BULAC

[도서관] [शिक्षा] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

INTRODUCTION AUX HUMANITÉS NUMÉRIQUES

Cycle recherche

Equipe Formation

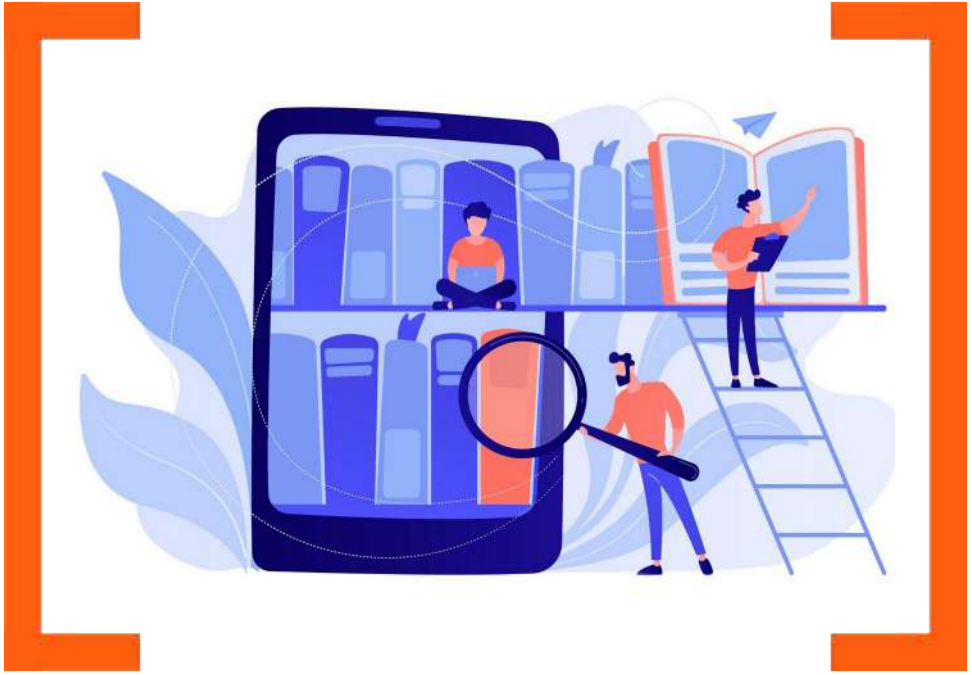
CC - BY - NC - SA

Le programme de la formation :

1. Les humanités numériques : périmètre et enjeux
   - ➔ *Définitions générales et enjeux*
   - ➔ *Exemples et apports pour la recherche en SHS*
2. Se lancer dans un projet d'humanités numériques : les bonnes pratiques
   - ➔ *Principes d'une bonne gestion des données*
3. Exemple d'utilisation des HN pour l'étude de manuscrits arabes
   - ➔ *Un exemple d'entraînement à la transcription automatique d'une graphie complexe*

BULAC
[도서관] [लिब्रा] [ספרייה] [ᵐᵉᵈᵃ]
Bibliothèque universitaire des langues et civilisations

1. LES HUMANITÉS NUMÉRIQUES : PÉRIMÈTRE, ENJEUX ET APPORTS



Crédits : Freepik (vectorjuice)

Parmi les chercheurs, absence de consensus :

Définition large : vision des humanités numériques comme « chapiteau » englobant toutes sortes de pratiques de recherche qui utilisent les technologies numériques



Crédits : Freepik (macrovector)

Définition restreinte :

Limitation du terme à des activités plus spécialisées “mobilisant de manière intensive des capacités calculatoires sur des jeux de données massifs pour produire des analyses spécifiques” (Pierre Mounier)



P. Mounier, *Les humanités numériques*, 2018

Crédits : Freepik (vectorjuice)

Qui est concerné par les humanités numériques ?



Crédits : Flaticon (Freepik)

1. Périmètre, enjeux et apports

Les humanités numériques : une "transdiscipline" concernant TOUTES les disciplines des sciences humaines :

- ➔ Sont une conséquence du "**tournant numérique**"
- ➔ **Ne font pas table rase du passé,** mais s'appuient sur l'héritage des humanités
- ➔ **Importent des méthodes, dispositifs et perspectives d'interprétation** liés au numérique dans le domaine des SHS

(« Manifeste des *digital humanities* », Paris THATCamp, 2010)

« Humanités » et « numérique », un drôle de mélange ?





Bref retour en arrière

Crédits : Flaticon (mikan933 ; ultimatearm)

1. Périmètre, enjeux et apports

Un projet pionnier dans l'histoire des humanités numériques : l'*Index Thomisticus* de Roberto Busa



Crédits : Wikimedia Commons ; Flaticon (Freepik)

1. Périmètre, enjeux et apports

Après guerre - Fin années 1970

Literary and Linguistic Computing

Années 1980

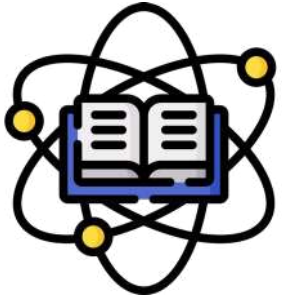
Humanities computing

Années 2000 à aujourd'hui

Digital Humanities + mouvement de la Science Ouverte







Crédits : Flaticon (Freepik)

Concrètement, qu'apportent les humanités numériques par rapport à une recherche "classique" ?



Crédits : Flaticon (Freepik)

Des apports pour la recherche en sciences humaines et sociales :

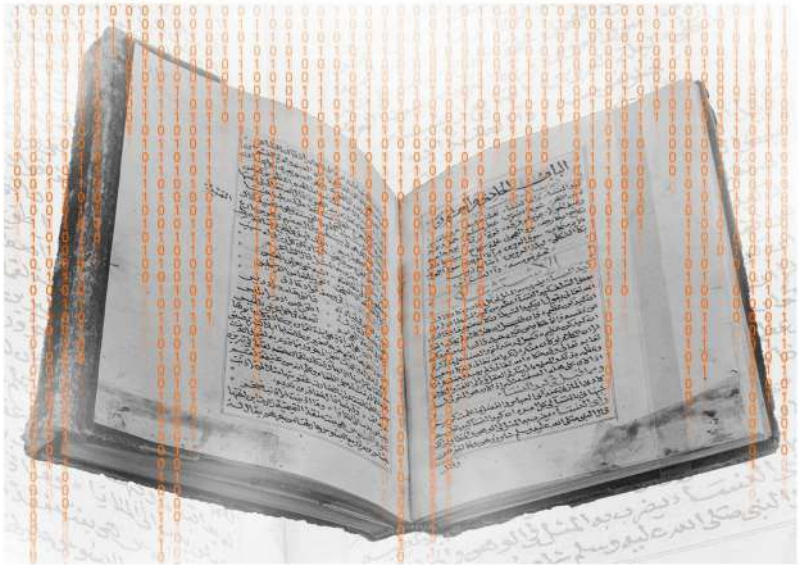
- Possibilité d'appliquer des traitements automatisés sur des corpus volumineux
- But : obtention de connaissances à partir du traitement et du filtrage de quantités massives de données (textuelles, sonores, iconographiques…) : fouille de données ou *data mining*
- Avantage de la lecture distante (*distant reading*) sur la lecture rapprochée (*close reading*) - Franco Moretti





Crédits : Flaticon (Freepik)

Apports des HN
Exemples

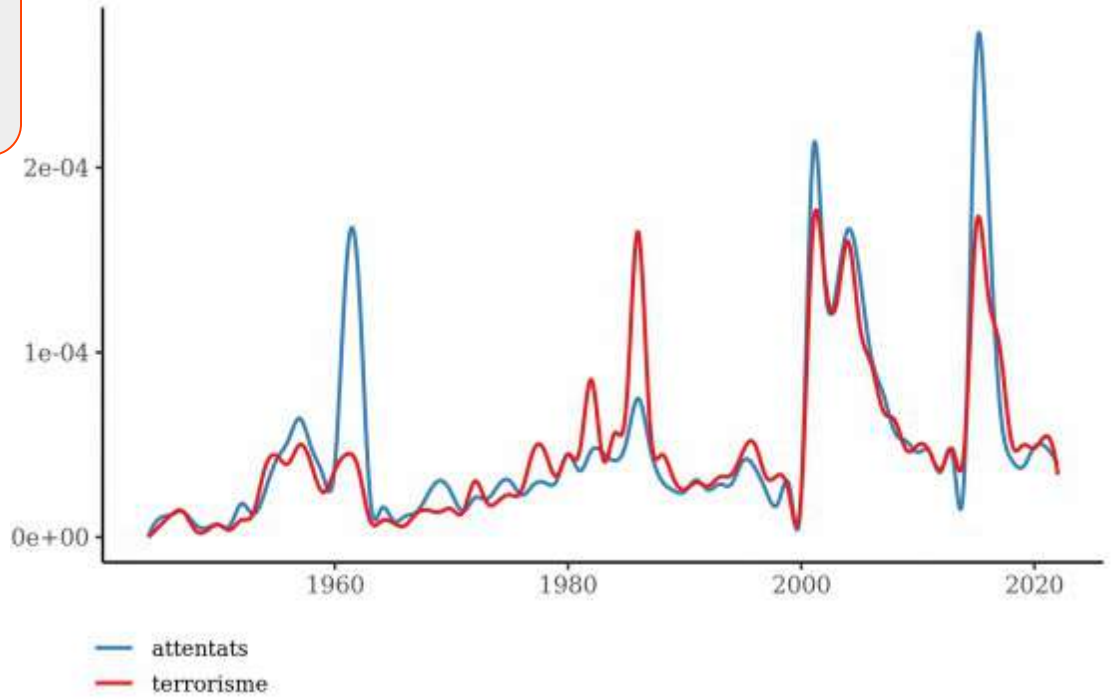


Crédits : M. Ruscio / BULAC

1. Périmètre, enjeux et apports

Analyse textométrique à partir de la fouille de texte

➔ Analyse statistique de la fréquence de deux termes dans les archives du *Monde* de 1944 à 2022 - outil : Gallicagram

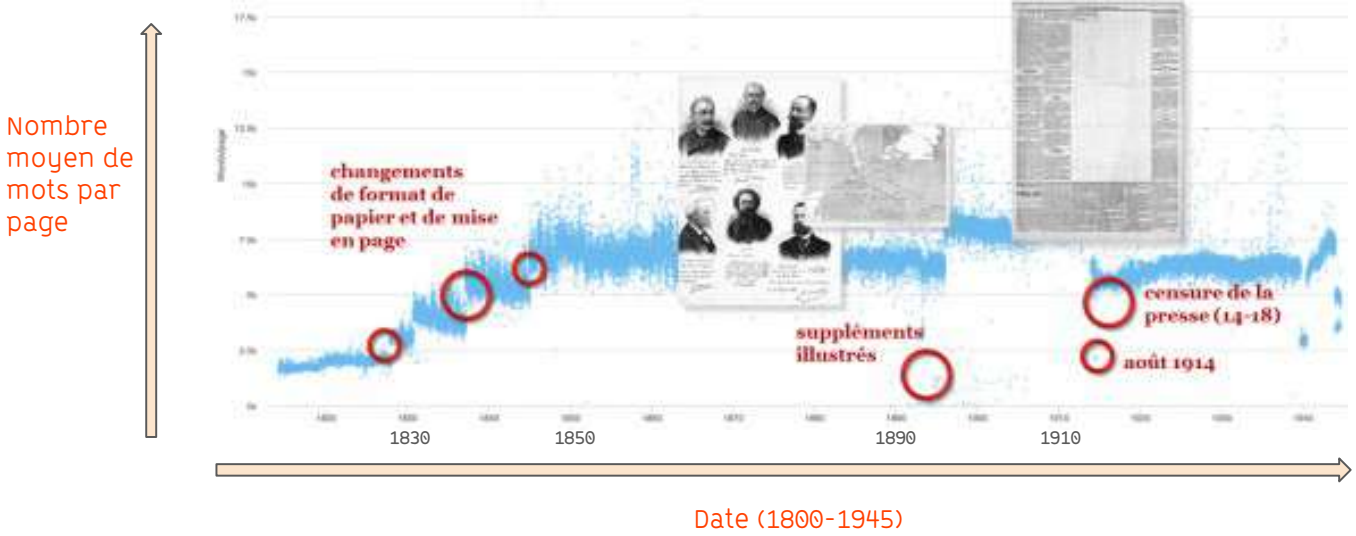


Source : https://journals.openedition.org/corpus/7944?lang=en

1. Périmètre, enjeux et apports

- **Formalisation** des modèles d'analyse : pour application au jeu complet de données
- **Visualisation** des résultats : facilite l'analyse de grands ensembles de données

➔ Analyse du nombre moyen de mots par page dans le *Journal des débats politiques et littéraires* (1814-1944)



Analyse : pourquoi ces fluctuations ?



