BULAC

[도서관] [शिक्षा] [ဘာသာစကား] [ᠰᡳᠯᡤᡳ]

Bibliothèque universitaire
des langues et civilisations

GÉRER SES DONNÉES DE RECHERCHE EN SHS

Enjeux et bonnes pratiques

Cycle Recherche

Printemps 2025

CC - BY - NC - SA

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

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

PAROLES DE CHERCHEURS …



Crédits : J. Stamp, Digitalbevaring.dk. CC BY 2.5.

BULAC
[עברית] [大学] [γλώσσες] [ཤེས་རིག]
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 ..."

Crédits : Jørgen Stamp, Digitalbevaring.dk. CC BY 2.5.

Source :

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

I. DÉFINITIONS ET TYPOLOGIES



Crédits : JørgenStamp, Digitalbevaring.dk.CC BY 2.5.

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)

BULAC
[עברית] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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

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

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*

quantitatives

qualitatives

produites

collectées

chaudes = en cours de traitement

froides = analysées

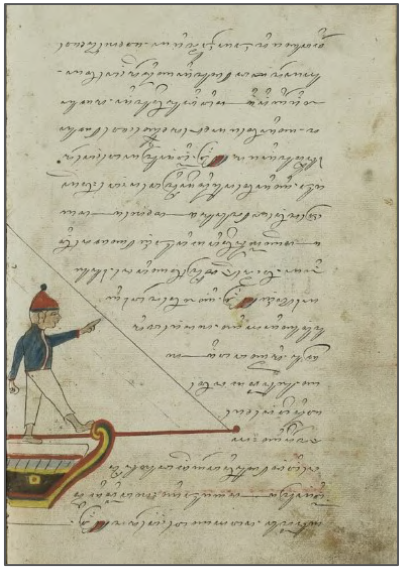
BULAC
[ספריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

1. Définitions et typologies

Mon corpus textuel



Mes photos d'archives



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



Autres données dérivées

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

Fichier Édition Affichage Insertion Outils Fenêtre Aide

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-Guimiliau
Marie-Louise Charran	Kermaria-Sulard
Philomène Cadoret	Bonen
Sarojini Naidu	Hyderabad

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

Mes enregistrements





Ma bibliographie

Crédits : Freepik ; Jazella ; G. Altmann ; M. Hassan / Pixabay

BULAC
[ספרים] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

1. Définitions et typologies

À quoi ressemble un corpus cohérent de données ?



8 min.



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 ? À quelle(s) condition(s) ?
- Dans quel(s) but(s) ces données sont-elles mises à disposition ?

zenodo

Anti-Jewish Pogroms in the Russian Empire



Matériaux de recherche - Thèse de Doctorat en Anthropologie sociale et ethnologie

Crédits : M. Hassan ; A. Hulsov / Pixabay

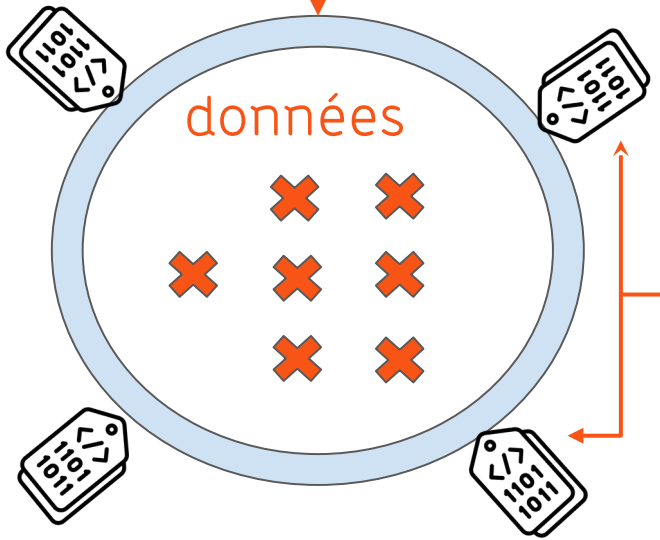
BULAC
[הפריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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



Collection

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)

Crédits : Flaticon (Juicy_Fish) ; Freepik (Smashicons)

BULAC

[הפריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

II. POURQUOI SE PRÉOCCUPER DE SES DONNÉES : LES ENJEUX



Crédits : J. Stamp, Digitalbevaring.dk. CC BY 2.5.

