

# Journée d'information et d'échanges sur la gestion des données de recherche en SHS

-

## Publier ses données sur la plateforme Nakala d'Huma-Num

Michel Jacobson

23-novembre-2021

# Nakala

## Un entrepôt de données

- Pour la publication de données en s'efforçant de mettre en œuvre les principes FAIR et les valeurs de la « science ouverte ».
- Un service mis en place et maintenu par Huma-Num (depuis 2015).
- Pour des déposants liés à l'ESR français en SHS
- Pour des réutilisateurs scientifiques ou autres



# Des annuaires

**Cat OPIDoR** : catalogue des services français dédiés aux données scientifiques hébergé par l'INIST. Compte 128 entrepôts de données en France. Hébergé par le CNRS.

## NAKALA

NAKALA allows research teams, who so request, to file their digital data (text files, sound, image, video) in a secure warehouse, with DOI, IIF and which ensures both data availability and quotability time. NAKALA is a full Web application and a repository for humanities and social sciences. It's created and powered in France by Huma-Num, the french infrastructure for digital humanities.

humanities and social sciences digital humanities humanities arts and humanities data warehouse

Less info

Repository URL

<https://www.nakala.fr>

Contact

[cogrid@huma-num.fr](mailto:cogrid@huma-num.fr)

Data Access

restricted or open

Persistent Identifier

DOI

hdl

Certificates

none

Data Upload

restricted: registration

Provider Type

dataProvider

Repository Type

disciplinary

**Repository Finder** : outil hébergé par DataCite pour chercher un entrepôt de données dans le répertoire Re3data (qui compte actuellement plus de 2700 référentiels pour les données de recherche).

## Nakala

NAKALA est un service mis en place par la TGIR Huma-Num. NAKALA est un **entrepôt de données de recherche**. Il s'agit d'une base de données destinée à accueillir, conserver, rendre visible et accessible les données de recherche. Il permet d'enregistrer des données, de les décrire en vue de les exposer et les rendre réutilisables. Ainsi, le dépôt de données dans NAKALA offre des services sur plusieurs étapes du cycle de vie des données, sur la préservation, la publication et la réutilisation.

Le service NAKALA offre **deux niveaux de préservation** :

- Un **niveau par défaut** qui est mis en pratique dès lors qu'une donnée est enregistrée dans NAKALA. La donnée est décrite, contextualisée et stockée de manière sécurisée.
- Un **niveau avancé** qui s'inscrit dans un partenariat avec le CINES. Dans ce circuit de dépôt avancé, la préservation à long terme est assurée par le CINES. Le CINES apporte des garanties supplémentaires telles que la valeur juridique du dépôt, l'intégrité du fichier, la migration des formats pour leur lisibilité, etc.

NAKALA permet à des équipes de recherche, qui en font la demande, de déposer leurs données numériques (fichiers texte, son, image, vidéo) dans un **entrepôt sécurisé** qui assure à la fois l'**accessibilité** aux données et leur **citabilité** (attribution d'**identifiant pérenne DOI**) dans le temps. NAKALA propose des services **interopérables** d'accès aux données elles-mêmes et de présentation des **métadonnées descriptives** (format standard **Dublin Core** étendu). Les données sont **stockées** et **sauvegardées** régulièrement sur les serveurs sécurisés de la TGIR Huma-Num au sein du centre de calcul de l'IN2P3-CNRS.

NAKALA offre une interface Web dont le dépôt de données a été entièrement pensé par et pour les chercheurs(es) avec l'objectif de faciliter la mise en œuvre des principes FAIR en SHS. À ce titre, NAKALA participe activement à la Science Ouverte.

