

ANF Huma-Num 2022

Maîtriser l'exposition des données entreposées dans Nakala

<https://anf2022-humanum.sciencesconf.org/>

Introduction aux services d'Huma-Num

Jeudi 24 novembre 2022
Helene.jouguet@huma-num.fr
Licence de réutilisation : cc-by-nc

« L'exposition des données
entreposées dans Nakala »

Exposition

Voir et accéder à l'ensemble notice de métadonnées + fichiers depuis Nakala

Voir et accéder aux fichiers déposés depuis Nakala

Consulter des données depuis un dispositif extérieur à Nakala

Voir et accéder aux notices de métadonnées enregistrées dans Nakala depuis l'entrepôt OAI

Autres...

La gestion des données de recherche

Enjeux

Incitations et politiques publiques

Cadre de mise en oeuvre

Economique : Coût de la mauvaise gestion des données
(3 Milliards d'euros pour la France)

Difficultés de gestion : mauvaises pratiques
de stockage, perte de données, données
inutilisables,
« gouffre de données non utilisées »

Enjeux

absence d'ouverture, impossibilité de
reproduire des résultats

positionnement des éditeurs commerciaux

Ressource :
<https://www.ouvrirlascience.fr/les-donnees-de-la-recherche-un-sujet-strategique/>

FRANCE
Incitations et
politiques
publiques

2013	BSN Bibliothèque scientifique numérique
2016	Loi pour une république numérique : obligation d'ouverture des données publiques, libre réutilisation des données de la recherche publique Définition des principes FAIR (à l'échelle internationale)
2018	Création du Comité pour la Science Ouverte (CoSO) Plan National Science Ouverte 1
2019	Obligation de rédaction de PGD et Appel Flash science ouverte par l'ANR
2020	« Pour une politique publique de la donnée », Rapport demandé par le Premier ministre au député EricBothorel, déc. 2020 Nomination d'Isabelle Blanc, administratrice ministérielle des données, des algorithmes et des codes sources du MESRI (CNRS) Plan Données de la recherche novembre 2020
2021	PNSO 2 « Politique des données des algorithmes et des codes sources Feuille de route 2021-2024 »
2022	Plateforme fédérée <i>RechercheDataGouv</i>

Ressource :

Violaine Rebouillat. Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs. Sciences de l'information et de la communication. Conservatoire national des arts et métiers - CNAM, 2019. Français. (NNT : 2019CNAM1254). (tel-02447653)

Principes FAIR

Findable a pour objectif de faciliter la découverte des données par les humains et les systèmes informatiques et requiert une description et une indexation des données et des métadonnées (Tout ce qui sert à décrire et caractériser une donnée, comme son titre, le nom de ses auteurs, sa taille, son format ou la date de sa création).

	NAKALA	ISIDORE	Dispositif web
F1 Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne	oui	oui	possible
F2 Les métadonnées décrivant les données sont riches	possible	possible	possible
F3 Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher	oui	oui	possible
F4 Les métadonnées spécifient l'identifiant de la donnée	oui	oui	possible

Principes FAIR

Accessible encourage à stocker durablement les données et les métadonnées et à faciliter leur accès et/ou leur téléchargement, en spécifiant les conditions d'accès (accès ouvert ou restreint) et d'utilisation (licence).

	NAKALA	ISIDORE	Dispositif web
A1 Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé.			
A1.1 Le protocole utilisé est ouvert (accessible à tous gratuitement), libre (sans licence restrictive d'utilisation) et peut être implémenté de manière universelle.	oui	oui	possible
A1.2 Le protocole utilisé permet l'accès par authentification et autorisation si besoin.	oui	-	possible
A2 Les métadonnées sont accessibles même quand les données ne le sont pas ou plus	oui	-	possible

Principes FAIR

Interoperable peut se décomposer en : téléchargeable, utilisable, intelligible, et combinable avec d'autres données, par des humains et des machines. Pour garantir cela, les (méta)données doivent respecter un schéma standard, commun et partagé pour garantir une harmonie. Ce principe préconise d'utiliser des standards comme par exemple les normes ISO, des réglementations pour structurer des informations géographiques, la retranscription de dates... et éviter les confusions entre les systèmes et la mauvaise interprétation des informations.

