

B U L A C

[도서관] [शिक्षा] [അക്കാദമിക്] [ሥልጣኔ]

Bibliothèque universitaire  
des langues et civilisations

# INTRODUCTION À LA GESTION DES DONNÉES DE LA RECHERCHE EN SHS

Printemps 2024

## Le programme de la formation :

**Introduction** : Des données en SHS ? Paroles de chercheurs

1. Les données de la recherche en SHS : définition(s) et typologie(s)
2. Pourquoi se préoccuper de ses données : les enjeux
3. Les principes et outils d'une bonne gestion des données
4. La bonne gestion en action
5. Conclusion et pistes d'approfondissement

# BULAC

[ספרייה] [大学] [γλώσσες] [අભ්‍යන්තර]

Bibliothèque universitaire  
des langues et civilisations

PAROLES DE  
CHERCHEURS ...



### Format

“J’ai toujours plaisir [...] à travailler sur des documents papiers [mais maintenant], je prends beaucoup de photos.”

### Définition

“Où est la limite entre un article et les données ?”

### Propriété

“Quand vous travaillez à plusieurs sur un projet, ces données appartiennent à qui ?”

### Reconnaissance

“Notre métier n’est pas de créer des bases de données pour créer des bases de données, c’est de publier...”



### Partage

“On ne va pas filer notre stock de données pour que d’autres les travaillent ...”

### Archivage

“j’ai perdu toutes les données de ma thèse ...”

# BULAC

[ספרייה] [大学] [γλώσσες] [අසැසි]

Bibliothèque universitaire  
des langues et civilisations

## I. DÉFINITIONS ET TYPOLOGIES



# 1. Définitions et typologies

“ **Enregistrements factuels** (chiffres, textes, images et sons) qui sont utilisés comme **sources principales** pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats** de recherche. (OCDE, 2007) ”

“ Ces données [...] peuvent être de nature chiffrée, descriptive ou visuelle, **brutes** ou **traitées** [...], **expérimentales**, d'**observation** ou générées par une **machine**. (Université de Melbourne) ”

## 1. Définitions et typologies

Données d'observation / de terrain  
= donnée "brutes"

*photo, enregistrement audio, inscription sur table funéraire ...*

Données expérimentales

*ex : dispositifs d'apprentissage en science de l'éducation*

quantitatives

qualitatives

produites

collectées

Données dérivées

Issues de sélections ou de traitements de données "brutes"

*bibliographie, données de fouille de texte ...*

Données de simulation  
générées par modèles informatiques

*modèles économiques, climatiques*

chaudes =  
en cours de  
traitement

froides =  
analysées

Données de référence et données sources

données de référence (= publiées) *corpus* de textes, d'images

données sources = *lettre manuscrite* ; brouillon d'écrivain ; pièce d'archive

## 1. Définitions et typologies

Mon corpus  
textuel



Mes photos  
d'archives



*Cycle de Panji*, (Schoemann II 4) 37r.,  
javanais, XIXe siècle, Berlin

Mes  
enregistrements



Autres données  
dérivées



Ma bibliographie

prelib.odt : Femmes nées au XIXe siècle avec leur ville de naissance

	NomPersonne	NomVille
	Anna Ropars	Bolazec
	Anne Le Bastard de Mesmeur	Quimper
	Beatrix Potter	Kensington
	Charlotte Guest	Uffington
	Elisa Méseray	Honfleur
	Jeanne Malivel	Loudéac
	Josèphe Daniélou (née Creignou)	Roscoff
	Lady Mond	Belle-Isle-en-Terre
	Madeleine Desroseaux	Rennes
	Marie-Anne Abgrall	Lampaul-Guililiou
	Marie-Louise Charran	Kermaria-Sulard
	Philomène Cadoret	Bonen
	Sarojini Naidu	Hyderabad

Données auteurs, projet sur littérature  
populaire bretonne, J.-B. Pressac



## 1. Définitions et typologies

À quoi ressemble un corpus cohérent de données ?

Consultez ces deux ensembles de données et observez :

- Quels **types** de données (= fichiers) sont rassemblés ?
- Qu'est-ce qui permet de **retrouver** ces ensembles à coup sûr (sans erreur 404 par ex.) ?
- Peut-on utiliser **librement** ces données ? A quelle(s) condition(s) ?
- Dans **quel(s) but(s)** ces données sont-elles mises à disposition ?



8 min.



Anti-Jewish Pogroms in the Russian Empire



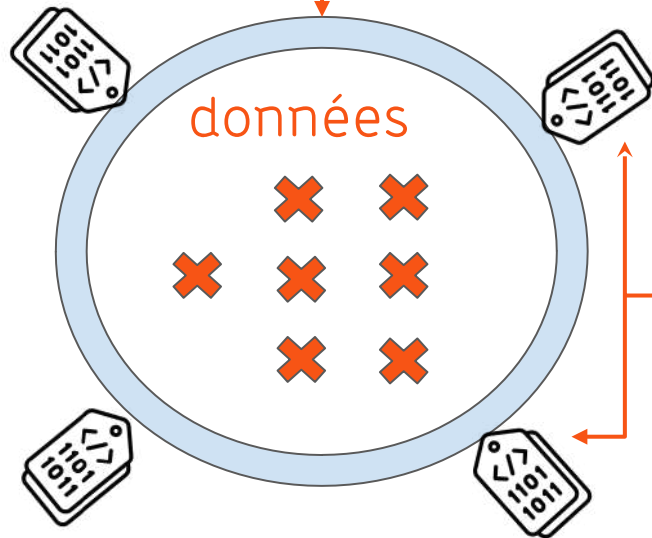
Bullaire de Valier

## 1. Définitions et typologies

## Jeu de données

Agrégation [...] de données brutes ou dérivées présentant une **certaine unité**, rassemblées pour former un **ensemble cohérent**. (DoraNum)

Parfois, "donnée" utilisé pour désigner tout un jeu



**Métadonnées**  
= données sur la donnée

informations qui permettent d'**identifier**, de **décrire** et de renseigner sur l'**origine** d'un objet numérique comme un jeu de données.