Domaines scientifiques :

Sciences Humaines & Sociales

Individus, marchés et organisations; Institutions, valeurs, environnement et espace; Le monde social, diversité, population; L'esprit humain et sa complexité; Cultures et production culturelle; L'étude du passé humain

Thématique et/ou mots clés :

Métadonnées Dublin Core Identifiant pérenne DOI Citation Entrepôt sécurisé

Type de données :

fichiers texte, son, image, vidéo

Communauté d'utilisateurs :

ESR français

Usagers et bénéficiaires :

Déposants : Enseignants-chercheurs, Chercheurs, Doctorants

Conditions d'usage :

L'accès à NAKALA se fait via HumanID, l'agrégateur d'authentifications d'Huma-Num qui permet d'utiliser eduGAIN, ORCID, HAL, Google, LinkedIn, Twitter pour s'authentifier et utiliser avec son compte universitaire l'ensemble des principaux services d'Huma-Num utilisant des interfaces Web.

Modèle économique :

Financé par l'État

Certification/Label :

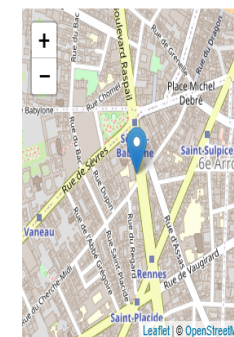
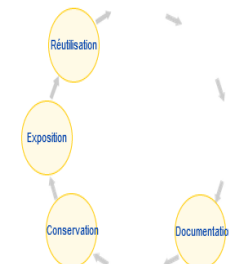
Conditions générales d'utilisation :

Nakala

Type de service	Entrepôt de données
Statut	En production
URL	<a href="https://www.nakala.fr">https://www.nakala.fr</a> , <a href="https://documentation.huma-num.fr/nakala/#utilisation-de-nakala">https://documentation.huma-num.fr/nakala/#utilisation-de-nakala</a>
Contact	<a href="mailto:nakala@huma-num.fr">nakala@huma-num.fr</a>
Localisation	Paris
Structure d'appartenance	Huma-Num
Tutelles	CNRS, Campus Condorcet, AMU
Identifiants	
re3data	<a href="#">R3JM1B</a>

### CYCLE DE VIE DES DONNÉES

Ce service intervient au cours des stades du cycle de vie suivants :



# Un entrepôt parmi d'autres

Une comparaison détaillée (36 critères) à l'instant t (août 2021) entre

- NAKALA
- Deux entrepôts multidisciplinaires : Figshare et Zenodo
- Un entrepôt disciplinaire : Cocoon

Télécharger la comparaison : [https://anf2021-humanum.sciencesconf.org/data/Comparaison\\_NAKALA\\_Figshare\\_Zenodo\\_aout\\_2021.ods](https://anf2021-humanum.sciencesconf.org/data/Comparaison_NAKALA_Figshare_Zenodo_aout_2021.ods)

# Quelles fonctionnalités ?

Les fonctionnalités sont celles d'un classique entrepôt de données

- Dépôt des données
- Description des données
- Conservation des données
- Recherche de données
- Diffusion des données
- Faciliter la réutilisation des données

# Le dépôts des données

- Données
  - Issues de recherches en SHS, liées ou non à des publications
  - De tous types [texte, audiovisuel, image, mesures, cartes, modèles 3D, etc.]
  - De tous formats
    - Privilégiez les formats archivables au CINES et testez leur validité (<http://facile.cines.fr>) avant leur dépôt.
    - Attention, tous les formats ne sont pas lisibles par les visionneuses embarquées dans Nakala (Cf l'article « Partager ses données avec NAKALA » sur <https://humanum.hypotheses.org/774>)
- Dépôts
  - Par l'interface web (<https://nakala.fr>) par simple glisser/déposer
  - Par l'utilisation de l'API (<https://api.nakala.fr/doc>)
  - Mono-fichier, multi-fichiers, fichier structuré (zip)

# La description des données

- Des métadonnées obligatoires
  - Type, titres, auteurs, date de création, licence
  - Parfois associés à des syntaxes, des référentiels ou des vocabulaires contrôlés

The screenshot shows a web form for entering metadata. A modal dialog is open for adding an ORCID. The background form includes fields for 'Type de dépôt' (set to 'vidéo'), 'Titre' (with language dropdowns for 'anglais' and 'français'), 'Auteurs' (with a tag 'Michel Jacobson'), 'Date de création' (set to '2021'), and 'Licence' (set to 'Creative Commons Attribution 4.0 International').

