BULAC

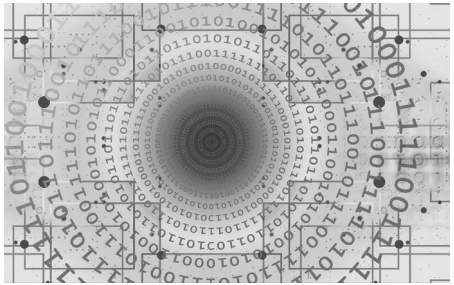
[도서관] [शिक्षा] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

DONNÉES DE LA RECHERCHE :
BIEN GÉRER LES DONNÉES DE SA THÈSE

Cycle doctoral

Automne 2026



CC - BY - NC - SA

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

Programme de la séance

1. Comprendre les données de la recherche

2. Connaître les bases d'une bonne gestion des données

3. Gérer ses données en action

Avant de commencer

Un petit tour de chauffe

Dans le *Chat*, notez :

- votre niveau
- votre discipline de recherche
- la ou les aires géographiques sur lesquelles portent vos recherches



Crédits : Pixabay

BULAC

[הפריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

1. COMPRENDRE LES DONNÉES DE LA RECHERCHE



Crédits : Gerd Altmann / Pixabay

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

1. Comprendre les données de la recherche

De quoi parle-t-on ?

> 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.

> Ces données [...] peuvent être de nature chiffrée, descriptive ou visuelle, brutes ou traitées [...], expérimentales, d'observation ou générées par une machine.

Crédits : OCDE (2007) / Université de Melbourne

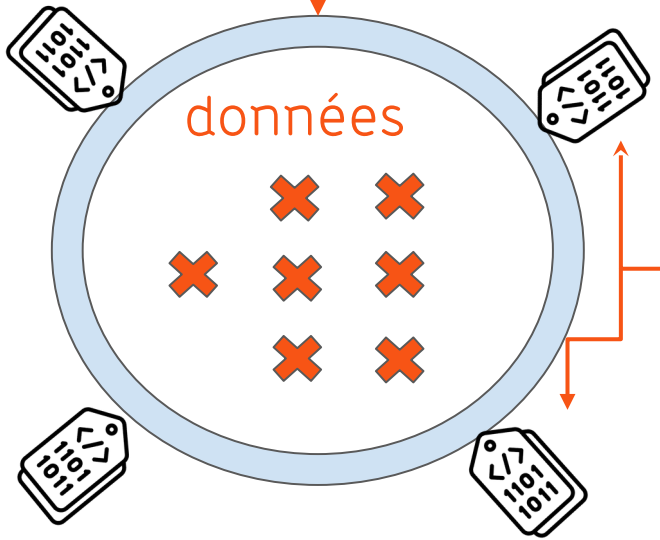
BULAC
[הפריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

1. Comprendre les données de la recherche

Jeu de données

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

Parfois, « donnée » est utilisé pour désigner tout un jeu



Exemples de jeux
n° 1 & n° 2

Métadonnées

Informations qui permettent d'identifier, de décrire et de renseigner sur l'origine d'un objet numérique comme un jeu de données (données sur la donnée).

Elles facilitent le partage et la réutilisation de la ressource.

Crédits : Flaticon (Juicy_Fish) / Freepik (Smashicons) / Doranum / Enssib

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

1. Comprendre les données de la recherche

Des réticences ?!

- « Des données ? je n'en ai pas ! »
- « Leur gestion est trop technique. »
- « Je vais me les faire voler si je les rends publiques. »
- « J'ai prévu des publications à partir de ces données, je ne peux pas les diffuser. »
- « Mon directeur de thèse s'oppose à la diffusion de mes données. »

en réalité…



Crédits : Passeport pour la Science ouverte / GTSO Couperin - 02.2021 / Freepik

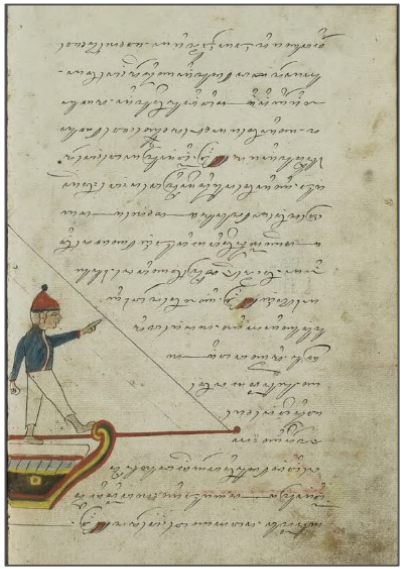
1. Comprendre les données de la recherche

