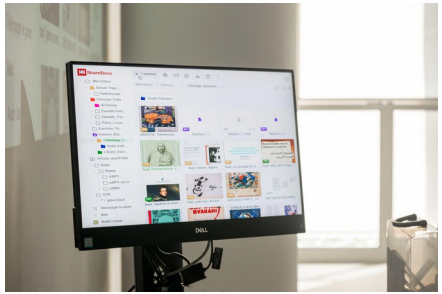
BULAC

[도서관] [शिक्षा] [ဘာသာစကား] [ሥልጣኔ]

Bibliothèque universitaire
des langues et civilisations

TRAITER, DIFFUSER SES DONNÉES DE RECHERCHE AVEC *SHAREDOCS* ET *NAKALA*

Mai 2025



Crédits: Maxime Ruscio/ BULAC

CC - BY - NC - SA

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

Plan de la séance

1. Les Données de la recherche... pour les nuls

2. Organiser et traiter ses fichiers avec Sharedocs

3. Déposer et diffuser ses données avec Nakala

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

1. LES DONNÉES DE LA RECHERCHE… POUR LES NULS

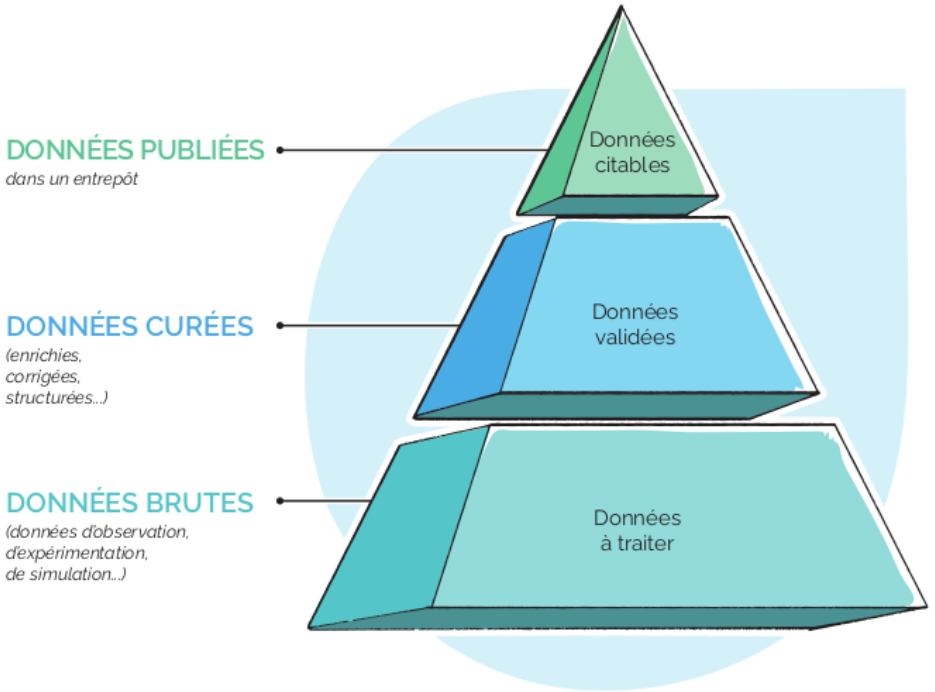


Crédits : Vlad Tchompalov / Unsplash, photos for everyone

1. Les données de la recherche… pour les nuls

Définition générale des "données de la recherche" (DoR)

… des informations numériques de divers types (observation, expérimentation, simulation,…) et nature (chiffres, textes, images, sons, etc.) collectées ou produites dans le cadre d'un projet de recherche ou d'activités scientifiques. Elles sont généralement regroupées dans un ensemble homogène que l'on appelle "jeu de données", et qui est susceptible d'être diffusé et ré-employé par l'ensemble d'une communauté scientifique.



Crédits : EPRIST /

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

1. Les données de la recherche… pour les nuls

… s'appuyer sur les principes FAIR , et notamment…

Facile à trouver 

Accessible 

Interopérable 

Réutilisable 

- ❑ … choisir des formats de fichier ouverts et standards, respecter certaines règles de nommage
- ❑ … décrire richement ses métadonnées, avec des vocabulaires contrôlés (Dublin Core) et des thésaurus…
- ❑ … attribuer des identifiants uniques et pérennes à ces jeux de données (DOI, etc.), ainsi que des licences ouvertes (Creative Commons ou autres)…
- ❑ … indiquer clairement la provenance des données et les déposer dans un entrepôt de confiance…
- ❑ … garantir l'accès ouvert aux métadonnées même si les données sont en accès fermé ou restreint.

