



HAL
open science

Découvrir les outils Huma Num pour la gestion de données scientifiques

Victoria Le Fournier, Florence Perret

► **To cite this version:**

Victoria Le Fournier, Florence Perret. Découvrir les outils Huma Num pour la gestion de données scientifiques. École thématique. MESHS, France. 2021, pp.63. hal-03672673

HAL Id: hal-03672673

<https://hal.science/hal-03672673>

Submitted on 19 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

DÉCOUVRIR LES OUTILS HUMA NUM POUR LA GESTION DE DONNÉES SCIENTIFIQUES



Victoria LE FOURNER
Florence PERRET

PRÉSENTATION DE L'ÉQUIPE ORGANISATRICE



Victoria Le Fournier

Ingénieure d'études chargée du
traitement des données scientifiques
Relais Huma-Num en Hauts-de-France

Contact : victoria.lefournier@univ-lille.fr



Florence Perret

Ingénieure d'études chargée des
humanités numériques

Contact : florence.perret@univ-lille.fr

FAIRE DE LA RECHERCHE EN SHS

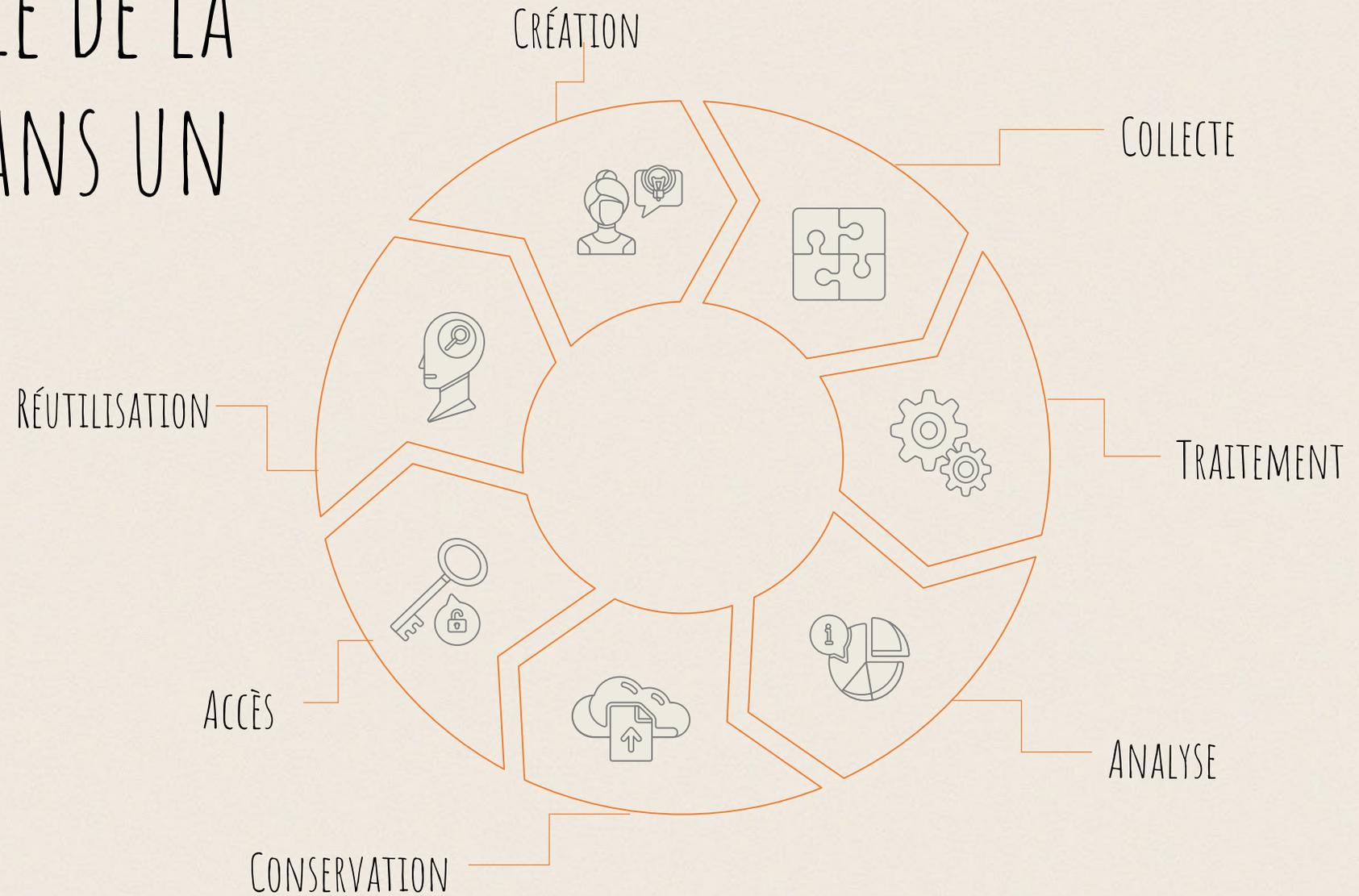
- Des données hétérogènes
- Des projets de recherche sur un temps long
- Des besoins précis, adaptés aux types de données

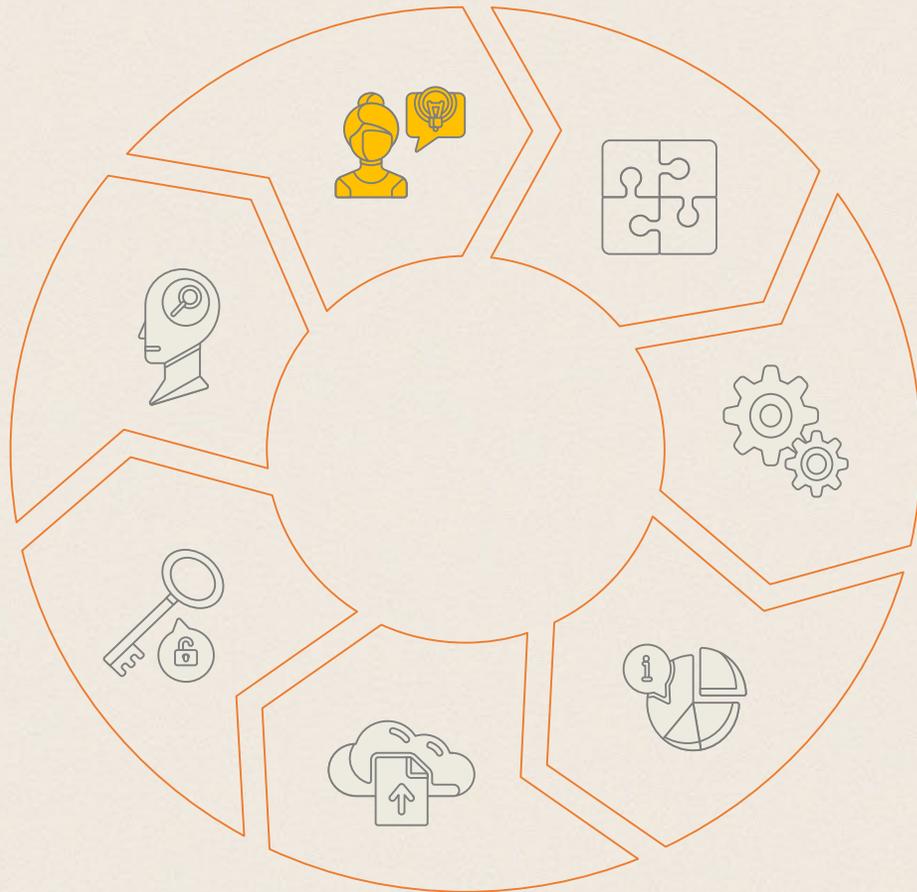




LE CYCLE DE VIE DES DONNÉES

CYCLE DE VIE DE LA DONNÉE DANS UN PROJET EN SHS





CRÉATION

À FAIRE

- Répartir le travail sur le temps du projet

BESOINS

- Logiciel d'organisation du travail d'équipe
- Logiciel de discussion interne au projet

