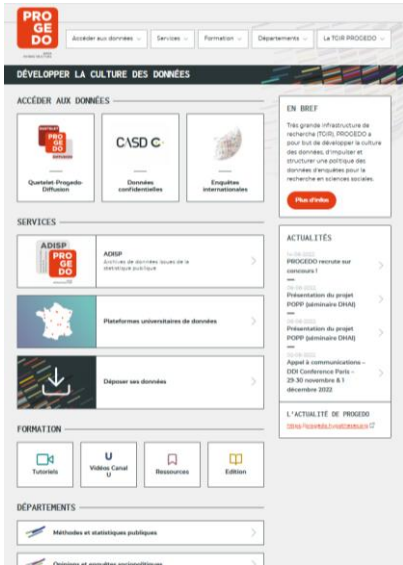


DATA  
INFRASTRUCTURE

# *Développer la culture des données*



Merci de votre attention



<https://www.proge>