



Huma-Num


la TGIR des humanités numériques

*L'infrastructure de recherche Huma-Num :  
pratiques de recherche en SHS et services  
numériques pour des données FAIR*



Aix-Marseille  
université

CAMPUS  
CONDORCET  
Paris-Aubervilliers

**STRATÉGIE NATIONALE  
DES INFRASTRUCTURES DE RECHERCHE**

## SCIENTES HUMAINES ET SOCIALES

### > Humanités numériques- Huma -Num



Huma-Num est une très grande infrastructure de recherche (TGIR) visant à faciliter le tournant numérique de la recherche en sciences humaines et sociales. Huma-Num met en œuvre un dispositif humain (concertation collective) et technologique en s'appuyant sur un réseau de partenaires et d'opérateurs...

### > PROduction et GEstion de DONnées-PROGEDO



Acteur central de la politique nationale sur les données en sciences humaines et sociales, PROGEDO soutient la réalisation de grandes enquêtes européennes, la mise à disposition des données françaises, le développement de la culture des données et la constitution de briques de compétences de niveau européen...

### > OpenEdition. Plateformes de publication électronique en sciences humaines et sociales-OpenEdition



OpenEdition est une infrastructure complète d'édition électronique au service de la communication scientifique en sciences humaines et sociales. Elle rassemble quatre plateformes complémentaires dédiées aux livres, aux revues, aux carnets de recherche et aux événements scientifiques...

### > Réseau national des Maisons des Sciences de l'Homme-RnMSH



Le Réseau national des MSH fédère actuellement 23 Maisons des Sciences de l'Homme réparties sur toute la France, et regroupées autour de cinq principes fondamentaux : l'Interdisciplinarité, l'Internationalisation, une Dynamique Interinstitutionnelle, l'Implantation territoriale et une Identité scientifique...

## INFORMATION SCIENTIFIQUE ET TECHNIQUE

### > Collections d'excellence pour la Recherche Persée-COLLEX-PERSEE



CollEx-Persée fédère la plateforme Persée, les bibliothèques de recherche, la BnF et les opérateurs en IST pour optimiser l'accès aux gisements documentaires par les chercheurs. Ses maîtres-mots sont : excellence, collecte des matériaux pour la recherche, visibilité des collections, numérisation de corpus, services...

### > Archive ouverte HAL (Hyper Article en Ligne)-HAL



HAL est la plateforme nationale française d'archives ouvertes, multidisciplinaire, utilisée par la communauté scientifique française. HAL est interconnectée aux grands dépôts d'archives ouvertes internationaux tels que PubMed Central ou ArXiv...

### > NUMEDIF (NUMérique pour l'ÉDition et la DIFFusion de la production scientifique)-NUMEDIF



L'infrastructure Numédif propose un ensemble d'outils et de méthodes pour l'édition structurée (« MÉTOPES ») permettant d'organiser production et diffusion papier et numérique dans un environnement normé à fort potentiel d'interopérabilité, sur le modèle du Single Source Publishing...



**« les IR doivent s'inscrire dans une dynamique de science ouverte et de maîtrise du cycle de vie des données qu'elle produit et utilise »**

# Les missions

1. Accompagner l'évolution des communautés SHS dans le contexte du numérique et de la science ouverte.
2. Mettre en œuvre une infrastructure pour les données de la recherche SHS (dans la logique des *FAIR principles* : *Findable, Accessible, Interoperable, and Reusable*) et en tenant compte des pratiques des SHS.
3. Participer à la construction des infrastructures internationales, en accompagnant les communautés SHS nationales, dans le contexte de l'European Open Science Cloud.



# L'accompagnement des projets de recherche

## Être accompagné pour :

*Être mis en réseau avec d'autres projets ou acteurs à différentes échelles*

*Obtenir des conseils méthodologiques et techniques*

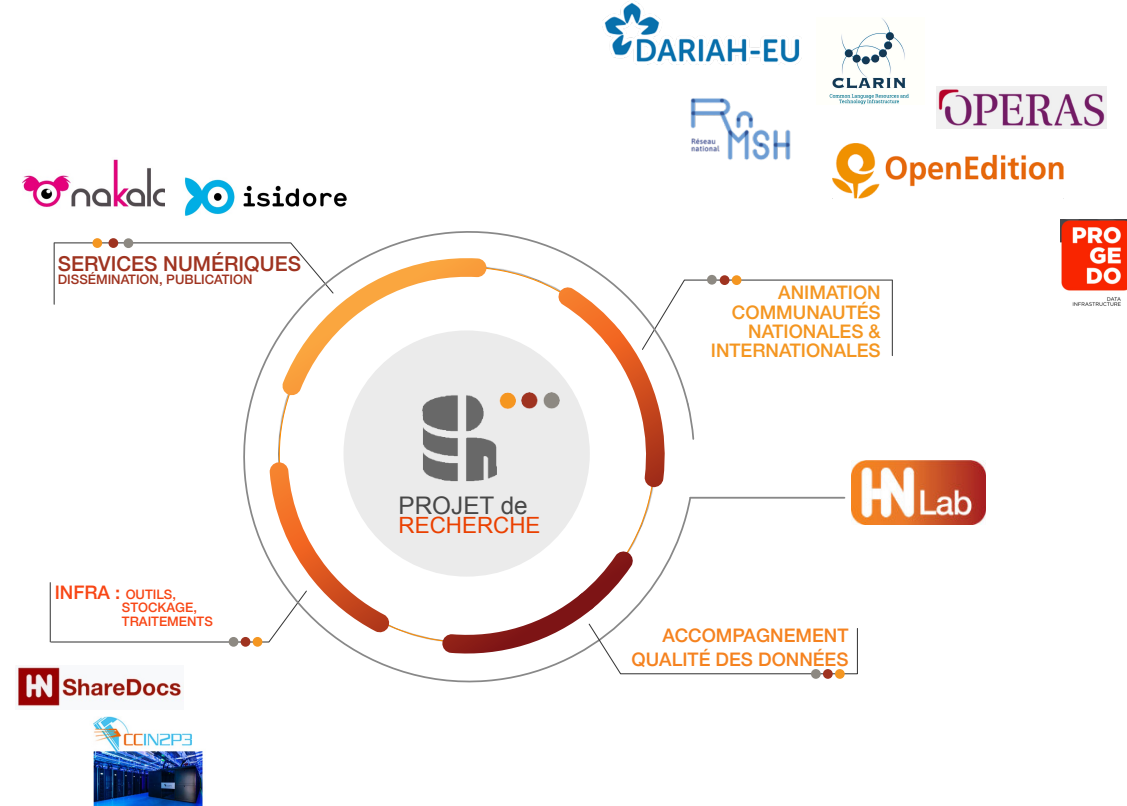
*Organiser, collecter, traiter, préserver, publier, réutiliser des données*