BULAC
[הפריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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

BULAC
[הפריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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

BULAC
[הפריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

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



Crédits : Jørgen Stamp, Digitalbevaring.dk. CC BY 2.5.

3. Principes et outils d'une bonne gestion



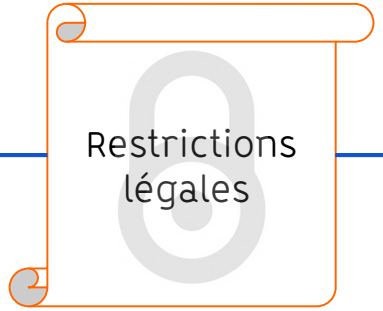
Partez sur de bonnes bases juridiques

Crédits : Jørgen Stamp, Digitalbevaring.dk. CC BY 2.5.

3. Principes et outils d'une bonne gestion

Gérer dans le respect du droit





- données assimilables à des **oeuvres de l'esprit**
- données à **caractère personnel**
- données relevant d'un **secret** légal
- données issues d'un partenariat **public / privé**

Crédits ; Freepik / Pixabay

BULAC
[עברית] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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 partenAB BLinePro**

⇨

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

Crédits : Jørgen Stamp, Digitalbevaring.dk. CC BY 2.5.

3. Principes et outils d'une bonne gestion

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

D'après : Cocaud, S. ; L'Hostis D, INRAE, 2019.

BULAC
[ספריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

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



DMP OPIDoR

Modèles

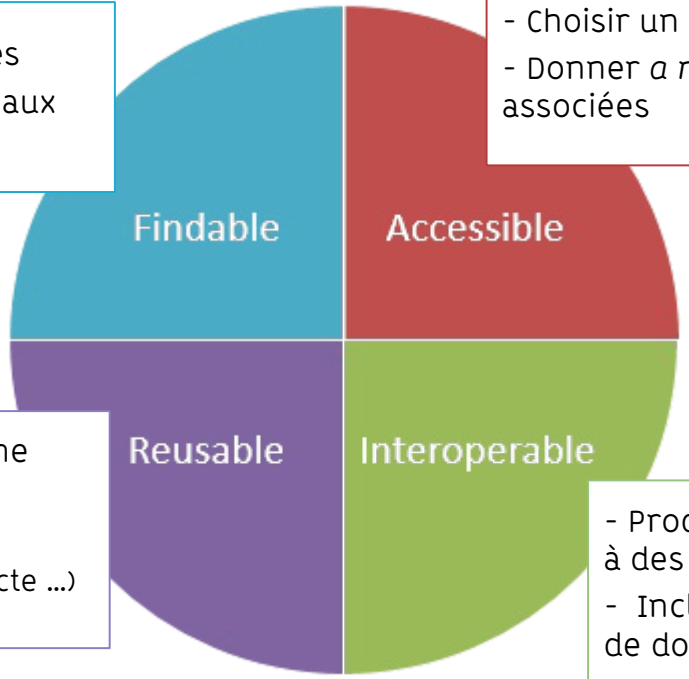
Crédits : M. Hassan ; A. Hulsov / Pixabay ;
vector tanks / Flatiocn