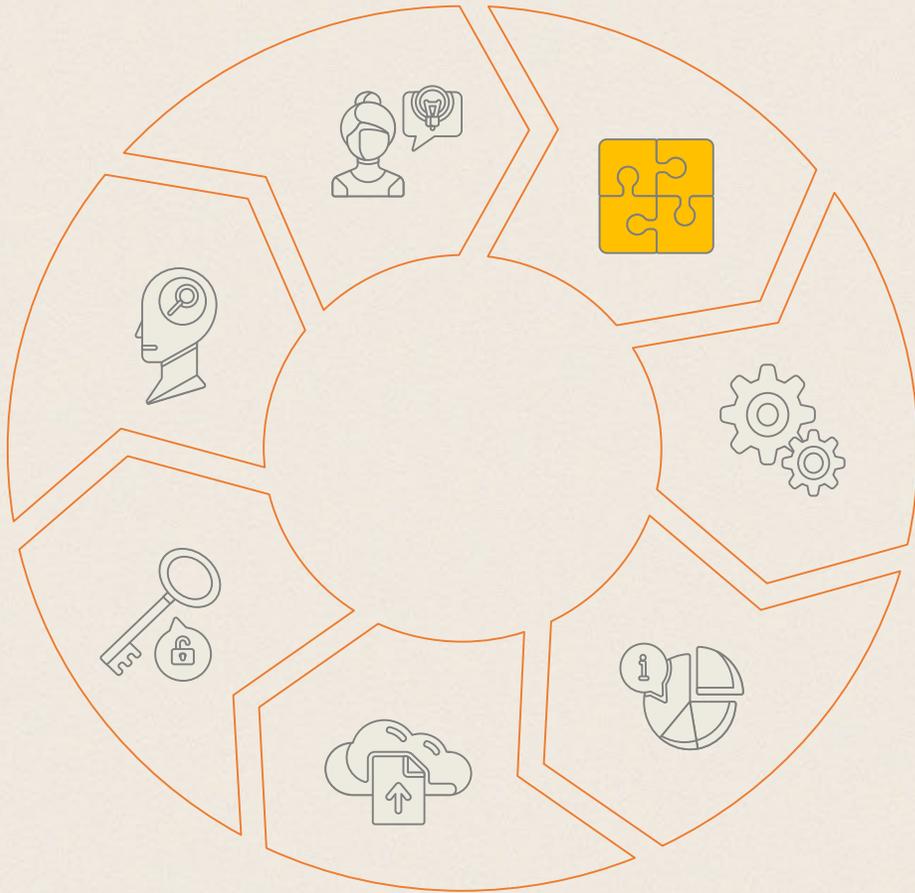
COLLECTE

À FAIRE

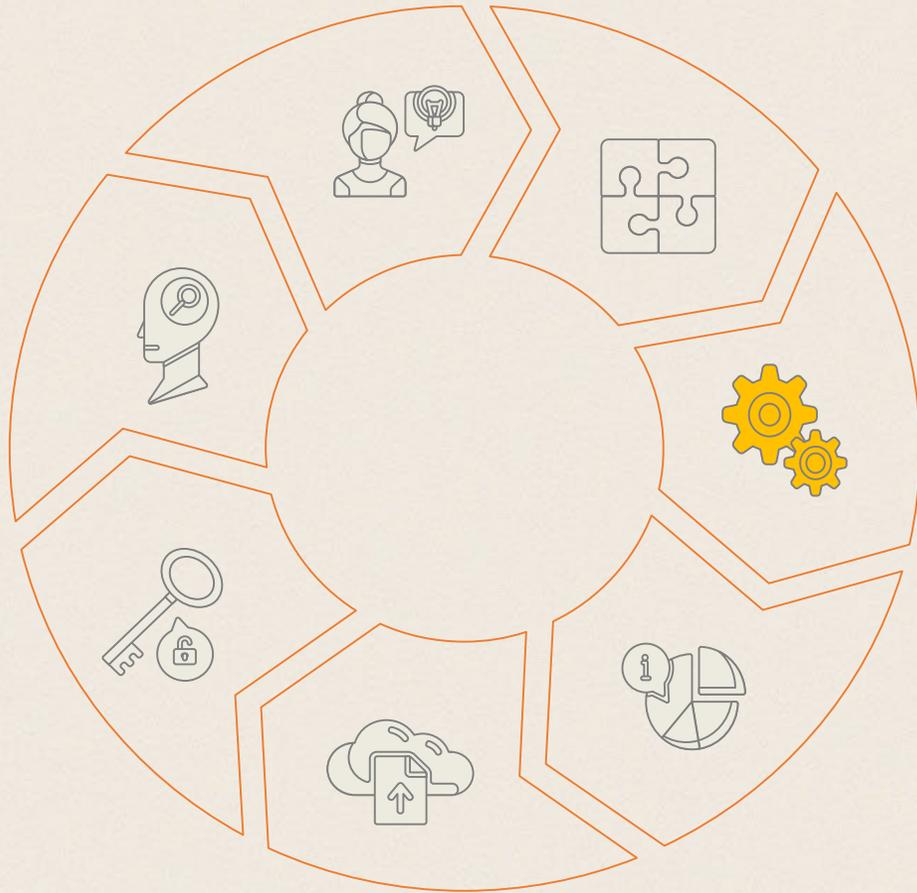
- Repérer les données (textes, images, enquêtes, bibliographie...)
- Créer le corpus

BESOINS

- Stocker les données collectées
- Voir les données collectées par les autres membres du projet



TRAITEMENT

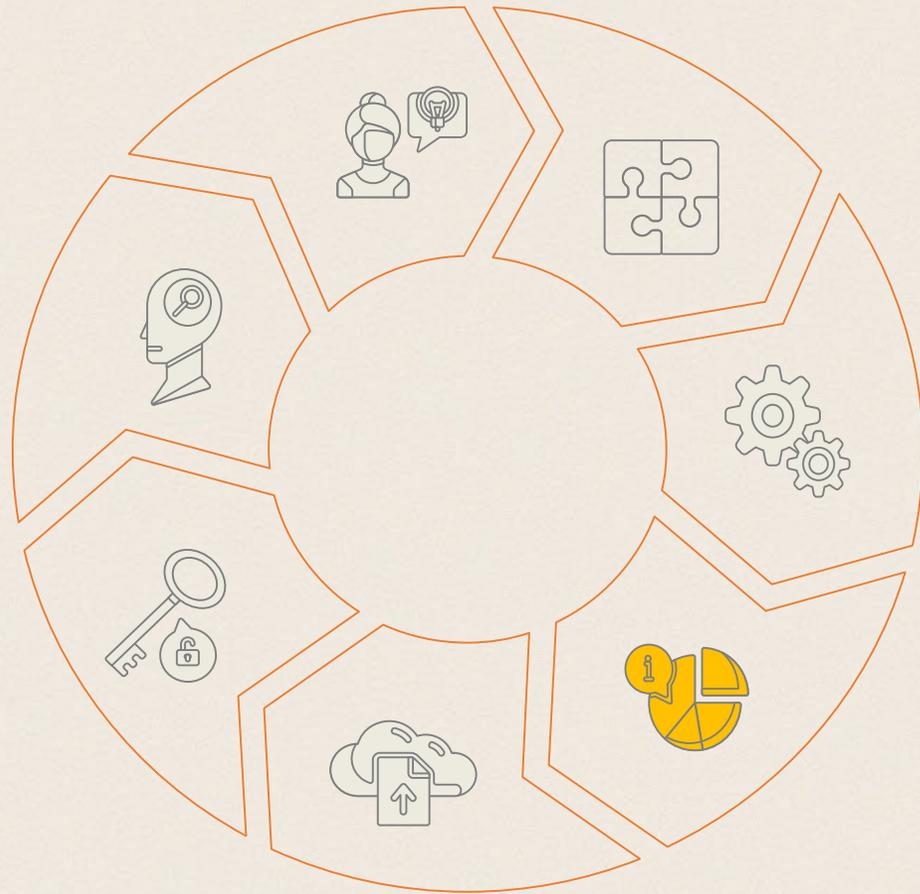


À FAIRE

- Nettoyer et aligner les données
- Créer de nouvelles données
- Rédaction de scripts