Rechercher dans NAKALA...

**Métadonnées principales**

Type de dépôt ⓘ  
vidéo

Titre ⓘ  
anglais ▼ My title  
français ▼ Mon titre

Auteurs  
✕ Michel Jacobson

Date de création ⓘ  
2021

Licence  
Creative Commons Attribution 4.0 International ▼

Prénom  
Machin

Nom  
Truc

Orcid ⓘ  
1234-1234-1234-1234

Annuler + Ajouter

# La description des données

- Des métadonnées complémentaires recommandées
  - Description, Langues, Mot-clés
- D'autres métadonnées
  - Toutes les rubriques du « Dublin-Core qualifié » + quelques éléments des ontologies foaf et owl

**Ajouter d'autres métadonnées** Optionnel

Propriété ▾	Type ▾	Pas d'information de langue ▾
<input type="text"/>		
Propriété		
dcterms:abstract		
dcterms:accessRights		
dcterms:accrualMethod		
dcterms:accrualPeriodicity		
dcterms:accrualPolicy		

Ajouter une métadonnée



# Conservation de données ?



## Donnée au statut « déposée »

- Données stockées dans les limites de : 1000 données et 100 Go
- Suppression possible
- Visibilité/accès limités aux utilisateurs ayant un droit d'accès dessus

## Donnée au statut « publiée »

- Pas de limite de nombre ou de volume annoncée
- Suppression impossible ou limitée aux cas de force majeure avec mise en œuvre de la suppression par un administrateur technique d'HN
- Toutes les métadonnées sont publiques et envoyées à Datacite pour la publication du DOI
- Les fichiers peuvent être librement accessibles ou soumis à un embargo dont le déposant précise le délai

# Recherche de données

- Par l'interface web, par l'API, ou par un service tiers
  - Les métadonnées ainsi que les contenus textuels des fichiers sont indexés et donc « cherchables »
  - Un service tiers, comme Isidore par exemple, peut proposer des recherches dans des informations enrichies

# Diffusion des données


Notice de la donnée « Landing-page » montre les métadonnées et les fichiers associées (lisible par une visionneuse).

La page de la notice est accessible via

- Nakala :  
[https://nakala.fr/\[identifiant\]](https://nakala.fr/[identifiant])
- Des « Resolvers »  
[https://doi.org/\[identifiant\]](https://doi.org/[identifiant])

## Nakala et IIIF


ID : 10.34847/nkl.44e39577


 Publiée

**Auteurs :** Adrien Desseigne et Laurent Capelli

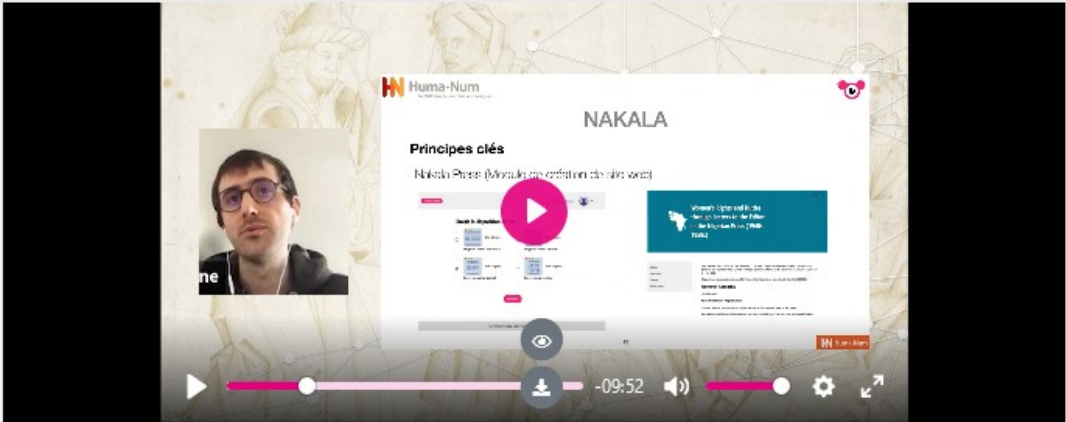
Présentation du service Nakala et de son API IIIF Image à l'occasion du Rendez-vous IIIF360 organisé le 24 mars 2021 par Biblissima et la TGIR Huma-Num.