*Me former à certaines notions et certains services ou outils*

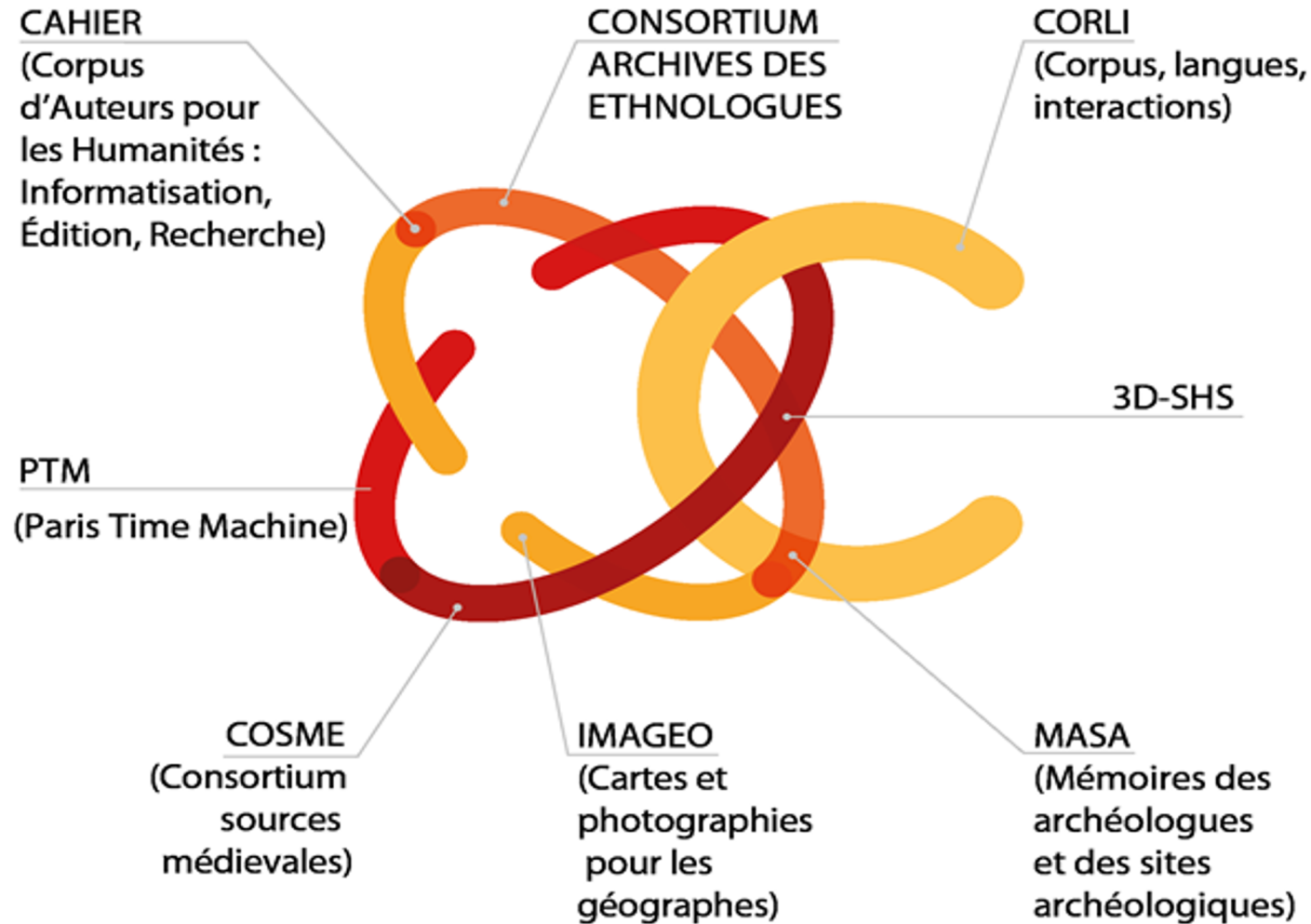
*Faire évoluer des services existants pour mieux les adapter à mes besoins*

*Partager des réflexions et développements autour d'objets de prospective*

...



## Des projets au sein des communautés de recherche : les consortiums HN



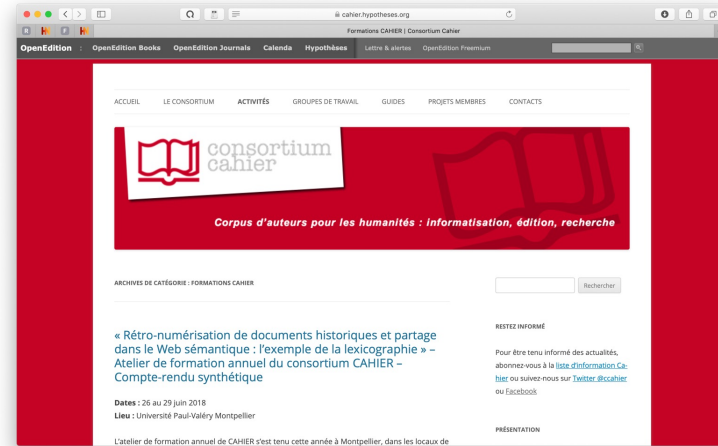
- 8 consortiums,
- 80-100 équipes de recherches qui couvrent des champs variés des SHS
- Un réseau de chercheurs, ingénieurs, bibliothécaires, archivistes, etc.
- Plateformes, outils, guides, formations,...