Vos données en SHS

Corpus textuel



Photos d'archives



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

Bibliographie

Enregistrements



Tableurs

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

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

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

1. Comprendre les données de la recherche

Les différentes typologies de données

Les données collectées

Données d'observation / de terrain / brutes
Collectées sur le terrain (photo, enregistrement audio, inscription épigraphique, etc)

Données de référence
Publiées et accessibles au grand public (corpus de textes, films, images)

Données sources
Non publiées mais accessibles dans des institutions publiques ou privées (lettre manuscrite, brouillon d'écrivain, pièce d'archive)

Chaudes
en cours de traitement

Froides
analysées

Les données produites

Données dérivées
Issues de sélections ou de traitements de données de terrain (bibliographie, statistiques, fouille de texte)

Données expérimentales
Résultent d'un dispositif mis en place pour répondre à une interrogation initiale

Données de simulation
Générées par modélisation informatique (modèles économiques, climatiques)

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

1. Comprendre les données de la recherche

Une incitation forte à l'ouverture des données…

« 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 scientifique grâce à traçabilité des données
- Favoriser la réutilisation (fouille …)
- Augmenter la visibilité des chercheurs
- Diminuer recherches redondantes et gérer de façon transparente les fonds publics
- Favoriser l'interaction science / société

Mais dans les limites du droit !

« Ouvert autant que possible, fermé autant que nécessaire »

Crédits : Commission européenne

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

1. Comprendre les données de la recherche

Les doctorants en SHS sont-ils obligés d'ouvrir leurs données ?

Pas d'obligation à ouvrir les données (ni sur Internet ni sur l'Intranet national des thèses), **mais de fortes incitations** :

- à bien **documenter** les jeux de données avec leurs métadonnées associées
- à les **rendre accessible** dès que possible, dans les limites posées par réglementation
- à **lier** les futures publications aux données et les déposer dans un entrepôt dédié

dans tous les cas

Respecter les règles en vigueur :