Crédits : Flaticon (Icon Home)

Source : https://bnf.hypotheses.org/208

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

1. Périmètre, enjeux et apports

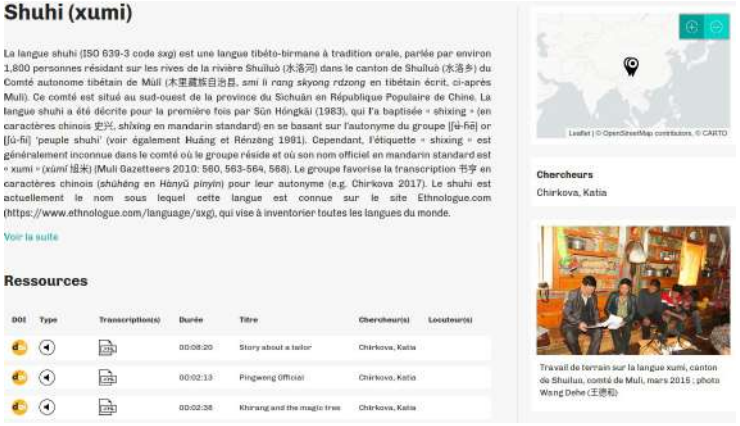
Analyse de corpus sonores : traitement & transcription

Pangloss

archive ouverte de langues en danger ou peu documentées



Exemple : La langue shuhi (xumi)

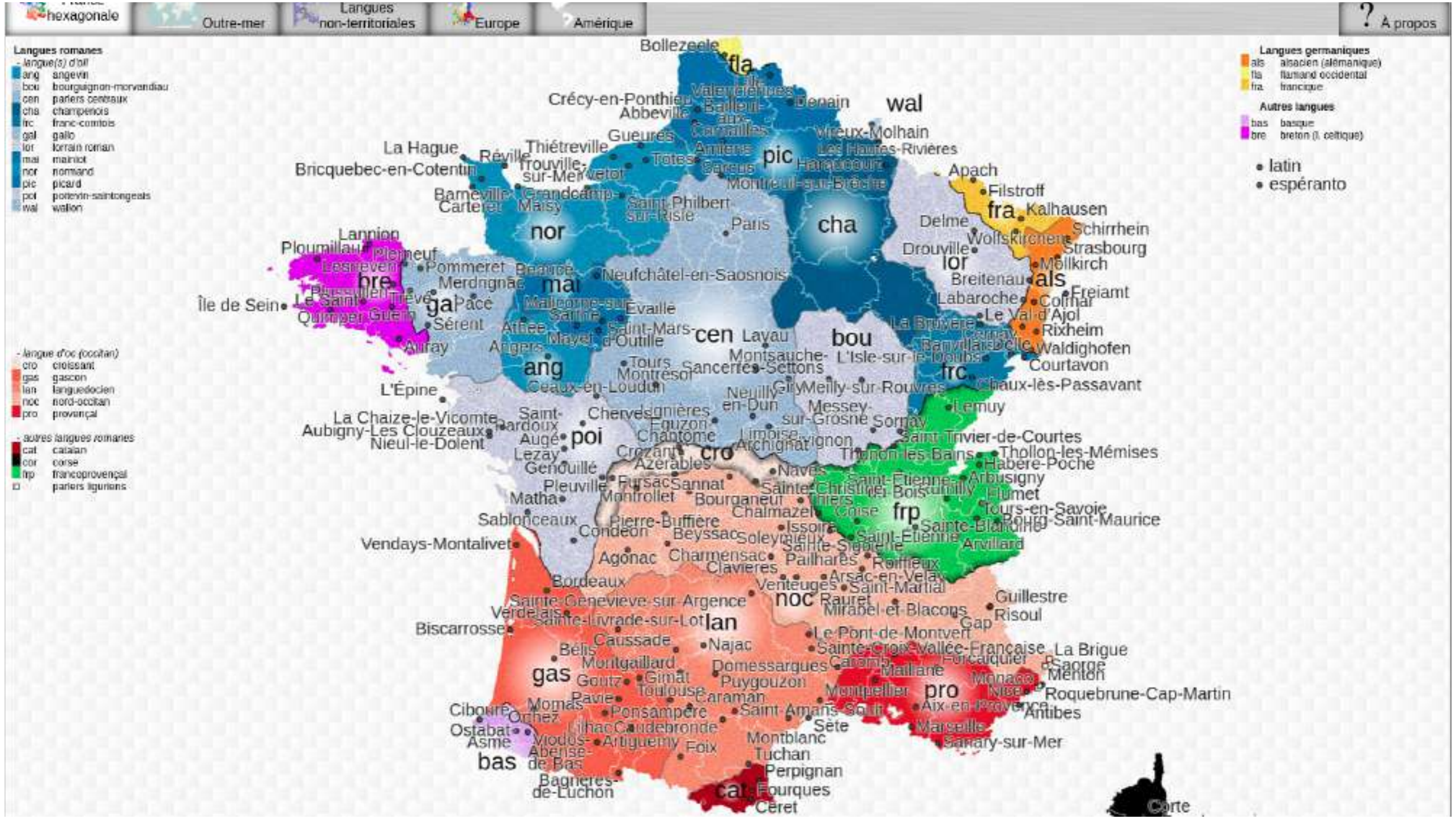


BULAC

Bibliothèque universitaire des langues et civilisations

1. Périmètre, enjeux et apports

- Valorisation par la visualisation : Atlas sonore des langues régionales de France : https://atlas.limsi.fr/

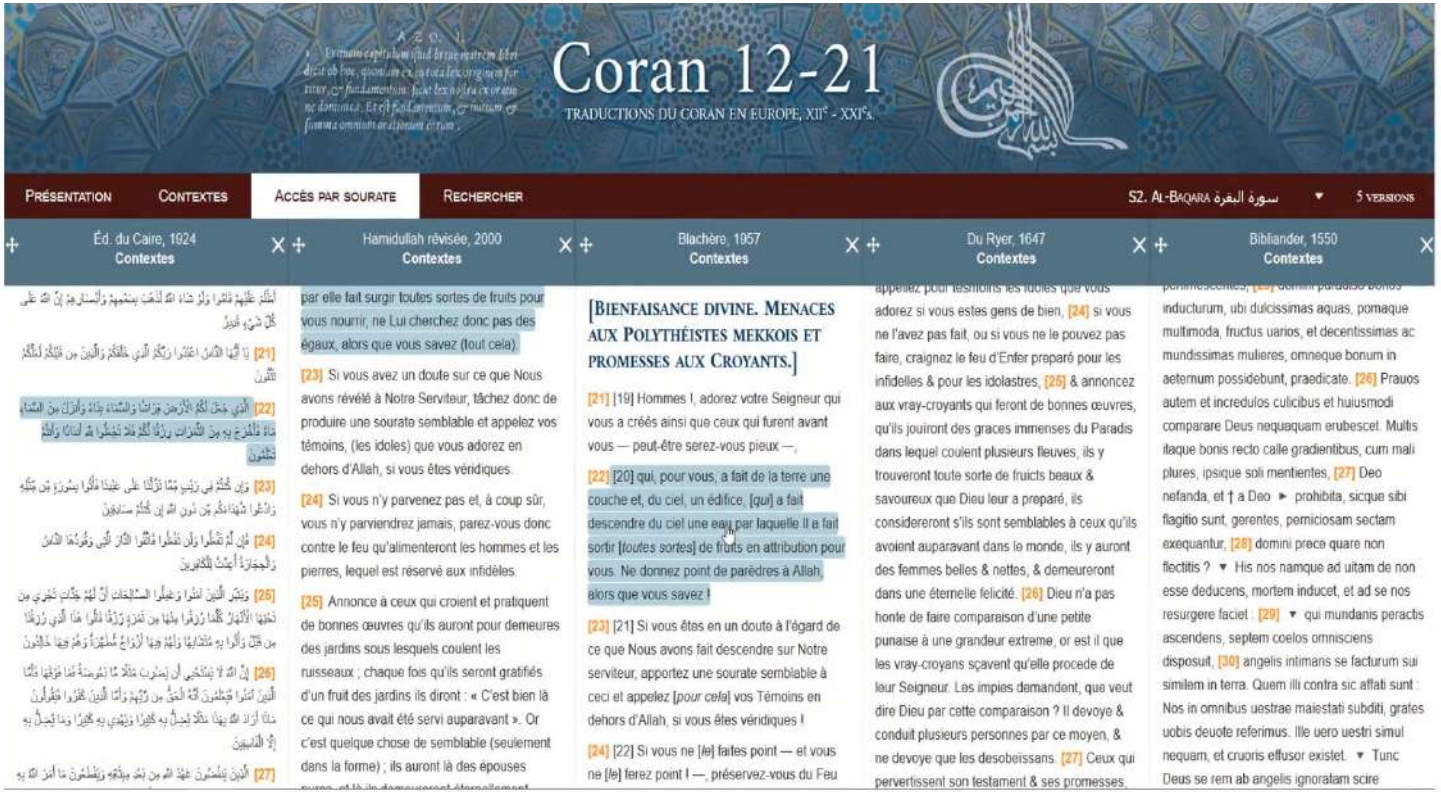


1. Périmètre, enjeux et apports

L'édition critique enrichie

Source : (https://coran12-21.org/fr/tutoriel - 3,35')

Lien
coran12-21.org



1. Périmètre, enjeux et apports

- Mise en évidence de la “textualité rayonnante” d’une œuvre par l’édition numérique

Texte numérisé, océrisé, structuré et enrichi



Mise en surbrillance automatique des sourates correspondantes (grâce à la structuration des données textuelles en XML-TEI)

Source : (https://coran12-21.org/fr/tutoriel - 3,35’)

Résumons :

Quels sont les apports des humanités numériques pour la recherche en sciences humaines et sociales ?