Elles facilitent le **partage** et la **réutilisation** de la ressource.

(adapté de : Enssib)

# BULAC

[ספרייה] [大学] [γλώσσες] [भाषा]

Bibliothèque universitaire  
des langues et civilisations

## II. POURQUOI SE PRÉOCCUPER DE SES DONNÉES : LES ENJEUX



## 2. Pourquoi se préoccuper de ses données ?

### Intérêts pour le chercheur et la collaboration scientifique

Pour lui-même

- Données bien décrites
- Données bien nommées
- Données bien ordonnées



Gain  
de temps  
d'efficacité



Crédits : M. Hassan / Pixabay  
DoraNum -CNRS

Pour ses  
collaborateurs

- Données bien stockées
- Données bien classées, faciles à retrouver
- Règles communes de traitement

Gain  
d'efficacité  
de sécurité

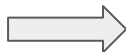
## 2. Pourquoi se préoccuper de ses données ?



### Des incitations à ouvrir les données

Une incitation forte à l'ouverture...

“L'accès **total et ouvert** aux données scientifiques **devrait devenir la norme** internationale pour l'échange des données scientifiques issues de la recherche financée **sur fonds publics**” (OCDE, 2007, [en ligne](#))



... car des bénéfices à la clé

- › Garantir **qualité** grâce à **traçabilité** des données
- › Favoriser **réutilisation** (fouille ...)
- › Augmenter **visibilité** des chercheurs
- › Diminuer recherches redondantes et gérer de façon **transparente** les fonds publics
- › Favoriser Interaction **science / société**

... mais dans les limites du droit fixé

“as open as possible, as closed as necessary”

*Commission européenne*



## 2. Pourquoi se préoccuper de ses données ?

### De l'incitation ... à l'injonction

#### EUROPE

**2019 : Directive  
2019/1204/UE**

= Principe de libre accès aux données de la recherche (sur fonds publics)

**2021 : Programme  
financement Horizon  
Europe**

= Obligation d'ouverture des données

#### FRANCE

Cadre juridique

**Loi pour une République numérique (2016)**

Mesures institutionnelles :

**Plan national pour la science ouverte 1 et 2 (2018 => 2024)**

- ★ Ouverture = principe par défaut
- ★ Appui à l'ouverture & partage (données & codes sources)

**Agence nationale de la recherche (ANR) :** Depuis 2019, obligation :

- Plan de gestion des données et
- Dépôts des publications issues du projet dans archive ouverte

# BULAC

[ספרייה] [大学] [γλώσσες] [අන්වීක්ෂණ]

Bibliothèque universitaire  
des langues et civilisations

## III. PRINCIPES ET OUTILS D'UNE BONNE GESTION DES DONNÉES



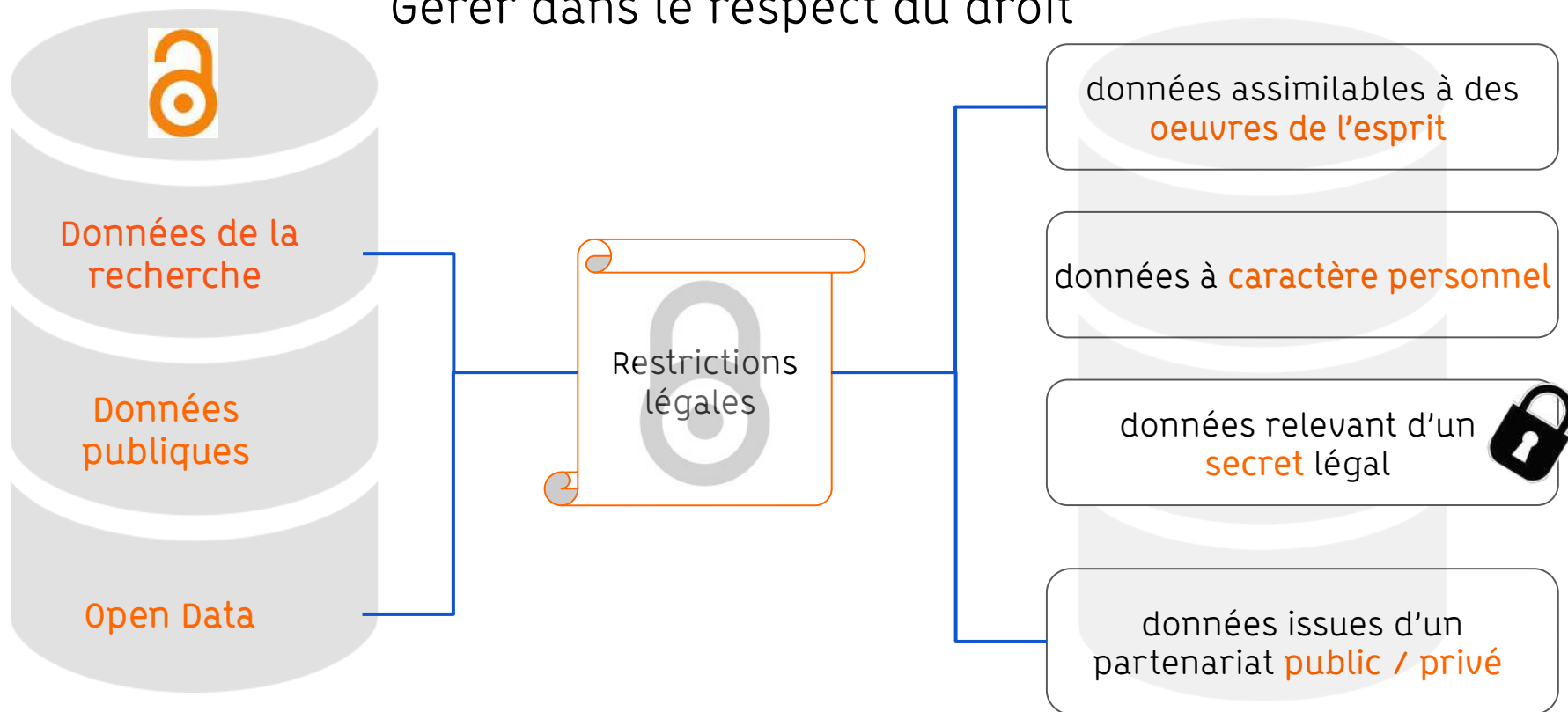
## 3. Principes et outils d'une bonne gestion