Crédits : Flaticon, rawpixel, vectorjuice / Freepik

1. Les données de la recherche… pour les nuls

Focus sur le vocabulaire Dublin Core (aide-mémoire)



Crédits : Jørgen Stamp (CC-BY-2.5) digitalbevaring.dk

Le Dublin Core (DC) est un format descriptif standardisé qui permet de documenter et de référencer simplement tout type de données. Il comprend un socle commun de 15 éléments simples, tous facultatifs et répétables.

Une forme étendue ou "qualifiée" du Dublin Core (DC Terms) propose d'autres éléments (près de 50!)

Dans Nakala, 5 champs de métadonnées équivalents au DC sont obligatoires*.

Voir le guide de description des données

Title*	Titre de la donnée
Creator*	Entité responsable de la création de la donnée
Subject	Thèmes du contenu de la donnée (*mots-clefs, voir thésaurus ou autres référentiels*)
Description	Présentation du contenu intellectuel de la donnée (*résumé*)
Publisher	Entité responsable de la diffusion de la donnée
Contributor	Entité responsable de contributions au contenu de la donnée
Date*	Date d'un événement dans le cycle de vie la donnée (*création, publication,…*)
Type*	Nature ou genre du contenu de la donnée (*dataset, event, image, interactive resource, moving image, physical object, software, sound, still image, text…*)
Format	Matérialisation physique ou numérique de la donnée (*medium, dimensions…*)
Identifier	Référence univoque à la donnée dans un contexte donné (*DOI ou autres*)
Source	Référence à une ressource dont la donnée décrite est dérivée (en tout ou partie)
Language*	Langue du contenu intellectuel de la donnée
Relation	Référence à une ressource apparentée (*références bibliographiques ou connexes*)
Coverage	Couverture spatio-temporelle (*position géographique, dates*)
Rights	Informations sur les droits associés à la donnée

Crédits : Morgane Mignon, URFIST Bretagne - Pays-de-Loire

Huma-Num : des services pour les DoR en SHS tout au long du cycle de vie des données

ORGANISATION

Des services pour organiser le travail collaboratif autour de vos données.

- ShareDocs
- GitLab
- Kanboard
- Mattermost

TRAITEMENT

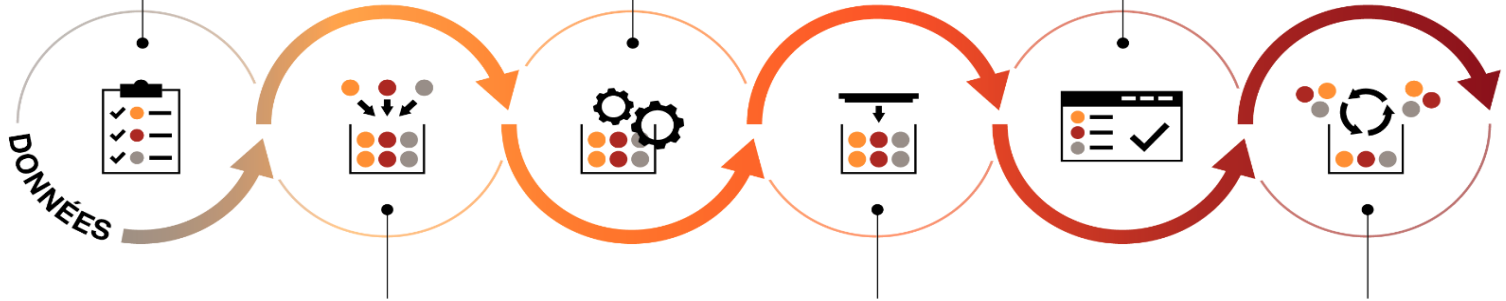
Des services et outils spécifiques pour le traitement et l'analyse de vos données.

- Calcul statistique et environnements R
- Logiciels d'enquête et d'analyse de données
- Reconnaissance de caractères
- Puissance de calcul (+ CC-IN2P3)
- ..sharedocs (conversion pdf, transcodage, OCR)

PUBLICATION

Vos données peuvent être publiées depuis Nakala sur le web et signalées dans Isidore, moteur de recherche pour les SHS.