BESOINS

- Logiciels en fonction du type de traitement
- Compétences techniques



ANALYSE

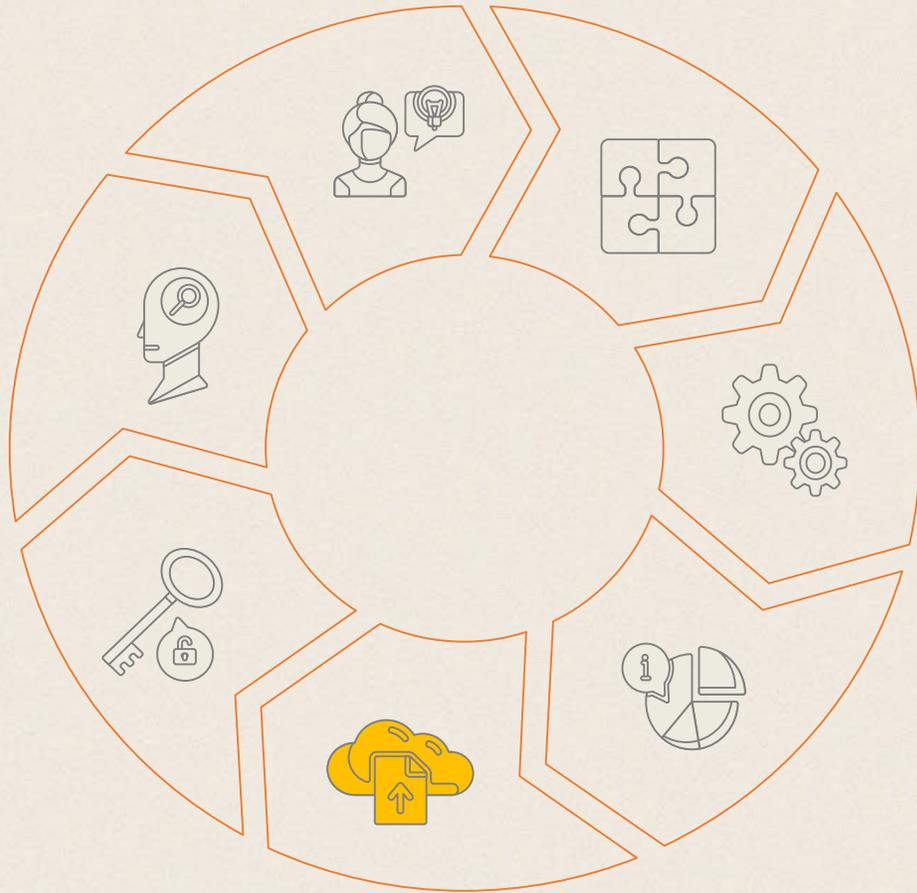
À FAIRE

- Interroger les données
- Ecrire les articles, thèses, chapitres...