Crédits : Flaticon (Leremy ; Juicy_Fish)

Les humanités numériques se déclinent de multiples façons dans l'activité du chercheur :

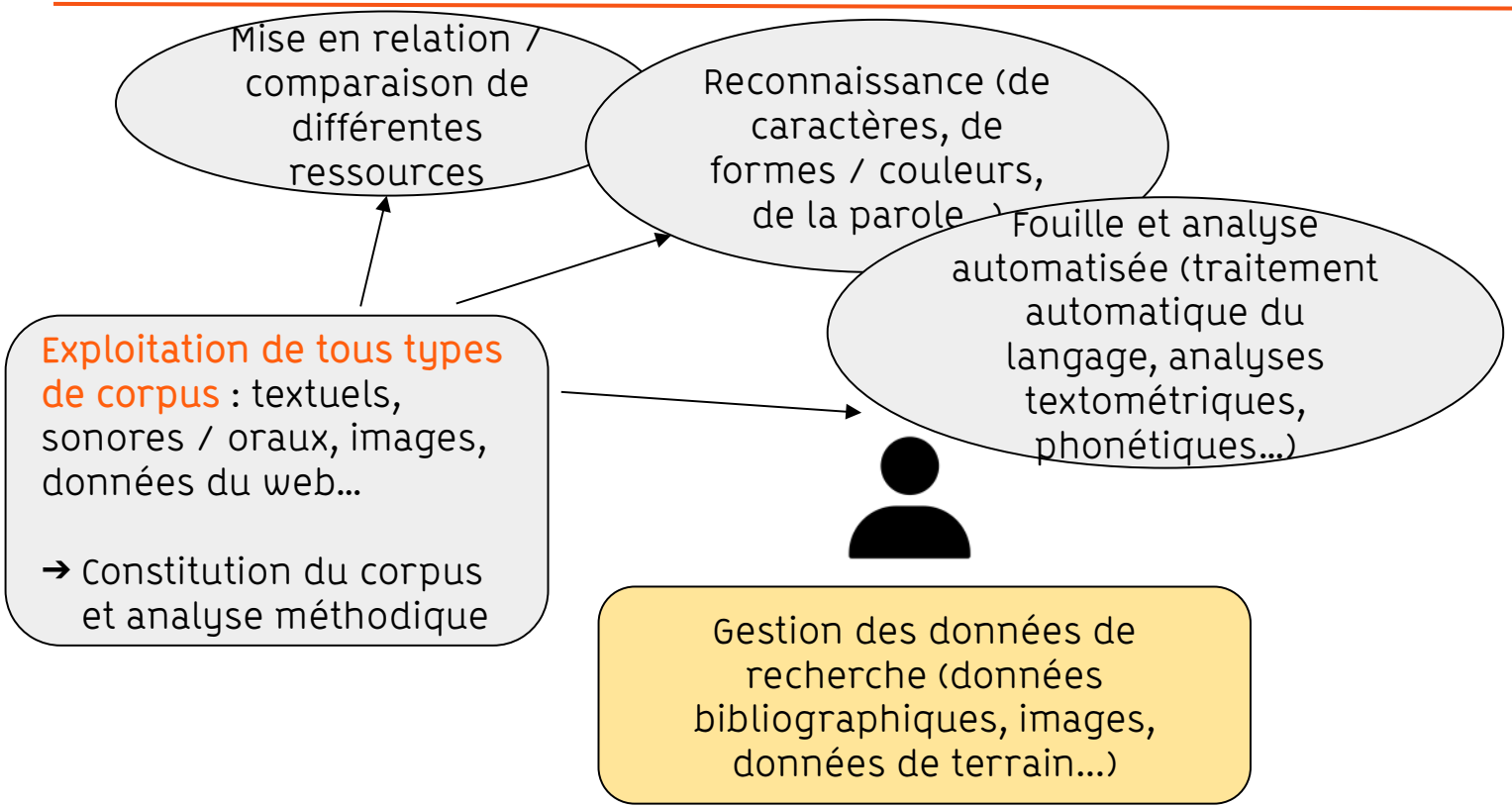


Gestion des données de recherche (données bibliographiques, images, données de terrain…)

Crédits : Flaticon (frans wahyu)

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

1. Périmètre, enjeux et apports



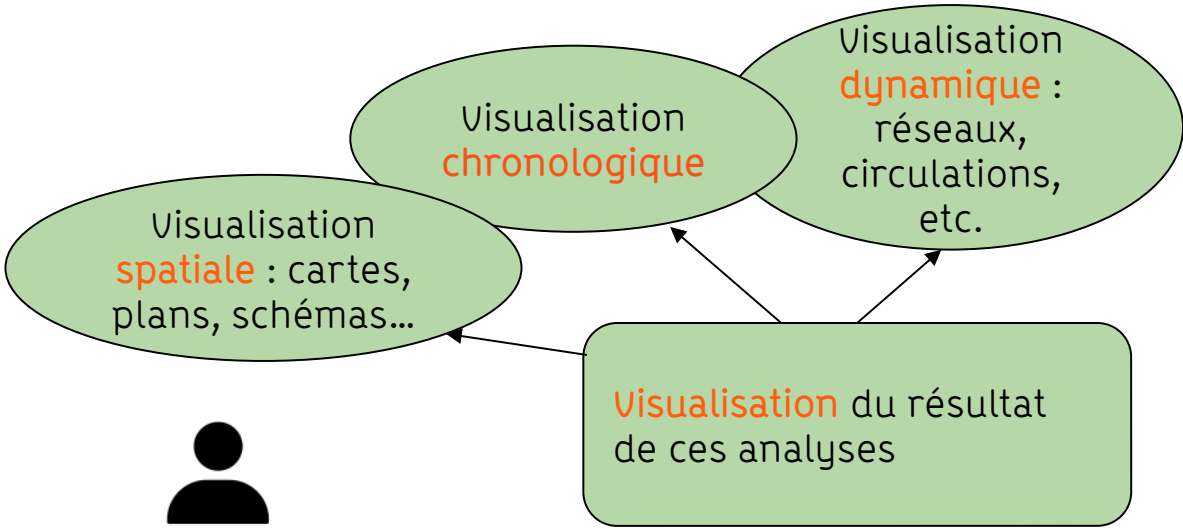
Crédits : Flaticon (frans wahyu)

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

1. Périmètre, enjeux et apports

Exploitation de tous types de corpus : textuels, sonores / oraux, images, données du web…

➔ Constitution du corpus et analyse méthodique



Gestion des données de recherche (données bibliographiques, images, données de terrain…)

Crédits : Flaticon (frans wahyu)

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

1. Périmètre, enjeux et apports

Exploitation de tous types de corpus : textuels, sonores / oraux, images, données du web…

➔ Constitution du corpus et analyse méthodique

Publication des résultats de manière accessible et ouverte en ligne

Édition numérique enrichie, carnet de recherche, etc.

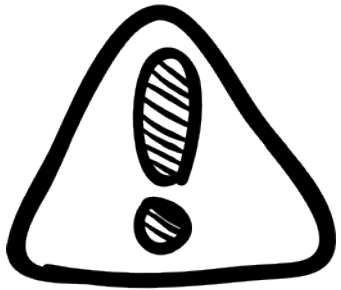
Visualisation du résultat de ces analyses



Gestion des données de recherche (données bibliographiques, images, données de terrain…)

Crédits : Flaticon (frans wahyu)

Mais attention, les humanités numériques ne permettent pas de tout faire !



Crédits : Flaticon (Freepik)

1. Périmètre, enjeux et apports

Un certain nombre de biais :



- l'effet « boîte noire » (Pierre Mounier)
- le risque d'un « oubli de l'oubli » (Milad Doueihi)
- tendance à se focaliser davantage sur l'outil que sur le besoin de recherche

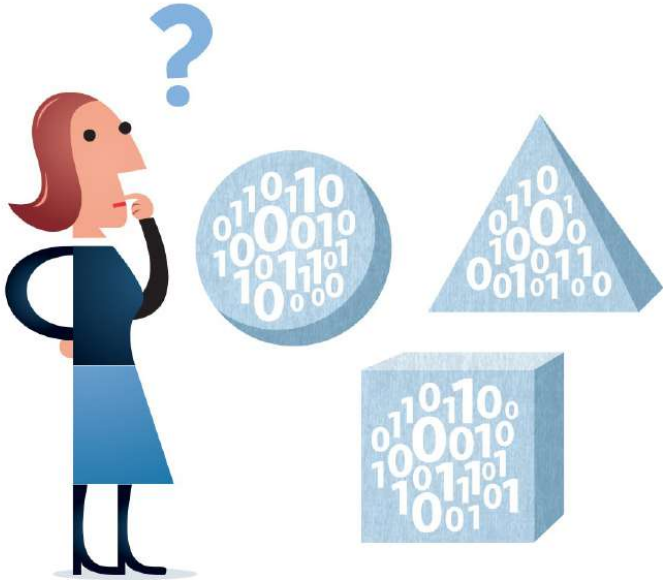


Une vigilance à conserver :

- un projet d'HN = toujours au service d'un besoin de recherche
- le numérique est un moyen et non une fin en soi
- les HN doivent garder un ancrage « humaniste » (M. Doueihi)

Crédits : Flaticon (srip)

Des questions ?



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

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

2. SE LANCER DANS UN PROJET D'HUMANITÉS NUMÉRIQUES : LES BONNES PRATIQUES



Crédits : Freepik (vectorjuice)

2. Humanités numériques : les bonnes pratiques

Au cours de ses recherches, un chercheur collecte, exploite et produit lui-même des données (textes, images, sons, chiffres, etc.)







Source : "La gestion des données et le cycle de vie des données de la recherche", CoSO,

Crédits : Flaticon

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

2. Humanités numériques : les bonnes pratiques

Au cours de ses recherches, un chercheur collecte, exploite et produit lui-même des données (textes, images, sons, chiffres, etc.)

Mais plus on accumule de données, plus leur gestion peut s'avérer complexe…



➔ De leur création à leur partage, il convient de savoir bien gérer ses données

Crédits : Flaticon

Source : "La gestion des données et le cycle de vie des données de la recherche", CoSO,

2. Humanités numériques : les bonnes pratiques

Mais avant de parler de bonne gestion des données

… il faut s'entendre sur ce qu'est une
donnée, une métadonnée et
un jeu de données



Crédits : Flaticon (Becris)

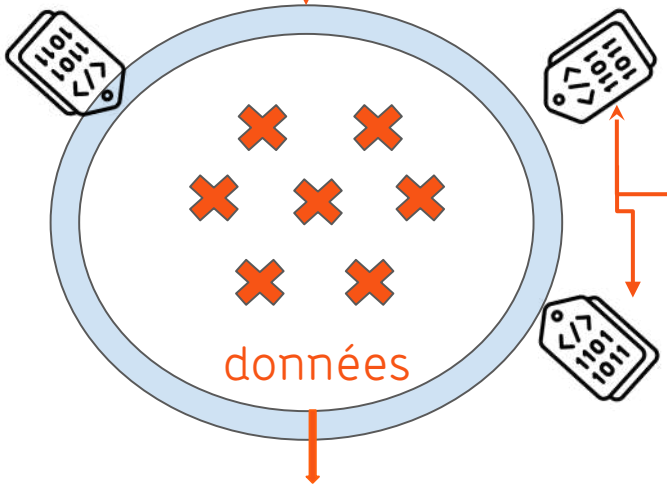
BULAC
[עברית] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

2. Humanités numériques : les bonnes pratiques

Que sont les données au coeur des HN ?