- Hébergement Web
- Machines Virtuelles
- Nakala
- Isidore



COLLECTE

Des services de stockage sécurisé pour la collecte et la création de vos données.

- ShareDocs
- Huma-Num Box

PRÉSERVATION

Huma-num vous accompagne pour le dépôt et la documentation de vos données dans Nakala, entrepôt pour les données en SHS.

- Nakala
- Huma-Num Box
- Préservation à long terme (+ CINES)

RÉUTILISATION

Vos données entreposées dans Nakala et signalées dans Isidore sont réutilisables.

- Portail web
- API
- Triplestore
- OAI-PMH

Crédits :

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

2. PRÉPARER SES DONNÉES AVEC SHAREDOCS



Crédits : Roman / Pixabay

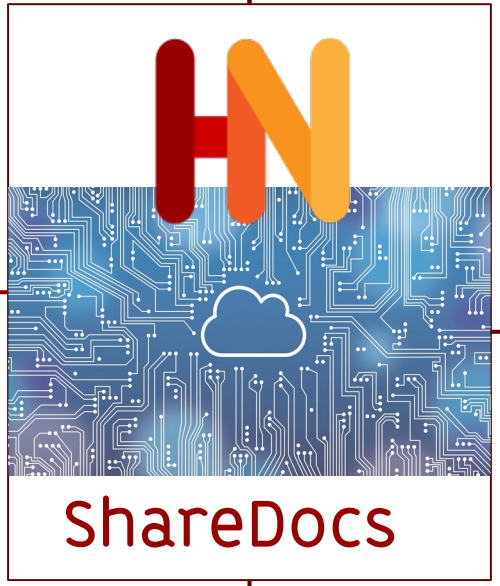
BULAC
[ספריה] [大学] [γλώσσες] [ቤተ-ፍጫ]
Bibliothèque universitaire des langues et civilisations

2. Préparer ses données avec Sharedocs

1. Créer, importer, visualiser et éditer ses fichiers

2. Organiser ses fichiers dans une arborescence de dossiers et collections et avec des mots-clefs

3. Traiter ses fichiers à l'aide d'outils embarqués (transcodage audio/vidéo, OCR, conversion de pdf…)



4. Partager ses fichiers (e-mails ou liens sécurisés)

5. Compléter les métadonnées, ajouter des notes et modifier le statut de ses DoR pour faciliter leur curation

Attention

Les métadonnées en Dublin Core appliquées aux fichiers Sharedocs sont internes à l'application et ne peuvent pas être exportées

Crédits : felicities / Freepik

Crédits : Roman, Pixabay / IR Huma-Num

BULAC
[ספריה] [大学] [γλώσσες] [ቤተ-ፍጥ]
Bibliothèque universitaire des langues et civilisations

2. Préparer ses données avec Sharedocs

30 minutes





Démonstration (pas-à-pas)

1. Télécharger le dossier d'exercice suivant https://page.hn/t6eu86 puis importer l'ensemble dans votre espace personnel Sharedocs (de préférence avant la formation).

2. Dans le répertoire "*Stockage_traitement_Sharedocs*", identifier les trois thématiques présentées, puis réunir les fichiers dans des sous-dossiers correspondant à ces thématiques.

3. Dans l'un de ces sous-dossiers, lister les possibles besoins de traitement sur un ou plusieurs fichiers qu'il contient. En fonction des besoins identifiés, marquer les fichiers avec des étiquettes fournies ou personnalisées.