BESOINS

- Faire les expériences
- Logiciels de traitement de texte collaboratifs

CONSERVATION



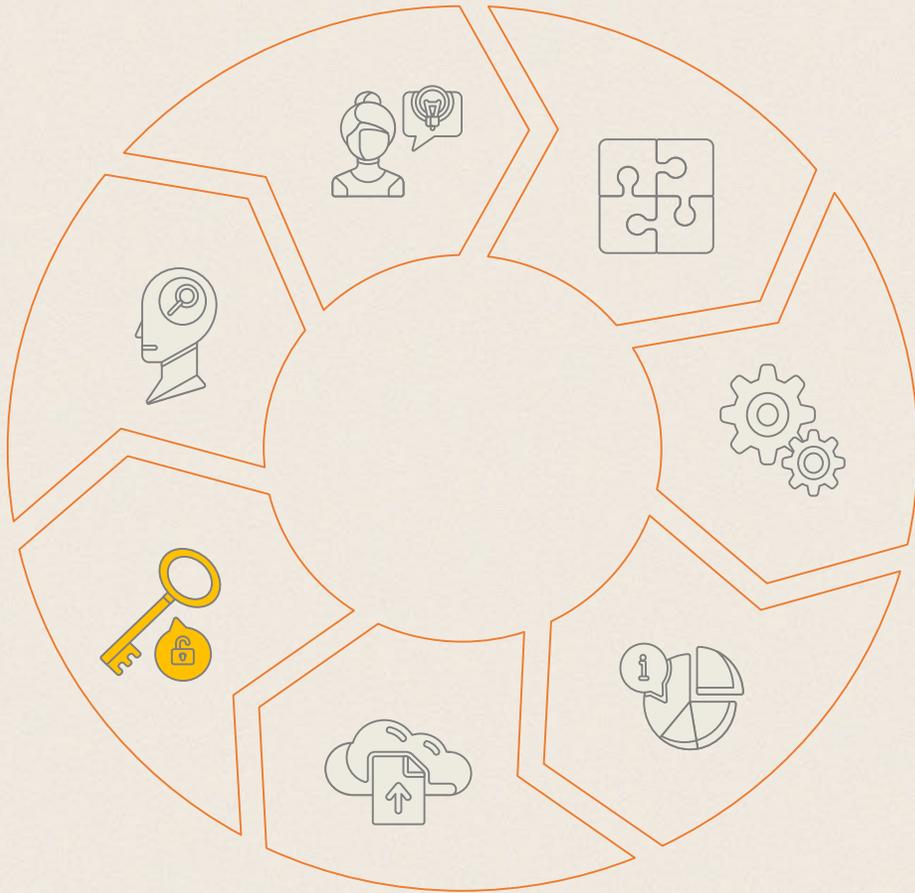
À FAIRE

- Stocker les données de manière pérenne

BESOINS

- Entrepôts de données

ACCÈS



À FAIRE

- Création d'un site dédié au projet
- Rendre visible le projet et les résultats

BESOIN

- Logiciel pour créer un site (CMS)
- Plateformes de publication

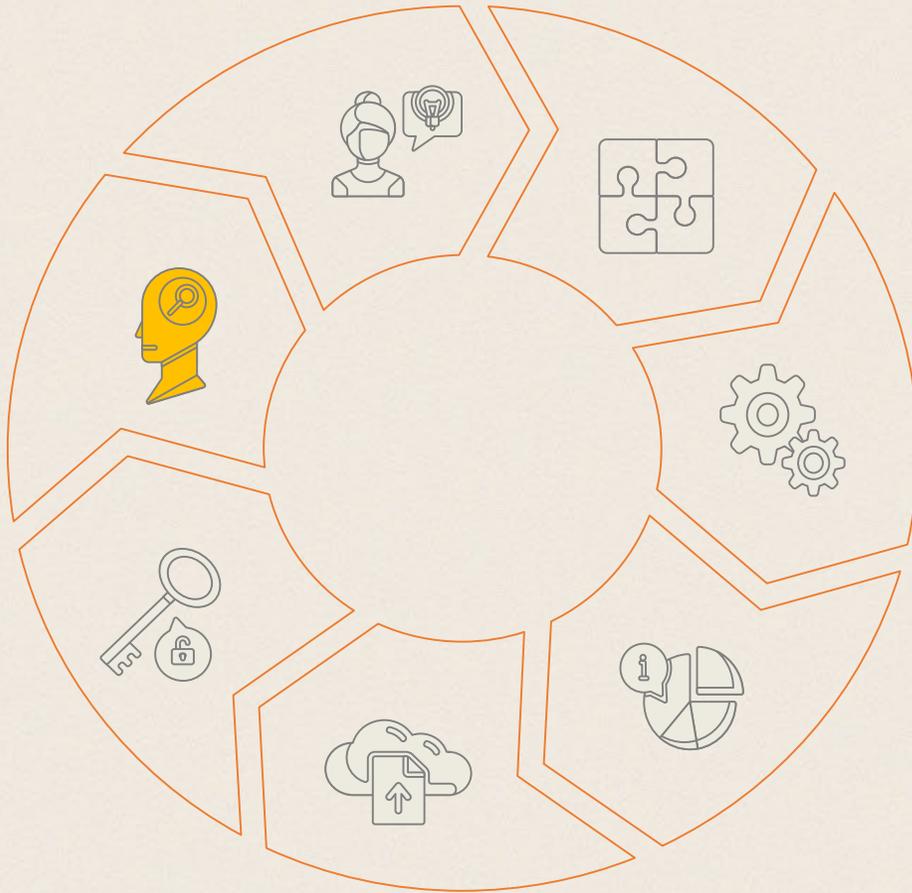
RÉUTILISATION

À FAIRE

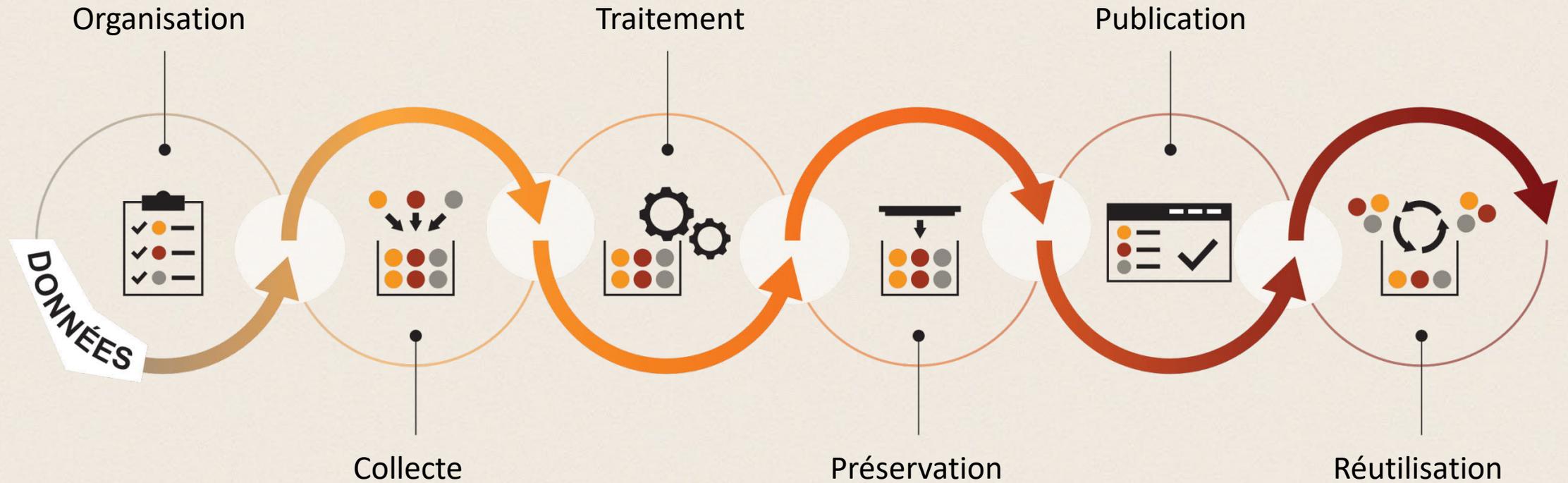
- Trouver données liées au projet
- Rendre lisible les données par d'autres

BESOIN

- Moteur de recherche
- Interopérabilité



LE CYCLE DE LA DONNÉE PAR HUMANA-NUM



« AUSSI OUVERTES QUE POSSIBLE, AUSSI FERMÉES QUE NÉCESSAIRE »

- Facile à trouver / Findable
- Accessible / Accessible
- Interopérable / Interoperable
- Réutilisable / Reusable





L'ÉCOSYSTÈME HUMA-NUM

SUIS-JE ÉLIGIBLE À L'OFFRE DE SERVICES?

Monde académique et projet scientifique validé par la TGIR Huma-Num.

Pour avoir un projet scientifique validé, plusieurs critères possibles :

- besoins spécifiques sur le traitement des données
- engagement sur l'interopérabilité des données de la recherche et des métadonnées associées
- démarche d'archivage à long terme des données.



PRODUIT D'UNE FUSION EN 2013



+ CORPUS (IR)



Huma-Num

la TGIR des humanités numériques

LES MISSIONS D'HUMA-NUM



DÉVELOPPER L'APPROPRIATION PAR LES
COMMUNAUTÉS SCIENTIFIQUES DU CYCLE
DE VIE DES DONNÉES NUMÉRIQUES



ORGANISER ET FACILITER LE TRAVAIL EN
ÉQUIPE ET LA GESTION DE PROJETS



PROPOSER DES SERVICES POUR LES DONNÉES
AU JUSTE NIVEAU ET AU BON MOMENT



TRAVAILLER COLLECTIVEMENT AUTOUR DES
DONNÉES



STABILISER L'ACCÈS AUX DONNÉES ET AUX
OUTILS



CO CRÉER DES SERVICES

BOTTOM UP



- Besoins des chercheurs
- Volonté d'une communauté de chercheurs
- Suggestion et accompagnement

TOP DOWN



- Adaptation des nouveautés technologiques
- Anticipation des besoins
- Comité scientifique



HUMA-NUM LAB

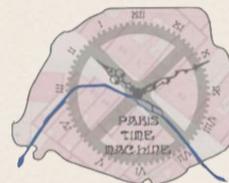
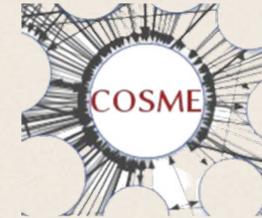


- Pôle d'accueil des chercheurs,
- des activités de recherche et développement,
- d'innovation et de valorisation

LES CONSORTIUMS

- Regroupement d'unités et équipes de recherche autour de thématiques et d'objets communs
- 10 consortiums labellisés
- Ancrage dans les Maisons de Sciences de l'Homme (MSH)
- Formation et expertises, création de services, guides de bonnes pratiques (usages de la 3D...)





LES MAISONS DES SCIENCES DE L'HOMME

- Participation à des groupes de travail spécifiques
- Connaissance de l'ensemble de l'écosystème et de la documentation
- Organisation de formations dédiées



VICTORIA LE FOURNER



LES SERVICES ET LES OUTILS

ACCÈS AUX SERVICES - HUMANID

Interface d'authentification regroupant les services Web d'Humanum que vous avez ouvert ou que vous pouvez demander à ouvrir.

 Votre assistant de recherche en Sciences Humaines et Sociales accéder	 Plateforme de stockage et de partage de fichiers (Web et clients WebDAV) accéder	 Partager, publier et valoriser vos données scientifiques accéder
 Un éditeur de texte simplifiant la rédaction et l'édition d'articles scientifiques en SHS accéder	 Plateforme de forge basé sur git Demander l'accès	 Service de discussion d'équipes Demander l'accès
 Analyse d'audience de sites Webs Demander l'accès	 Réduire et pérenniser vos URL accéder	 Kanboard, Outil de gestion de projet Demander l'accès



Sur le site web d'Huma-Num

HN Huma-Num
la TGIR des humanités numériques

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

Sélectionner votre mode de connexion interne ou externe

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

Service Central d'Authentification du CCSD

Entrez votre identifiant et votre mot de passe.

HAL archives-ouvertes.fr

Prévenez-moi avant d'accéder à d'autres services.

[Login oublié](#) | [Mot de passe oublié](#)

Vous n'avez pas de compte

[Aide](#)

Remplir les champs nécessaires

HN Huma-Num
la TGIR des humanités numériques

Nous n'avons pas trouvé de compte Huma-Num pour cet identifiant externe. Merci de

- 1) vous créer un compte
- OU
- 2) connectez-vous avec votre compte HumanID et associez le à cet identifiant externe sur la page "Mes Identifiants"

Nom

Prénom

Mail

Captcha

[Retour au portail](#)

INTERFACE UNE FOIS LA CONNEXION FAITE :

Les services **Isidore** et **Stylo** sont immédiatement disponibles.