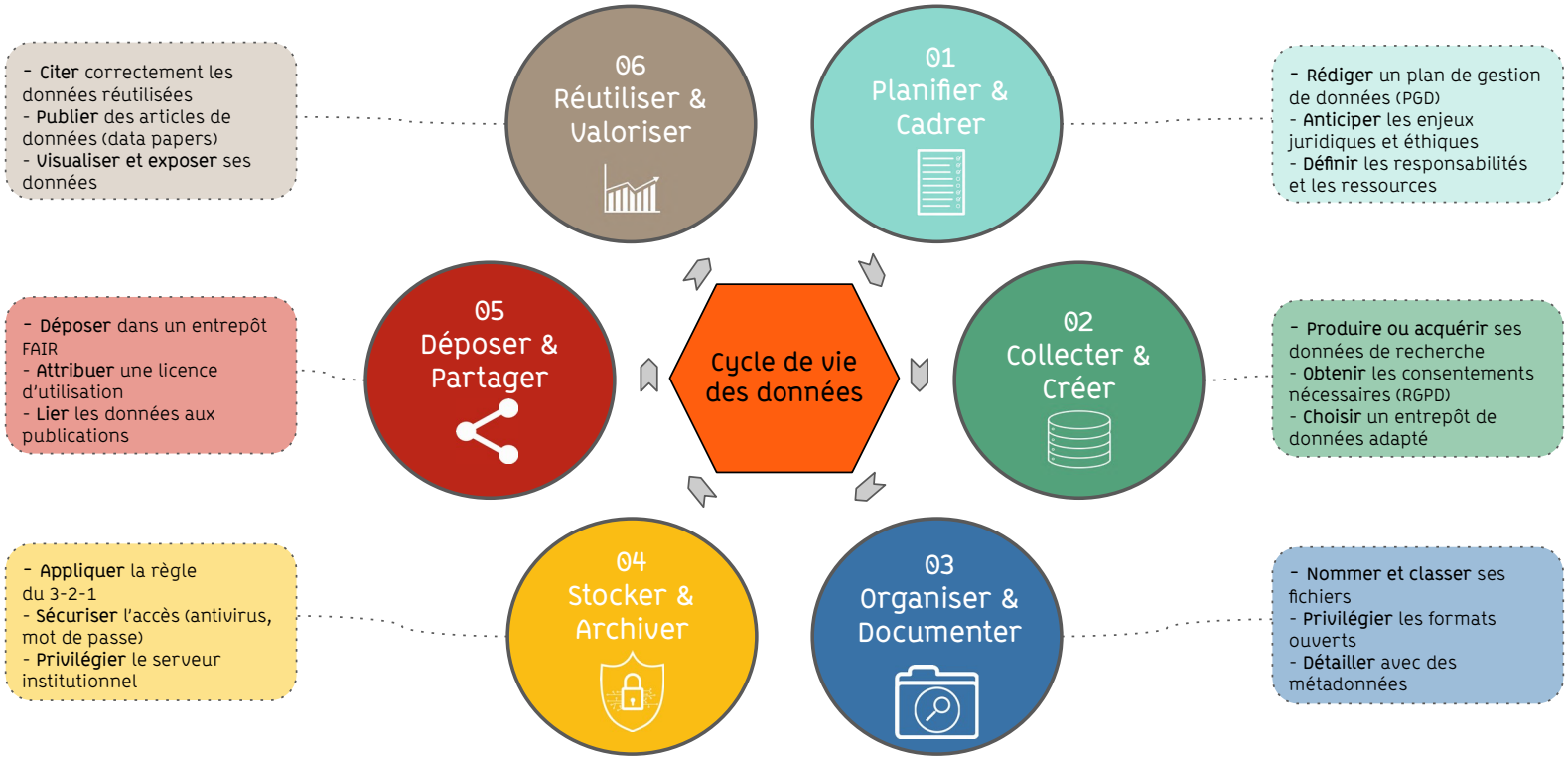
- la **législation** relative à l'usage et à la réutilisation des données
- les règles de citation
- les exigences de certains éditeurs pour accepter des publications ou des financeurs pour des projets collectifs post-doctoraux

Crédits : Paonkz / Flaticon

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

1. Comprendre les données de la recherche

Une préoccupation tout au long du cycle de vie des données



Crédits : BULAC / Pixabay

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

2. CONNAÎTRE LES BASES D'UNE BONNE GESTION DES DONNÉES

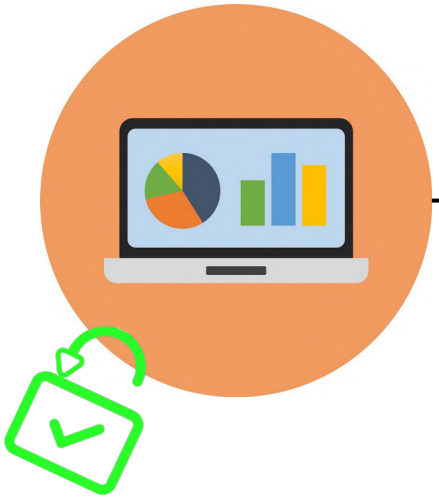


Crédits : Wikimedia Commons

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

2. Connaître les bases d'une bonne gestion des données

Gérer ses données dans le respect du droit



Application des restrictions légales



- données assimilables à des œuvres de l'esprit
- données à caractère personnel
- données relevant du 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