	NAKALA	ISIDORE	Dispositif web
I1 Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.	oui + possible	oui + possible	possible
I2 Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR.	oui + possible	oui + possible	possible
I3 Les données et les métadonnées incluent des liens vers d'autres (méta)données.	possible	oui	possible

Principes FAIR

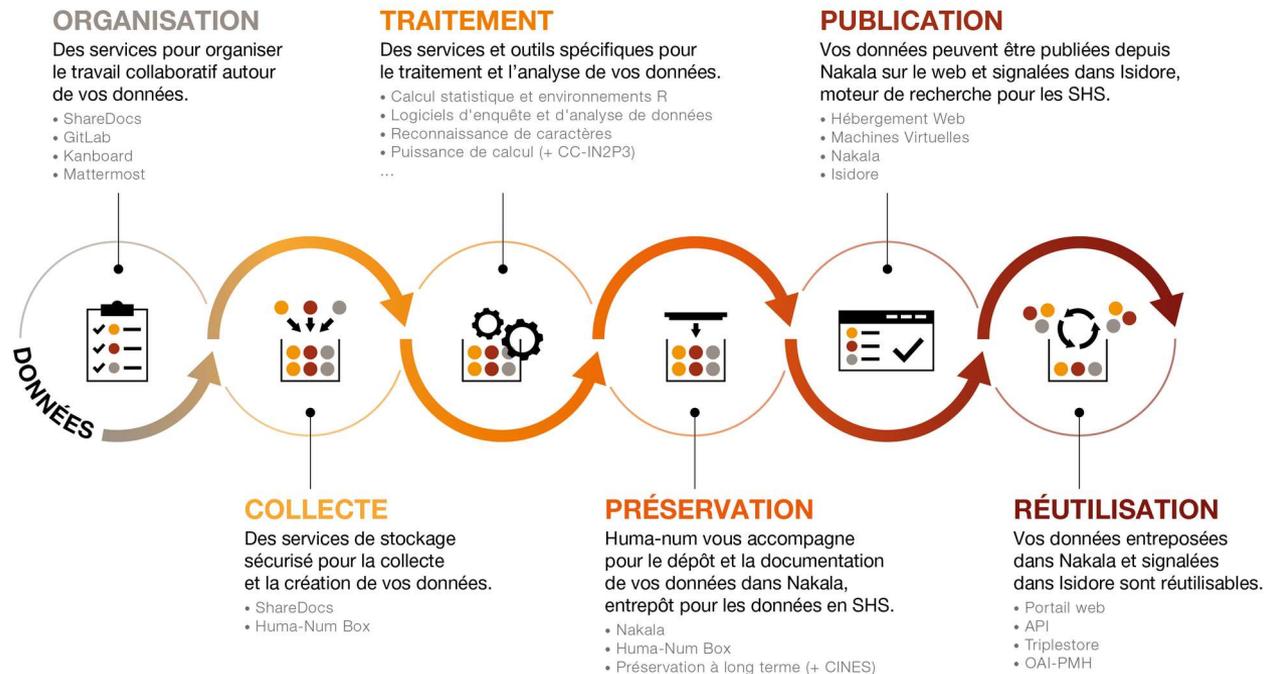
Reusable met en avant les caractéristiques qui rendent les données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction/transparence de la science). Il vise en premier lieu à rendre tous les résultats vérifiables.

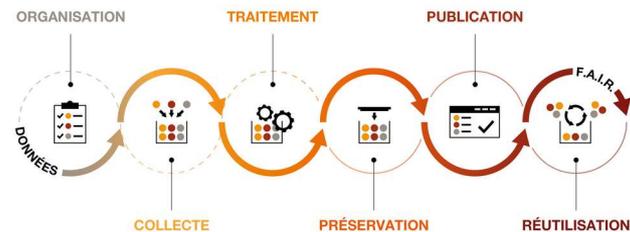
	NAKALA	ISIDORE	Dispositif web
R1 Les données et les métadonnées ont des attributs multiples et pertinents.			
R1.1 Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible.	oui	possible	possible
R1.2 Les données et les métadonnées sont associées à leur provenance.	possible	possible	possible
R1.3 Les données et les métadonnées correspondent aux standards des communautés indiquées.	possible	possible	possible

Nakala dans les services d'Huma-Num



DES SERVICES POUR LES DONNÉES EN SCIENCES HUMAINES ET SOCIALES





Des données brutes aux données F.A.I.R.