3. Principes et outils d'une bonne gestion

Appliquer les principes FAIR

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

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



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

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

Source : Université Paris-Saclay / Ifremer

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

5 minutes

Pause ... questions



Crédits : Freepik

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

IV. LA BONNE GESTION DES DONNÉES EN ACTION



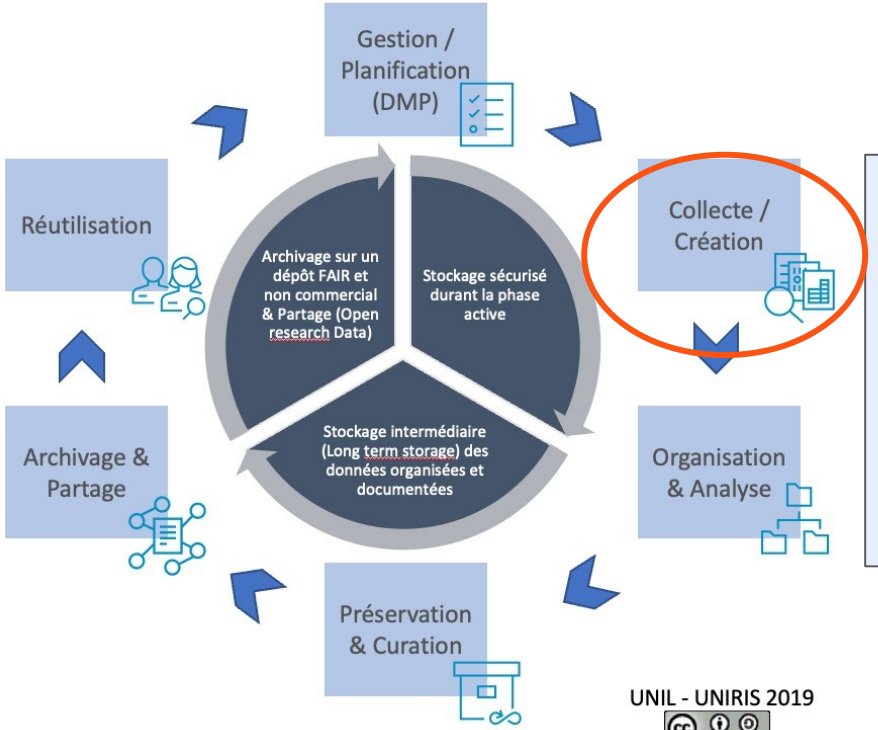
Crédits : Jørgen Stamp, Digitalbevaring.dk. CC BY 2.5.

4. Bonne gestion 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

CC BY SA

BULAC
[ספרים] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

4. La bonne gestion en action : collecter

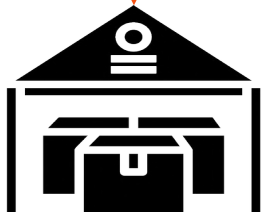
La collecte

Sur le terrain



Données à caractère personnel ? Obtenir [consentement]

Entrepôt de données



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

Trouver un entrepôt adapté à ses besoins



Google Dataset Search [lien]

Cat OPIDoR [lien]

re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES [lien]

Crédits : Freepik (Phatplus, IconGeek26, Faji Rama)