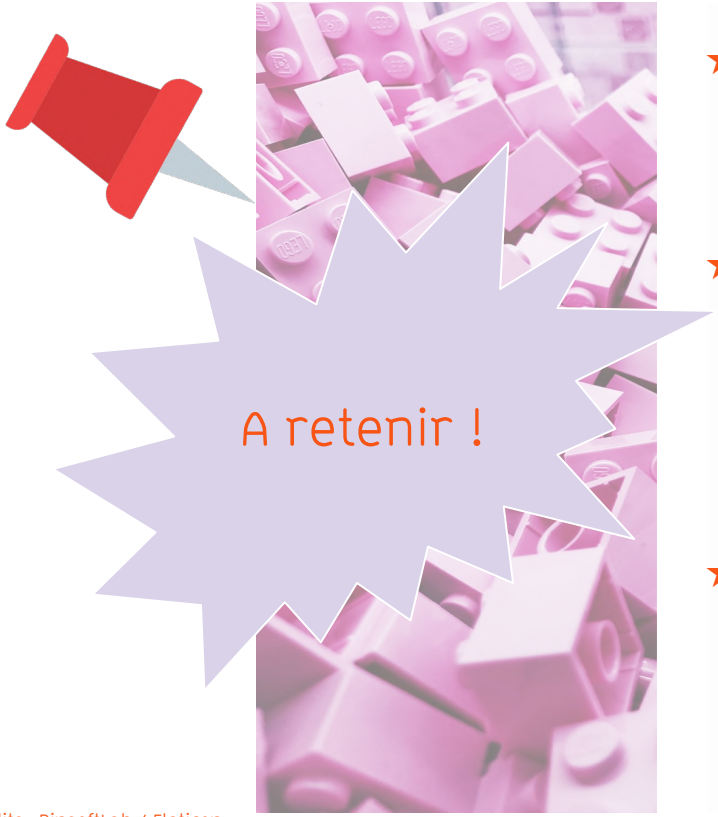
4. Certains fichiers ne sont pas utilisables en l'état et doivent être rejetés. Selon vous, lesquels ? (Marquer avec l'étiquette correspondante).

5. Test sur l'une des fonctionnalités de traitement… par exemple, choisir un des fichiers mp3, et lancer sa conversion au format texte (format medium).

Crédits : Flaticon

Crédits : Ryan Quintal / Unsplash

2. Préparer ses données avec Sharedocs



Crédits : DinsoftLab / Flaticon

- Une donnée est composée d'un seul ou plusieurs fichiers, ainsi que d'une "enveloppe" de métadonnées.
- Un jeu de données est un ensemble contextualisé de données plurielles, organisé par exemple sous la forme d'un tableur ou d'une arborescence de fichiers, compréhensible par des machines (exemples 1, 2, 3)
- Une bonne pratique pour le dépôt d'un jeu de données consiste aussi à les accompagner d'un fichier README, afin d'en faciliter la compréhension par les humains.

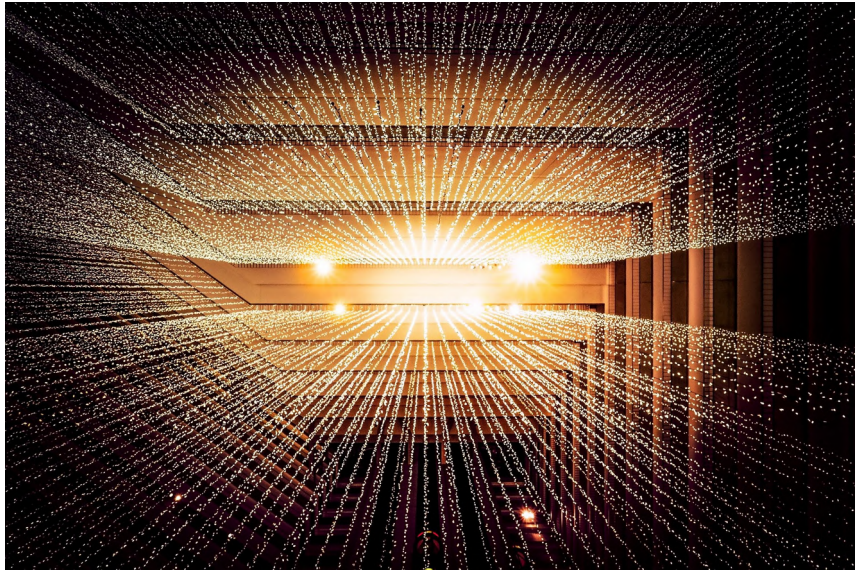
Crédits : Ryan Quintal / Unsplash

BULAC

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

3. DIFFUSER & EXPOSER SES DONNÉES AVEC NAKALA



Crédits : Joshua Sortino / Unsplash

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

3. Exposer et diffuser ses données

1. Déposer (et visualiser) des DoR et métadonnées associées

Ces DoR peuvent être "créées" (privées) ou "publiées" (publiques)

2. Regrouper ses DoR dans des collections cohérentes

(privées ou publiques)

3. Gérer, via des listes d'utilisateurs,

les droits d'accès à ses DoR & collections



4. Créer un site d'exposition

à partir de ses collections publiques via *Nakala_Press*… ou en demander la diffusion via Isidore

… et un tableau de bord

pour visualiser son activité
(statistiques de dépôts, de publications, capacité de stockage, etc.).