Il faut demander à accéder à **ShareDocs**, **Nakala**, **GitLab** et **Mattermost**

HN HumanID

Mes services Mes identifiants Mon mot de passe Mes informations

VL

Gérer mes services Huma-Num



Votre assistant de recherche en Sciences Humaines et Sociales

accéder



Plateforme de stockage et de partage de fichiers (Web et clients WebDAV)

accéder



Partager, publier et valoriser vos données scientifiques

accéder



Un éditeur de texte simplifiant la rédaction et l'édition d'articles scientifiques en SHS

accéder



Plateforme de forge basé sur git

Demander l'accès



Service de discussion d'équipes

Demander l'accès

Paramètres du compte personnel

SERVICES ADDITIONNELS

KB



GitLab



SERVICES TIERS HÉBERGÉS

Outils et plateformes dont la conception, la maintenance et la documentation sont assurés par d'autres instances qu'Huma-Num :

The logo for Opentheso, featuring the word "Opentheso" in white text on a green rectangular background.The logo for Voyant Tools, featuring a red owl icon and the text "Voyant Tools" in white on a blue rectangular background.The logo for Stylo, featuring the word "Stylo" in white italicized text between two white underscores on a black rectangular background.The logo for HEURIST, featuring a purple globe icon and the word "HEURIST" in dark blue text.

DES OUTILS CRÉÉS PAR HUMANA-NUM



- Une équipe de développeurs au sein d'Huma-Num
- Adapter des outils existants

TYPES DE STOCKAGE

Les besoins

- Stocker des données hétérogènes : base des données, fichiers xml, photos d'archives...
- Documenter les données grâce aux métadonnées
- Collaborer
- Archiver

Les différentes réponses :

Stockage « outillé » pour les SHS : [ShareDocs](#)

Stockage des données documentées : [Nakala](#)

Archivage et conservation des données : [Huma-Num Box](#)

SHAREDOCS

Gestionnaire de fichiers, pouvant être utilisé via un navigateur web, un client WebDAV ou un logiciel de synchronisation de fichiers.

- stocker,
- échanger,
- partager,
- travailler sur des données de type fichiers





ShareDocs

COMME OUTIL DE STOCKAGE ET DE COLLABORATION

→ Adapté pour stocker, échanger, partager, travailler sur des données de type fichiers (ensemble de photos, de textes transcrits, etc) pendant la durée du projet.

- Notion de versioning des fichiers
- Accès par le navigateur web
- Capable à l'intérieur de chaque dossier de faire des groupes de recherche et d'attribuer des droits différents

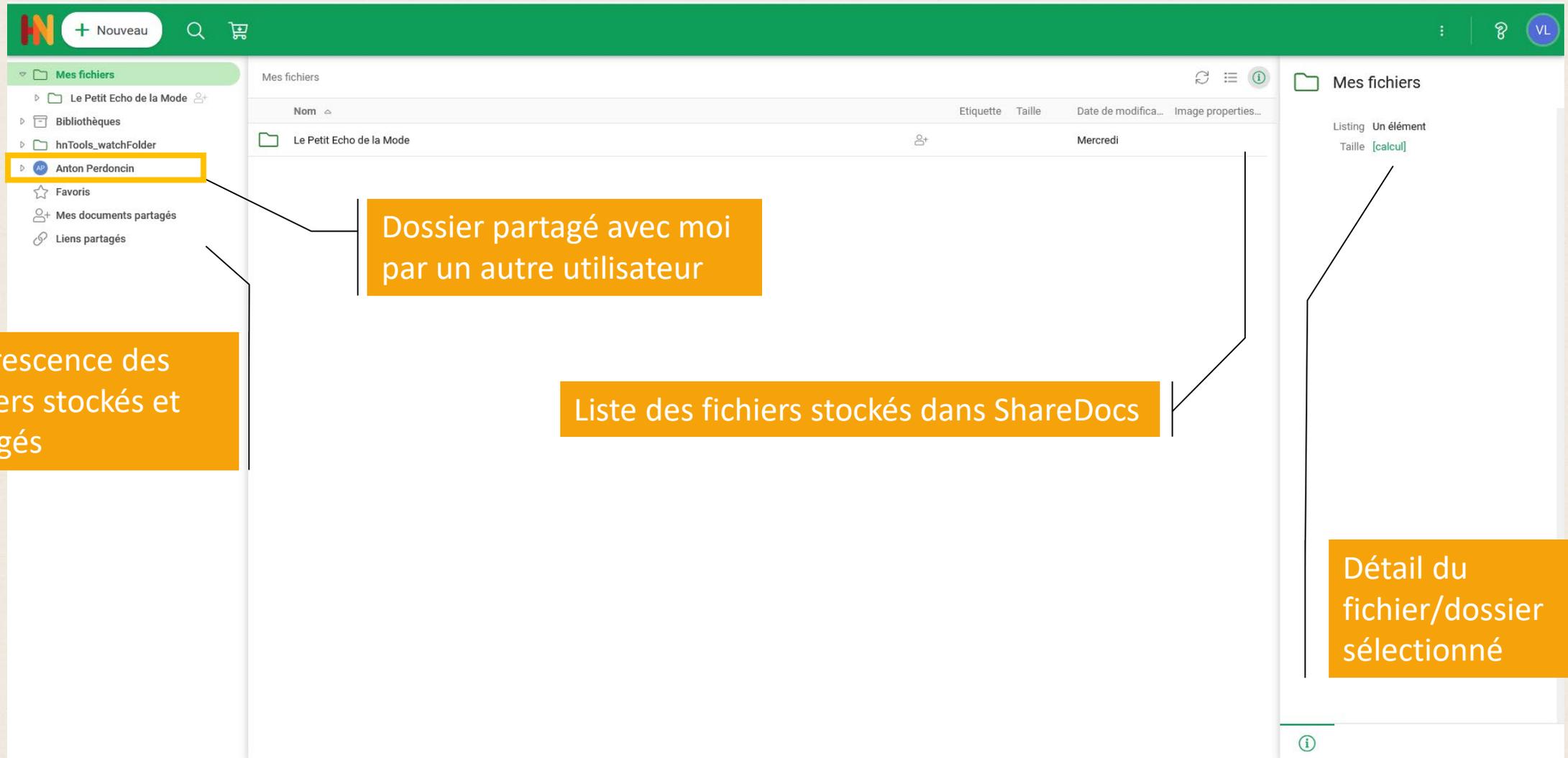
Exemple : permet de préparer des fichiers pour une édition en ligne ou bien une diffusion dans NAKALA