BULAC
[עברית] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

4. La bonne gestion en action : collecter



Crédits : Freepik

Gros plan : entrepôts de données en SHS



disciplinaire SHS

lien nakala

lien cessda

par types de données

lien COCON COllections de COrpus Oraux Numériques

lien Pangloss

mediHAL image, audio, vidéo lien

institutionnels

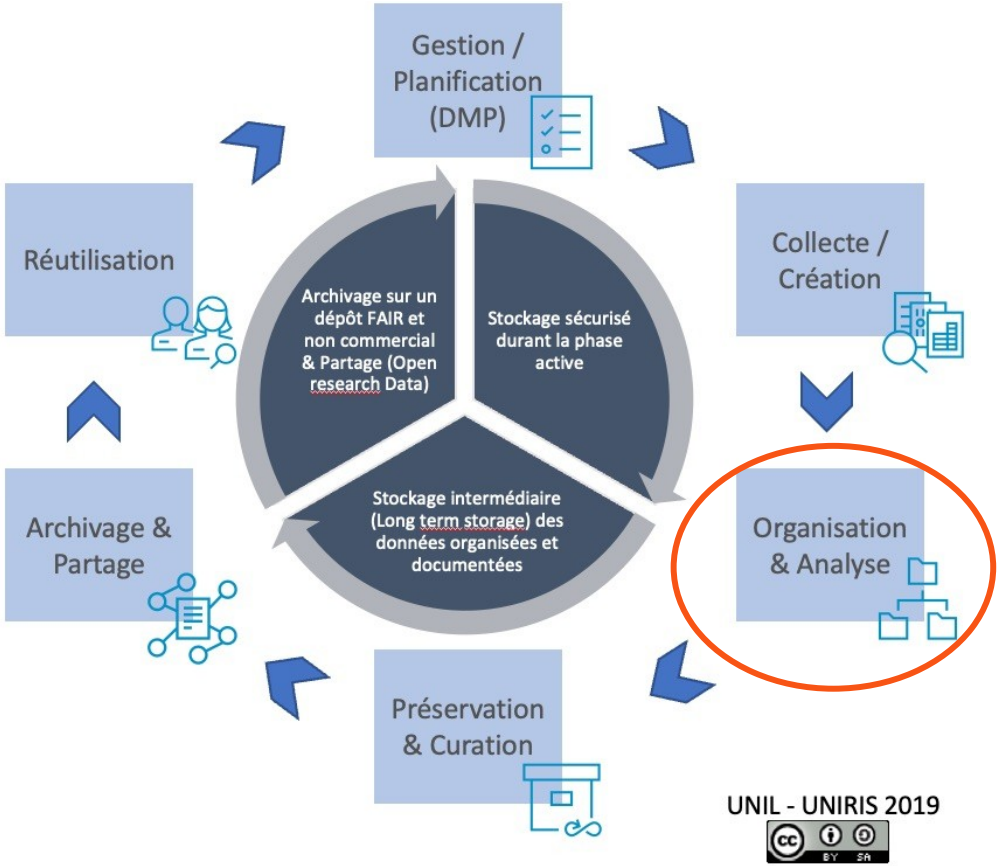
lien Didómena

pluridisciplinaire - généraliste

lien RÉPUBLIQUE FRANÇAISE Liberté Égalité Fraternité recherche.data.gouv.fr

lien zenodo

3. La bonne gestion en action : organiser & traiter



UNIL - UNIRIS 2019

4. La bonne gestion en action : organiser & traiter

Organiser et traiter

projet collectif =
plan de classement
dossiers

Nommage des fichiers

- noms descriptifs, courts
- exclure : ? ! : / ; # \
- pas d'espace => bulac_test_1
- format date : AAAAMMJJ
- n° de version : 01 ; 02 ; 10

V01 ; V02 ; V10

documenter

avec

métadonnées

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 qu

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

Crédits : Freepik

BULAC
[עברית] [大学] [γλώσσες] [ཨ་རུ་རིན]
Bibliothèque universitaire des langues et civilisations