Nakala gère les données multilingues

7000 langues disponibles
+ possibilité de rajouter de nouvelles langues pour chaque champ ou de spécifier l'alphabet en dupliquant le champ de métadonnées dédié *(dc:language)*

Crédits : Joshua Sortino / Unsplash
IR Huma-Num

The Royal Library and the Danish National Archives / Jørgen Stamp (CC-BY-2.5) /

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

3. Exposer et diffuser ses données

45 minutes



A vous de jouer !



1. Rendez-vous dans la base https://test.nakala.fr/, puis connectez-vous à l'aide des identifiants et mots de passe fournis sur la page d'accueil.

2. Dans le dossier *"Déposer_Exposer_Nakala"*, choisir un dossier de fichiers, puis faire un dépôt en renseignant les métadonnées obligatoires ou optionnelles à l'aide des fichiers de documentation (ReadMe ou tableur).

3. Dupliquer les métadonnées de contenu (titre, sujet, description) dans une seconde langue de votre choix

4. Créer une collection publique intitulée "Exercice BULAC"

Le champ "Auteur" du formulaire renvoie généralement au créateur de la donnée numérique, qui est généralement le déposant lui-même.

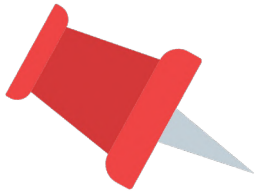


Crédits : Greg Rosenque / Unsplash

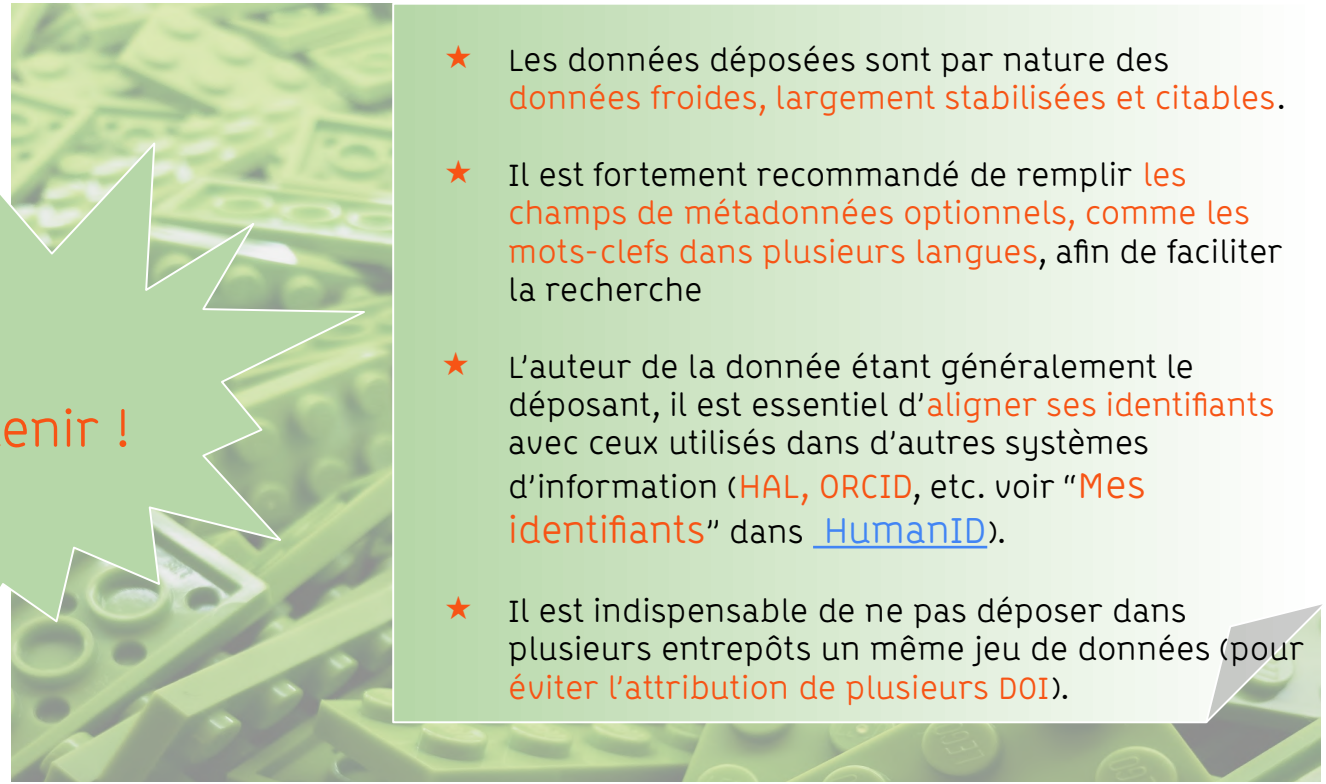
Crédits : Flaticon, felicities/Freepik

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

3. Exposer et diffuser ses données



A retenir !



- ★ Les données déposées sont par nature des données froides, largement stabilisées et citables.
- ★ Il est fortement recommandé de remplir les champs de métadonnées optionnels, comme les mots-clefs dans plusieurs langues, afin de faciliter la recherche
- ★ L'auteur de la donnée étant généralement le déposant, il est essentiel d'aligner ses identifiants avec ceux utilisés dans d'autres systèmes d'information (HAL, ORCID, etc. voir "Mes identifiants" dans HumanID).
- ★ Il est indispensable de ne pas déposer dans plusieurs entrepôts un même jeu de données (pour éviter l'attribution de plusieurs DOI).

Crédits : DinsoftLab / Flaticon

Crédits : Greg Rosenque / Unsplash

BULAC
[ספריה] [大学] [γλώσσες] [ศิลปะ]
Bibliothèque universitaire des langues et civilisations

3. Exposer et diffuser ses données

1. Déposer (et visualiser) des DoR et métadonnées associées

Faire la distinction entre…

nakala

… un dépôt public…

Donnée publiée

Donnée publiée sous embargo