2. Connaître les bases d'une bonne gestion des données

données assimilables à des œuvres de l'esprit

→ Cadre légal : Droits de la propriété intellectuelle
Démarches :
- obtenir autorisation pour collecte, traitement et partage
- mention obligatoire de l'auteur

données à caractère personnel

→ Cadre légal : Règlement général de protection des données
Démarches :
- obtenir consentement pour collecte, traitement + partage (exemples Panthéon-Sorbonne et Inalco) et anonymiser les données

données relevant du secret légal

→ Cadre légal : Code des relations entre le public et l'administration
Démarches :
- occulter les données sensibles sans dénaturer le contenu du document, sinon ne pas les diffuser

données issues d'un partenariat public / privé

→ Cadre légal : Si données financées à + de 50% par acteur privé, exclusion du principe d'ouverture par défaut
Démarches :
- conclure en amont un « accord de consortium »

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

2. Connaître les bases d'une bonne gestion des données

Appliquer les principes FAIR à ses données

Findable



Permettre à toute personne intéressée d'avoir connaissance des données en interrogeant n'importe quel moteur de recherche, sans devoir explorer précisément la plateforme sur laquelle elles sont conservées.

- Décrire précisément les métadonnées associées
- Attribuer un **identifiant pérenne** aux jeux de données

Accessible



Une fois identifiées par un individu ou un moteur de recherche, les données doivent pouvoir être obtenues (éventuellement sous certaines conditions).

- Choisir un entrepôt **indexé par les moteurs de recherche**
- Donner accès *a minima* aux **métadonnées** associées

Interoperable



Faire en sorte que vos données puissent être utilisées conjointement à d'autres jeux de données ou par d'autres interfaces techniques que celles que vous avez vous-même utilisées.

- Utiliser des formats de fichiers ouverts **non-propriétaires** et des vocabulaires de métadonnées lisibles par différentes machines
- Associer **une documentation** aux données (description, contexte de collecte…)

Reusable



Indiquer très précisément comment et par qui vos données pourront être réutilisées.

- Attribuer aux jeux de données **une licence** fixant les droits d'usage
- Préciser la façon de **citer** les données

Source : Wikimedia Commons / Urfist de Paris

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

2. Connaître les bases d'une bonne gestion des données

Traiter ses données d'après leur cycle de vie

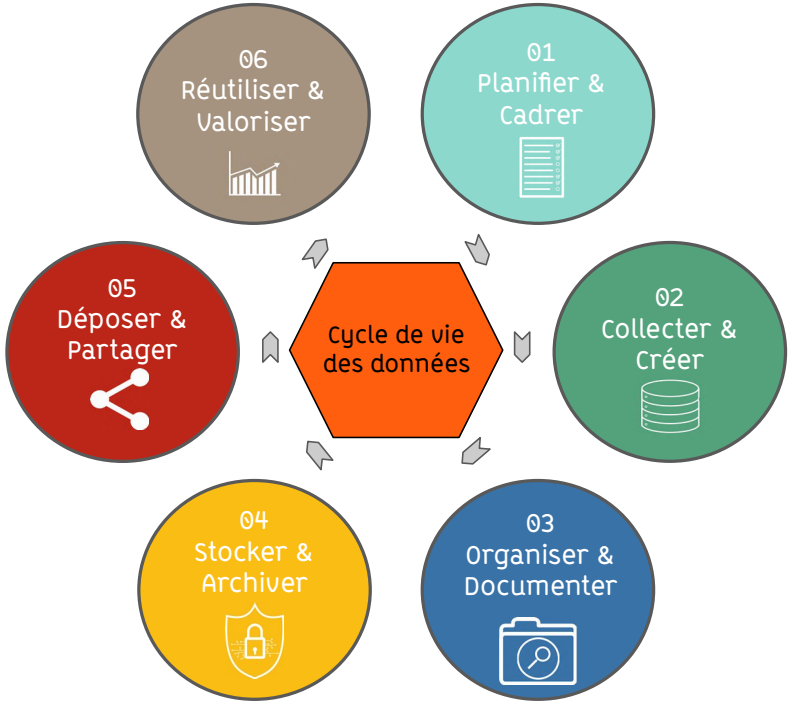
Votre projet

- la thèse et les futures publications associées

Vos objectifs

- bien traiter vos données pour une réutilisation ultérieure
- pouvoir lier vos jeux de données et vos publications qui les citent
- rendre accessible tout ou partie de vos données pour augmenter votre visibilité de chercheur et stimuler la recherche dans votre discipline

Exemple :
N. Richard, Matériaux de recherche - Thèse de Doctorat en Anthropologie sociale et ethnologie (EHESS)



BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

3. GÉRER SES DONNÉES EN ACTION



Crédits : Gerd Altmann / Pixabay

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

3. Gérer ses données en action

01 Planifier & Cadrer

Le plan de gestion de données (PGD)

C'est quoi ?