Jeu de données

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



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)

Parfois, terme "donnée" désigne tout un jeu

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

Crédits : Flaticon (Juicy_Fish) ; Freepik (Smashicons)

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

2. Humanités numériques : les bonnes pratiques

Bien gérer ses données à
TOUTES les étapes du **cycle de vie** de la donnée



Collecte



Description



Stockage



Traitement



Mise en accès



Archivage

Crédits : Flaticon (mikan933 ; Marz Gallery ; DinosoftLabs ; Freepik ; Smashicons)

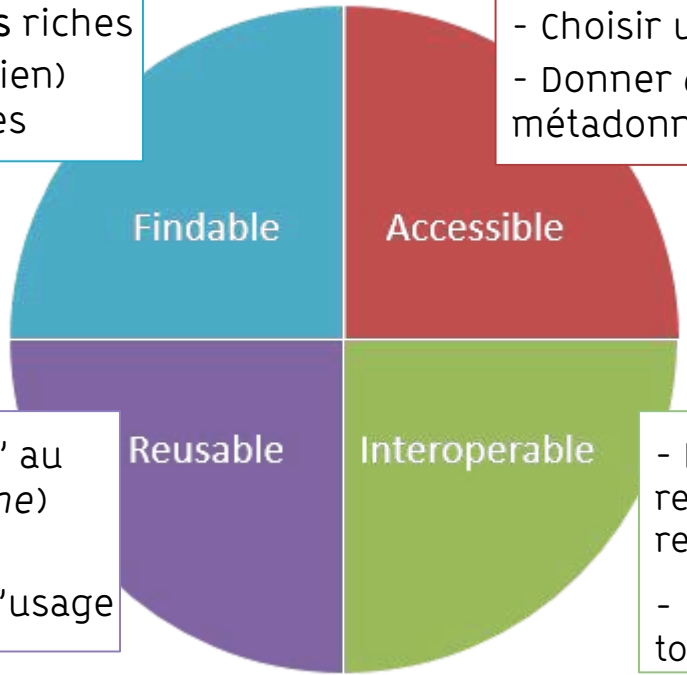
BULAC
[ספריה] [大学] [γλώσσες] [ཤེས་རིག]
Bibliothèque universitaire des langues et civilisations

2. Humanités numériques : les bonnes pratiques

Gérer ses données selon les **principes FAIR**

- Produire des **métadonnées** riches
- Attribuer un **identifiant** (= lien) pérenne aux jeux de données

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



- Fournir un "**mode d'emploi**" au jeu de données (fichier *readme*)
- Attribuer aux données une **licence** spécifiant les droits d'usage

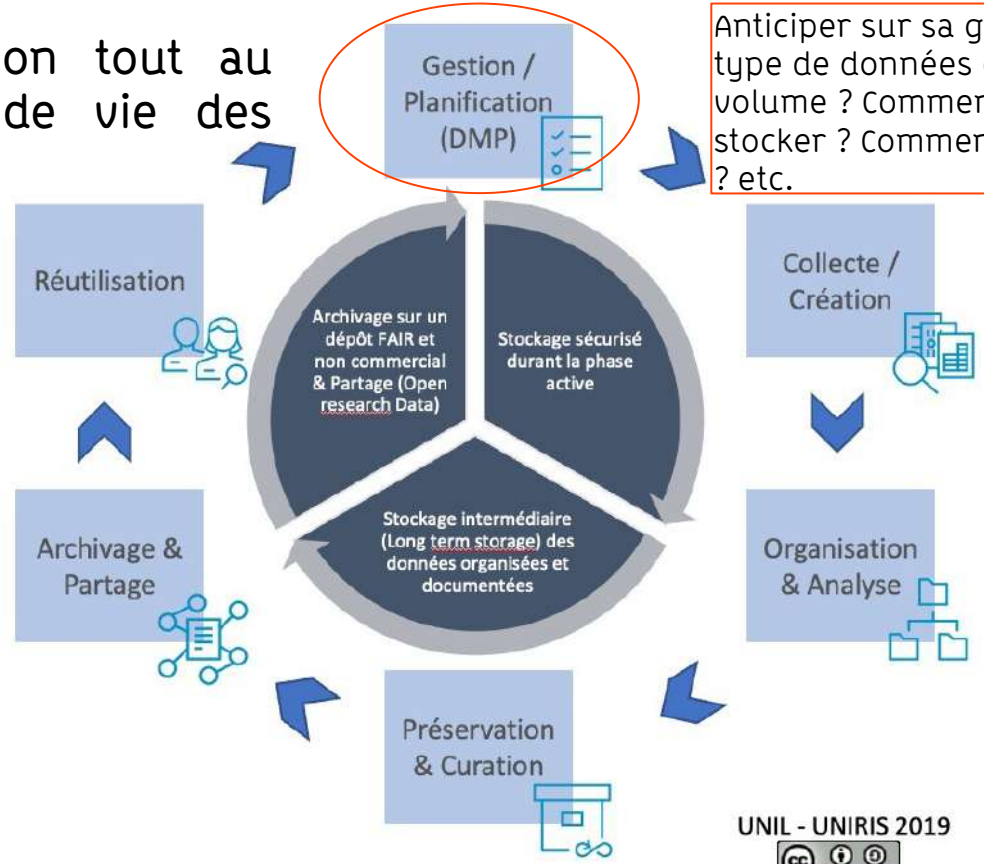
- **Métadonnées** de qualité = repérables par moteurs de recherche
- **Formats ouverts** = lisibles par toutes les machines

Source : Université Paris-Saclay /

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

2. Humanités numériques : les bonnes pratiques

Une préoccupation tout au long du cycle de vie des données



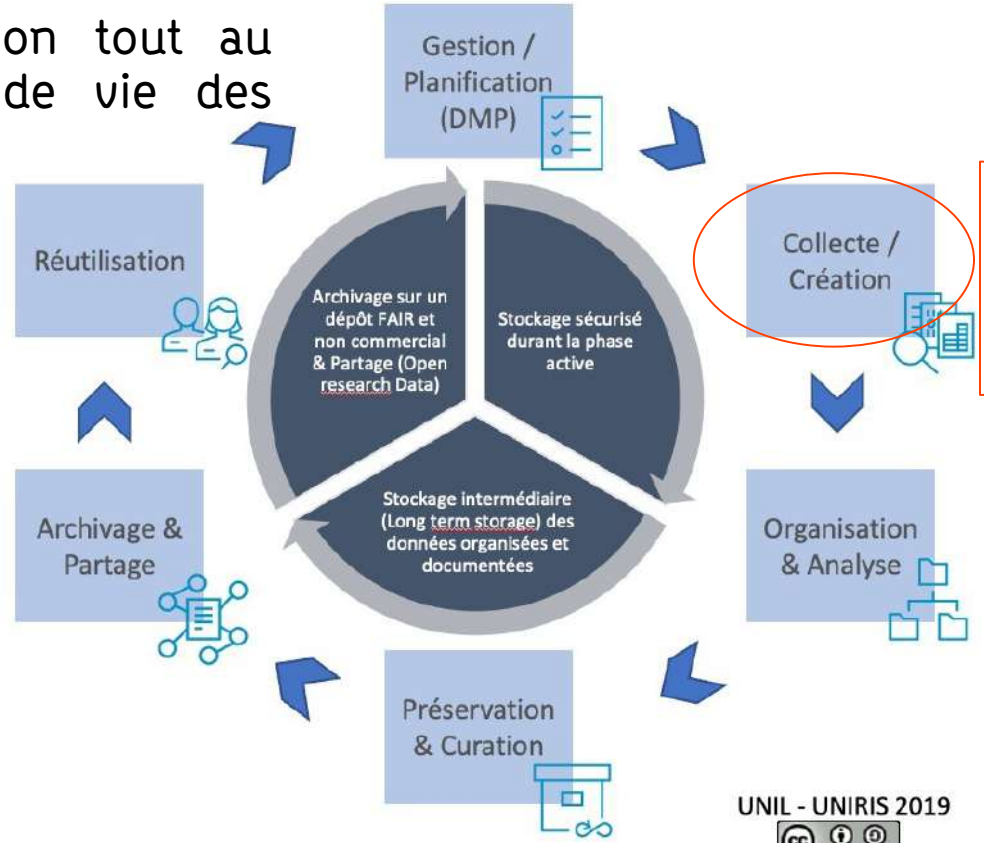
Anticiper sur sa gestion des données : Quel type de données collectées/produites ? Quel volume ? Comment les décrire ? Comment les stocker ? Comment les partager ? Les archiver ? etc.

UNIL - UNIRIS 2019

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

2. Humanités numériques : les bonnes pratiques

Une préoccupation tout au long du cycle de vie des données



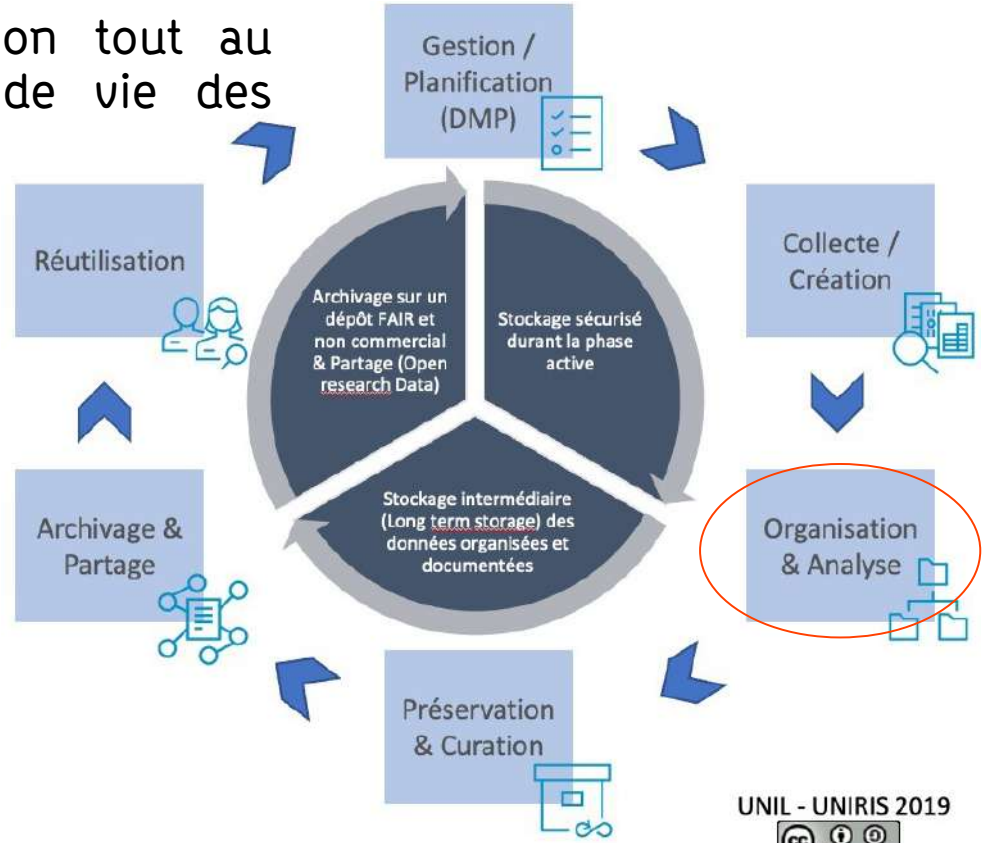
Documenter le processus : comment les données ont-elles été collectées ? A quelle date ? Qui a contribué à leur collecte ? etc.
=> rôle des MÉTADONNÉES

