

Module 2 : Le Dépôt dans NAKALA



Adrien Desseigne Pôle ACCES

ANF-2021 Gérer ses données en SHS avec les services et outils proposés par la TGIR Huma-Num 16 septembre 2021



Déroulé du module 2



- Les entrepôts de données de recherche
 - L'entrepôt de données NAKALA
- TP Déposer un jeu de données dans NAKALA









Qu'est-ce qu'un entrepôt de données de recherche?

Un entrepôt de données de recherche (Research Data Repository ou Data Repository) est un service autour des données permettant :

- leur dépôt
- leur description
- leur conservation
- leur recherche
- leur diffusion en vue de leur réutilisation











Différents types d'entrepôts de données

Il existe de nombreux entrepôts de données que l'on peut classer selon plusieurs critères.

Par quelle type d'organisation sont-ils portés ?

-organisation publique : <u>Zenodo</u> (collaboration entre l'infrastructure européenne <u>OpenAIRE</u> et le <u>CERN</u>)

-organisation à but non lucratif : Dryad

organisation commerciale : <u>Figshare</u> (développé par la société <u>Digital Science</u>)









Différents types d'entrepôts de données

Il existe de nombreux entrepôts de données que l'on peut classer selon plusieurs critères.

Pour quel type de données ?

-pluridisciplinaires : Zenodo, Figshare, Dryad -thématiques ou disciplinaires : <u>GenBank</u> (séquences génétiques), <u>Cocoon</u> (corpus de parole), <u>ArkeoGIS</u> (données archéologiques)











Pourquoi déposer ses données dans un entrepôt ?

Déposer ses données dans un entrepôt présente plusieurs avantages :

- Conservation des données dans un environnement sécurisé
- -Meilleure visibilité des données et accès facilité pour les moteurs de recherche
- Interopérabilité des données grâce à l'utilisation de standards de métadonnées
- Citation du jeu de données facilitées par son identifiant pérenne



- -Gestion des modalités de partage des données par l'attribution de licences de diffusion
- -Respect des recommandations des financeurs et institutions sur l'ouverture des données
- -Reproductibilité de la recherche, intégrité et validation scientifique améliorées
- -Valorisation des données par leur réutilisation dans de nouvelles études







Comment choisir son entrepôt de données ?

- Type de données que vous allez déposer
- -Recommandations de votre institution, des financeurs de votre projet ou de la revue dans laquelle vous publiez
- -Reconnaissance dans votre discipline et par la communauté scientifique



Pour sélectionner votre entrepôt, vous pouvez prendre en compte les critères suivants :

-Choix des licences disponibles

- Conservation des données

- Champs de métadonnées disponibles

-Autres fonctionnalités (visionneuses, dépôt par lots, versionnage des dépôts,...)









Comment choisir son entrepôt de données ?

Plusieurs annuaires permettant de sélectionner des entrepôts de données en fonction de ces critères :

pour le partage des données de recherche. Re3data compte actuellement plus de 2700 entrepôts.

- <u>Repository Finder</u> : outil hébergé par DataCite permettant de trouver un entrepôt de données dans le répertoire Re3data.
- de 127 entrepôts de données en France. Hébergé par le CNRS.



- <u>Re3Data</u> : répertoire d'entrepôts de données créé en 2012 par le consortium international <u>DataCite</u> qui œuvre

- <u>Cat OPIDoR</u> : catalogue des services français dédiés aux données scientifiques. Cat OPIDoR présente une liste









L'entrepôt de données NAKALA

Qu'est-ce que NAKALA ?

Un entrepôt qui a pour vocation la publication de données issues de projets de la "<u>science ouverte</u>".