Facile à trouver • Accessible • Interopérable • Réutilisable

Avant et au début du projet

Mettre à l'abri les données et documents
S'organiser en équipe

Travail sur les données

Travailler collectivement
Traiter, transformer
Organiser

Publier

Avec préservation :

Dépôt des données
Archivage

Sans préservation :

Publication web

Signaler

ISIDORE

Autre...

D'autres services qui bénéficient de l'infrastructure d'Huma-Num pour produire, enrichir et signaler des données de recherche

Pour enrichir et normaliser les descriptions

OpenTheso
logiciel libre de gestion de thésaurus multilingue
<https://opentheso.huma-num.fr/>

Pour créer des bases de données

Heurist
<https://heurist.huma-num.fr>

Pour signaler des sources dans un service qui réalise de l'enrichissement et de la catégorisation de métadonnées



<https://isidore.science/>

Services pour les données de recherche

Les fonctions

Stocker

Archiver ?

Décrire

Identifier

Exposer / Donner accès

Requêter

Lire

Ecouter

Visionner

Spatialiser

Gérer les droits d'accès aux fichiers

Des objectifs
communs

Favoriser la découverte / Disséminer

Faciliter l'accès

Permettre la réutilisation

Assurer la conservation sur le long terme

D'autres outils/entrepôts pour les données
de recherche en SHS

Ils couvrent tout ou partie des fonctions
possibles

Ils sont spécialisés dans certaines fonctions
(visualisation par exemple)

Services pour
les SHS en
lien avec
l'infrastructure
Huma-Num

<https://www.huma-num.fr/les-consortiums-hn/>

Conservatoire National des Données 3D
<https://3d.humanities.science/>
Consortium 3D SHS

Cocoon <https://cocoon.huma-num.fr>
Collections de CORpus Oraux Numériques
Site de dépôt et de diffusion spécialisé dans le langage oral

Navigae <https://www.navigae.fr/>
Plateforme de recherche et de visualisation cartographique

La Fabrique numérique du passé <https://arscan.opendatasoft.com>
Signalement et accès aux jeux de données géohistoriques

ArkeoGIS <https://arkeogis.org/>
Partage et interrogation de données spatialisées sur le passé

D'autres...

https://cat.opidor.fr/index.php/Entrep%C3%B4t_de_donn%C3%A9es

Services pour
les données
de la
statistique et
les données
sensibles

IR* PROGEDO

Données d'enquêtes pour la recherche en sciences
sociales

Quetelet-Progedo-Diffusion

Archives des Données Issues de la Statistique Publique (ADISP)
Datalab de l'Institut national d'études démographiques (Ined)

<https://www.progedo.fr/>

CASD

Centre d'accès sécurisé aux données
Données des Ministères de la justice, de l'éducation, de
l'agriculture, des finances.
Données entreprise et de Big Data.

<https://www.casd.eu/>

Données de santé

Statut des données

→ Délégué à la Protection des Données (DPD)

- Recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé
- Données impliquant une certification Hébergeur de Données de Santé (HDS)

<https://www.cnil.fr/fr/sante>



Lancement en 2014 :
pour répondre aux besoins d'accès
persistant et interopérable aux données
numériques

conçu en 2 ensembles de briques :

- accès aux données
- présentation des métadonnées

Les services proposés par NAKALA

Accès et citation
par un identifiant pérenne
(handle)



Stockage sécurisé

Accès permanent



Services d'accès aux données



Exposition dans un
Triple Store RDF

Accès via un
entrepôt OAI-PMH



Services interopérables
de présentation des métadonnées

Ces services sont proposés en partenariat avec le CC-IN2P3.