UNIL - UNIRIS 2019

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

2. Humanités numériques : les bonnes pratiques

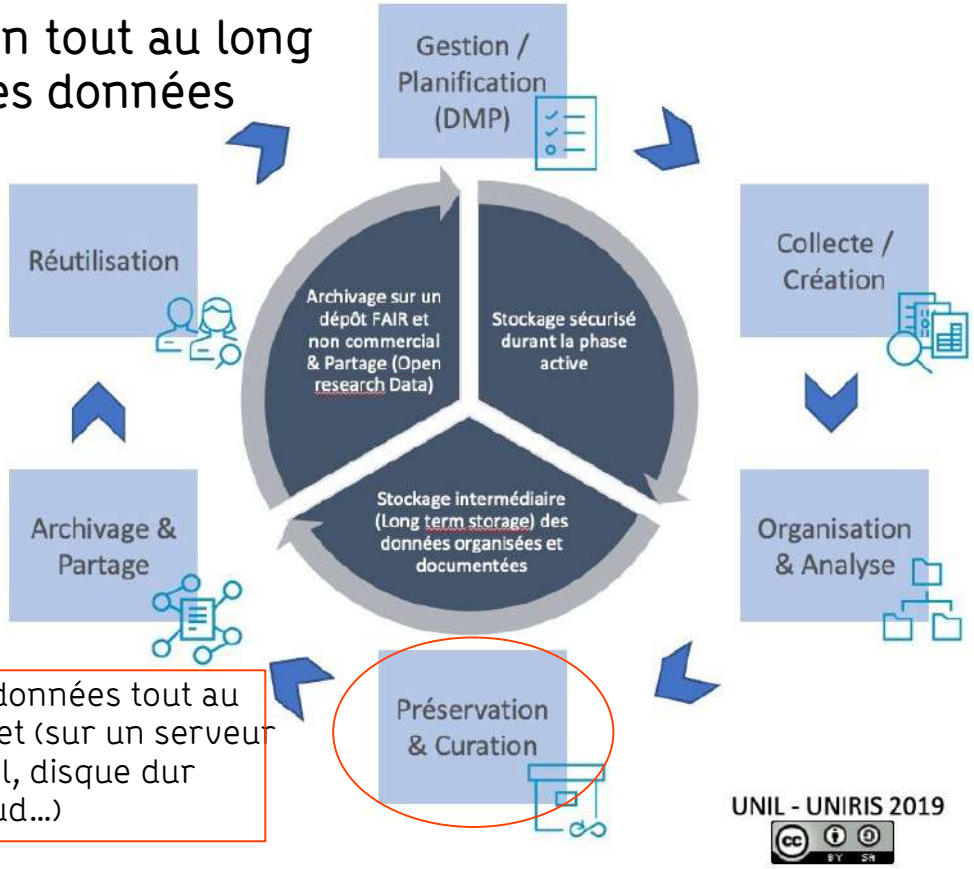
Une préoccupation tout au long du cycle de vie des données



Documenter tout le processus de recherche ; Bien structurer ses données => formats ouverts favorisant l'interopérabilité

2. Humanités numériques : les bonnes pratiques

Une préoccupation tout au long du cycle de vie des données

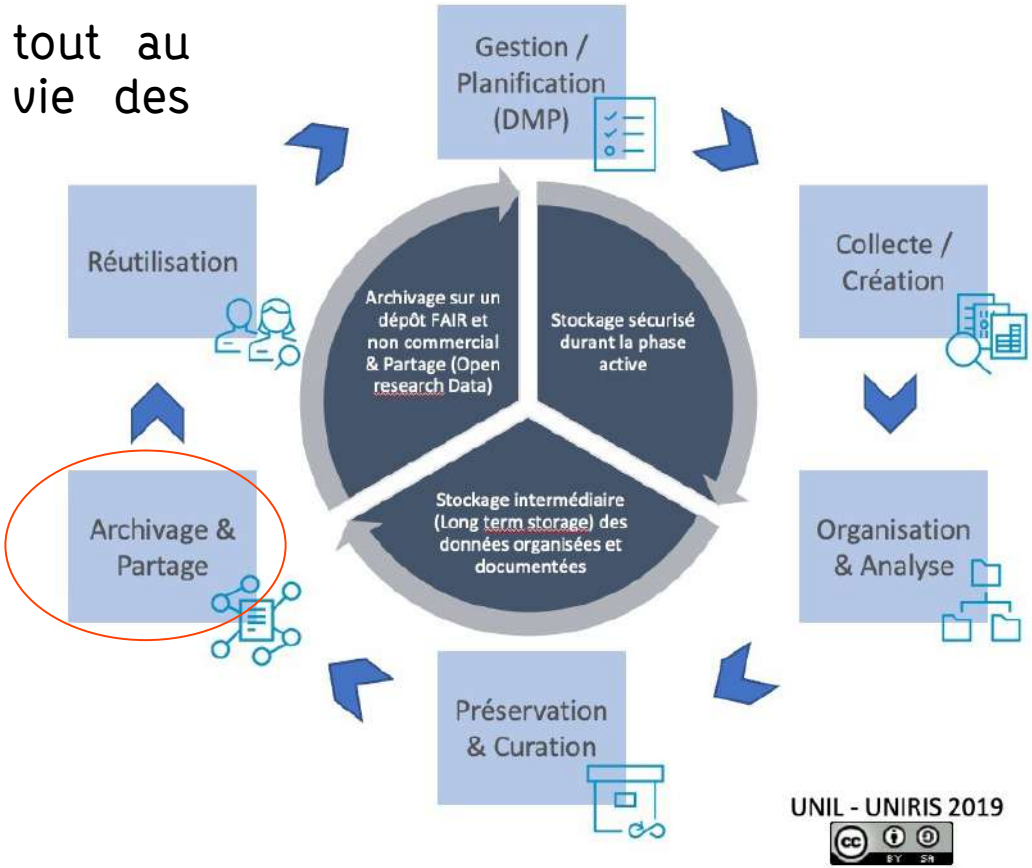


Stocker ses données tout au long du projet (sur un serveur institutionnel, disque dur externe, cloud…)

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

2. Humanités numériques : les bonnes pratiques

Une préoccupation tout au long du cycle de vie des données



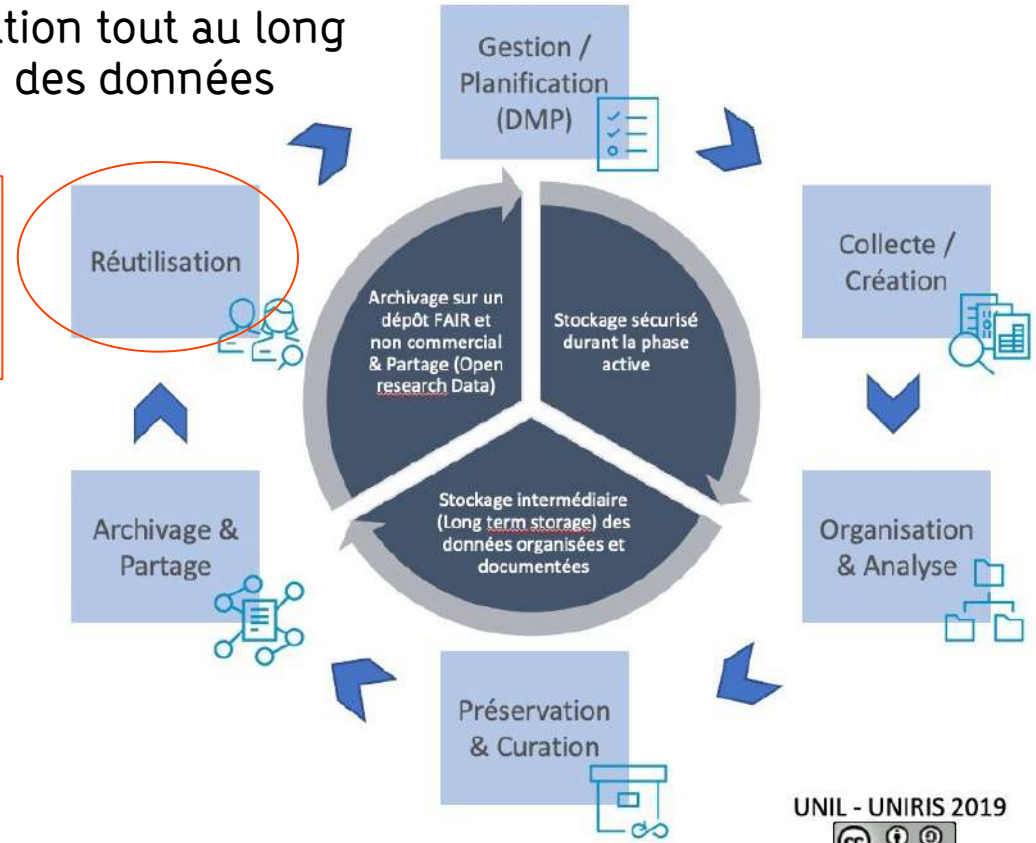
Déposer ses données dans un entrepôt de confiance où elles seront conservées, décrites et prêtes à être partagées

UNIL - UNIRIS 2019

2. Humanités numériques : les bonnes pratiques

Une préoccupation tout au long du cycle de vie des données

Veiller à la possibilité de réutilisation / citabilité du jeu de données produit

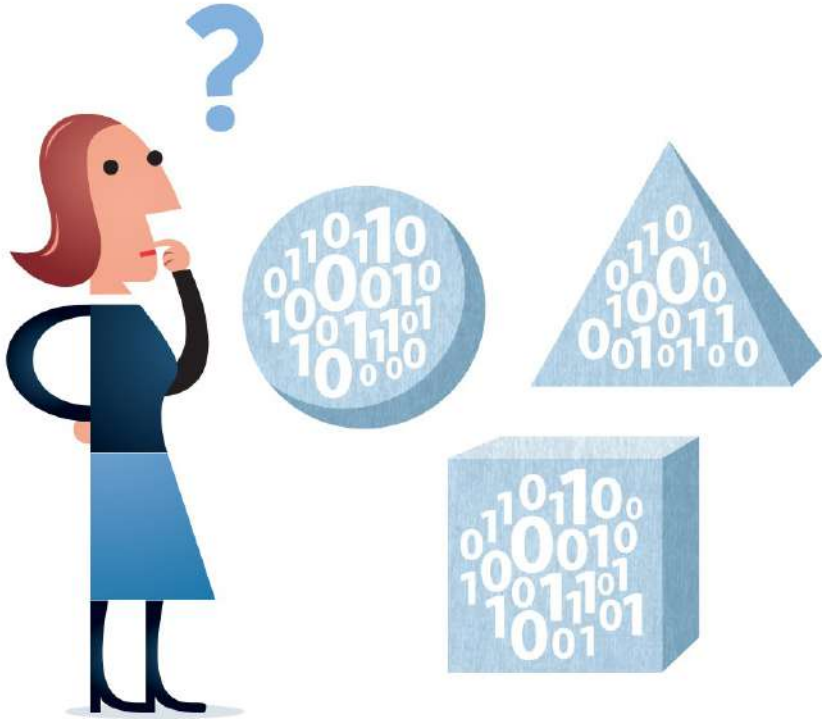


UNIL - UNIRIS 2019

CC BY SA

2. Humanités numériques : les bonnes pratiques

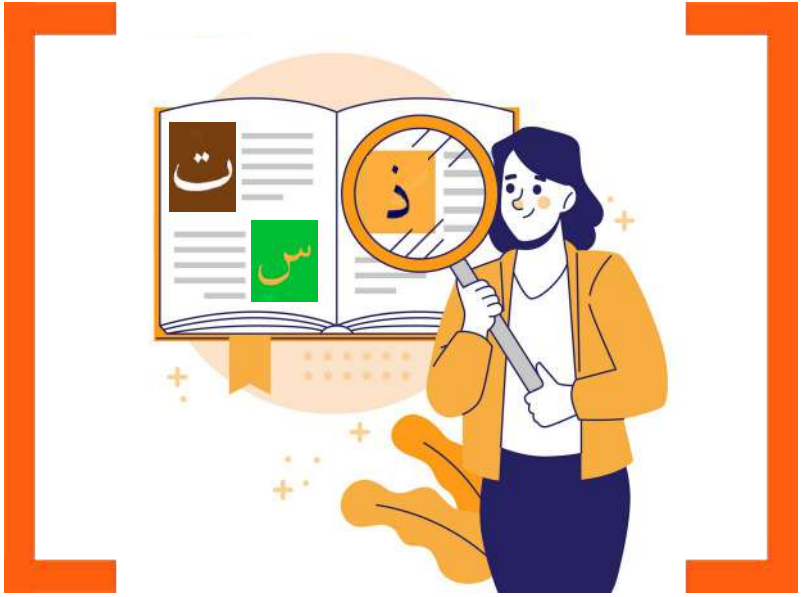
Des questions ?