« 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. »

Tutos et exemples :
- Le PGD en bref (Doranum)
- Le PGD pas à pas (Doranum)
- Tuto de prise en main de DMP OPIDoR (Doranum)
- Modèles de PDG (DMP OPIDoR)

Que doit y figurer ?

- Financement
- Responsabilités
- Modalités de collecte
- Documentation des données
- Stockage et sécurisation
- Plan d'archivage
- Aspects juridiques et éthiques
- Propriété intellectuelle
- Règles d'accès et de partage

En doctorat, c'est obligatoire ?

Pas d'obligation pour le doctorant à ce jour, mais une assurance pour l'avenir de sa recherche.

Se familiariser avec un PGD, c'est un plus à valoriser auprès de :
- son comité de suivi individuel
- son jury de thèse
- ses futurs recruteurs pour des post-docs

Exemple du programme ANR LiPoL :
Enrichir la recherche sur la littérature populaire du Proche-Orient et sur le « moyen arabe » à partir de l'édition du *Roman de Baybars*.

- Le PGD du projet

Crédits : Doranum / Cocaud, S. / L'Hostis D. / INRA, 2019

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

3. Gérer ses données en action

02 Collecter & Créer

Collecter ses données

En en créant soi même



Produire ou acquérir ses données directement selon leur type : de terrain, de référence, sources (cf. partie 1)

Pour les données à caractère personnel, pensez à obtenir le consentement !

En réutilisant des jeux de données existants



Les différentes typologies d'entrepôts de données :

- Généralistes
- Disciplinaires
- Par type de données

Pour trouver le sien, il faut utiliser des répertoires spécialisées !

En trouvant l'entrepôt adapté à ses besoins



Google Dataset Search lien

Cat OPIDoR lien

re3data.org lien
REGISTRY OF RESEARCH DATA REPOSITORIES

Crédits : Freepik (Phatplus, IconGeek26, Faji Rama) / Pixabay

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

3. Gérer ses données en action

02 Collecter & Créer

Focus sur les entrepôts de données en SHS

Disciplinaire

Par types de données

Institutionnel

Généraliste

nakala lien

cessda lien

COCOON COllections de COrpus Oraux Numériques lien

mediHAL image, audio, vidéo lien

Pangloss lien

Didómena lien

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

zenodo lien

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

3. Gérer ses données en action

03 Organiser & Documenter

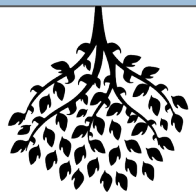
Organiser ses données

Bien nommer ses fichiers

- Noms descriptifs courts
- Exclure : ? ! : / ; # \
- Usage de l'underscore : bulac_test_1
- Format date : AAAAMMJJ
- N° de version : 01 ; 02 ; 10
  V01 ; V02 ; V10

Établir un plan de classement des dossiers

Une arborescence claire des fichiers permet de mieux partager ses dossiers dans le cadre d'un projet collaboratif.



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

LibreOffice

Un format peut être « propriétaire » mais ouvert… comme le PDF !