Partez sur de bonnes bases  
juridiques



## Gérer dans le respect du droit



### 3. Principes et outils d'une bonne gestion

données assimilables à  
oeuvres de l'esprit



**Quel cadre légal** : Droits de propriété intellectuelle

**Quelles démarches** :

- obtenir autorisation pour collecte, traitement et partage
- mention obligatoire de l'auteur

données à caractère  
personnel



**Quel cadre légal** : RGPD

**Quelles démarches**

- obtenir [consentement](#) (collecte, traitement, partage) - [lien](#)
- anonymisation

données dans le cadre  
d'un projet partenarial



**Quel cadre légal** : Si données financées à + de 50% par acteur privé → pas d'ouverture par défaut

**Quelles démarches** : pour tout type de partenariat → "accord de consortium"

## 3. Principes et outils d'une bonne gestion

---



Adoptez  
les bonnes pratiques

### Établir un Plan de Gestion des Données (*Data Management Plan*)

“Document synthétique [...] qui explique pour chaque jeu de données **comment seront gérées les données** d'un projet, depuis leur création ou collecte **jusqu'à leur partage ou archivage**”  
(Doranum)

[ressources](#)

[Pas-à-pas PGD](#)

Quels financements ?

Répartition des responsabilités

Règles d'accès et de partage des données

Modalités de collecte/production de données

À qui revient la propriété des données ?

*What's your plan ?*

Modalités de documentation des données

Données sensibles / personnelles = quel traitement ?

Quel plan d'archivage pour les données ?

Modalités de stockage et sécurisation des données

## 3. Principes et outils d'une bonne gestion

À quoi ressemble un  
Plan de Gestion des  
Données ?



5 min.



Parcourez rapidement ces deux PGD :

- Observez la typologie des rubriques
- Identifiez une à deux différences structurelles entre les deux PGD



PGD du projet "Ce qui nous concerne"

PGD du projet "SourcEncyMe"

La bonne adresse



Modèles

## Appliquer les principes FAIR

- **Métadonnées** précises, complètes
- Attribuer un **identifiant pérenne** aux données / jeux de données

Findable

- Choisir un **espace de stockage** accessible
- Donner *a minima* accès aux **métadonnées** associées

Accessible

- Attribuer aux jeux de données une **licence** fixant les droits d'usage
- Associer une **documentation** aux données (description, contexte collecte ...)

Reusable

- Produire des métadonnées adossées à des **référentiels**
- Inclure des **liens** vers d'autres jeux de données

Interoperable

# BULAC

[ספרייה] [大学] [βιβλιοθήκη] [सुप्रसिद्ध]

Bibliothèque universitaire  
des langues et civilisations

5 minutes

Pause ... questions



# BULAC

[ספרייה] [大学] [γλώσσες] [අන්දම]

Bibliothèque universitaire  
des langues et civilisations

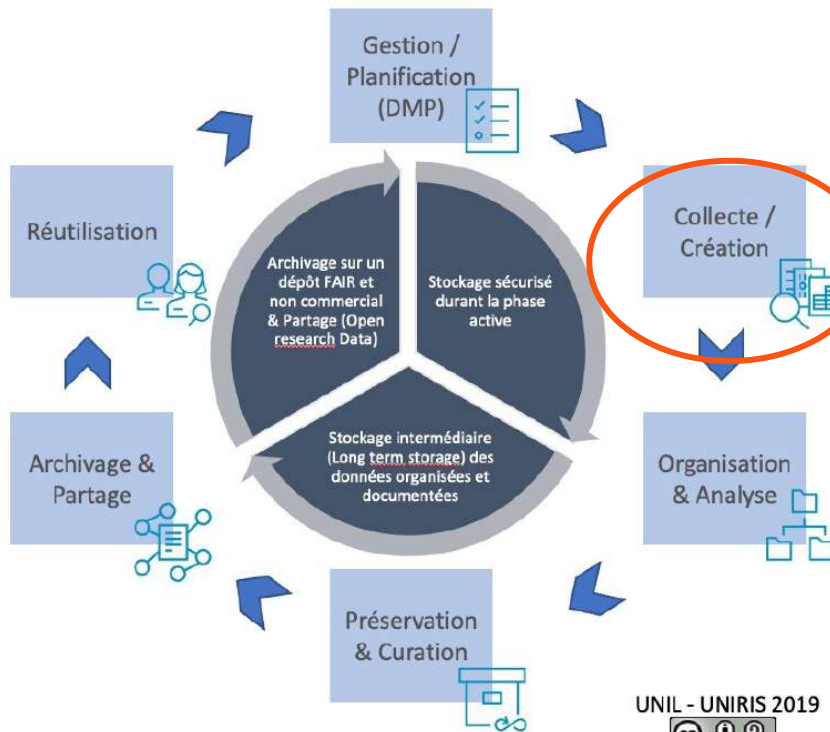
## IV. LA BONNE GESTION DES DONNÉES EN ACTION





### Une bonne gestion tout au long du cycle de vie des données

“Ensemble des étapes de gestion, de conservation, de diffusion et de réutilisation des données scientifiques liées aux activités de recherche”  
(Cirad)



6 étapes

- Planification
- Collecte / Création
- Organisation / analyse
- Stockage intermédiaire
- Archivage / partage
- Réutilisation

## 4. La bonne gestion en action : collecter

### La collecte

Sur le terrain

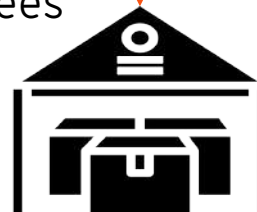


Données à caractère personnel ? Obtenir [consentement](#)

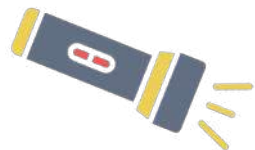
Entrepôts ...