4. La bonne gestion en action : organiser & traiter

Les métadonnées

embarquées

enrichies

Utiliser un standard de métadonnées

Dublin Core

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

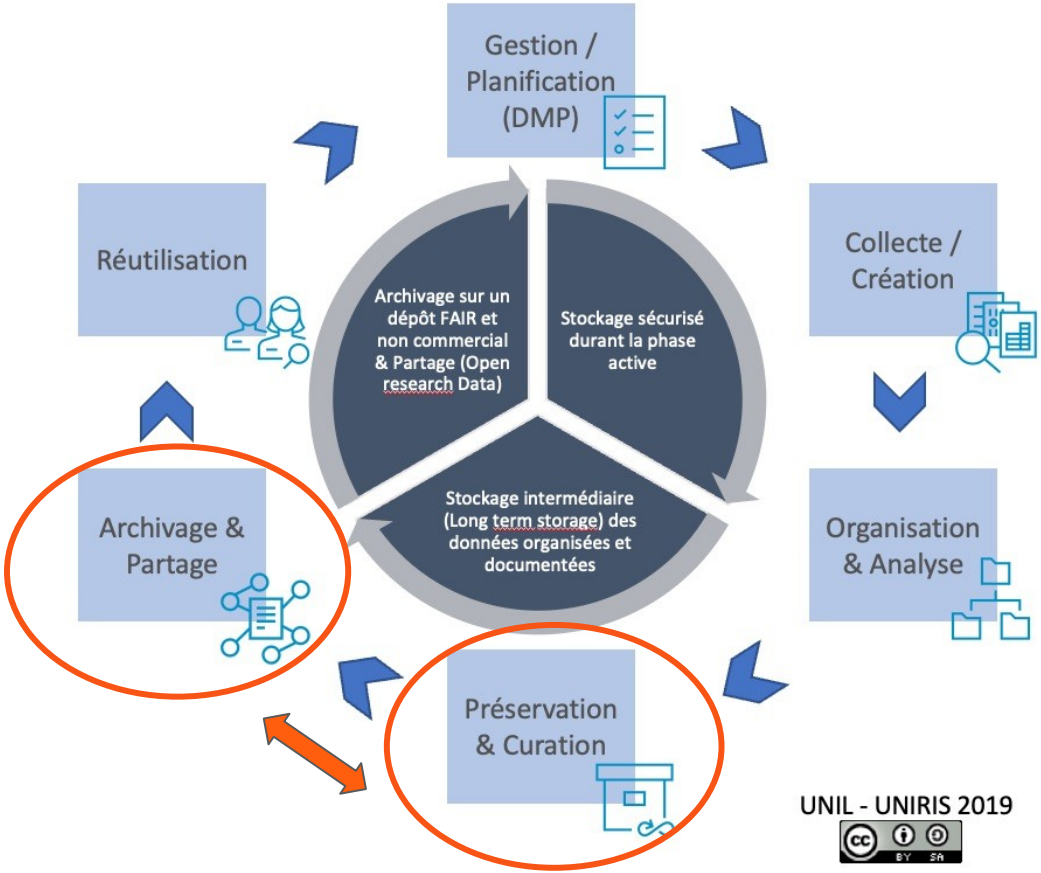
vocabulaires contrôlés

GeoNames
Lien

PACTOLS

data.bnf.fr
Lien

3. La bonne gestion den action : stocker & partager



UNIL - UNIRIS 2019

BULAC
[עברית] [大学] [γλώσσες] [ཨ་མདོ་རིག]
Bibliothèque universitaire des langues et civilisations

4. La bonne gestion en action : stocker



stockez !

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
= À 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 !

Crédits : Pixabay

BULAC
[ספריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

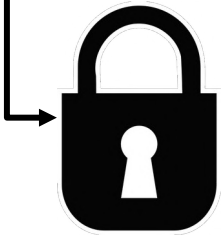
4. La bonne gestion den action : stocker & partager

Crédits : Jcomp / Freepik ; Pixabay



Partage restreint

Ouverture



- données sous embargo pour exploitation

- données sous droits

- fin de l'embargo

- données répondant aux critères de la Loi 2016



BULAC
[ספרים] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

4. La bonne gestion en action : partager

6. Lier données aux publications

- Option 1:
- Option 1+ :

exemple

1.
Partager … toutes les données ?



2. Rappels : *souvenez-vous …*

- Droits & autorisations obtenus ?
- Principes FAIR !