Crédits : Freepik

3. Gérer ses données en action

03
Organiser &
Documenter

Documenter ses métadonnées selon un standard

Embarquées

Dublin Core

01. Titre nom des données	09. Format (donnée, jeu de données)
02. Auteur (producteur données)	10. Identifiant
03. Sujet ou mots-clés	11. Source
04. Description (contenu)	12. Langue
05. Éditeur (resp. diffusion)	13. Relation (avec d'autres données)
06. Contributeurs	14. Couverture spatio-temporelle
07. Date (création des données)	15. Licence (gestion des droits)
08. Type (catégorie, nature)	

Enrichies

vocabulaires contrôlés

GeoNames Lien

Lien

data.bnf.fr Lien

Crédits : Freepik

3. Gérer ses données en action

04
Stocker &
Archiver

Stocker et archiver ses données

On stocke pendant le projet

Pourquoi	Préserver les données pendant toute la durée du projet
Comment	Méthode du 3 2 1 : 3 copies sur 2 supports différents ET au moins et 1 copie à distance
Où	++ Serveur ou cloud institutionnel + Support externe (disque dur) ≃ Ordinateur (risque technique) ≃ Cloud privé (risque de sécurité)
Sécurité	Antivirus, mot de passe

On archive à l'issue du projet

Pourquoi	Permettre l'accessibilité sur le long terme de certains types de données - Tout archiver ? Faire un tri !
Comment	À planifier dès le début du projet (coûts !) À mettre en œuvre une fois le projet achevé
Qui	- Scientifiques - Informaticiens - Archivistes et documentalistes - Juristes
Où	Le Cines en passant par Huma-Num

Crédits : Pixabay

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

3. Gérer ses données en action

05 Déposer & Partager

Déposer ses données dans un entrepôt…

Les étapes du dépôt

01. Sélectionner les données
Dépôt des données froides uniquement

02. Vérifier et préparer
Droits, formats ouverts, FAIR

Findable Accessible Interoperable Reusable

03. Choisir l'entrepôt
Disciplinaire, institutionnel, généraliste

disciplinaire +++
institutionnel ++
généraliste +

04. Attribuer un identifiant pérenne
DOI, Orcid

05. Fixer les conditions d'utilisation
Licence Creative Commons, Etalab

06. Lier les données aux publications
Deux options possibles

- Option 1:
- Option 1+ :

exemple

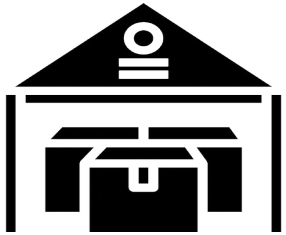
Avec Data Paper

Crédits : Freepik / Pixabay

3. Gérer ses données en action

05
Déposer &
Partager

… et définir leurs modalités de partage



Ouvert

Les données sont librement accessibles et consultables, voire téléchargeables sur leur entrepôt de dépôt.

Depuis 2016, principe de l'*open data* pour les données publiques.

À la fin de l'embargo, les données deviennent automatiquement accessibles.

Restreint

Les données apparaissent avec la mention « Privé » mais les métadonnées associées restent accessibles.

Dans le PGD, spécifier les catégories de personnes pouvant prétendre à l'accès aux données avec la mise en place dispositif technique de restriction d'accès.

Crédits : Jcomp / Freepik / Pixabay

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

3. Gérer ses données en action

06 Réutiliser & Valoriser

Donner une seconde vie à ses données

Valoriser

Publier des « articles de données » (Data paper)

Journal of Open Humanities Data — Lien

Data & Corpus — La revue des données en SHS

Réutiliser

Reproduire, vérifier des résultats

Réutiliser des données uniques

Les répliquer dans un autre champ d'étude

Créer des corpus

Exposer & visualiser

Structurer les données de manière visuelle pour :

- Faciliter l'analyse
- Faciliter la compréhension

Avec les outils appropriés :