- généralistes
- (pluri)disciplinaires
- par type de données

Entrepôt de données



Trouver un entrepôt adapté à ses besoins



Google

[lien](#) Dataset Search



[lien](#)

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

[lien](#)



Crédits : Freepik

### Gros plan : entrepôts de données en SHS



disciplinaire  
SHS

[lien](#)



[lien](#)



par types de  
données

[lien](#)



institutionnels

[lien](#)



pluridisciplinaire -  
généraliste

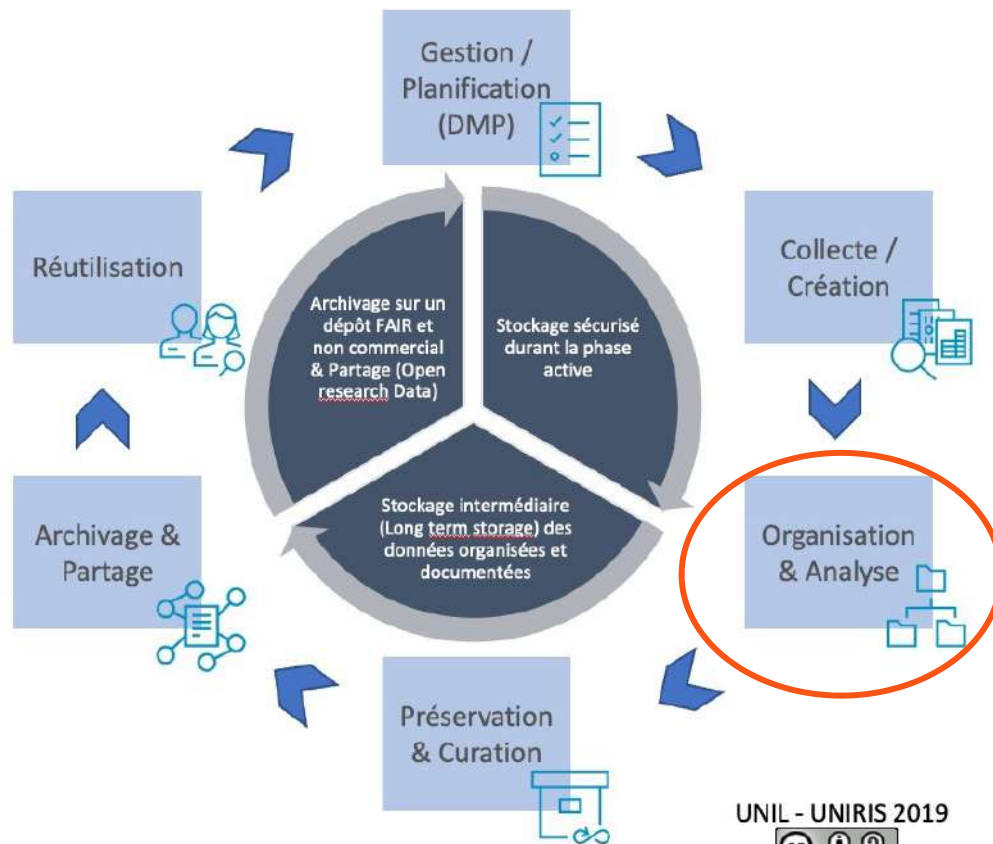
[lien](#)



[lien](#)



## 3. La bonne gestion en action : organiser & traiter



## 4. La bonne gestion en action : organiser &amp; traiter

## Organiser et traiter

## Nommage des fichiers

- noms descriptifs, courts
- exclure : ? ! : / ; # \
- pas d'espace => bulac\_test\_1
- format date : AAAAMMJJ
- n° de version : 01 ; 02 ; 10  
U01 ; U02 ; U10

projet collectif =  
plan de classement  
dossiers



## Privilégier les formats ouverts

.odt ou .txt	plutôt que .docx
.csv	plutôt que .xls
.jpg ou .png	plutôt que .tif
.wav ou .flac	plutôt que .mp3
.mkv	plutôt que .mp4



un format peut être  
"propriétaire"...mais ouvert  
ex : pdf

## 4. La bonne gestion en action : organiser & traiter



### Les métadonnées



embarquées

Utiliser un **standard** de métadonnées

enrichies

vocabulaires  
contrôlés



[Dublin Core](#)



- |   |   |
|---|---|
| 1. <b>Titre</b> nom de la/des donnée(s) | 09. Format (donnée / jeu de do.)        |
| 2. <b>Auteur</b> (producteur données)   | 10. Identifiant                         |
| 3. Sujet ou mots-clés                   | 11. Source                              |
| 4. Description (contenu)                | 12. Langue                              |
| 5. Editeur (resp. diffusion)            | 13. Relation (avec autres données)      |
| 6. Contributeurs                        | 14. Couverture spatio-temporelle        |
| 7. <b>Date</b> (création des données)   | 15. <b>Licence</b> (gestion des droits) |

[GeNames](#)