**Fichiers**

 Pres\_NakalaIIIF\_2...0525.mp4

 Vignette\_Pres\_Nak...0525.png

**Visualisation**



ID : 10.34847/nkl.44e39577/183338b67696814249bc496aa784e37043d89410

Url d'intégration : <https://formation-api.nakala.fr/embed/10.34847/nkl.44e39577/183338b67696814249bc496aa784e37043d89410>

Url de téléchargement : <https://formation-api.nakala.fr/data/10.34847/nkl.44e39577/183338b67696814249bc496aa784e37043d89410>

## Mots-clés

OpenSeadragon IIIF NAKALA Biblissima Campus Condorcet

## Licence

## Citer

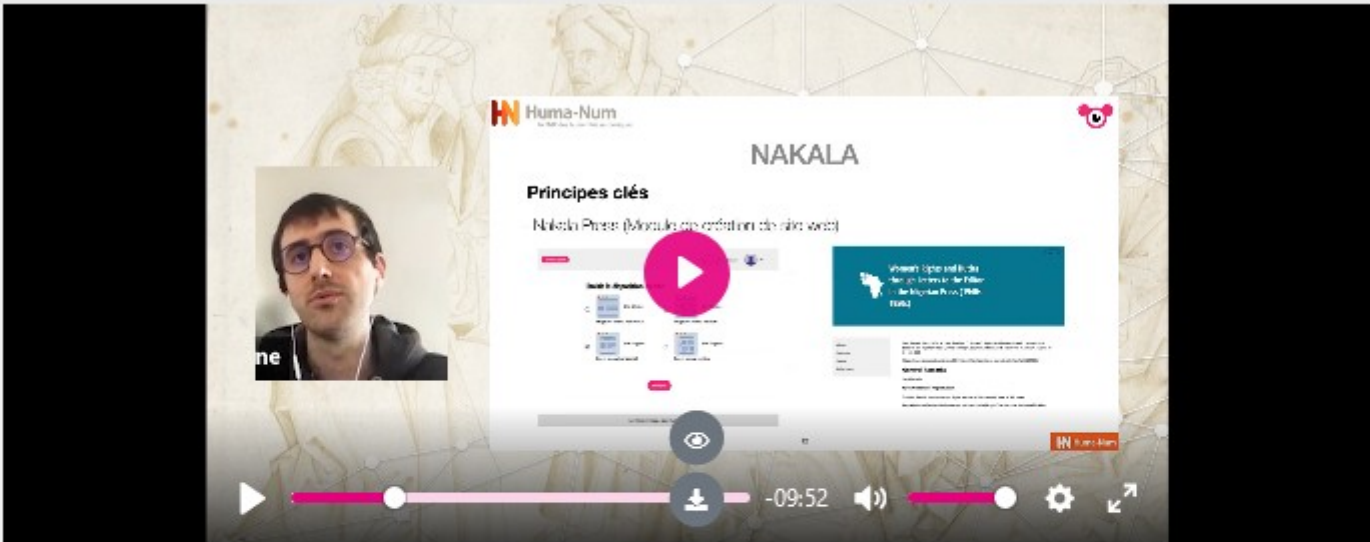
Desseigne, Adrien et al. (2021) «Nakala et IIIF» [Conference] NAKALA. <https://doi.org/10.34847>

# Diffusion des données

Les fichiers d'une donnée sont accessibles via.