Evolution majeure en 2020

Mise en place d'interfaces web de dépôt et de consultation plus conviviales

Déposez vos données

Déposer les fichiers ici, coller ou [navigateur](#)

Métadonnées principales du dépôt

Type de dépôt (O) [Liste déroulante]

Titre (O) [Texte] +

Statut [Liste déroulante] [Anonyme]

Date de création (O) [Texte] [Historique]

Statut [Liste déroulante]

Informations complémentaires (optionnel)

Description (O) [Texte] +

Notes (O) [Texte] +

Langues (O) [Texte] +

Ajouter d'autres métadonnées (optionnel) [Ajouter une métadonnée]

Relations vers d'autres données publiées dans NAKALA (optionnel) [Ajouter une relation]

CNRS-CFEETK 82408. Karnak, KIU 1027 / chapelle blanche XIIe dynastie / Sesostris Ier

ID : 112800a31e873

Auteur : Anonyme



Mots-clés

8011025 <http://ark.huma-num.fr/ark:/61902/3Q011025>

Licence

Creative Commons Attribution 4.0 International (CC BY 4.0)

> [Afficher la liste complète des métadonnées](#)

Collections

Fonds CFEETK - USR 3172 du CNRS

Projet Karnak - Index global des inscriptions des temples de Karnak (Labex ARCHEMEDE - ANR-11-LABX-0932-01)

Citer

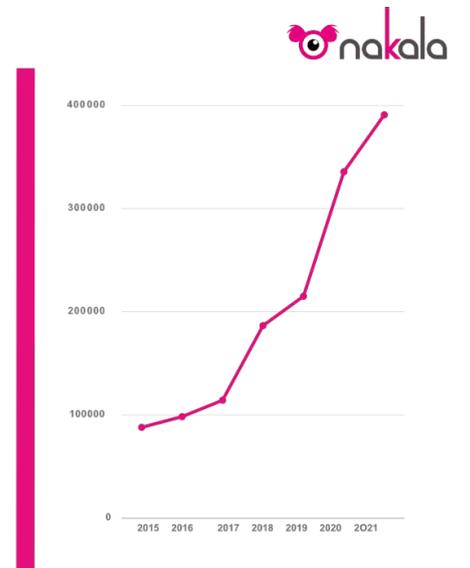
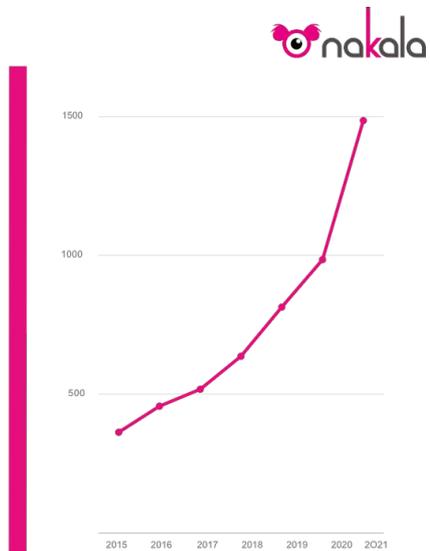
Fils d'Auteurs (2020) «CNRS-CFEETK 82408. Karnak, KIU 1027 / chapelle blanche XIIe dynastie / Sesostris Ier» [Image] NAKALA, <https://naka.org/112800a31e873>

Partager



Fonction	Nakala < 2020
Accès et citation par identifiant pérenne	(Handle) - DOI
Stockage sécurisé et accès permanent	Infrastructure Huma-Num (Centre de Calcul de l'IN2P3)
Accès via entrepôt OAI-PMH	Endpoint : https://api.nakala.fr/oai2
Accès via API	Endpoint : https://api.nakala.fr/doc Recherche, gestion des données et des collections, gestion des groupes et des utilisateurs, consultation des vocabulaires
Exposition dans un triplestore RDF	Endpoint : https://www.nakala.fr/sparql ! Données de décembre 2020 : En cours de mise à jour