Pas de limite de stockage
Attribution automatique d'un DOI
La publication est définitive

Une donnée publiée et ses métadonnées sont accessibles par tou·te·s depuis le moteur de recherche de Nakala. La donnée peut être référencée dans Isidore et publiée sur un site web.

Seules les métadonnées sont accessibles pendant une période d'embargo temporaire ou indéfinie. Leur intégration dans une collection publique ou un site Web demeure possible.

Crédits : Morgane Mignon, URFIST Bretagne - Pays-de-Loire / IR Huma-Num

…un dépôt privé…

Donnée "créée"

État transitoire avant publication
Suppression possible
Espace de stockage limité à 100 Go

La donnée et ses métadonnées sont accessibles seulement pour les utilisateurs ayant des droits d'administration, d'édition ou de lecture.
Elles sont invisibles dans le moteur Nakala



Crédits : storyset / Freepik

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

3. Exposer et diffuser ses données

2. Regrouper ses DoR dans des collections cohérentes (publiques ou privées)

Les quatre points importants pour les collections sont…

- ❑ L'organisation : pas de hiérarchie entre les collections
- ❑ Le statut des contenus : une collection publique ne peut contenir que des données publiées
- ❑ La gestion des accès : un gestionnaire, des administrateurs, des éditeurs, et des lecteurs de collections privées
- ❑ La publication et l'exposition : Les collections publiques sont accessibles par tout le monde depuis le moteur de recherche Nakala, mais pas les collections privées.

La publication Web est liée à une collection, et non à un compte-déposant !

Crédits : Morgane Mignon, URFIST Bretagne - Pays-de-Loire / IR Huma-Num



Crédits : storyset / Freepik

Exemple 1

Exemple 2

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

3. Exposer et diffuser ses données

3. Gérer, via des listes d'utilisateurs, les droits d'accès à ses DoR & collections

La liste permet une saisie plus rapide des partages, et une gestion évolutive des membres d'un projet

nakala



Pour créer une liste…

1. Aller dans "Listes" et "Créer une nouvelle liste"
2. Renseigner le titre et ajouter les différents utilisateurs qui disposent d'un compte Huma-Num
3. Pour chaque utilisateur, attribuer un rôle (administrateur ou membre)

Comme évoqué plus haut, vous pouvez aussi gérer vos partages plus finement au niveau d'une donnée ou d'une collection.