Un service mis en place et maintenu par la TGIR Huma-Num (depuis 2015).



recherche en SHS en s'efforçant de mettre en œuvre les principes FAIR et les valeurs de







- Plateforme de dépôt de fichiers (tout type de formats)

*0 *	Rechercher des données dans Nakala Q Déposer TN
-	Déposez vos données
Tableau de bord	
🗁 Données	
Collections	
	Déposer les fichiers ici, coller ou naviguer
🍪 Listes	
🖲 Sites web	





Déposez vos données

	Déposer les fichiers ici, coller ou naviguer					
✓ Termine	é					
md	README.md	4.6 kB				
pdf	59539.pdf	121.9 kB				
jpg	001 003 037_35a.jpg	591.6 kB				
mp3	5860.mp3	7.0 MB				
			Cliquez sur un fichier ou plusieurs fichiers pou			
			pouvoir paramétrer son accessibilité et éditer s description.			











- Ajout de métadonnées (Dublin Core, mais vocabulaire extensible sur demande)

Type de depot 🕐	
Quel est le type de votre donnée?	•
Titre ⑦	
Pas d'information de langue 👻	+
Auteurs	
	🗆 Anonyme
Date de création ③	
	🗆 Inconnue
Licence	



cf. ANF Module 2

Aiouter d'autres	métadonnées Optionnel		
Propriété	• Туре	 Pas d'information de langue 	•
Propriété			/
dcterms:abstract	Ajouter u	une métadonnée	
dcterms:accessRights			
dcterms:accrualMethod			
dcterms:accrualPeriodicity			
dcterms:accrualPolicy	onnée avec d'au	tres utilisateurs Optionnel	
determstalternative			





























- Gestion fines des droits d'accès aux dépôts, collections et fichiers

Chercher un utilisateur ou une lis	te		•
Utilisateur Nakala #1	0- ×	Utilisateur Nakala #2	• • ×
			C administrateu
			✓ editeur ✓ lecteur



Déposer les fichiers ici, coller ou naviguer										
🗸 Terminé										
			5860.	mp3	5					
✓ mp3 5860.mp3		7.0 MB	5000.	mps						
			Date	de vi	sibil	ité				
			23/	/03/	202	1		Ľ		Jamais
			mars	s 202 [.]	1 -			\uparrow	\downarrow	
			L	м	м	J	v	s	D	
			1	2	3	4	5	6	7	
			8	9	10	11	12	13	14	
			15	16	17	18	19	20	21	(+ En
			22	23	24	25	26	27	28	
			29	30	31	1	2	3	4	
			5	6	7	8	9	10	11	











- Attribution d'identifiants pérennes (<u>Handle</u> à l'origine, <u>DOI</u> depuis 2020)



Resolve a DOI Name

Type or paste a DOI name, e.g., 10.1000/xyz123, into the text box below. (Be sure to enter all of the characters before and after the slash. Do not include extra characters, or sentence punctuation marks.)

10.34847/nkl.aeaaty7f



















cf. ANF Module 2

- Moteur de recherche sur les collections, données, métadonnées et contenu des fichiers

n	adagascar	Q	Fr 🔹 Se conn
7 résultats pour " ma	adagascar "		
Filtres Article de journal × 2014 ×			
	Filtrer par : Type - L	icense • Années •	Trier par : Pertinence -
Contraction for political power Contraction for political power Contraction for political power Contraction Contracti	<page-header></page-header>		
AU's role in peaceful competition for political po Anonyme 25/09/2020	Achievements realized by Taita Taveta during th Anonyme 25/09/2020	Kenya elected president of regional Catholic scri Anonyme 25/09/2020	It's a dream come true Anonyme 24/09/2020
	WORLD	9	
Nyanza MPs ahead of ODM polls	Seven dead as western Madagascar hit plague	Kenya`s liberal democratic state is alding terrori	

À propos	T	Contact	I	API	I	OAI	I	Mentions légales
Service développé par Huma-Num								













L'entrepôt de données NAKALA

Principes clés

- Expositions des données : OAI-PMH

OAI 2.0 Nakala Request Results

Identify | ListSets | ListMetadataFormats | ListIdentifiers | ListRecords

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option.

Datestamp of response 2021-03-23T08:57:31Z

Request URL https://api.nakala.fr/oai2

verb GetRecord

Request was of type GetRecord.