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

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

3. EXEMPLE D'UTILISATION DES HUMANITÉS NUMÉRIQUES POUR L'ÉTUDE DE MANUSCRITS ARABES



Crédits : Freepik (NHJenny)

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

3. Les HN pour l'étude de manuscrits arabes

Reconnaissance automatique de l'arabe maghrébin manuscrit

- **L'origine du projet** (lien):
  - ➔ Reconnaissance automatique d'écritures manuscrites en arabe *maghribī*
  - ➔ Collaboration entre une chercheuse (N. Lucas), une association (Calfa) et la BULAC



- **Une visée double :**
  - ➔ Objectif de valorisation : Faciliter l'exploitation de ces manuscrits par les chercheurs
  - ➔ Objectif technique et expérimental : Tester un outil HTR (*Handwritten Text Recognition*) sur une graphie non latine

Crédits : Freepik (NHJenny)

3. Les HN pour l'étude de manuscrits arabes

Comment s'y est pris cette équipe de chercheurs / développeurs ?

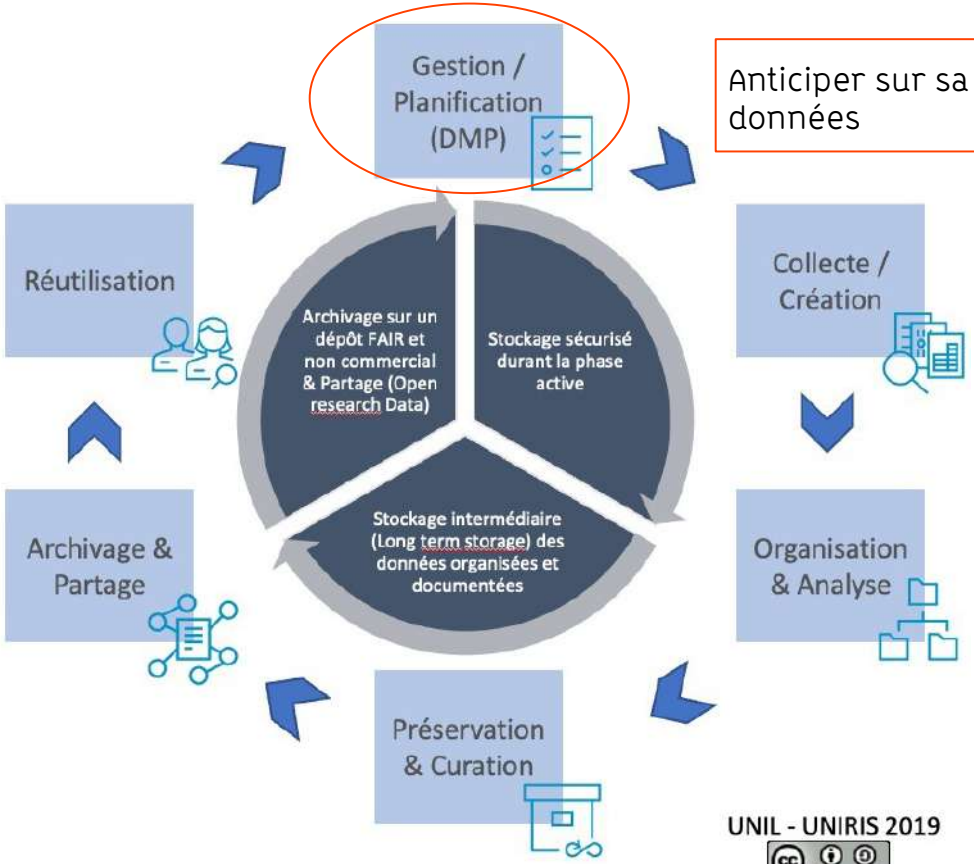


Reprenons les étapes importantes…

Crédits : Flaticon (Freepik)

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

3. Les HN pour l'étude de manuscrits arabes



Anticiper sur sa gestion des données

UNIL - UNIRIS 2019

3. Les HN pour l'étude de manuscrits arabes

Étape 1 (a) : La planification



- Formuler une **question de recherche**, identifier **objectifs** et **besoins** :
  - *Est-il possible d'entraîner un outil d'HTR à reconnaître une graphie non latine ?*
- Constituer un **groupe projet** : expertises **scientifiques** et **techniques** :
  - *linguistes, codicologues, spécialistes du Maghreb, informaticiens …*
- Établir le **cahier des charges** : objectif ? outil(s) utilisé(s) ? choix techniques ?
  - *300 pages ; transcription harmonisée ; Calfa vision pour traitement et annotation*
- Prévoir **stockage, archivage** et **réutilisation** du jeu de données produit :
  - Rendre ouvert et disponible en ligne pour permettre réutilisation

Crédits : Freepik (vectorjuice)

3. Les HN pour l'étude de manuscrits arabes

Étape 1 (b) : Veiller à la faisabilité du projet



- Se fixer un **jeu de données** *(dataset)* réaliste
  - Ici : 300 images annotées
- Prévoir **une durée et un format adapté**
  - Format hackathon (plusieurs séances sur quelques mois)
- Identifier les **difficultés potentielles**
  - Grande variété d'écritures, mais peu de données. Comment parvenir à de bons résultats ?

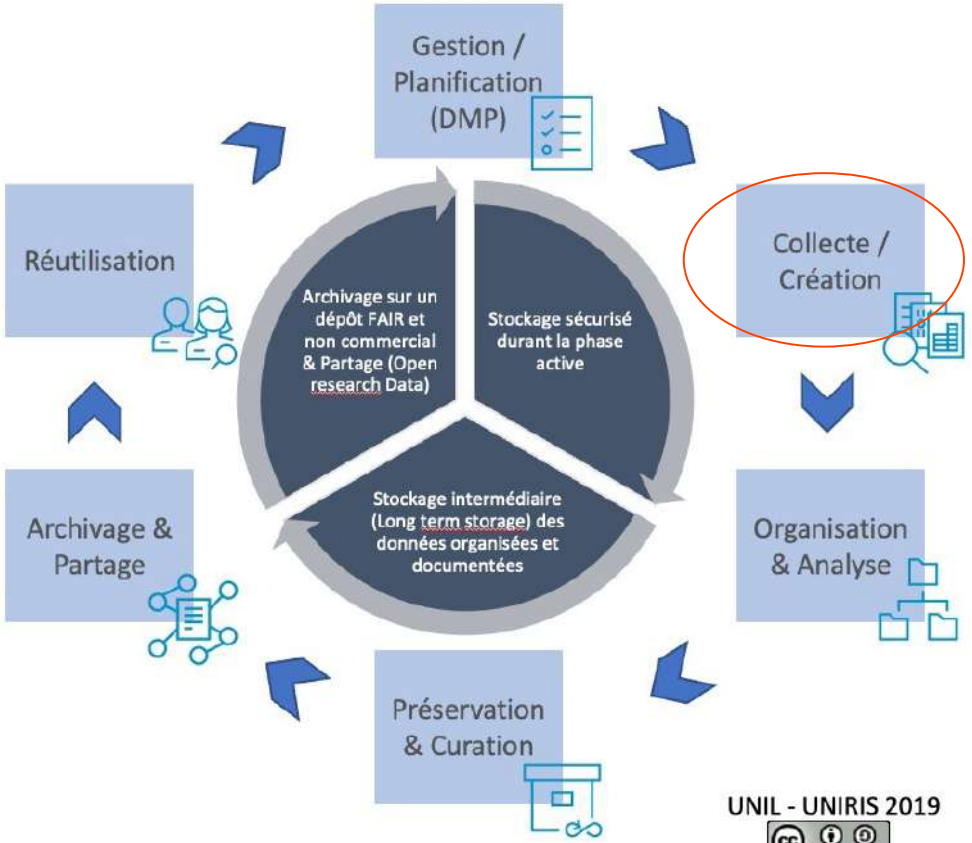


Source : BINA, MS.ARA.609, شرح الغرة البيضاء

Crédits : Freepik (storyset)

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

3. Les HN pour l'étude de manuscrits arabes

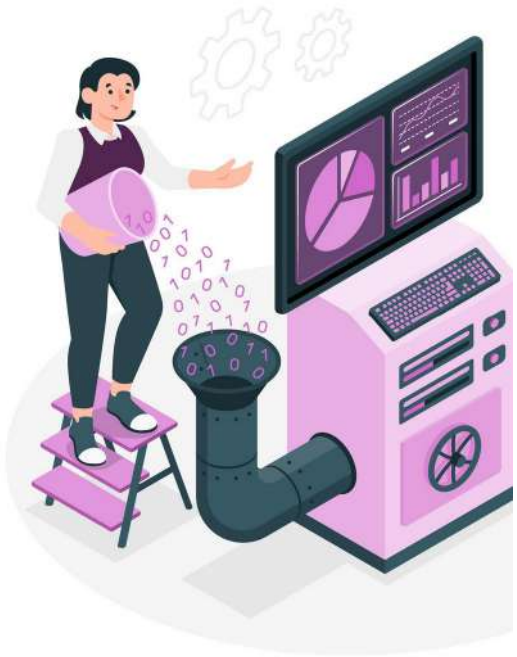


Documenter le processus

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

3. Les HN pour l'étude de manuscrits arabes

Étape 2 : La collecte



Crédits : Freepik (storyset)

- Constituer son jeu de données / corpus selon des **critères précis**
  - *graphies répandues MAIS variété des écritures, mises en page et sujets*
- **Documenter** la collecte et **référencer** son jeu de données à l'aide de **métadonnées**
  - *3 manuscrits (BULAC) : 2 sur la BiNA + 1 numérisé à la demande*