## Exemples de réalisation des consortiums HN

Plateforme Transcrire



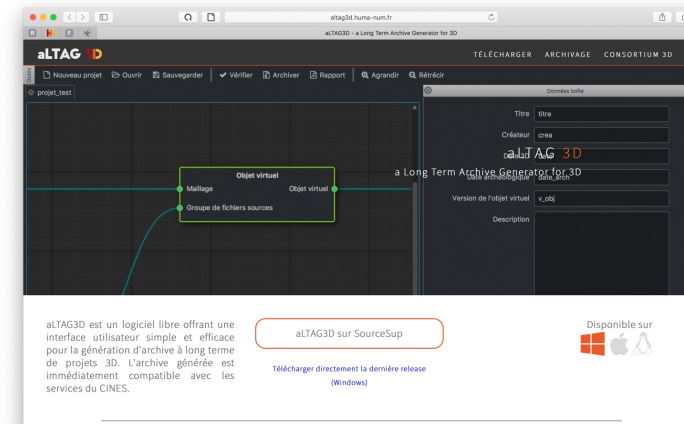
[transcrire.huma-num.fr](http://transcrire.huma-num.fr)



Formations du consortium Cahier  
<https://cahier.hypotheses.org/category/evenements/formations>

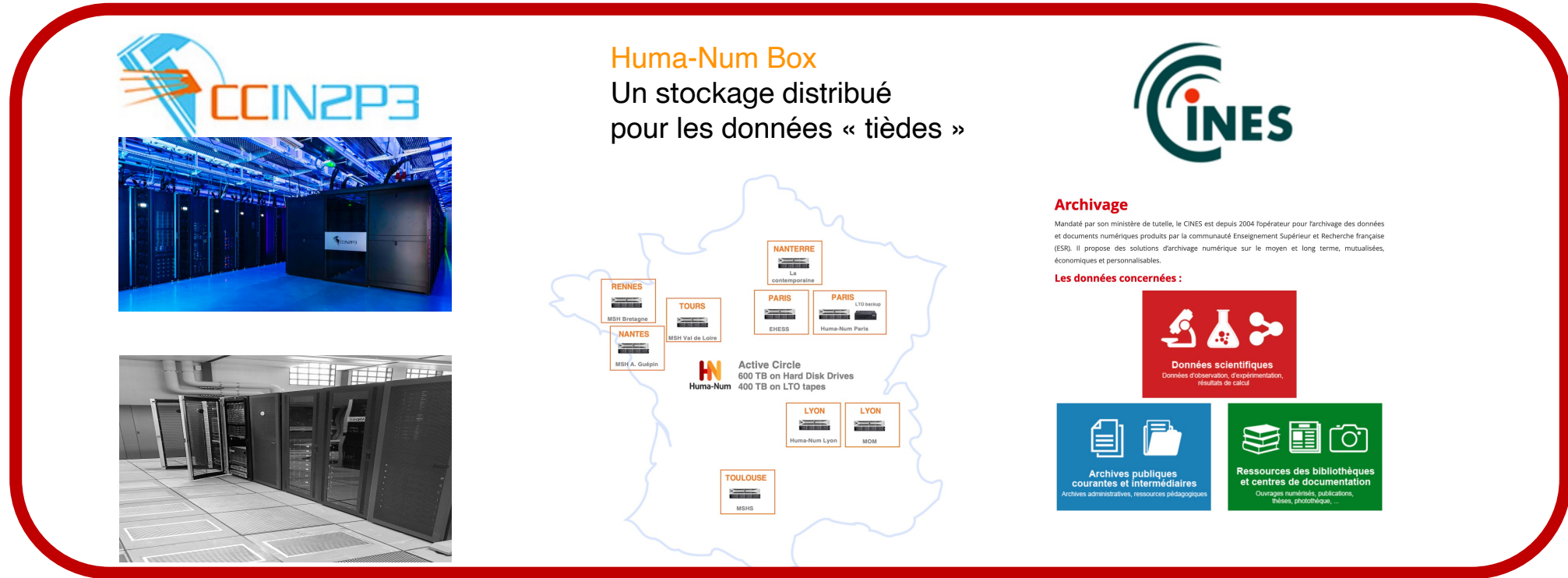


Guide juridique et éthique - consortiums, MSH, communautés...  
voir [ethiquedroit.hypotheses.org/1937](http://ethiquedroit.hypotheses.org/1937)



Outil d'archivage des données 3D SHS/Patrimoine :  
consortium 3D + Programmes européens  
[altag3d.huma-num.fr](http://altag3d.huma-num.fr)

# Une infrastructure pour la préservation, le calcul et la sécurité



**CCIN2P3**

**Huma-Num Box**  
Un stockage distribué pour les données « tièdes »

**INES**

**Archivage**  
Mandaté par son ministère de tutelle, le CINES est depuis 2004 l'opérateur pour l'archivage des données et documents numériques produits par la communauté Enseignement Supérieur et Recherche française (ESR). Il propose des solutions d'archivage numérique sur le moyen et long terme, mutualisées, économiques et personnalisables.

**Les données concernées :**

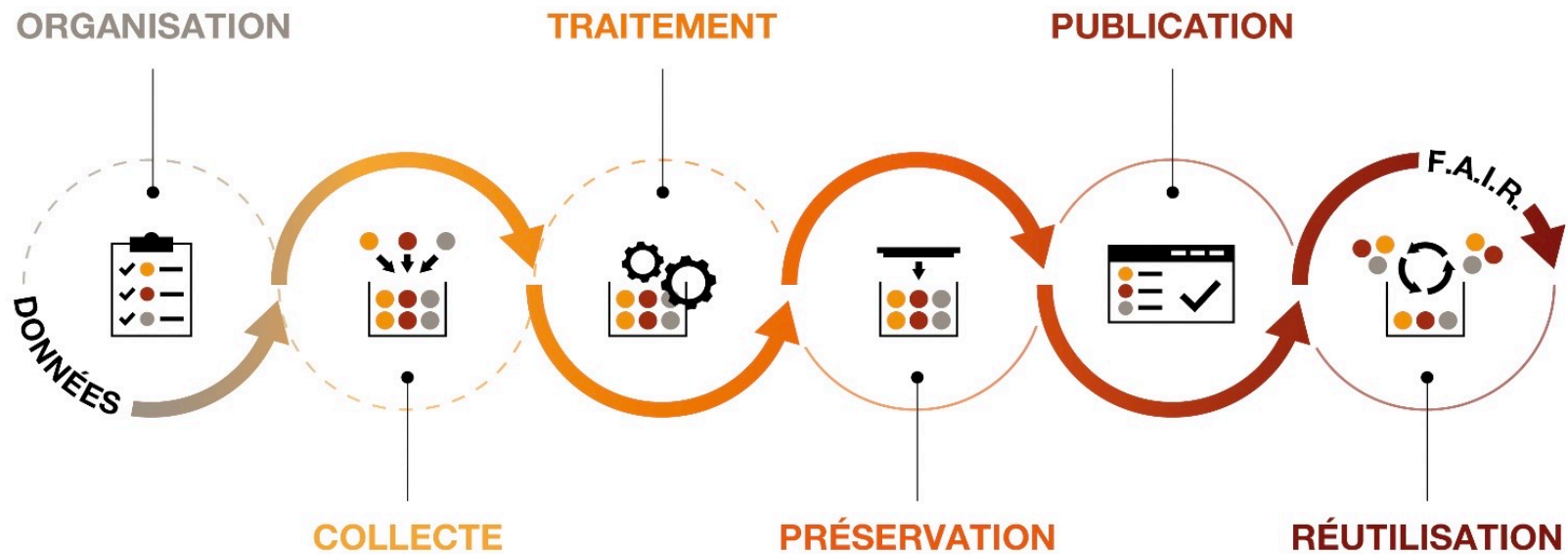
- Données scientifiques**  
Données d'observation, d'expérimentation, résultats de calcul
- Archives publiques courantes et intermédiaires**  
Archives administratives, ressources pédagogiques
- Ressources des bibliothèques et centres de documentation**  
Ouvrages numérisés, publications, thèses, photothèque, ...