OAI Record: oai:nakala.fr:hdl_11280_751da5ed

OAI Record Header

OAI Identifier oai:nakala.fr:hdl_11280_751da5ed oai_dc formats

Datestamp 2020-01-21

setSpec hdl_11280_650b406b Identifiers Records setSpec hdl_11280_64d48dce Identifiers Records

Dublin Core Metadata (oai dc)

Resource Identifier	11280/751da5ed
Resource Identifier	https://nakala.fr/11280/751da5ed
Resource Identifier	https://api.nakala.fr/data/11280/751da5ed/0aa01d28d5c8415986a0466c4546e1b68335a429
Title	Vignette PDF
Date	2019
Rights Management	CC-BY-4.0
Resource Type	http://purl.org/coar/resource_type/c_c513
Author or Creator	IRIHS
Resource Type	Image
Relation	11280/a5439de7



cf. ANF Module 6











- Expositions des données : RDF

Virtuoso SPAR	QL Query Editor	
Default Data Set Name	e (Graph IRI)	About Namespace Prefixes Inference Rules Macros RI
Query Text		
select distinct ?C	oncept where {[] a ?Concept} LIMIT 100	
(Security restrictions of this Results Format: Execution timeout:	server do not allow you to retrieve remote RDF data, see <u>details</u> .) HTML	d, see <u>details</u>) are ignored)
Options:	 Strict checking of void variables Strict checking of variable names used in multiple clause Log debug info at the end of output (has no effect on so Generate SPARQL compilation report (instead of executed) 	es but not logically connected to each other me queries and output formats) ting the query)
(The result can only be sen Run Query Reset	t back to browser, not saved on the server, see <u>details</u>)	
	Copyright © 2021 OpenLink Virtuoso version 08.03.3314 on Linux (x86_64-generic-	<u>Software</u> -linux-glibc25), Single Server Edition



cf. ANF Module 6















L'entrepôt de données NAKALA

Principes clés

- Expositions des données : API REST



Nakala - API 🚥 [Base URL: api.nakala.local:81

Cette API vous permet de créer, gérer et administrer vos dépôts sur la plateforme NAKALA

Déposez vos données, exposez vos données, valorisez vos données, etc.

search Recherche données et collections de Nakala \sim /authors/search Récupération des auteurs associés aux données de Nakala 1 /search Recherche des données Nakala GET \sim datas Gestion des données **a** /datas/{identifier} Récupération des informations sur une donnée /datas/{identifier} Modification des informations d'une donnée î /datas/{identifier} Suppression d'une donnée DELETE 1 POST /datas Dépôt d'une nouvelle donnée





cf. ANF Module 3 et 4

OpenAP

tas Gestion	n des données	\sim				
GET /da	/datas/{identifier} Récupération des informations sur une donnée					
Retourne l'enser statut de la liste des me liste des co métadonné	emble des informations relatives à la donnée a donnée létadonnées ollections contenant la donnée ées sur les fichiers associés à la donnée					
Parameters		Try it out				
lame	Description					
dentifier * ^{req}	Identifiant de la donnée					
	identifier - Identifiant de la donnée					
version nteger (query)	version de la donnée					
	version - version de la donnée					
	nat					
netadata-form tring <i>query)</i>	format des metadonnees					



Authorize 🔒





- Nakala Press (Module de création de site web)

prévisualiser		fr 🗸 Accueil 🖌 🖌
C	hoisir la disposition du site	
С	Site pleine	Site pleine
	largeur + menu horizontal	largeur + menu vertical
۲	Site largeur	Site largeur
	fixe + menu horizontal	fixe + menu vertical
	Envoyer	

Service développé par Huma-Num



cf. ANF Module 5

en 🤜

Women's Rights and Duties through Letters to the Editor in the Nigerian Press (1940s-1950s)

About Sources Years

References

See: Panata, Sara. 2020. « 'Dear Readers...': Women's Rights and Duties through Letters to the Editor in the Nigerian Press (1940s-1950s) ». Sources. Materials & Fieldwork in African Studies nº 1:141-198.

https://www.sources-journal.org/86 | https://halshs.archives-ouverte s.fr/halshs-02865251/

General Remarks

Sara Panata

Open Access and Reproduction

To date, there is no open-access digital archive of the materials used in this paper.