3. Choisir un entrepôt de données

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

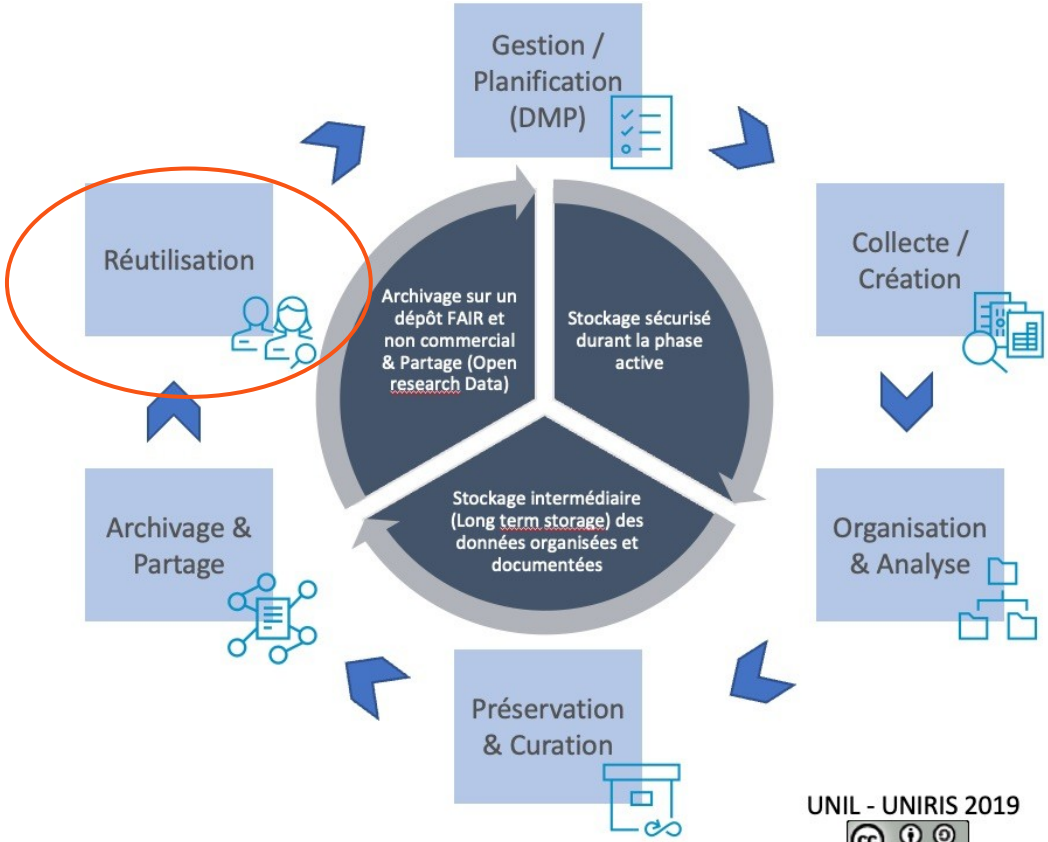
4. Attribuer des identifiants pérennes

5. Fixer conditions d'utilisation

Licence CC :

Crédits : Freepik / Flaticon (Icon home - Design circle)

4. La bonne gestion en action : réutiliser & valoriser



UNIL - UNIRIS 2019

BULAC
[הפריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

4. La gestion en action : réutiliser & valoriser

Valoriser

Publier des
"articles de données"
(Data papers)



Journal of Open Humanities Data

Lien

Réutiliser



- Reproduire / Vérifier résultats
- Répliquer dans autre champ d'étude
- Réutiliser données uniques
- Créer des corpus

Exposer & Visualiser

Structurer les données de manière visuelle :