**Map of France showing Huma-Num Box locations:**

- RENNES**: MSH Bretagne
- NANTES**: MSH A. Guépin
- TOURS**: MSH Val de Loire
- NANTERRE**: La contemporaine
- PARIS**: EHESS
- PARIS**: Huma-Num Paris (LTO backup)
- LYON**: Huma-Num Lyon
- LYON**: MDM
- TOULOUSE**: MSHS
- Active Circle**: 600 TB on Hard Disk Drives, 400 TB on LTO tapes

Projet sur la sécurité des données sensibles

## Des services pour la gestion des données



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

Facile à trouver • Accessible • Interopérable • Réutilisable



# Des données FAIR ?

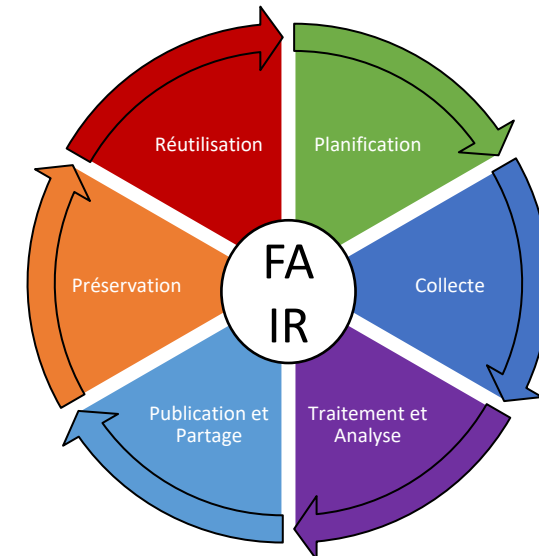
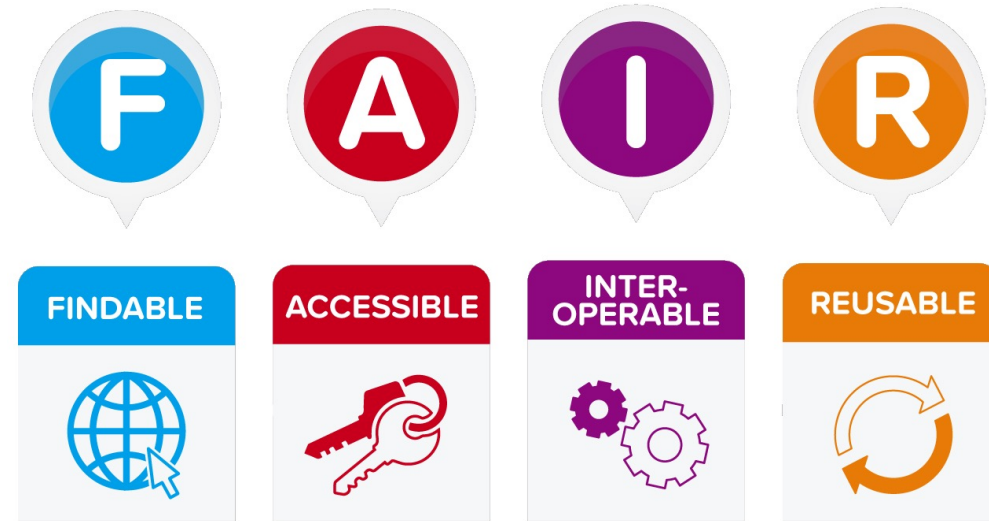
## Constat :

- Le taux de réutilisation des données (scientifiques) est extrêmement faible ;
- Objectifs :
  - Améliorer les infrastructures pour permettre aux machines de retrouver les données, et aux chercheurs de les réutiliser ;
  - Promouvoir des principes concis et mesurables – bonnes pratiques pour la gestion des données.

*Chacun des 13 principes se décline en un ensemble de caractéristiques que doivent présenter les données et les métadonnées pour faciliter leur découverte et leur utilisation par les hommes mais aussi par les machines*

- Trois types d'entités:
  - Données (ou tout objet numérique),
  - Métadonnées (informations sur l'objet numérique),
  - L'infrastructure de support (ex. moteur de recherche, plateforme de stockage, etc.).

***Un outil pour la science ouverte. Les données au cœur d'un process scientifique***





Faciliter la découverte des données par les humains et les systèmes informatiques et requiert une description et une indexation des données et des métadonnées.

FINDABLE



- **F1** Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne.
- **F2** Les métadonnées décrivant les données sont riches.
- **F3** Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher.
- **F4** Les métadonnées spécifient l'identifiant de la donnée.



Intégrer les données téléchargeables, utilisables, intelligibles, et combinables avec d'autres données, par des humains et des machines.

INTER-OPERABLE



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



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).

ACCESSIBLE



- **A1** Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé.
  - **A1.1** Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle.
  - **A1.2** Le protocole utilisé permet l'authentification et l'autorisation si besoin.
- **A2** Les métadonnées sont accessibles même quand les données ne le sont plus.