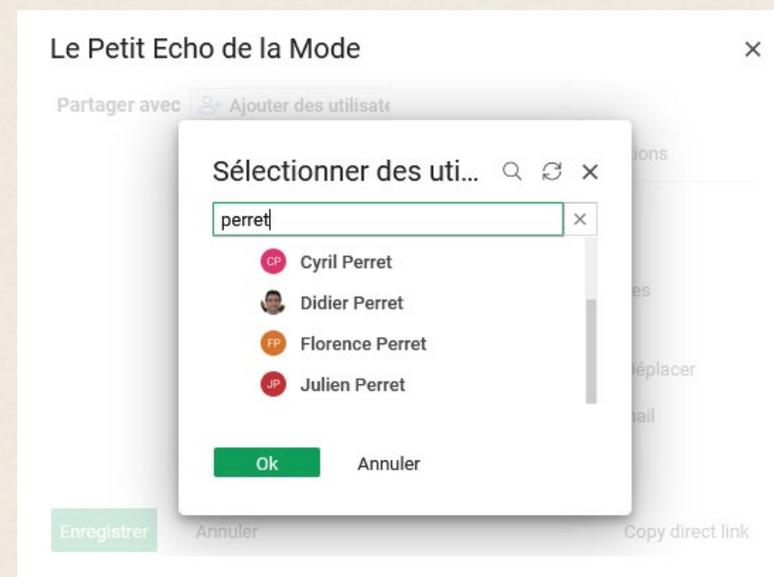
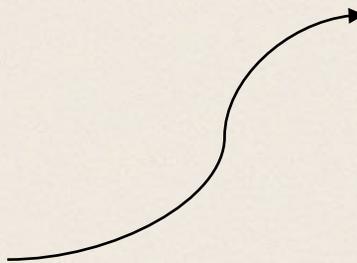
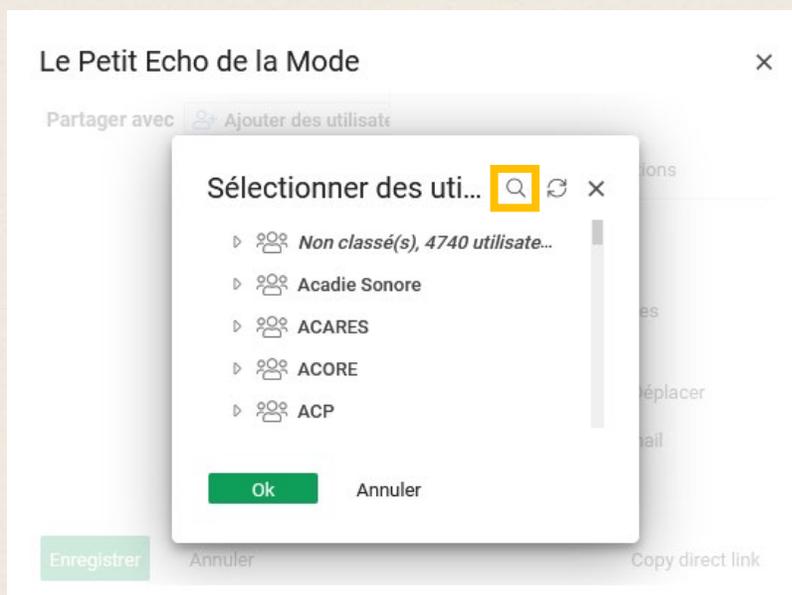
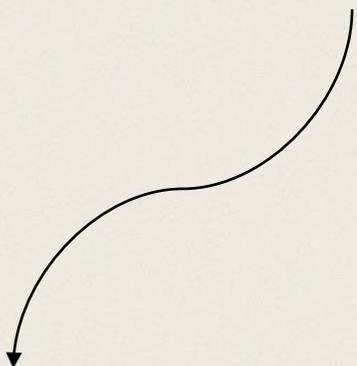
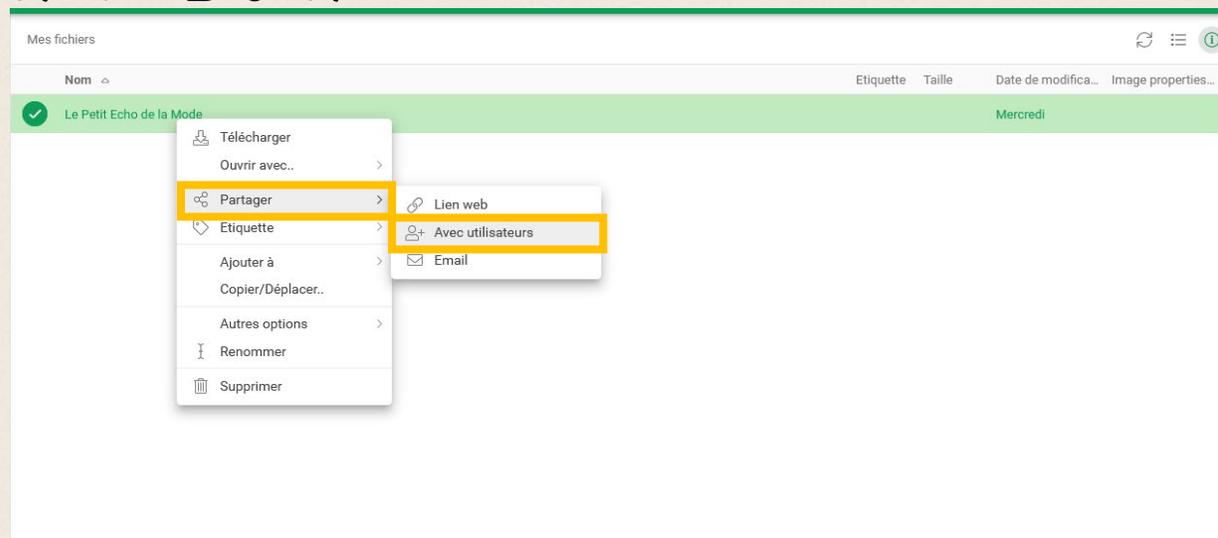
HN ShareDocs



INTERFACE DE SHAREDOKS



AJOUTER UN COLLABORATEUR



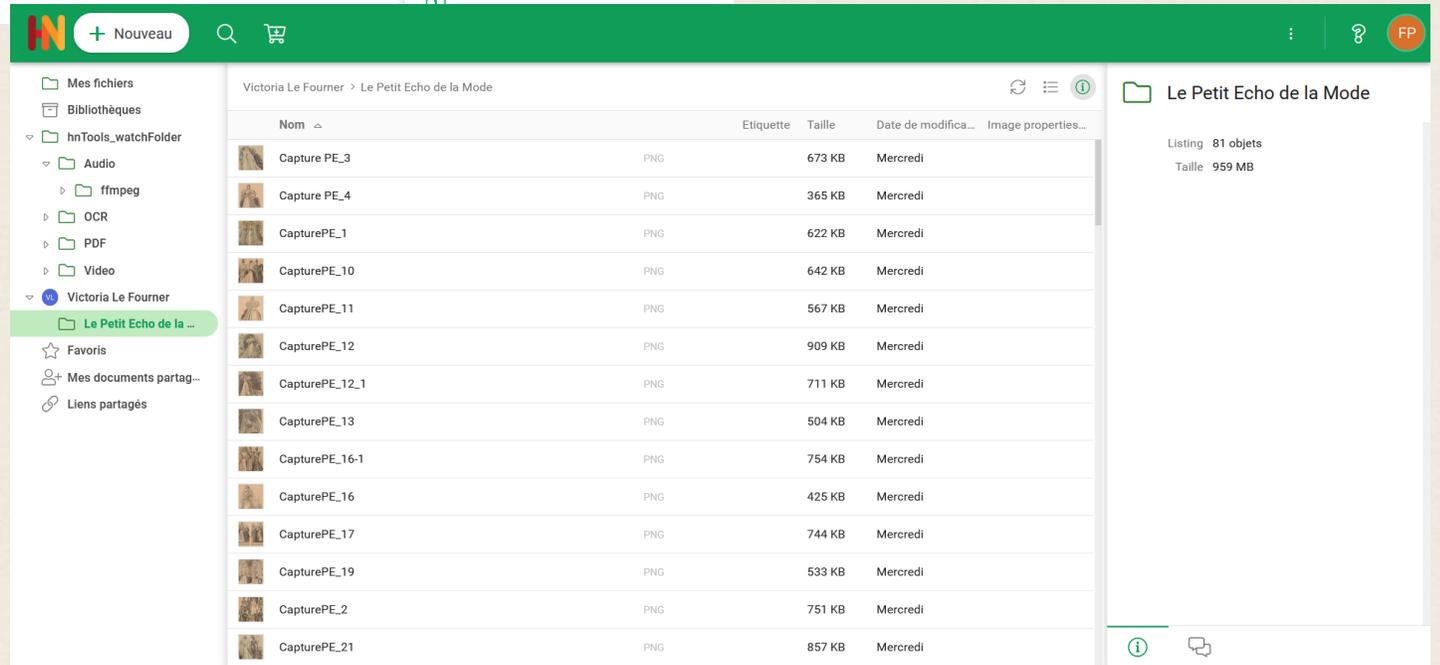
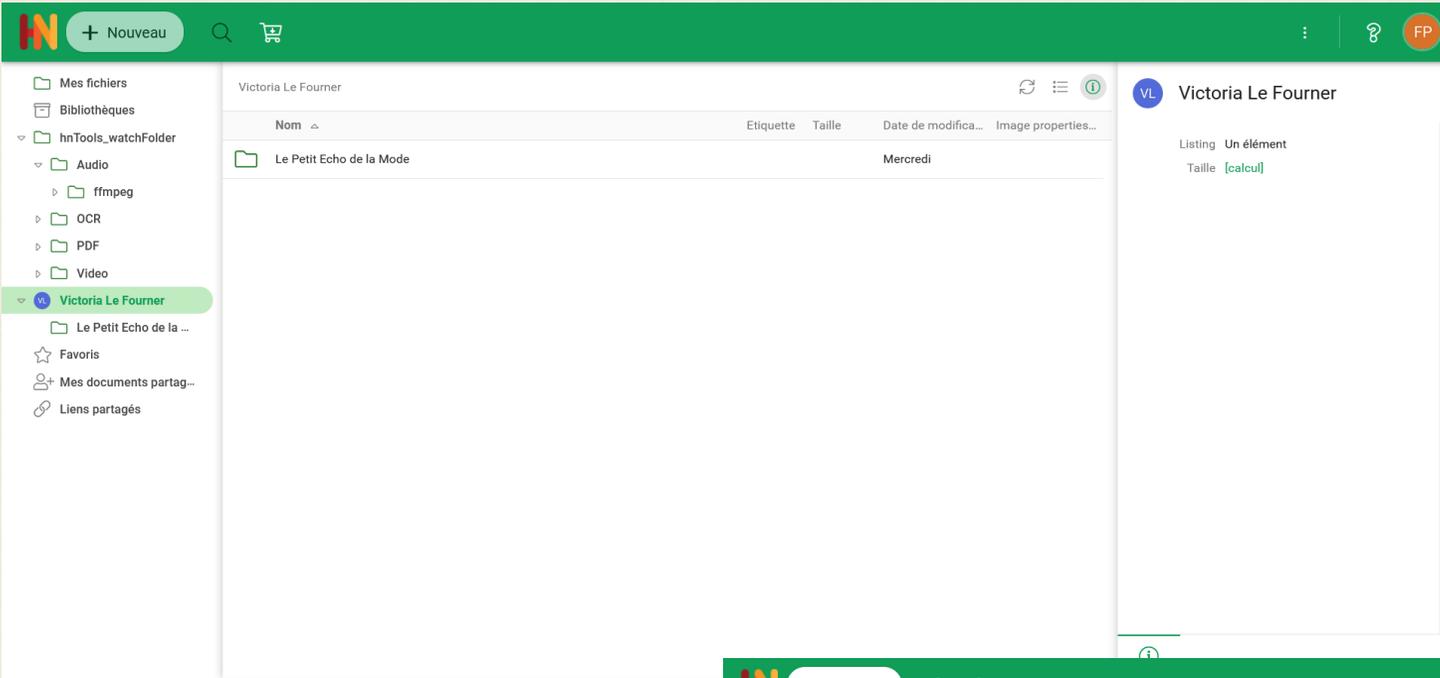
CONFIGURER SES DROITS