Reproductions of the archival materials are from the holdings of the National Archives of Ibadan.











L'entrepôt de données NAKALA

Comparaison avec d'autres entrepôts

Nous vous proposons une comparaison détaillée à l'instant t (août 2021) entre NAKALA et :

-Deux entrepôts multidisciplinaires : Figshare et Zenodo

-Un entrepôt disciplinaire : <u>Cocoon</u>

Télécharger la comparaison : <u>https://anf2021-humanum.sciencesconf.org/data/</u> Comparaison NAKALA Figshare Zenodo aout 2021.ods











L'entrepôt de données NAKALA

Informations utiles

- Explorer : <u>https://nakala.fr</u>
- Tester : <u>https://test.nakala.fr</u>
- Documentation : https://documentation.huma-num.fr/nakala
- Demander un compte : <u>https://humanid.huma-num.fr</u>
- Contact : <u>nakala@huma-num.fr</u>











Mise en pratique

<u>Objectif</u> : Récupérer dans Sharedocs le jeu de données nettoyé à l'issue du module 1 de l'ANF et effectuer son dépôt dans la version de test de NAKALA.

Les données seront déposées dans des collections reprenant à plat l'arborescence des dossiers de Sharedocs. Nous attribuerons des droits d'administration sur ces données et ces collections à plusieurs membres d'une même équipe à travers l'utilisation des listes d'utilisateur. Les métadonnées à renseigner pour chaque dépôt sont disponibles dans le tableur associé au jeu de données. On s'arrêtera aux 10 premières données.







Ressources à consulter avant la mise en pratique

- num.fr/nakala-guide-de-description
- Mode d'emploi de l'interface de dépôt : <u>https://rhn2020.sciencesconf.org/data/</u> program/presentation_deposer_dans_nakala.pptx_BrunoVAnderaert.pdf



- Guide de description de ses données dans NAKALA : <u>https://documentation.huma-</u>















Astuce 1 : par où commencer ?

Pour déposer ce jeu de données, vous pouvez procéder dans cet ordre :

- Créer d'abord la ou les listes d'utilisateurs souhaitées.
- créées à l'étape 1.
- déjà créées à l'étape 2.



2. Créer les collections avec leurs métadonnées et attribuer les droits aux listes déjà

3. Déposer les données une à une en renseignant les métadonnées pendant le chargement des fichiers et en attribuant les droits aux listes créées à l'étape 1. Vous pourrez également associer directement chaque dépôt aux collections que vous aurez











Astuce 2 : gérer les droits avec des listes d'utilisateurs

Vous pouvez attribuer des droits sur une donnée ou une collection à n'importe quel utilisateur ayant un compte dans NAKALA.











Astuce 2 : gérer les droits avec des listes d'utilisateurs

- Cela permet une gestion fine des droits, mais présente plusieurs inconvénients :
- -Pour chaque nouvelle donnée, vous devrez saisir de nouveau l'ensemble des droits pour chaque utilisateur.
- -Pour ajouter ou supprimer un nouvel utilisateur sur un jeu de données, vous devrez modifier les droits de chacune des données et des collections auxquelles vous voudrez lui donner accès.







Astuce 2 : gérer les droits avec des listes d'utilisateurs

Pour éviter ces inconvénients, vous pouvez passer par l'utilisation des listes

Ajouter un utilisateur ou une liste d'utilisateurs	
Chercher un utilisateur ou une liste	



Modifier la liste d'utilisateurs





Huma-Num



Astuce 2 : gérer les droits avec des listes d'utilisateurs

Cela permet une gestion simplifiée des droits :

- Pour chaque nouvelle donnée, vous pourrez saisir uniquement la ou les listes concernées et non l'ensemble des utilisateurs.

-Pour ajouter ou supprimer un nouvel utilisateur sur un jeu de donnée, vous pourrez simplement modifier les membres de la liste utilisée sur le jeu de données sans modifier une à une les données et les collections.