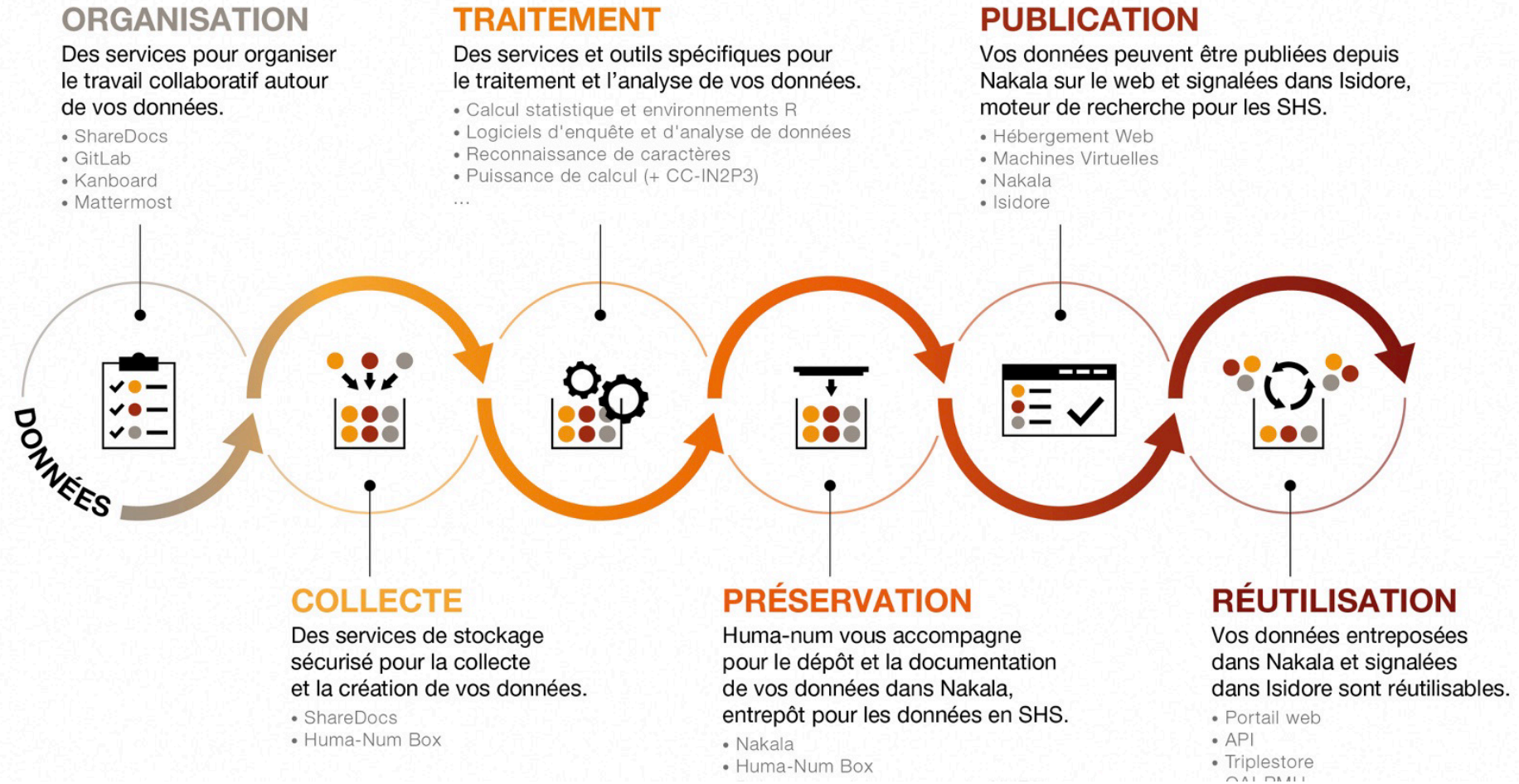
Rendre les données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction et/ou transparence de la science).

REUSABLE



- **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.
  - **R1.2** Les données et les métadonnées sont associées à leur provenance.
  - **R1.3** Les données et les métadonnées correspondent aux standards des communautés indiquées.

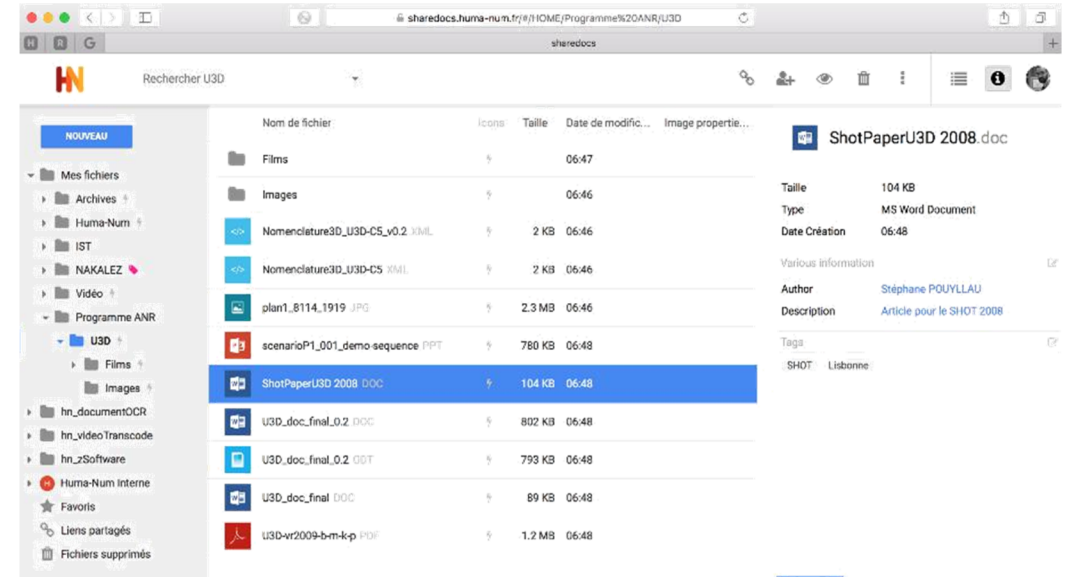
# Des services pour les données en SHS



# Des services classiques ...

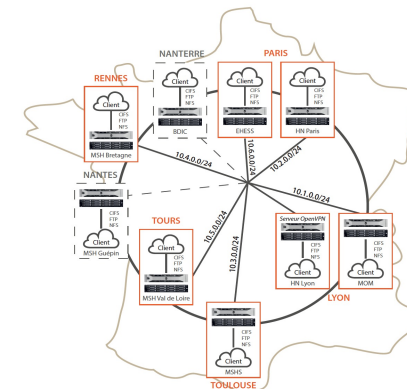
## **ShareDocs : le stockage outillé**

- Stockage des données en mode « dropbox » (mais sur les serveurs d'Huma-Num, en France)
- Partage, édition, traitement des données
- Un espace de travail collaboratif, sécurisé.