The screenshot shows a sharing dialog box titled "Le Petit Echo de la Mode" with a close button (X) in the top right corner. Below the title, it says "Partager avec" followed by a plus icon and the text "Ajouter des utilisat". A user named "Florence Perret" is listed with a red "X" and the word "Supprimer" next to it. Below the user list, there are two tabs: "Permissions" (selected) and "Options". Under the "Permissions" tab, there is a list of permissions, each with a checked checkbox: "Télécharger", "Lire les commentaires", "Ajouter des commentaires", "Envoi", "Supprimer, Renommer, Déplacer", and "Partage par lien-web/email". At the bottom left, there are two buttons: "Enregistrer" (green) and "Annuler". At the bottom right, there is a link that says "Copy direct link".

Plusieurs autres utilisateurs peuvent être ajoutés

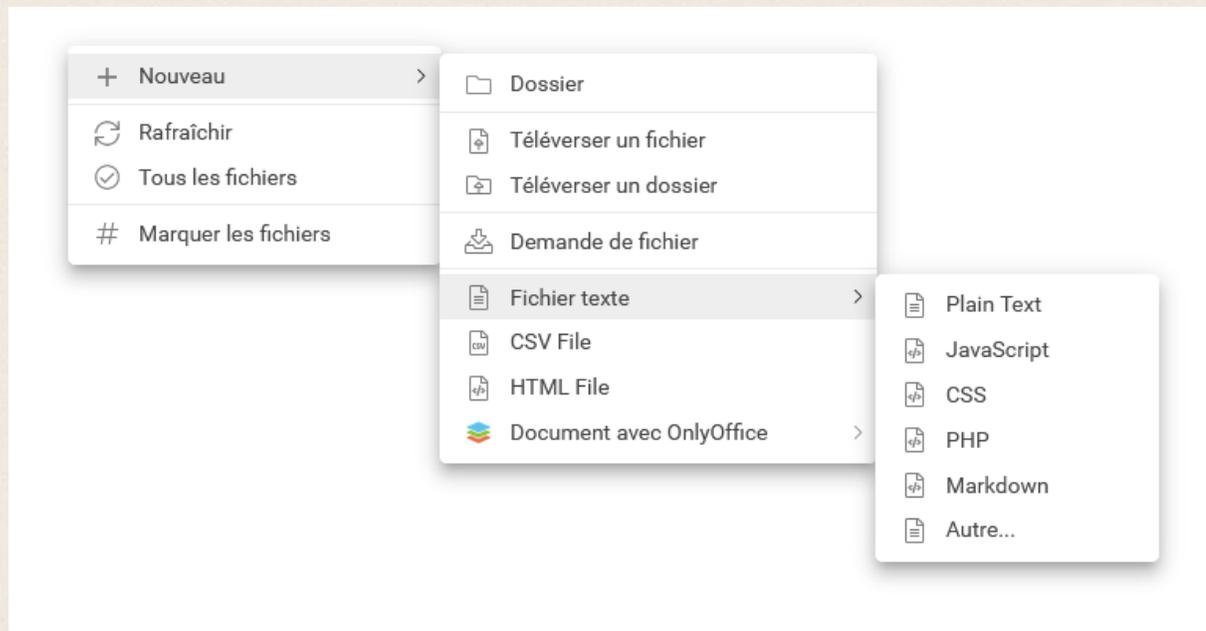
Possibilité de moduler les droits

VUE DU COLLABORATEUR

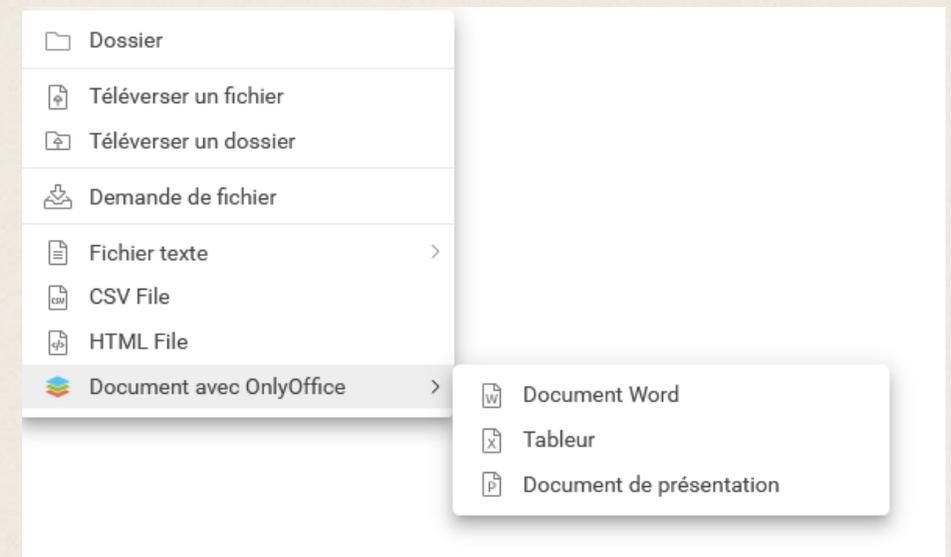


SHAREDOCS COMME AIDE AU TRAITEMENT

- Possibilité d'éditer les données en collaboration
- Possibilité de créer des documents directement sur ShareDocs