Astuce 3 : quel nom donner à mes listes d'utilisateurs ?

Pour vous y retrouver lorsque vous gérez plusieurs listes, vous pouvez par exemple indiquer dans le nom de la liste le rôle qu'assureront les membres de cette liste sur vos données.

Créer une nouvelle lis	te d'utilisateurs	
Titre		
HUMA-NUM (Administrateu	rs)	
	Adrien Desseigne	Propriétaire
	Laurent Capelli	×



ltre		
HUMA-NUM (Lecte	urs)	
	Adrien Desseigne	Propriétaire







Astuce 3 : quel nom donner à mes listes d'utilisateurs ?

Ce sera plus facile ensuite d'attribuer les bons droits aux bonnes listes ou pour ajouter ou supprimer des personnes dans les bonnes listes.





		and all and the set of the			
le	ction	avec d'autres util	isateurs	Optionne	2
•		Propriétaire			
					•
	•••	HUMA-NUM (Lecteurs)	•	• X	





Astuce 4 : gérer à plat une arborescence de collection

Voici une proposition pour reproduire à plat dans NAKALA une arborescence de dossiers et de fichiers du jeu de données



Arborescence sur l'ordinateur



ons	< Partagé	es avec moi			+ Créer une nouvelle c	ollection
	Q	Filtrer par :	Statut 🝷	Année de création 👻	Trier par : Date de dépôt (décroissa	inte) 🔻
				Date de création	Statut	
				07/09/2021	public	:
				15/09/2021	public	:
				15/09/2021	public	:

Collections à plat dans NAKALA pour chaque dossier







Astuce 4 : gérer à plat une arborescence de collection

Voici une proposition pour reproduire à plat dans NAKALA une arborescence de dossiers et de fichiers du jeu de données



Créée le 07. La collection

Arborescence sur l'ordinateur



D : 10 ID : 10 Créée le La collect	JMA-N .34847/nk 07/09/202 tion d'HUM	UM (l.428909a3 11 IA-NUM	le Publique					
								≡
			Filtrer par :	Туре 👻	Licence 🝷	Années 🔻	Trier par : Date de dép	ôt (décroissante) 🔻
			Auteurs				Date de dépôt	
	т	data_1	Pierre Martin				14/09/2021	÷
	т	data_2	Pierre Martin,Jea	n Dupont			14/09/2021	:
		data_3	Jean Dupont				14/09/2021	:
	⊲))	data_4	Pierre Martin				14/09/2021	:
		data_5	Jean Dupont				14/09/2021	:

Collection « racine » qui contient toutes les données



Huma-Num



Astuce 4 : gérer à plat une arborescence de collection

Voici une proposition pour reproduire à plat dans NAKALA une arborescence de dossiers et de fichiers du jeu de données



id : 1	SIDORE 10.34847/1	nkl.9a7bd992	Publique													
Créée	le 15/09/2	021														
						NA ID : 10.	KALA 34847/nkl.8	845f8581	🕑 Pul	blique						
			Filtrer par :	Туре 🝷	Licence 🔻	Créée le	15/09/2021									
	-		Auteurs													
	Т	data_2	Pierre Martin,Je	ean Dupont												
		data_3	Jean Dupont							Filtrer p	par :	Туре 🔻	Licence 🝷	Années 👻	Trier par : Dat	e de dépôt (décroissante) 🔻
											Auteurs			Date de dépô	t	
							41)	data_4			Pierre Ma	artin		14/09/2021		÷
								data_5			Jean Dup	ont		14/09/2021		:

Arborescence sur l'ordinateur



Collections « enfants » pour chaque sous-dossier

(ex.: https://nakala.fr/10.34847/nkl.a597284v)







Le dépôt dans NAKALA

Astuce 5 : forme auteur sans nom, prénom

Lorsque vous renseignez un nouvel auteur dans NAKALA, vous devez saisir au minimum son nom et son prénom :