## **Le réseau des Huma-Num Box**

La Huma-Num Box est un service de stockage sécurisé pour la préservation de volumes importants de données (plusieurs centaines de To), par réplication et distribution sur plusieurs sites en France.



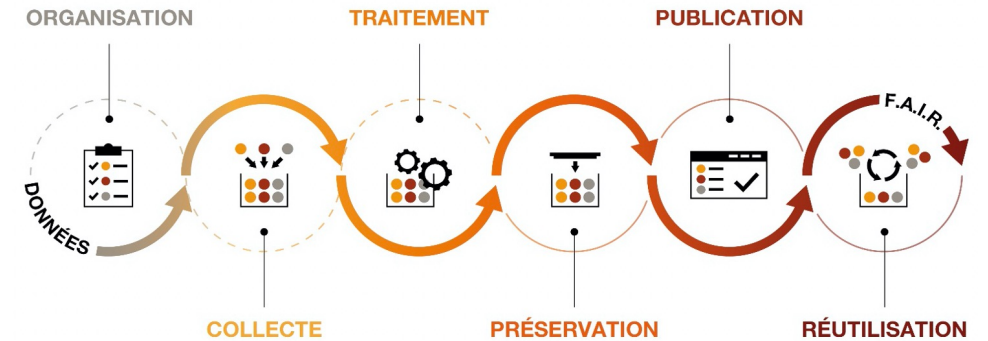
# Des services classiques ...

## ***Infrastructure physique depuis 2013***

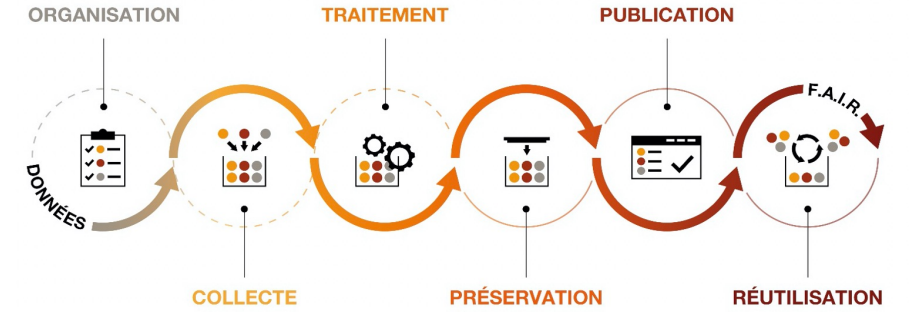
- Cluster Web, VM
- Puissance de calcul (CPU, GPU,...)

## ***Accompagner la collecte et les traitements des données :***

- OCR-HCR
- Licences (Oxygen, Maxqda, Sphinx, Qgis, Arcgis, Photoscan,...)
- Serveur R avec environnement logiciel R Studio Pro
- Serveur FileMaker Pro
- Gitlab
- ...



...aux services pour les données à forte valeur ajoutée pour préserver, signaler et publier



Préservation  
à  
long terme

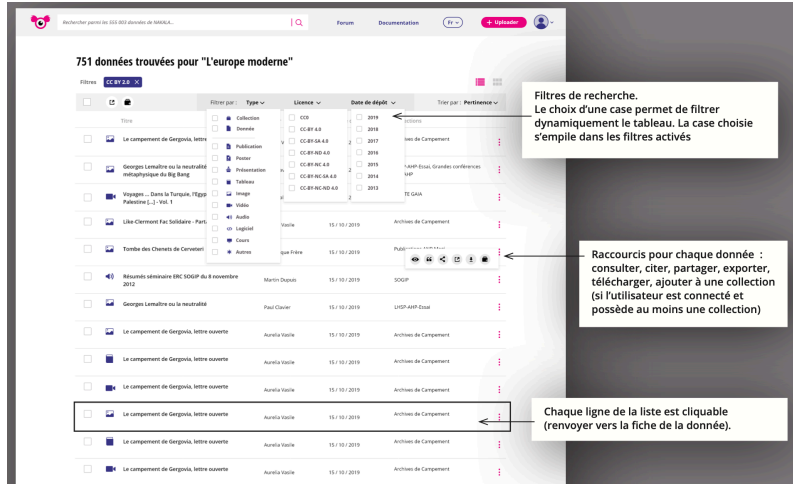
...aux services pour les données à forte valeur ajoutée :



NAKALA, créé en 2013, est un entrepôt de données pour les SHS opéré par d'Huma-Num permettant à des équipes de recherche de déposer tous types de données numériques (fichiers textes, sons, images, vidéos, objets 3D, etc.) documentées dans un entrepôt sécurisé afin de les partager.

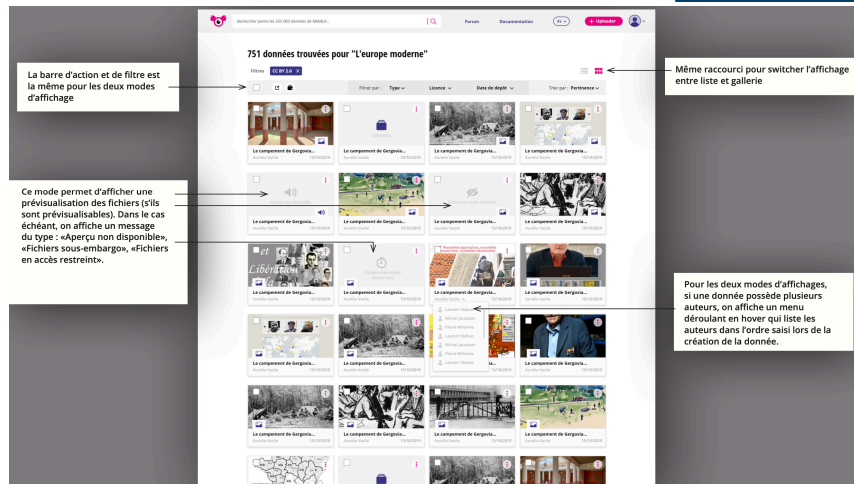
## Principes clés

- Donnée publique (publication) / Donnée privée (dépôt)
- Dépôt simple / Dépôt par lot | Donnée unitaire / Donnée multiple
  - ➡ chaque objet est identifié par un DOI : collection, donnée, version, fichier
  - ➡ embargo sur les fichiers
  - ➡ métadonnées obligatoires + facultatives (modèle extensible)
- Regroupement de n'importe quelles données au sein de Collections
- Partage de droits sur les données et/ou collections
- Consultation, API : REST (in/out), SPARQL, IIIF et OAI (out), Intégration (OEmbed)
  - Possibilité d'être indexé et catalogué par ≠ moteur de recherche : de Google data search au futur entrepôt national des données en passant par les moteurs européens (VLO, etc.) et ISIDORE.