- Leur URL :  
[https://api.nakala.fr/data/\[identifiant\]/\[identifiant\\_fichier\]](https://api.nakala.fr/data/[identifiant]/[identifiant_fichier])
- La visionneuse  
[https://api.nakala.fr/embed/\[identifiant\]/\[identifiant\\_fichier\]](https://api.nakala.fr/embed/[identifiant]/[identifiant_fichier])  
(eg. pour intégration dans un blog [hypothes.is](https://hypothes.is/)...)
- Le protocole IIIF (pour les images)  
[https://api.nakala.fr/iiif/\[identifiant\]/\[identifiant\\_fichier\]/{region}/{taille}/{rotation}/{qualité}.{format}](https://api.nakala.fr/iiif/[identifiant]/[identifiant_fichier]/{region}/{taille}/{rotation}/{qualité}.{format})

Visualisation



ID : 10.34847/nkl.44e39577/183338b67696814249bc496aa784e37043d89410


Url d'intégration : <https://formation-api.nakala.fr/embed/10.34847/nkl.44e39577/183338b67696814249bc496aa784e37043d89410>

Url de téléchargement : <https://formation-api.nakala.fr/data/10.34847/nkl.44e39577/183338b67696814249bc496aa784e37043d89410>

# Diffusion des données

## Visionneuse.

- Fichier image : OpenSeadragon
- Fichier CSV : DataTables
- Fichier audio ou vidéo : Plyr
- Fichier PDF : PDF.js
- Fichier Markdown
- Archive Zip
- Fichier de code (XML, HTML, JSON...)

 **CNRS-CFEETK 322. Karnak, KIU 5437 / Mur extérieur de Thoutmosis III, Scene, enceinte ouest, porte ouest passage nord de la porte nord-ouest de l'enceinte en grès : accomplir la libation et l'encensement XIXe dynastie / Sethi Ier**

ID : 11280/111643de  Publiée

Auteur : A. Bellod

Fichier	Visualisation
 322.jpg	
ID : 11280/111643de/b11ddf74a574b13ab04d790ebfd6d643db96eaeed	

# Diffusion des données

- Un outil intégré « Nakala-Press » pour l'éditorialisation de ses collections dans un site web

The screenshot displays the Huma-Num Nakala website interface. At the top, the Huma-Num logo is on the left, and the text "Collection Huma-Num Nakala" is in the center. A search bar on the right contains the text "Rechercher une donnée". Below the header, a navigation menu on the left lists: "Accueil", "La TGIR", "La collection", and "Autour de Victor Hugo". The main content area shows "16 Résultats de recherche" with a filter for "Image". Below this, there are several widgets and collection previews. The widgets include "SUGGESTIONS ISIDORE" with links to "L'information : une nouvelle culture ?", "Patrimoines du Maghreb à l'ère numérique", "Grands équipements documentaires et diffusion des savoirs", and "Anna-Maria Tiersen-Markel"; "ISIDORE widget - module wordpress - v..."; "ISIDORE widget - 2.0"; "Logo NAKALA"; "Widget Isidore - vue globale"; "Mazarinum" with a "Capture d'écran de Mazarinum"; "BYTE Magazine - Avril 1981"; "Isidore" with a "Capture d'écran de Mazarinum dans IS..."; and "Isidore Suggestions: Suggestion...". Each widget or preview includes a thumbnail image and a date of deposit.

# Diffusion des données

- Deux protocoles pour récupérer ou accéder aux métadonnées
  - OAI-PMH. Par exemple pour référencer sa/ses collections dans Isidore ou Gallica...
  - SPARQL-Endpoint (métadonnées exposées en rdf et publiées en mode « web de données ». Service interrompu depuis la nouvelle version. Doit être remis en place.
- Des métadonnées embarquées dans les notices et récupérées par exemple par « Google Data Search » qui cherche à relier les données à des articles dans Google Scholar

# Faciliter la réutilisation

- Des identifiants pérennes (DOI) utilisables pour citer les ressources
- Des licences associées obligatoirement
- Un modèle de description normalisé (Dublin-Core)
- Des formats de fichier recommandés
- Référencement possible



Merci pour votre écoute