SMAKALA		
Prénom	Nom	
Pierre	Dupont	ŗiç
Orcid ⑦		
Annuler	+ Ajouter	











Astuce 5 : forme auteur sans nom, prénom

Si vous n'êtes pas en mesure de distinguer le nom du prénom (ex: nom de labo), nous vous conseillons de cocher la case « anonyme » et de renseigner une nouvelle métadonnée dcterms:creator avec la valeur souhaitée :



Solution déconseillée



Anonyme + Ajouter d'autres métadonnées Optionnel dcterms:creator • xsd:string • français • TGIR HUMA-NUM	Auteurs									
+ — Ajouter d'autres métadonnées Optionnel dcterms:creator • xsd:string • français • TGIR HUMA-NUM						🗹 Anonyme				
+ Ajouter d'autres métadonnées Optionnel dcterms:creator xsd:string français TGIR HUMA-NUM										
Ajouter d'autres métadonnées Optionnel dcterms:creator xsd:string français TGIR HUMA-NUM X X X 			+							
dcterms:creator • xsd:string • français • TGIR HUMA-NUM • • • • • • •	— Ajouter d'autres métadonnées Optionnel									
TGIR HUMA-NUM	dcterms:creator	•	xsd:string	•	français		•			
	TGIR HUMA-NUM						1	×		
Ajouter une métadonnée			Ajouter u	ne métadonné	e					

Solution provisoire à privilégier

(nouvelles formes et rôles auteurs dans une future version de NAKALA)



Huma-Num



Astuce 6 : date de création hors format AAAA-MM-JJ

Lorsque vous renseignez une date de création dans NAKALA, vous devez saisir une date du type 2021, 2021-09 ou 2021-09-16 :

	Date de création ⑦	Au format YYYY-MM-DD, YYYY- MM ou YYYY		
2021-09-10	2021-09-16		🗌 Inconnue	









Astuce 6 : date de création hors format AAAA-MM-JJ

Lorsque vous n'êtes pas en mesure de respecter ce format, vous pouvez cocher la case « inconnue » et renseigner une nouvelle métadonnée dcterms: created avec la valeur souhaitée. Si possible, appuyez-vous sur le format EDTE qui sera prochainement intégré à NAKALA.

Date de création ⑦										
		Inconnue								
	+									
— Ajouter d'autres métadonnées Optionnel										
dcterms:created 🗸	xsd:string •	Pas d'information de la	ngue 🔻							
2004-06-~01/2004-06-~20				×						
	Ajouter une métadonnée	2								



Interval

For Level 2 portions of a date within an interval may be designated as approximate, uncertain, or unspecified.

- `2004-06-~01/2004-06-~20' Example 1 An interval in June 2004 beginning approximately the first and ending approximately the 20th
- `2004-06-XX/2004-07-03' Example 2 An interval beginning on an unspecified day in June 2004 and ending July 3.

cf. documentation Extended Date/Time Format (EDTF)







Astuce 7 : ajout d'un mot-clé issu d'un référentiel

Vous pouvez ajouter des mots-clés dans plusieurs langues. Une autocomplétion est proposée sur les mots-clés des référentiels utilisés dans ISIDORE. Pour le moment seule la chaîne de caractère est stockée en base. Pour renseigner une URI particulière, vous devez passer par l'ajout d'une autre métadonnée en sélectionnant le type approprié (ex: dcterms:URI) :

Mots-clés 🕐		
français	•	
× interopérabilité		
espagnol 🔹	×	
× interoperabilidad		
anglais 🔹	×	(+)
× interoperability	^	\bigcirc

Ajout de mots-clés multilingues sans référence à un référentiel particulier



dcterms:subject •	dcterms:LCSH 🗸	Pas d'information de langue 🔹
http://id.loc.gov/authorities/subjects/sh94007902		
		/_
dcterms:subject 🔹	dcterms:URI 🔹	Pas d'information de langue 🔹
http://data.bnf.fr/ark:/12148/cb13168215h		
		/_
Ajouter une métadonnée		

Ajout de concepts issus des référentiels LCSH et RAMEAU















huma-num.fr humanum.hypotheses.org @Huma_Num

TGIR HUMA-NUM CNRS UAR 3598 54 bd Raspail 75006 Paris

