



19<sup>ème</sup>

ÉCOLE INTER-ORGANISMES

QUALITÉ ET RESPONSABILITÉ SOCIÉTALE

EN RECHERCHE ET EN ENSEIGNEMENT SUPÉRIEUR

du 13 au 15 Septembre 2021

Transformation des pratiques et des usages  
dans la gestion des données

**PRO  
GE  
DO**

DATA  
INFRASTRUCTURE

Sébastien Oliveau, directeur de l'IR\* PROGEDO

# 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.

# OUVRIR

# LA SCIENCE !

Rendre les données

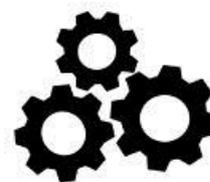
**F**  
Findable



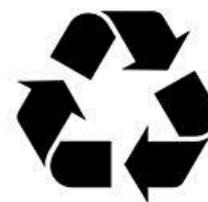
**A**  
Accessible



**I**  
Interoperable



**R**  
Reusable



# OUVRIR

# LA SCIENCE !



Rendre les données aussi ouvertes que possible...

...les garder aussi fermées que nécessaires

# 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

## Application

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

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

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

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

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

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

## Méthodes

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

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

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

## Exemples



# 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*

# FAIR et qualité des données



## Standard de métadonnées

**Dublin Core**

**DDI**

## Protocoles standards décharge

**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**