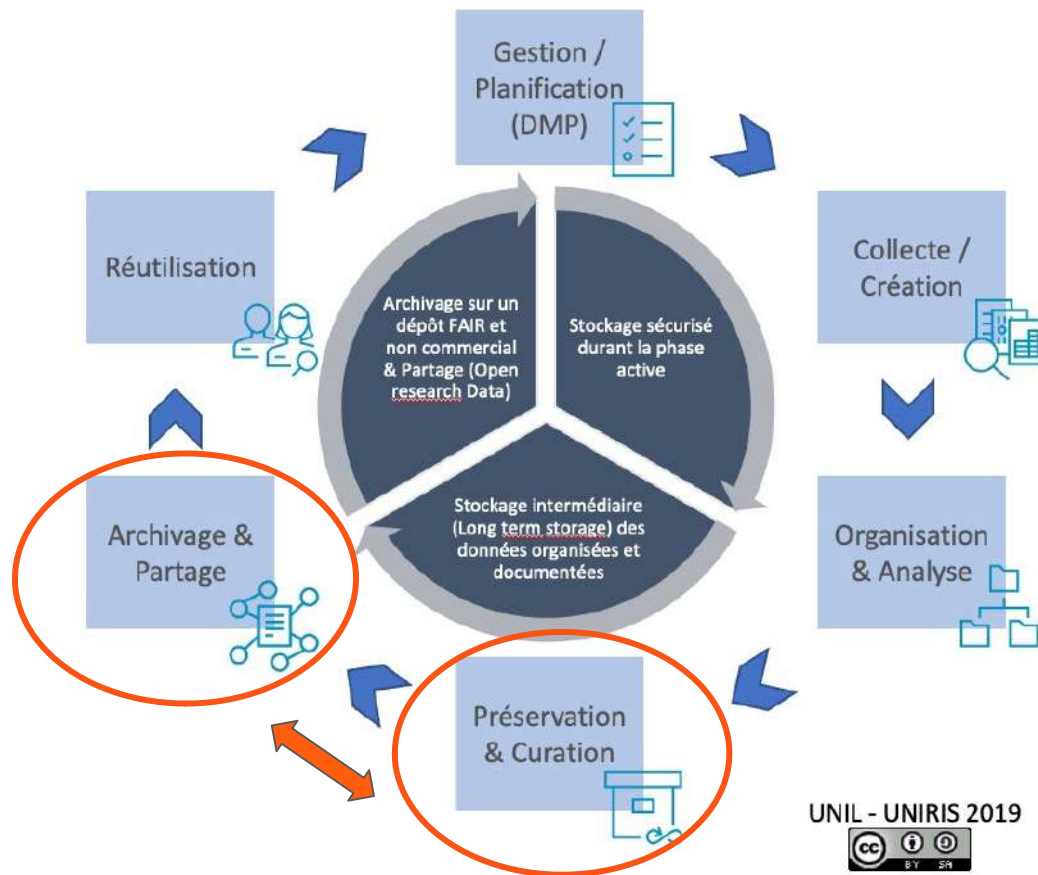
[Lien](#)

[PACTOLS](#)

[data.bnf.fr](#)

[Lien](#)

## 3. La bonne gestion des actions : stocker & partager



## 4. La bonne gestion en action : stocker

stockez !

Quelle  
différence ?

archivez !

Stockez vos données

= Pendant le projet

**Pourquoi**

Préserver pendant **tout le projet**

**Comment**

Méthode du **3 2 1**

**3** copies sur  
**2** supports différents ET au moins et  
**1** copie à distance

**Où**

++ serveur institutionnel  
+ support externe (DDur)  
≈ ordinateur (risque technique)  
≈ cloud (risque de sécurité)

**Sécurité**

antivirus, mot de passe

Archivez vos données

= A l'issue du projet

**Pourquoi**

- préserver certains types de données
- **accessibilité** sur long terme
- > Faire **un tri** !

**Comment**

- prévoir dès planification (**coûts** !)
- mise en oeuvre : une fois projet **achevé**