➔Faciliter l'analyse
➔Faciliter la compréhension

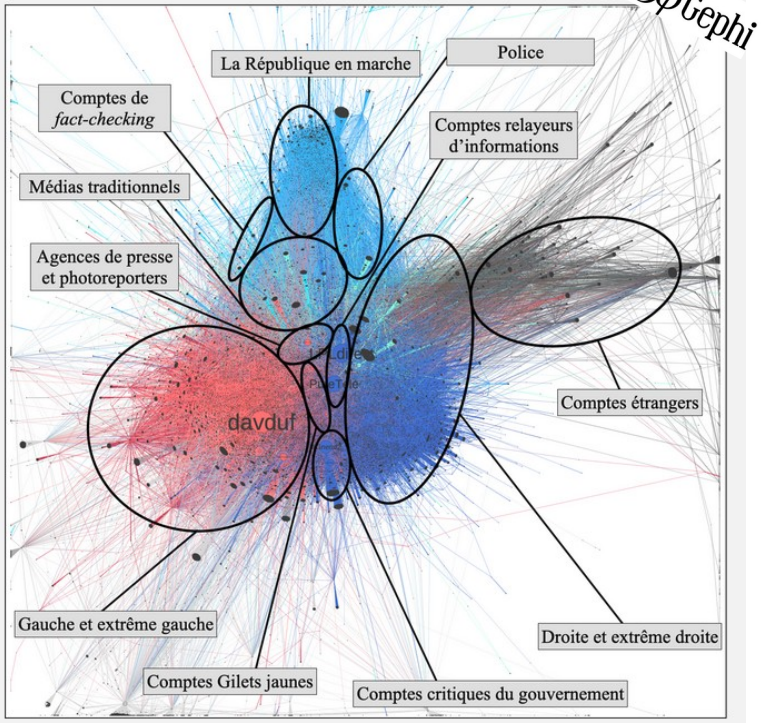
Avec les outils appropriés

Gephi
Khartis
omeka
RAWGraphs

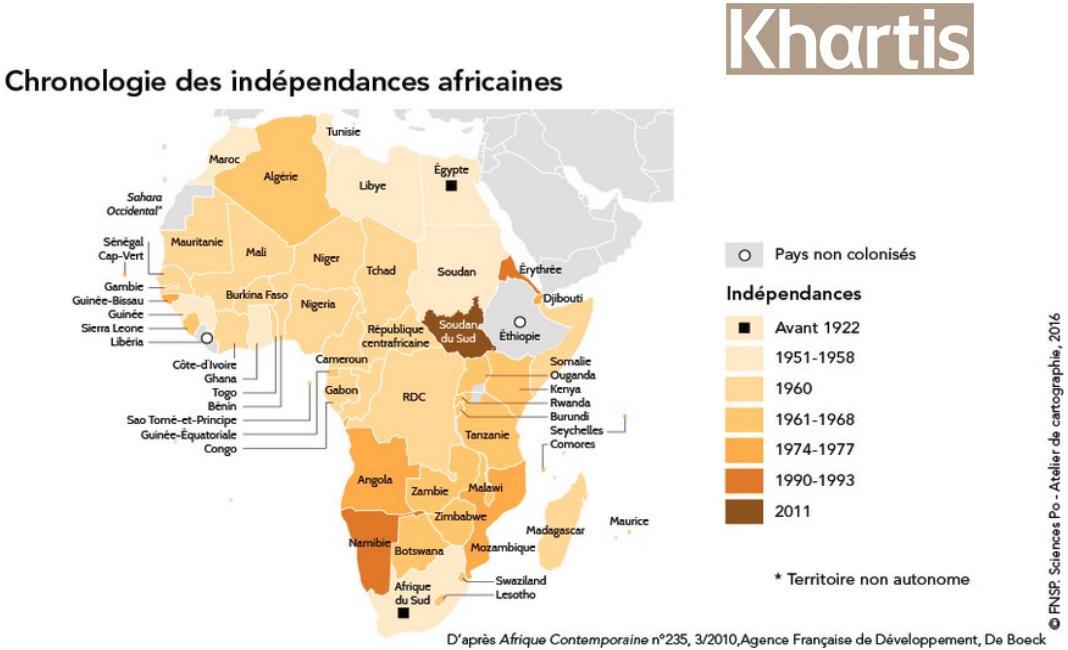
Crédits : M. Hassan, Pixabay / Freepik ; Pixabay