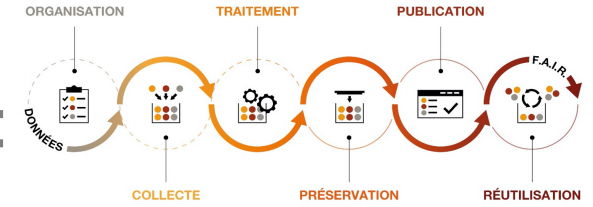
## NAKALA en résumé :

- Chaque donnée a un identifiant pérenne de type DOI et si la donnée est composée de plusieurs fichier(s) chacun aura également un DOI
- La gestion des partages est beaucoup plus souple (édition, lecture, publique etc.)
- Les métadonnées descriptives sont toujours du *dcterms* mais pourra être étendu à d'autres vocabulaires
- Chaque donnée a une page de présentation (*landing page*)
- Un moteur de consultation et de recherche est intégré
- Un module de présentation des collections « éditorialisées » est présent (*NAKALA Press*)
- <https://www.nakala.fr>





...aux services pour les données à forte valeur ajoutée :



## ISIDORE pour signaler, accéder et enrichir les données



Votre assistant de recherche  
en Sciences Humaines et Sociales

Documents ▾ Rechercher dans les 6 220 016 documents de ISIDORE... | Q

[Recherche avancée](#)

**ISIDORE, créé en 2009, est un moteur de recherche permettant d'accéder en ligne aux documents et aux données numériques en sciences humaines et sociales.**

- Publications
- Événements scientifiques
- Corpus et fonds d'archives
- Jeux de données
- ...

**+ de 10 millions de données signalées et reliées entre elles avec des référentiels scientifiques**

**+ d'1,3 million d'utilisateurs/an / + de 2000 comptes d'enseignant·e·s-chercheur·e·s et doctorant·e·s en SHS**

**+ de 9000 « sources » de données collectées dans le monde**

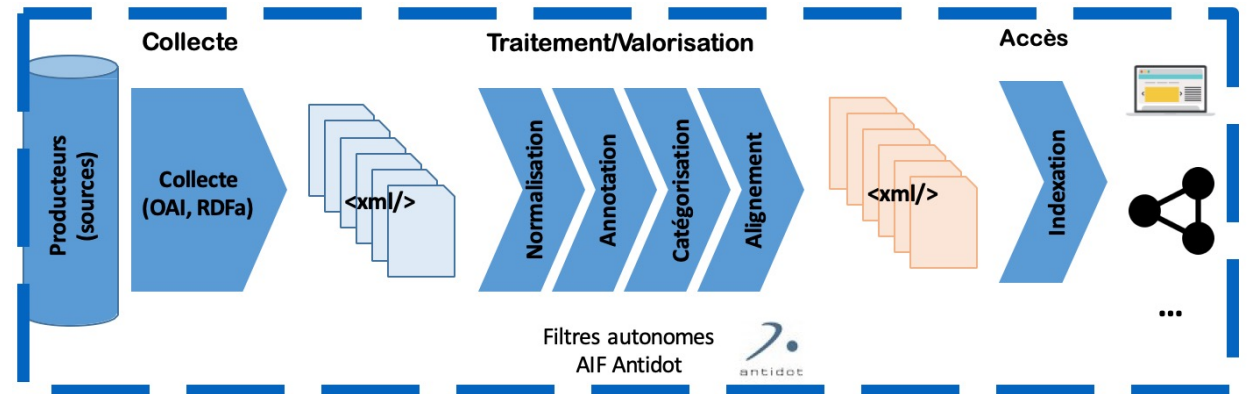


Un assistant de recherche et un réseau social académique en SHS, sur <https://isidore.science> fondé sur les technologies du *machine learning* et du web sémantique

Collecter  
Traiter

Enrichir  
Classer  
Catégoriser

Accéder





Collections ▾ Rechercher dans les 9 389 collections de ISIDORE...

## Mes auteurs ont récemment publié

**De la poussière à la lumière bleue. Émotions, récits, gestes de l'arch...**

Caroline Muller et al. 2021

Articles

HAL

**De la poussière à la lumière bleue**

Caroline Muller et al. 6 juin 2021

Articles

Histoire

**Issues and special features of a research**

Christian Ducrot et al.