**Qui**

- scientifiques
- informaticiens
- archivistes et documentalistes
- juristes

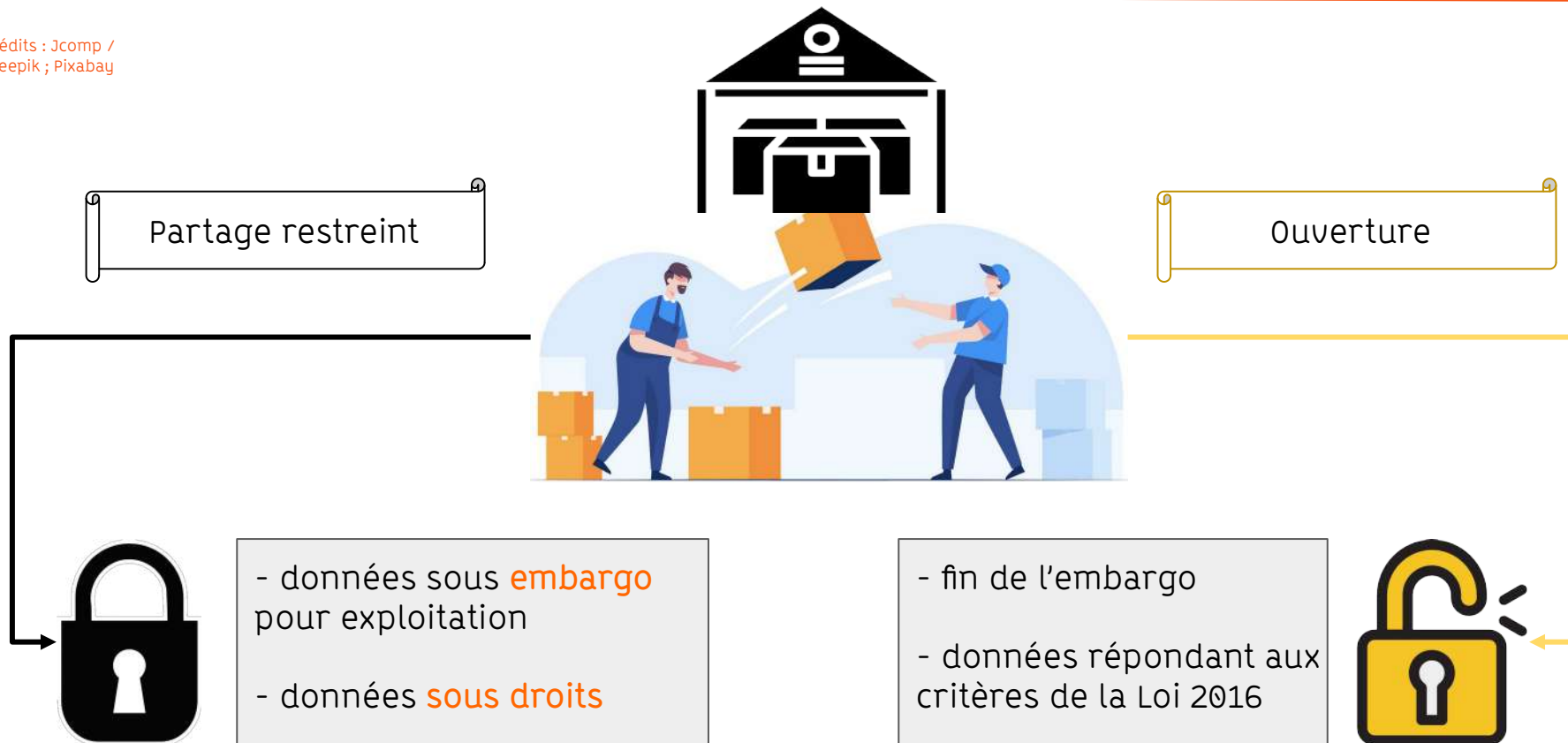
**Où**

- Le [Cines](#)
- => Huma-Num recommandé !



## 4. La bonne gestion des actions : stocker & partager

Crédits : Jcomp /  
Freepik ; Pixabay



## 4. La bonne gestion en action : partager

### 6. Lier données aux publications

- Option 1:



- Option 1+ :



exemple

### 5. Fixer conditions d'utilisation



Licence CC :  
[Mode d'emploi](#)



### 4. Attribuer des identifiants pérennes



1.

Partager ...  
toutes les données ?



### 2. Rappels : souvenez-vous ...

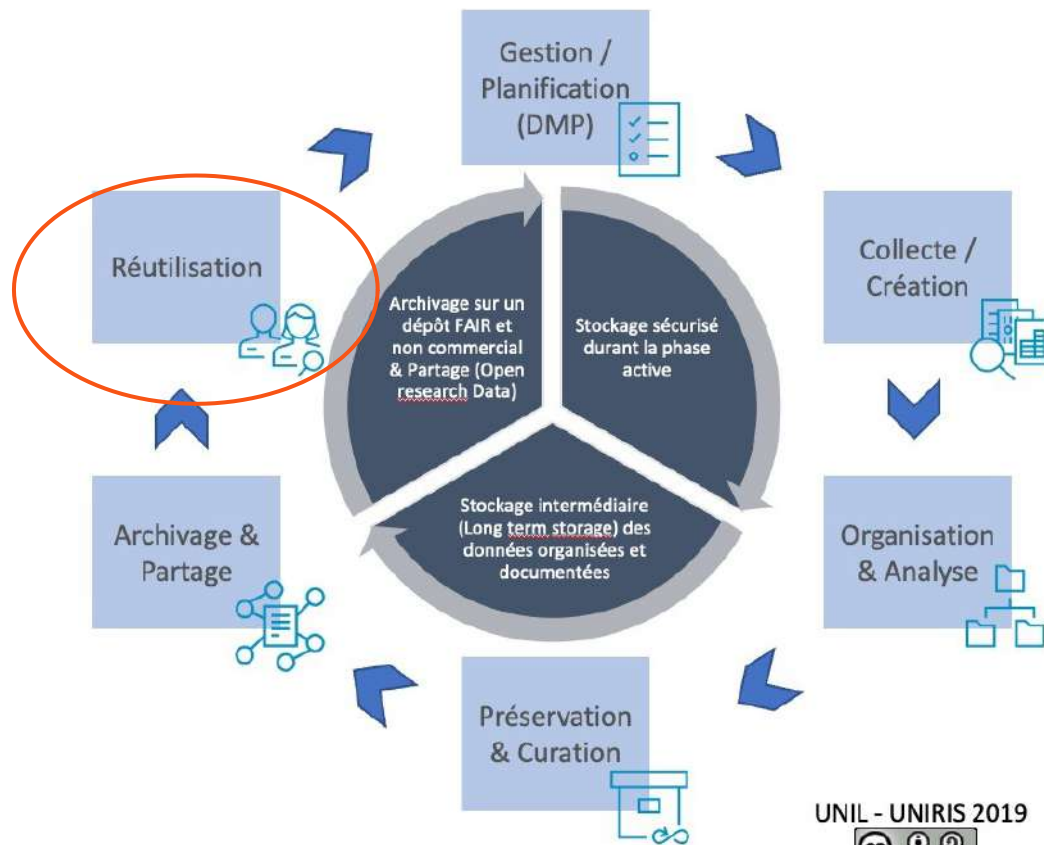
- Droits & autorisations obtenus ?
- Principes FAIR !