Crédits : Morgane Mignon, URFIST Bretagne - Pays-de-Loire / IR Huma-Num

Crédits : storyset / Freepik

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

3. Exposer et diffuser ses données

4. Créer un site d'exposition
à partir de ses collections publiques via *Nakala_Press*…

Un site Web est une vitrine, et il permet aussi de réfléchir à la structuration et à l'intelligibilité de ses données.

Pour créer un site Web sur Nakala…

1. Demander l'accès à Matomo via son Human-ID *(pour intégrer l'analyse d'audience Web)*
2. Trouver la collection publique à exposer, puis "Créer une site NAKALA_PRESS"
3. Choisir le préfixe du site. *Exemple : https://monprojet.nakala.fr*
4. Paramétrer le site Web à partir du *dashboard* ainsi généré (langues, apparence générale,…).
5. Activer le site Web, puis l'organiser et le structurer en créant des pages Web qu'il est possible de hiérarchiser (pages "parents" et "enfants" pour créer une arborescence à deux niveaux).

Crédits : Morgane Mignon, URFIST Bretagne - Pays-de-Loire / IR Huma-Num



Articles de synthèse et *data papers sont* d'autres moyens de valoriser simplement ses données.
Exemple

Crédits : storyset / Freepik

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

3. Exposer et diffuser ses données

… ou adosser un site déjà existant à Nakala
à partir de l'API (*Application programming Interface*)

Environ
70 requêtes actionnables via 5 verbes distincts
permettent d'utiliser les fonctionnalités de Nakala à partir d'une autre entité (format d'échange JSON)



Exemple d'usage de clef API avec l'application Mynkl (dépôt par lots)

1. Récupérer la clef d'API dans Nakala, puis se connecter à Mynkl à l'aide de cette clef.

2. Télécharger le modèle de tableur.csv qui va permettre de décrire et déposer les données de votre lot, puis renseignez-le. Attention, aux 7 colonnes obligatoires, parmi lesquelles les 5 colonnes "*Nakala terms*"

3. Téléverser le tableur correctement renseigné, ainsi que les fichiers correspondants aux lignes complétées (une ligne par donnée).

Cliquer sur "Je découvre la démo" pour visualiser le tutoriel complet de dépôt par lots !

Crédits : storyset / Freepik
Morgane Mignon
URFIST Bretagne - Pays-de-Loire / IR Huma-Num

3. Exposer et diffuser ses données

10 minutes



A vous de jouer !

1. Dans la base https://test.nakala.fr/, créer une liste d'utilisateurs (utiliser les noms d'utilisateurs fictifs qui apparaissent par auto-complétion) et appliquer cette liste à la collection publique créée précédemment.

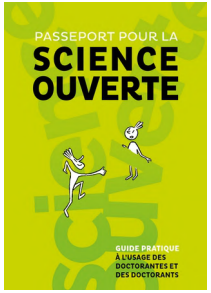
2. Commencer le paramétrage d'un mini site Web à l'aide du bouton *Nakala Press* (il n'est pas possible de finaliser l'opération dans la base de test).

Crédits : Greg Rosenque / Unsplash

Crédits : Flaticon, felicities/Freepik

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

Pour aller plus loin…



Des ressources utiles sur la bonne gestion et la diffusion de ses données de recherche

Doranum
une plateforme d'autoformation complète aux DoR

DMP Opidor
un site d'accompagnement pour créer son propre PGD

UAR2999 distam

Le passeport pour la science ouverte

Cat OPIDoR

Wiki collaboratif des services dédiés aux DoR selon les étapes de leur cycle de vie

DISTAM, agora numérique des études aréales

CoopIST
COOPERER EN INFORMATION SCIENTIFIQUE ET TECHNIQUE

Documentation Nakala

Documentation Sharedocs

HN ShareDocs

Crédits : IR Huma-Num

nakala

Introduction générale aux DoR

Crédits : Coop IST - Cirad

Un Tuto@Maté complet sur l'utilisation de Nakala et Nakala_Press

B U L A C

[ספריה] [大学] [γλώσσες] [ཤེས་རིག]

Bibliothèque universitaire
des langues et civilisations

Crédits : *Carte ethnographique de l'Europe, Juozas Gabrys* (Exposition "Kaléidoscope ukrainien" / BULAC).

services-recherche@bulac.fr

assistance@huma-num.fr
cogrid@huma-num.fr
nakala@huma-num.fr
isidore@huma-num.fr