Projet de certification en cours d'instruction

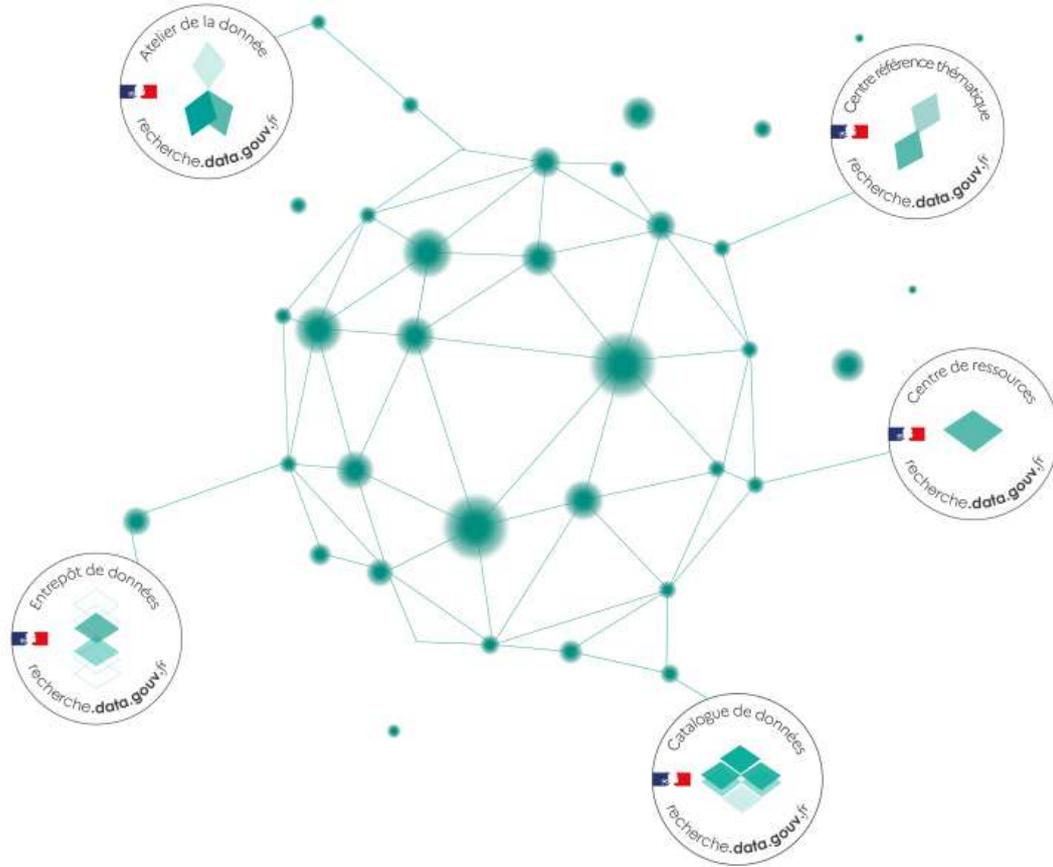


	2015	2016	2017	2018	2019	2020	2021
Volume total dans NAKALA en Tétra Octets	0,7 To	1,2 To	1,7 To	non mesuré	2,8 To	3,64 To	5,1 To

La structuration du paysage national

L'écosystème *Recherche Data Gouv*

<https://recherche.data.gouv.fr/fr>



RECHERCHE DATA GOUV



Module RDG	Interactions existantes
Catalogue	Nakala sera moissonné dans le catalogue (prévision 2023)
Ateliers de la donnée	Réseau des correspondants dans les MSH, services accompagnement dans les universités et les établissements
Centres de références thématiques	Huma-Num centre de référence thématique : productions (guides, standards communautaires, livres blancs, formations) des Consortiums-HN
Centres de ressources	Dorandum, Opidor, Urfist

<https://documentation.huma-num.fr/>

Les services d'Huma-Num sont, hors exceptions, gratuits pour les projets (mais ils coûtent).

Comment contacter l'IR* Huma-Num :

Vous souhaitez nous faire part de besoins concernant votre projet ou demander directement l'ouverture de services :

- contactez par mail l'équipe Huma-Num : **cogrid@huma-num.fr**
- les demandes d'accès à une partie des services sont directement accessibles sur humanid (ShareDocs, GitLab, Matomo, Mattermost, NAKALA, ISIDORE, Stylo, Kanboard) : <https://humanid.huma-num.fr/>.

L'équipe prend soin d'analyser les demandes de manière collégiale sur les aspects scientifiques, documentaires, informatiques afin d'apporter une réponse adaptée à vos besoins tout en permettant de rationaliser et mutualiser les moyens mis en oeuvre. Selon la nature de la demande le temps de réponse peut varier entre 1 jour et 2 à 3 semaines.

- Vous rencontrez un problème dans l'utilisation d'un service délivré par Huma-Num : contactez assistance@huma-num.fr
- Pour toute autre question sur l'IR* Huma-Num : contactez contact@huma-num.fr

Quelles sont les conditions pour utiliser les services d'Huma-num : Huma-Num délivre un ensemble de services et outils pour les données produites dans les projets de recherche en Sciences Humaines et Sociales : <https://www.huma-num.fr/presentation/#conditions>