### 3. Choisir un entrepôt de données

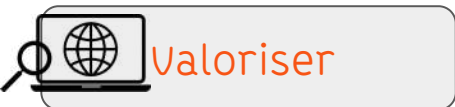


- Entrepôt
- disciplinaire +++
  - institutionnel ++
  - généraliste +

## 4. La bonne gestion en action : réutiliser & valoriser



## 4. La gestion en action : réutiliser & valoriser

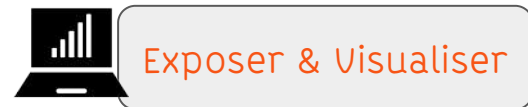
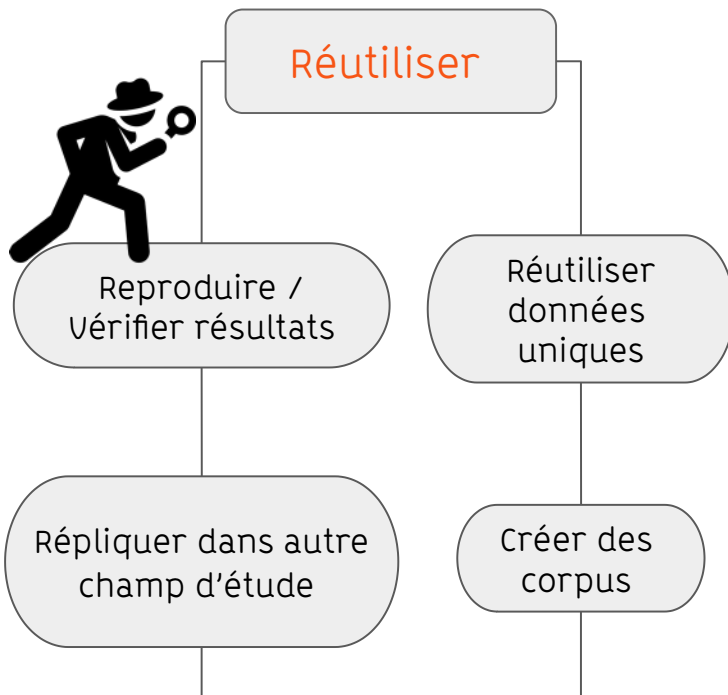


Publier des  
"articles de données"  
(Data papers)



Journal of Open  
**Humanities Data**

[Lien](#)



Structurer les données de  
manière **visuelle** :

- Faciliter l'analyse
- Faciliter la compréhension

Avec les **outils** appropriés

Gephi

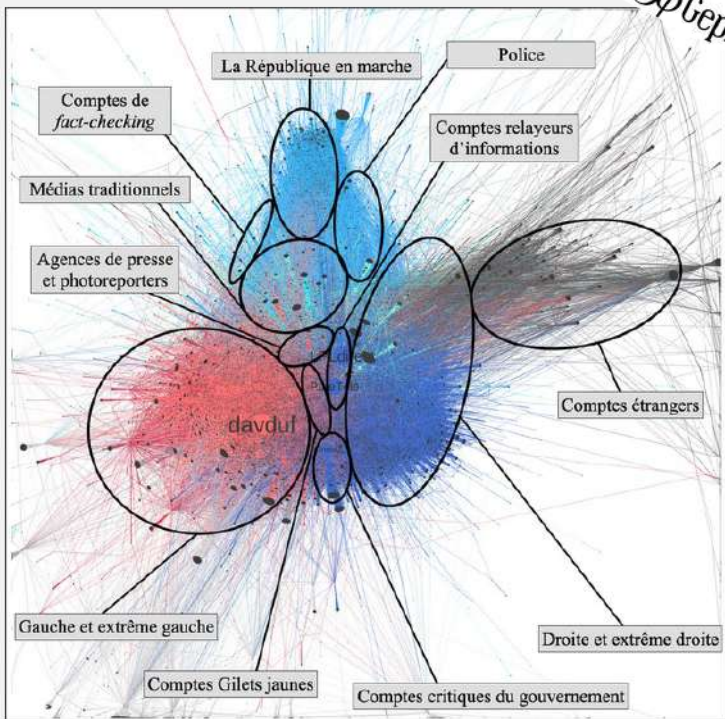
Khartis



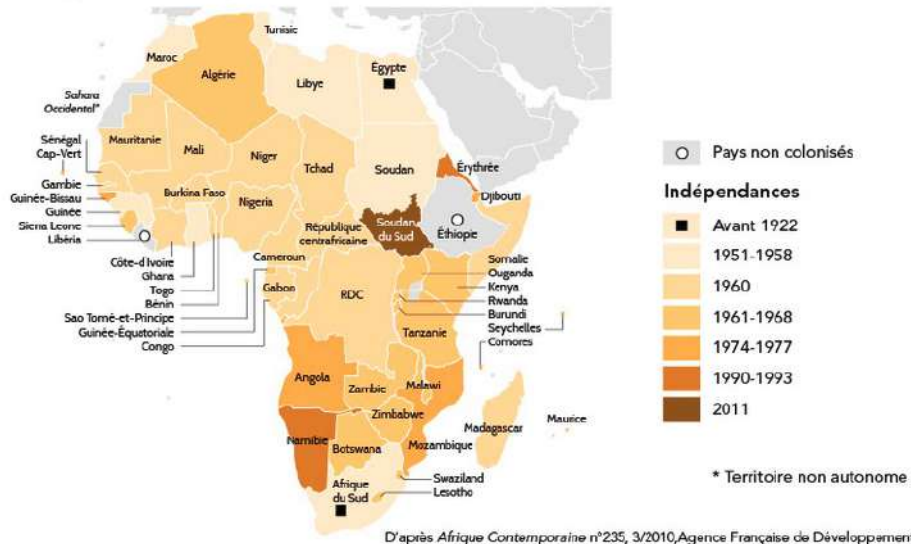
RAWGraphs

## 4. La gestion en action : réutiliser & valoriser

**FIGURE 3. COMMUNAUTÉS QUI S'INVESTISSENT AUTOUR D'IMAGES SUR TWITTER, À PROPOS DU MAINTIEN DE L'ORDRE, PENDANT LE MOUVEMENT DES GILETS JAUNES**