MS.ARA.609



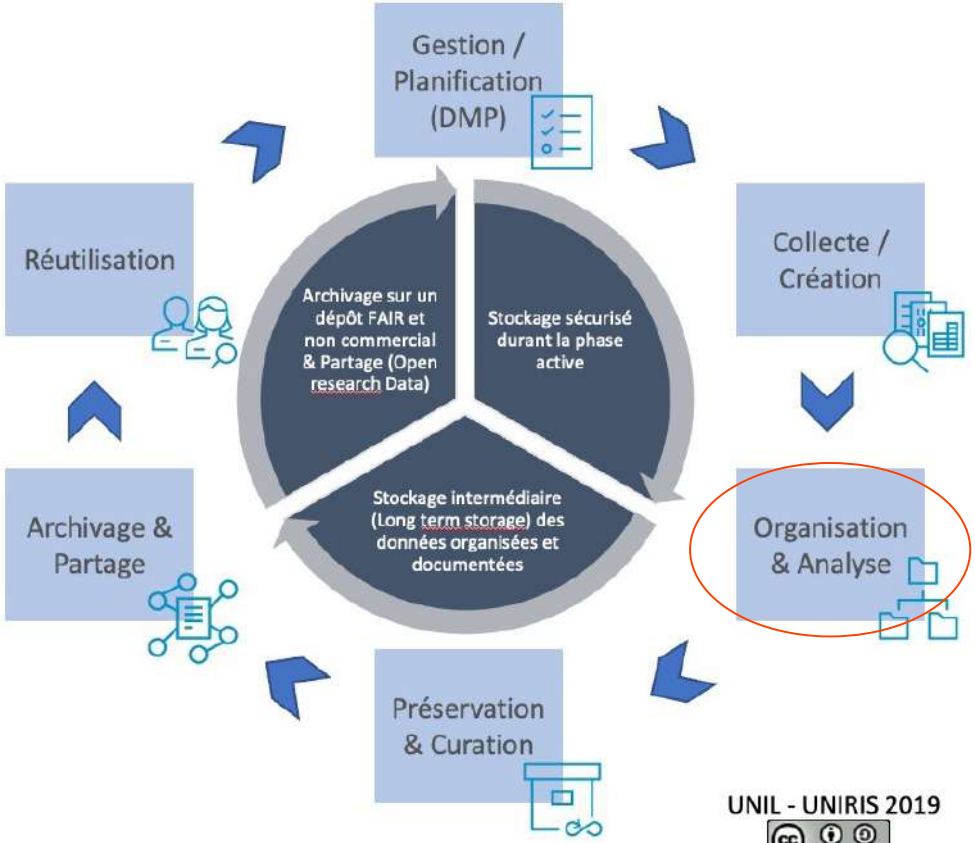
MS.ARA.1977



MS.ARA.417

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

3. Les HN pour l'étude de manuscrits arabes



Documenter le processus de recherche ; structurer ses données

3. Les HN pour l'étude de manuscrits arabes

Étape 3 : L'organisation des données (a)…



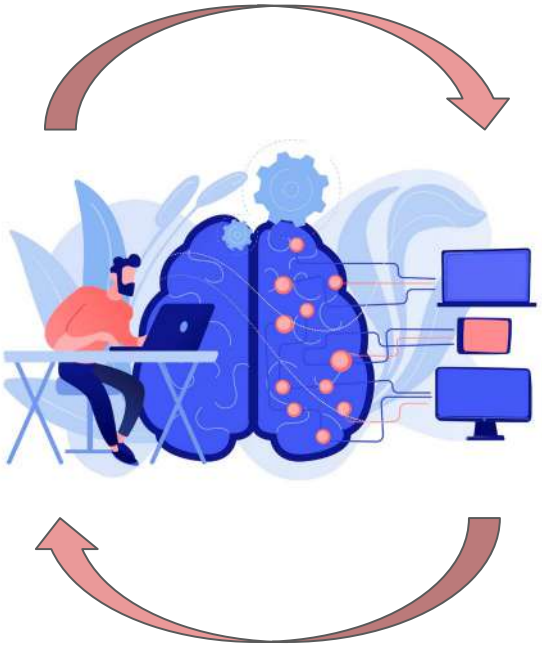
- Organiser ses fichiers :
  - choix des formats
  - règles de nommage
  - version des fichiers
  - classement des dossiers
- Privilégier les formats ouverts
- Documenter ses données avec des métadonnées (descriptives, de gestion, de préservation)

Pour bien organiser vos métadonnées, pensez aux outils de gestion des données de recherche comme **Zotero** ou **Tropy** !

Crédits : Freepik ; Flaticon

3. Les HN pour l'étude de manuscrits arabes

… ainsi que le traitement et l'analyse (b)



Entraînement des machines *(machine learning)*

- Entraînement du modèle de reconnaissance à partir d'une 1ère transcription **manuelle**
- Versement des corpus à transcrire dans *Calfa Vision* : 1ère transcription **automatique**

Vérification du résultat, correction des erreurs

- Relecture des transcriptions automatiques (*indispensable avant la reconnaissance des caractères*)

Affinage et spécialisation progressive (*fine tuning*)

- But : améliorer les performances du modèle de reconnaissance (ici arabe *maghribī*)

Crédits : Freepik (vectorjuice)

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

3. Les HN pour l'étude de manuscrits arabes

La reconnaissance se fait en plusieurs temps

- ❖ Analyse de la **mise en page** du manuscrit
  - Segmentation des différentes zones de la page *(texte principal, lignes, colonnes, images …)*
- ❖ Récupération du texte et **transcription**
  - Transcription de l'image numérisée en texte *(algorithmes de prédiction)*
- ❖ **Spécialisation** progressive du modèle et **correction** *a posteriori*
  - Affinage de la reconnaissance de la mise en page, puis nettoyage ou enrichissement du texte transcrit



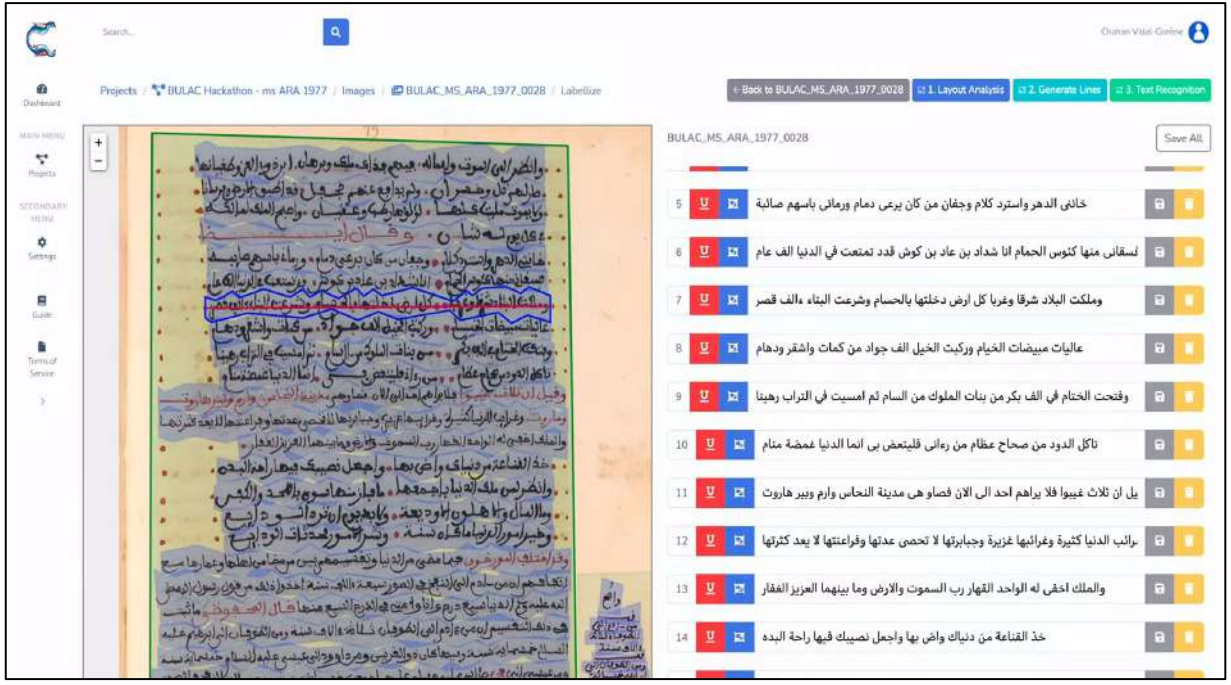
La spécialisation progressive des modèles a conduit à un gain de temps de **75%** dans la vérification de la mise en page.

Source : "Intelligence artificielle et khaṭṭ maghribī". N. Lucas ; C. Vidal-Gorène, BULAC, 8 juin 2021

3. Les HN pour l'étude de manuscrits arabes

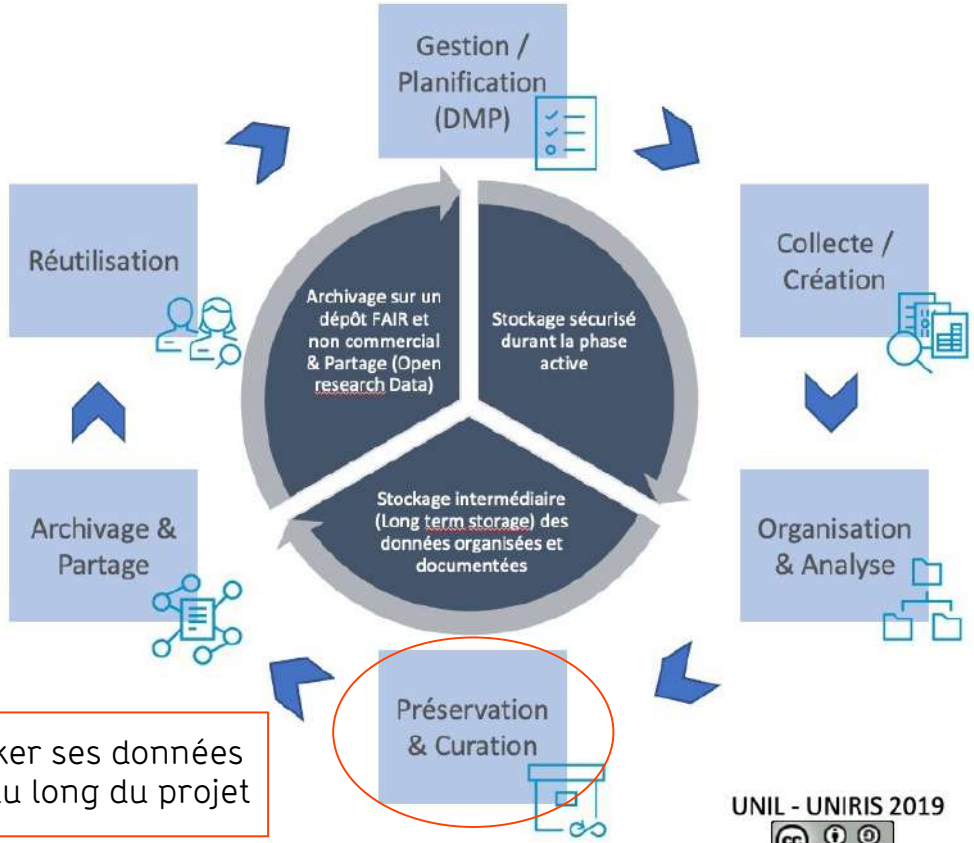
Reconnaissance des zones de texte, extraction des lignes et transcription :





Sources : "Two Examples of Analysis of Textual Document in Oriental and Under-Resourced Languages", 2021 / "Intelligence artificielle et khaṭṭ maghribī". 2021