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

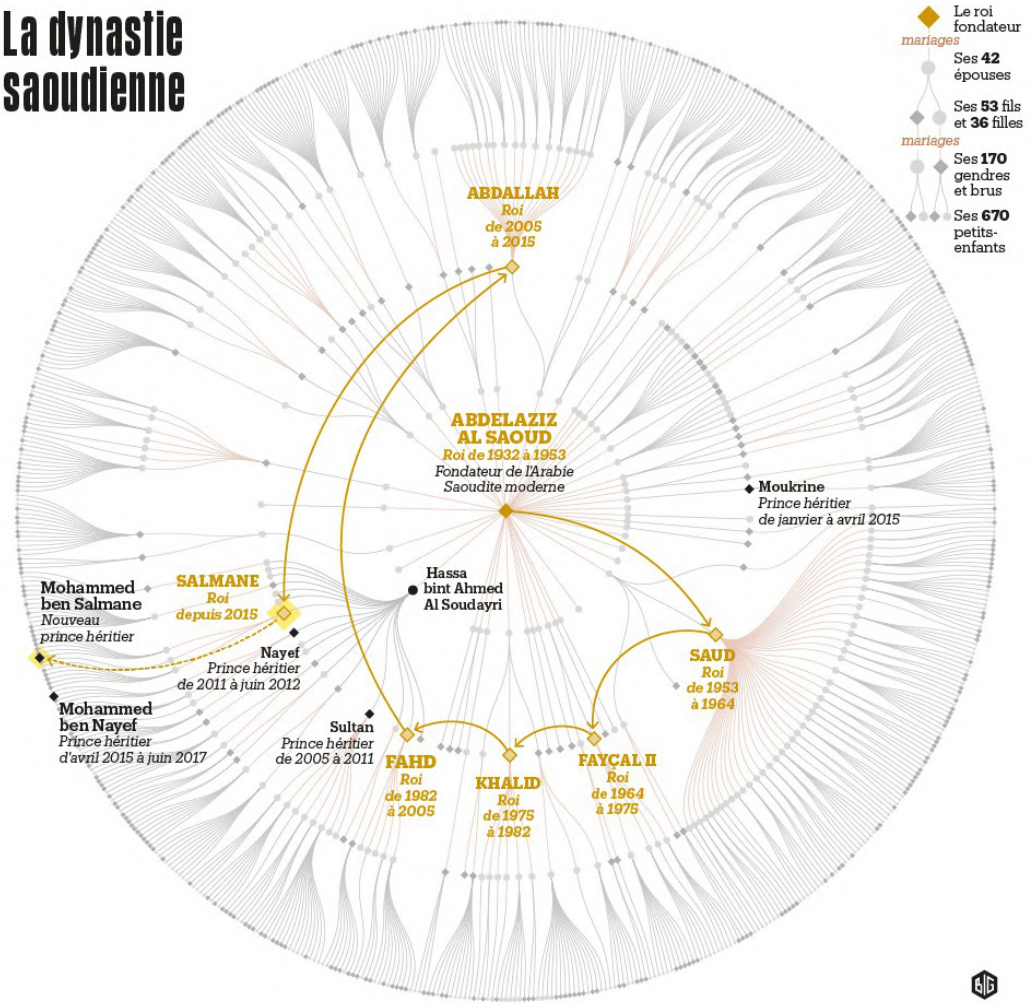


Edouard Bouté, *La mise en visibilité des forces de l'ordre sur Twitter pendant le mouvement des Gilets jaunes*, 2021



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.



Auteur : Agence BIG / Réalisé avec RawGraph

Graphique réutilisé dans un article de *Libération* du 21.06.2017

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

POUR CONCLURE … ET CONTINUER À SE FORMER



Crédits : Unsplash, photos for everyone

BULAC
[ספרים] [大学] [γλώσσες] [ཤེས་རིག]
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
- Guide juridique de l'entrepôt de données recherche.data.gouv.fr
- Université Aix-Marseille - LibGuides - les données de la recherche
- Cécile Arènes, Rédiger un plan de gestion des données, Sorbonne Université, 2023, Zenodo

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
- Véronique Ginouvès (entretien) :
  Les bonnes pratiques dans le cadre juridique et éthique pour la diffusion des données de la recherche, Doranum (chaîne Canal U)
- La documentation d'Huma Num sur l'entrepôt Nakala : https://documentation.huma-num.fr/nakala/
- Traitement et visualisation des données : les ressources de Sciences Po
  - › Medialab Tools : https://medialab.sciencespo.fr/en/tools/
    - › Ressources de l'Atelier de cartographie de ScPo :

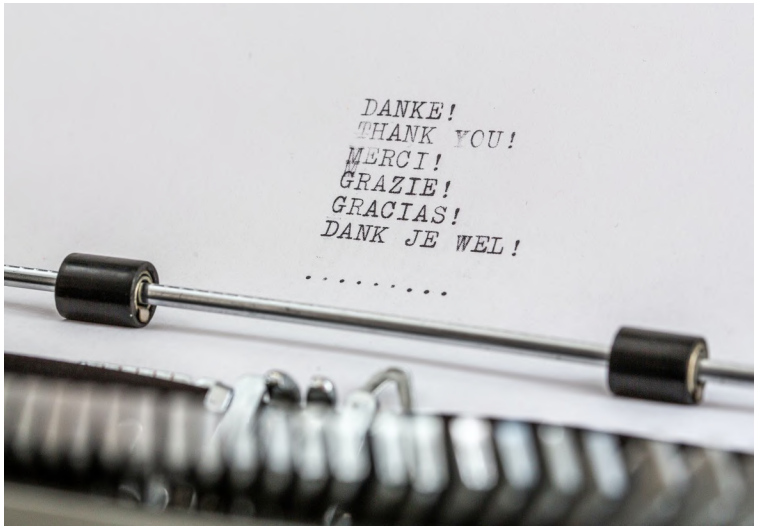
BULAC

[הפריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

services-recherche@bulac.fr

formation-publics@bulac.fr



Crédits : Wilhelm Gunckel / Unsplash