**Chronologie des indépendances africaines**



# Khartis

## 4. La gestion en action : réutiliser & valoriser

BULAC

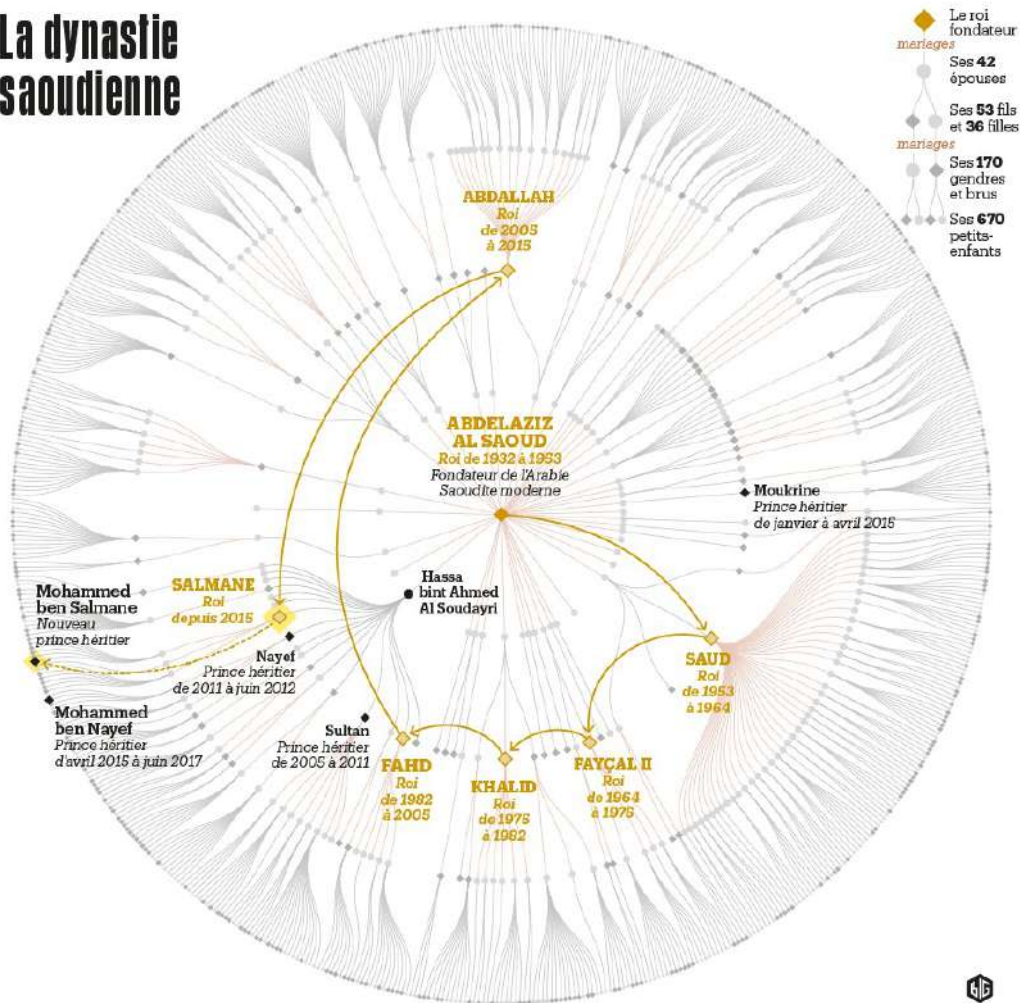
[בּוּלָאק] [大学] [βυλάκος] [ቅዕዳር]

Bibliothèque universitaire  
des langues et civilisations

## La dynastie saoudienne

### The Saudi dynasty

The complexity of the royal family of Saudi Arabia was a challenge. We chose this visualisation over the classic family tree that would have occupy too much space.



Graphique réutilisé dans un article de *Libération* du 21.06.2017

Auteur : Agence BIG / Réalisé avec RawGraph



# BULAC

[ספרייה] [大学] [βιβλιοθήκη] [අධ්‍යාපන]

Bibliothèque universitaire  
des langues et civilisations

POUR CONCLURE ... ET  
CONTINUER À SE FORMER



## Ressources généralistes :

- Doranum : <https://doranum.fr/> ; [Parcours interactif](#) sur la gestion des données de la recherche
- Université Aix-Marseille - LibGuides - les [données de la recherche](#)
- Les questions à se poser pour la diffusion des données, Doranum, [tutoriel en ligne](#)
- Comité pour la Science ouverte (CoSo) : - [Partager les données liées aux publications scientifiques](#)  
CoSo, [Guide d'application](#) de la Loi pour une République numérique pour les données de la recherche

## Ressources ciblées (SHS) :

- Coso [La diffusion numérique des données en SHS](#) (2018)
- InSHS, *Les sciences humaines et sociales et la protection des données à caractère personnel - Guide pour la recherche*, février 2021, [https://www.inshs.cnrs.fr/sites/institut\\_inshs/files/pdf/Guide\\_rgpd\\_2021.pdf](https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/Guide_rgpd_2021.pdf)
- La documentation d'Huma Num sur l'entrepôt Nakala : <https://documentation.huma-num.fr/nakala/>
- Carnet de recherche du consortium [DISTAM](#) - Agora numérique des études aréales
- Les ressources de Sciences Po (traitement et visualisation des données) :  
Medialab Tools : <https://medialab.sciencespo.fr/en/tools/>  
Ressources de l'Atelier de cartographie de ScPo : <https://www.sciencespo.fr/cartographie/ressources/>



# BULAC

[ספרייה] [大学] [βιβλιοθήκη] [අධ්‍යාපන]

Bibliothèque universitaire  
des langues et civilisations



[formation-publics@bulac.fr](mailto:formation-publics@bulac.fr)