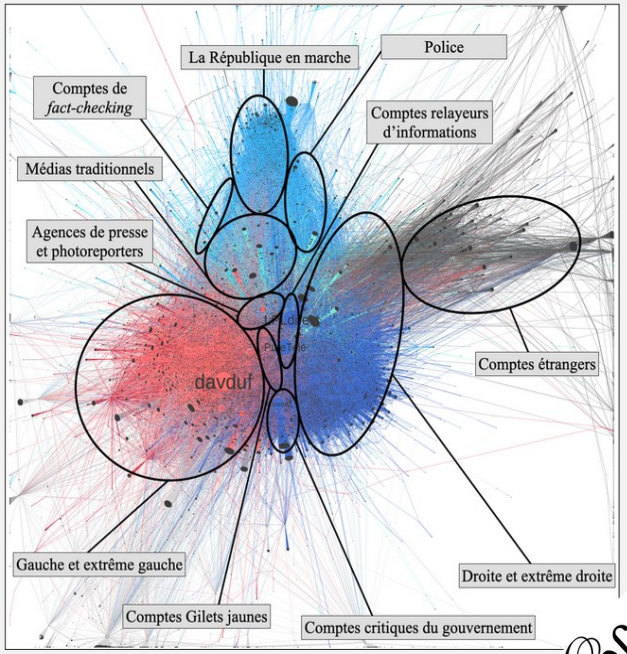
Gephi — Khartis — RAWGraphs — omeka

Crédits : M. Hassan, Pixabay / Freepik / Pixabay

3. Gérer ses données en action

06
Réutiliser & Valoriser

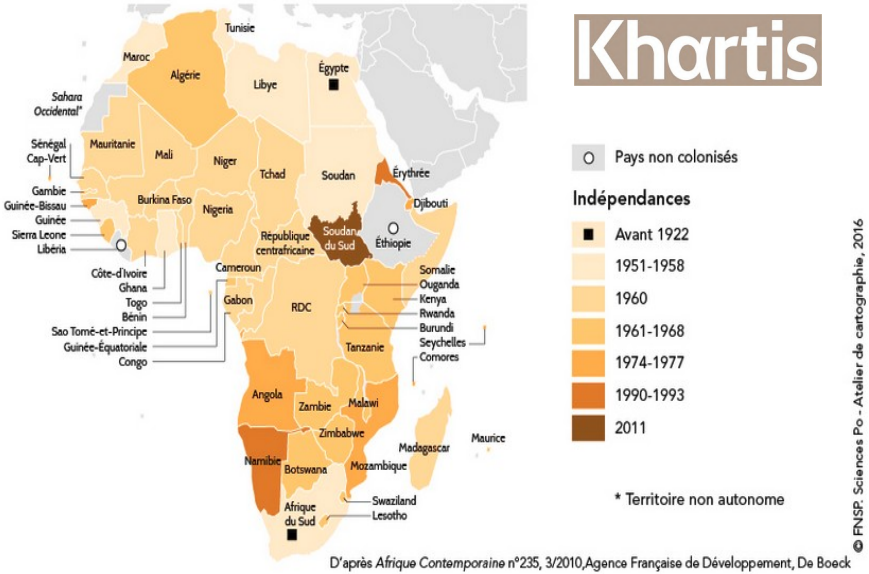
FIGURE 3. COMMUNAUTÉS QUI S'INVESTISSENT AUTOUR D'IMAGES SUR TWITTER, À PROPOS DU MAINTIEN DE L'ORDRE, PENDANT LE MOUVEMENT DES GILETS JAUNES