3. Les HN pour l'étude de manuscrits arabes



Stocker ses données tout au long du projet

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

3. Les HN pour l'étude de manuscrits arabes



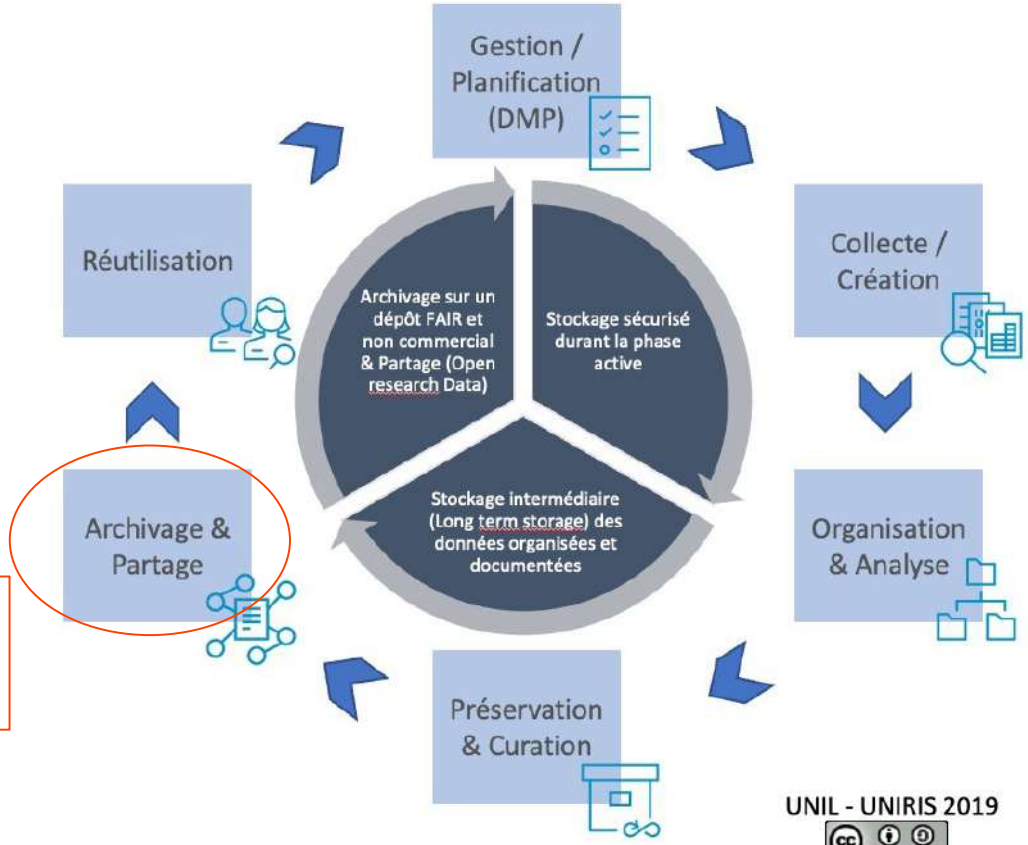
⚠ Stocker ≠ Archiver !

Stockez vos données	
Pourquoi	Pour garantir l'exploitation des données tout au long du projet
Qui	- chaque chercheur - l'équipe, de manière collective
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 externe) ≃ ordinateur (risque technique) ≃ cloud (risque de sécurité)
Sécurité	antivirus, mot de passe, cryptage

Crédits : Freepik (storyset) ; Flaticon (Tempo_doloe)

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

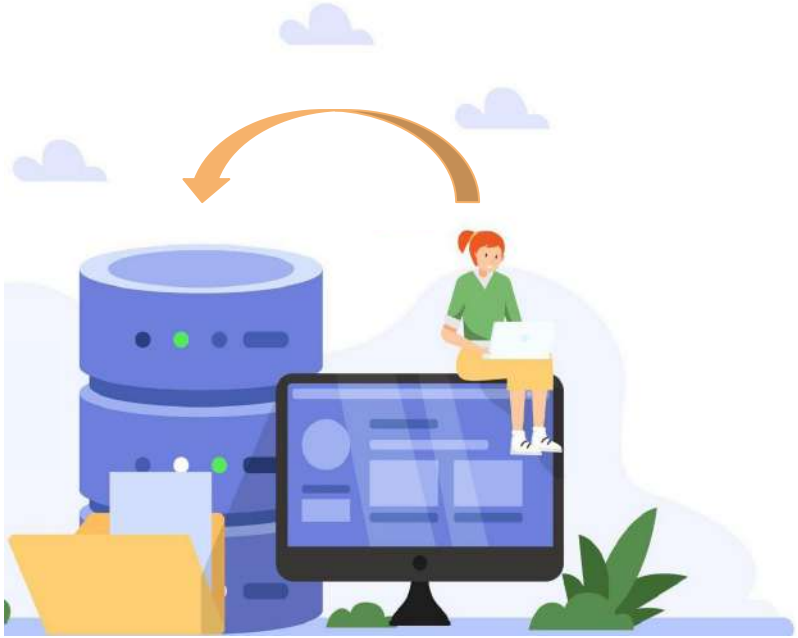
3. Les HN pour l'étude de manuscrits arabes



Déposer ses données dans un entrepôt

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

3. Les HN pour l'étude de manuscrits arabes



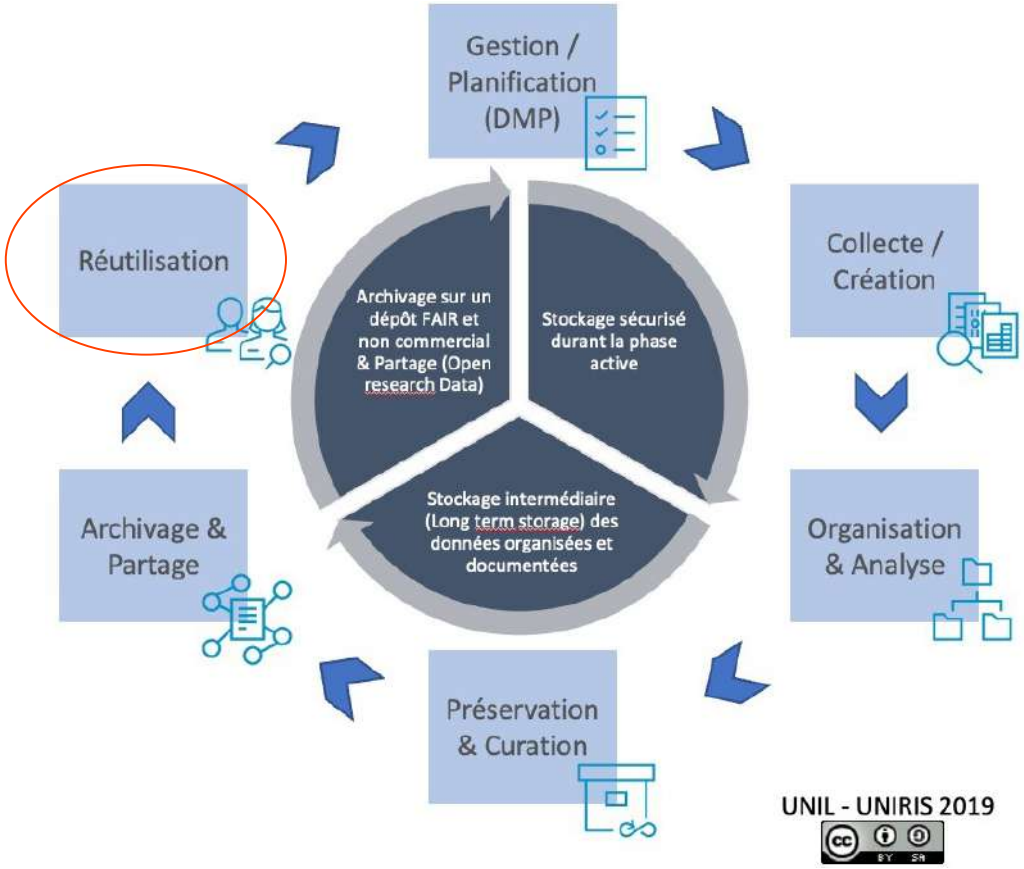
Crédits : Freepik

Archivez vos données	
Pourquoi	- conserver à long terme les données - garantir accessibilité et intelligibilité - Tout conserver ? Fixer des critères !
Quand / Comment	- prévoir dès planification projet (coûts !) - modèles d'archivage (ex. OAIS)
Qui	- scientifiques - informaticiens - archivistes et documentalistes - juristes
Où	- SHS : l' IR* Huma-Num (avec Cines) - Format d'archivage des données ? Voir FACILE

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

3. Les HN pour l'étude de manuscrits arabes

Veiller à la possibilité de réutilisation / citabilité du jeu



UNIL - UNIRIS 2019

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

3. Les HN pour l'étude de manuscrits arabes



Pensez "Science Ouverte" !



- Démarche d'*open data* encouragée : dépôt des données dans entrepôt et ouverture
  - But = Leur **RÉUTILISATION** : pour de nouveaux projets de recherche
    Jeu de données - hackhaton 2021:
    https://github.com/calfa-co/rasam-dataset
  - Meilleure gestion de l'argent investi
- Rôle crucial des infrastructures (en France, Huma-Num avec Nakala ; recherche.data.gouv)

=> Convergence Humanités numériques / Science Ouverte

Crédits : Freepik (user24531006) / Flaticon (rukanicon)

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

POUR CONCLURE… À VOUS D'ENTRER EN PISTE !



Crédits : Freepik (macrovector)

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

Pour aller plus loin…

- De Wilde M. ; Gillet F. *et. al.*, ***Introduction aux humanités numériques : méthodes et pratiques***, Louvain-La-Neuve : De Boeck, 2016, (https://www.cairn.info/introduction-aux-humanites-numeriques-methodes--9782807302150.htm)
- Mounier Pierre, ***Les humanités numériques : Une histoire critique***, Paris, Éditions de la Maison des sciences de l'homme, https://books.openedition.org/editionsmsh/12006 : https://books.openedition.org/editionsmsh/12006
- "**Les humanités numériques en question - Entretien avec Pierre Mounier**", Université d'Artois : https://artoistv.univ-artois.fr/video/4591-les-humanites-numeriques-en-question-entretien-avec-m-pierre-mounier/
- Puren Marie, "**La lecture distante : introduction et exemples d'application**", Master Outils et humanités numériques, 2020, ffhal-03152747f : https://hal.science/hal-03152747/document
- "**La gestion des données et le cycle de vie des données de la recherche**", CoSO, https://www.canal-u.tv/chaines/ouvrirlascience/la-gestion-des-donnees-et-le-cycle-de-vie-des-donnees-de-la-recherche
- Moreux Jean-Philippe, "**Approches innovantes pour la presse ancienne numérisée : fouille et visualisation de données**", in *Carnet de la recherche à la Bibliothèque nationale de France*, 30 décembre 2016 : https://bnf.hypotheses.org/208
- Azoulay Benjamin, Benoît de Courson, Clara de Courson, Laurent Vanni et Étienne Brunet, "**Gallicagram : les archives de presse sous les rotatives de la statistique textuelle**", *Corpus* [Online], 24 | 2023 : https://journals.openedition.org/corpus/7944?lang=en)
- Wisniewski Guillaume, Alexis Michaud, Benjamin Galliot, Laurent Besacier, Séverine Guillaume, et al., "**Ouvrir aux linguistes « de terrain » un accès à la transcription automatique**", 2èmes journées scientifiques du Groupement de Recherche Linguistique Informatique Formelle et de Terrain (LIFT), 2020, pp.83-94 : hal-03047148
- "**Corpus oraux, linguistique et humanités numériques**", blog d'Huma-Num, Hypotheses : https://humanum.hypotheses.org/184
- Lucas Noëmie, ***OCR/HTR et graphie arabe - Les manuscrits à l'heure de la reconnaissance automatique des écritures***, Cahier du GIS n°3, Avril 2022 : https://hal.science/hal-03822459
- "**Intelligence artificielle et khaṭṭ maghribī. Résultats d'un hackathon pour la reconnaissance de texte automatique de l'arabe manuscrit**", présentation de Noëmie Lucas et Chahan Vidal-Gorène, BULAC, 8 juin 2021 : https://bulac.hypotheses.org/32837

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

formation-publics@bulac.fr



Crédits : Wilhelm Gunckel / Unsplash