Articles

Histoire, Philosophie et Sociologie de

## Mes derniers documents à lire

**La réduction des risques à distance : un programme adapté et efficace...**

Magally Torres-Leguizamon et al. 2020

Articles

**Pandémie de la COVID-19 et `ibsupply chains`/ib mondiales : repenser u...**

Valérie Rabassa 2020

Articles

**Les usagers de drogues durant la pandémie de Covid...**

Miguel Velazquez

Articles

Sociologie Science politique

Fr



Thème

SP Mon profil

Se déconnecter

Mes bibliothèques

Mon historique

Mes requêtes

Mes auteurs

Mes alertes

# Un service de préservation à long terme des données



# HumanID : un compte utilisateur/trice pour tous les services d'HN

nom d'utilisateur HumanID

Mot de passe

*Réinitialiser mon mot de passe*

**HN** se connecter avec HumanID

*première visite sur HumanID ?*

**HN** Créer un compte HumanID


*connectez-vous avec un compte externe*


 se connecter avec eduGAIN

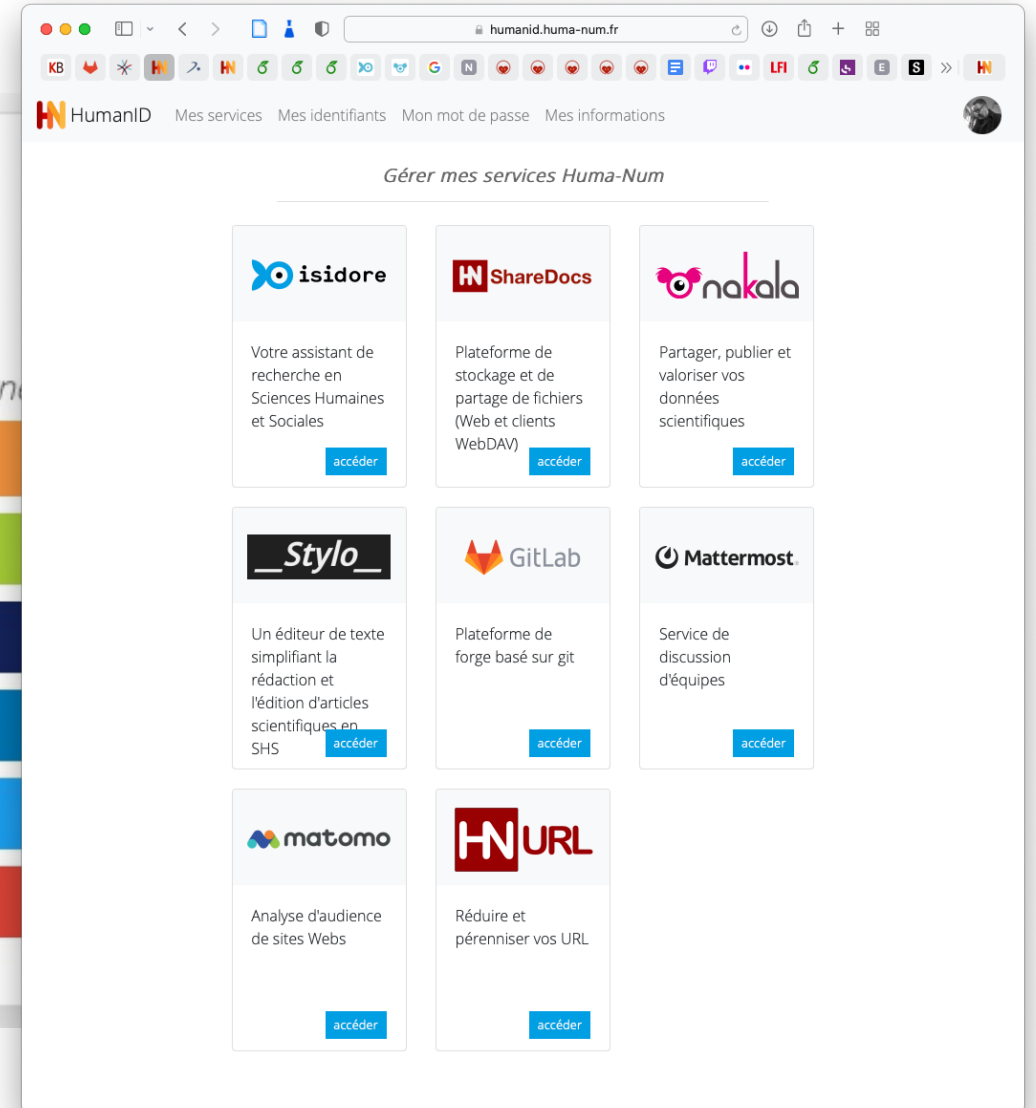
 se connecter avec ORCID

**HAL** se connecter avec HAL

 se connecter avec LinkedIn

 se connecter avec Twitter

 se connecter avec Google



humanid.huma-num.fr

HumanID Mes services Mes identifiants Mon mot de passe Mes informations

Gérer mes services Huma-Num

- isidore**  
Votre assistant de recherche en Sciences Humaines et Sociales  
[accéder](#)
- ShareDocs**  
Plateforme de stockage et de partage de fichiers (Web et clients WebDAV)  
[accéder](#)
- nakala**  
Partager, publier et valoriser vos données scientifiques  
[accéder](#)
- \_Stylo\_**  
Un éditeur de texte simplifiant la rédaction et l'édition d'articles scientifiques en SHS  
[accéder](#)
- GitLab**  
Plateforme de forge basé sur git  
[accéder](#)
- Mattermost**  
Service de discussion d'équipes  
[accéder](#)
- matomo**  
Analyse d'audience de sites Webs  
[accéder](#)
- HNURL**  
Réduire et pérenniser vos URL  
[accéder](#)

## Le HN Lab

### Objectifs

- La **production de connaissances** pour favoriser le développement des humanités numériques
- **L'accompagnement des disciplines SHS** dans leurs appropriations du numériques
- La **définition ou l'amélioration de services numériques** pour une discipline en vue de leur transfert vers les services de la TGIR Huma-Num

### Actions

- Accueillir des chercheur·e·s et des ingénieurs de recherche en **résidence**
- Participer à des programmes de recherche pour définir les besoins **d'accompagnement méthodologique des communautés SHS**
- Développer et faire développer des **preuves de concept** (POC)



# *Le Huma-Num Lab et ses partenaires*

- Data Lab de la BnF
- MSH Région
- Condorcet
- Labex PP
- ...



Catalogue des services du BnF DataLab



Maison DES Sciences de l'Homme EN Bretagne



# Des communautés et des infrastructures : du national à l'international

 Consortiums  
 Réseau national MSH  
 OpenEdition  
 perse  
 IR METOPES  
 CollEx-Persée  
 CCSD  
 CENTRE POUR LA COMMUNICATION SCIENTIFIQUE DIRECTE  
 La Science partagée  
 PROGE DO  
 DATA INFRASTRUCTURE  
 UMIFRE  
 Unités mixtes des instituts français de recherche à l'étranger  
 ResEFE



**Projets de recherche**



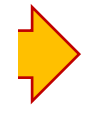
**Partenariats**



**Veille stratégique**



**GT, etc.**



 DARIAH-EU  
 CLARIN  
 Common Language Resources and Technology Infrastructure  
 RDA  
 RESEARCH DATA ALLIANCE EUROPE  
 FR NATIONAL NODE  
 OPERAS  
 open scholarly communication in the european research area for social sciences and humanities  
 EOSC-Pillar  
 Coordination and Harmonisation of National & Thematic Initiatives to support EOSC  
 Triple  
 ARIADNE plus  
 SSHOC  
 social sciences & humanities open cloud  
 EUROPEAN OPEN SCIENCE CLOUD  
 GO FAIR  
 ADHO  
 ALLIANCE OF DIGITAL HUMANITIES ORGANIZATIONS  
 Humanistica





# Huma-Num

la TGIR des humanités numériques

## Infos et contact

[www.huma-num.fr](http://www.huma-num.fr)

<https://humanum.hypotheses.org>

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

[humanum-diffusion@listes.huma-num.fr](mailto:humanum-diffusion@listes.huma-num.fr)

[@Huma Num](#)

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

Via les acteurs-relais

- Contacter la MSH la plus proche
- Annuaire public des correspondant.e.s HN (*bientôt disponible sur le site web*)