Gephi

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

Chronologie des indépendances africaines



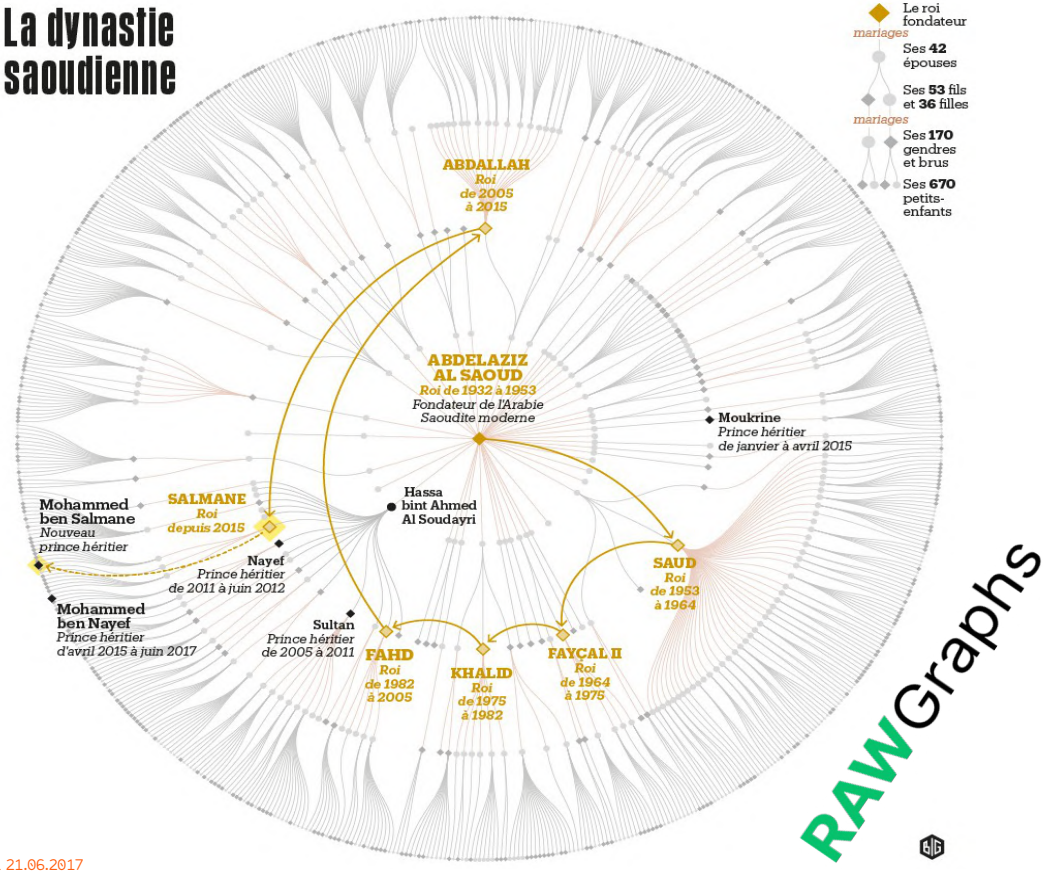
D'après *Afrique Contemporaine* n°235, 3/2010, Agence Française de Développement, De Boeck

© FNSP, Sciences Po - Atelier de cartographie, 2016

3. Gérer ses données en action

06
Réutiliser &
Valoriser

« *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.* »
(Agence BIG)



RAWGraphs

Crédits : 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 : SCY / Pixabay

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 :

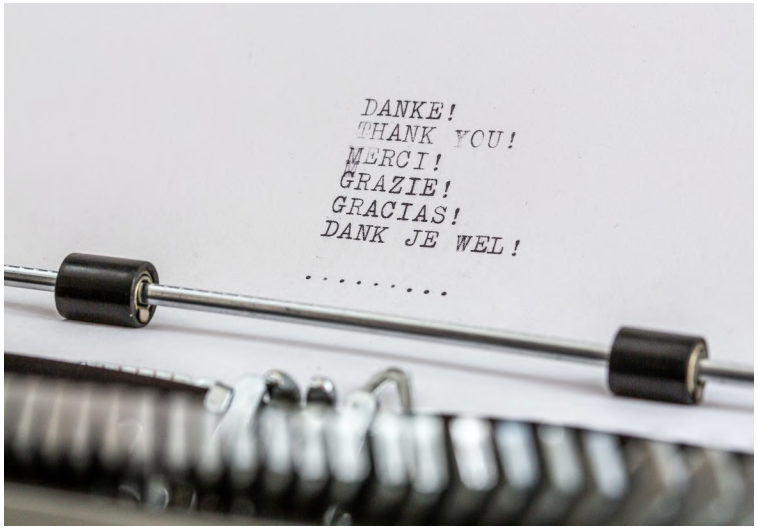
- Coso, La diffusion numérique des données en SHS (2018)
- InSHS, Les sciences humaines et sociales et la protection des données à caractère personnel - Guide pour la recherche, février 2021, https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/Guide_rgpd_2021.pdf
- Véronique Ginouvès (entretien) :
  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 : https://www.sciencespo.fr/cartographie/ressources/

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

formation-publics@bulac.fr
services-recherche@bulac.fr



Crédits : Wilhelm Gunckel / Unsplash