Création de type de fichiers très divers



OnlyOffice implémenté

CONVERSION/TRANSCODAGE AUDIO ET VIDÉO

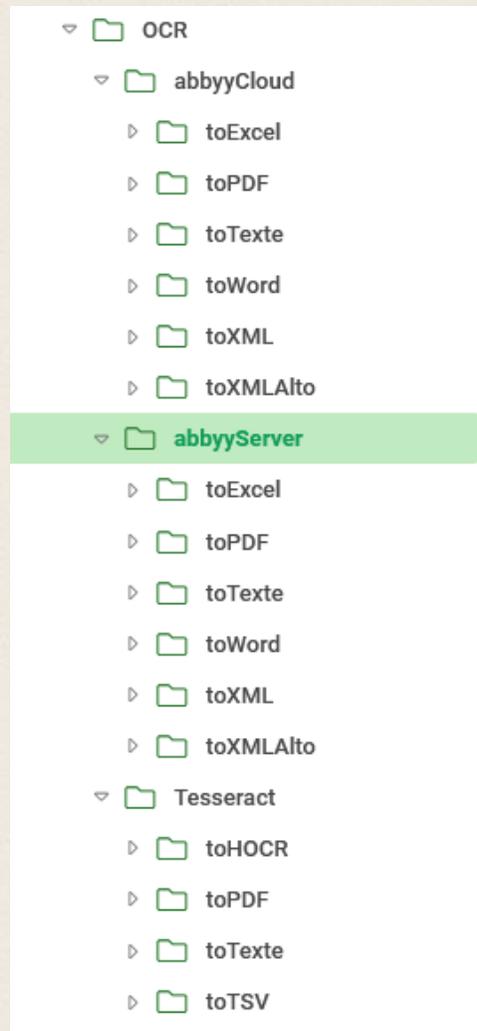
Les presets pour l'audio sont :

- toMP3 : transcodage en MP3 sans options
- toWAV : transcodage en WAV (codage PCM / Pulse Code Modulation) sans option
- toMP4_aac-lc : conteneur MP4 avec codec audio AAC (Low Complexity)

Pour la vidéo les presets généraux sont :

- To MP4_h264 : conteneur MP4 avec codec vidéo h264, codec audio AAC
- toMP4_h265 : conteneur MP4 avec codec h265, codec audio AAC
- toMP4_h264_720p : conteneur MP4 avec codec h264, codec audio AAC, 720 lignes max de hauteur
- toMP4_h264_1080p : conteneur MP4 avec codec vidéo h264, codec audio AAC, 1080 lignes max de hauteur

OPTICAL CHARACTER RECOGNITION (OCR)



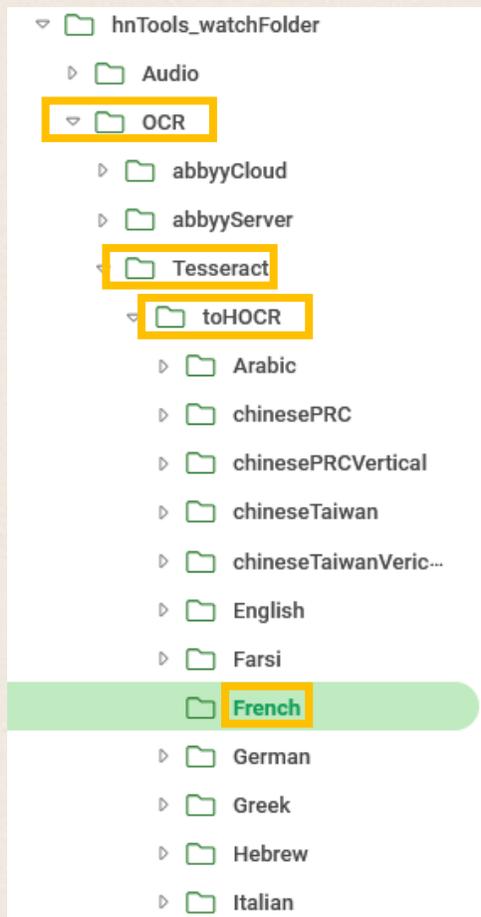
AbbyCloud est fait pour traiter rapidement des petits documents. Les fichiers ne doivent pas excéder 30Mo. Au-delà il faut utiliser **Abby Fine Reader serveur**.

900 pages par utilisateur pour AbbyCloud et Abby Fine Reader

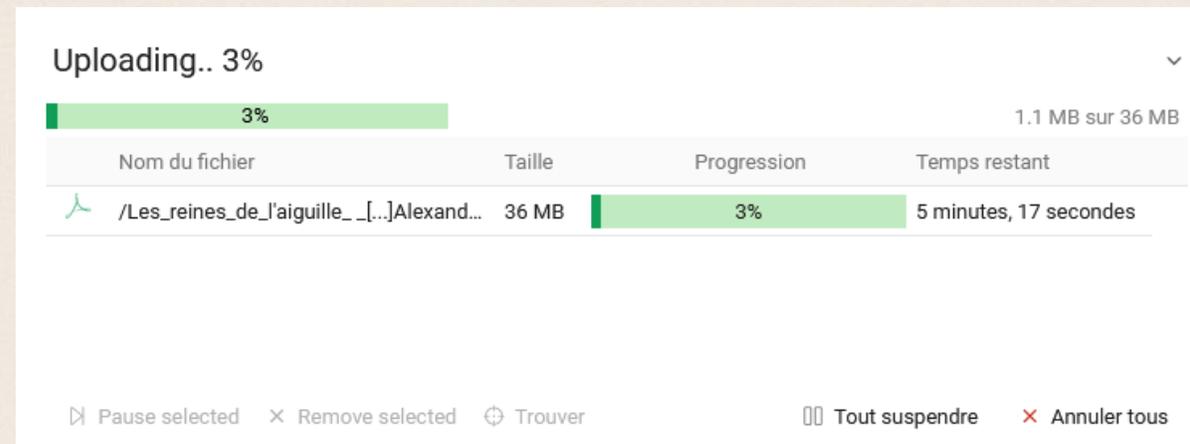
Tesseract ne pouvant lire des fichiers PDF, une conversion PDF vers TIFF est automatiquement faite. L'OCR Tesseract est open source, il n'y a pas de quota dans le nombre de fichiers.

L'ensemble des outils ne peuvent pas traiter de document crypté ou faisant l'objet d'une protection ou d'une restriction d'accès.

SHAREDOCS, AIDE À L'OCR ET PASSAGE AU TEXTE



- Glisser-Déposer dans le dossier son fichier à OCRiser
- Chargement du fichier sur ShareDocs
- Traitement automatique une fois placé dans le dossier



Attention : ShareDocs ne montre pas de signe qu'il est en cours de traitement !

HN-Tools : job information

```
-----  
--- INPUT  
-----  
File name      : Les_reines_de_l'aiguille_[...]Alexandre_Arsène_bpt6k6378008n.tif  
File path     : vlefournier/OCR/Tesseract/toHOCR/French  
File size     : 4095.35mo  
File extension : tif  
File type     : image/tiff  
File hash     : 30b8be89ef9a8fcde9bd9fb900e400a3189a20c3d7aad5aa508c413c  
File date     : 1620064013  
Submit by    : Le Fournier Victoria (victoria.lefournier@univ-lille.fr)  
-----  
--- OUTPUT  
-----  
Output file   : vlefournier/OCR/Tesseract/toHOCR/French/Les_reines_de_l'aiguille_[...]Alexandre_Arsène_bpt6k6378008n_hnOCR  
Tool         : OCR  
Engine      : Tesseract  
Preset 1    : toHOCR  
Preset 2    : French  
Execution time : 3mn 6.518s  
-----  
--- RESULT  
-----  
State       : Completed  
-----
```

For more information about this job, you can reply to this email

TGIR HUMA-NUM

Mail reçu une fois la tâche terminée

Résultat à récupérer à côté du PDF en TIF

hnTools_watchFolder > OCR > Tesseract > toHOCR > French

Nom	Etiquette	Taille	Date de modifica...	Image properties...
 Les_reines_de_l'aiguille_[...]Alexandre_Arsène_bpt6k6378008n	PDF	35.8 MB	19:43	
 Les_reines_de_l'aiguille_[...]Alexandre_Arsène_bpt6k6378008n	TIF	4 GB	19:46	

NAKALA

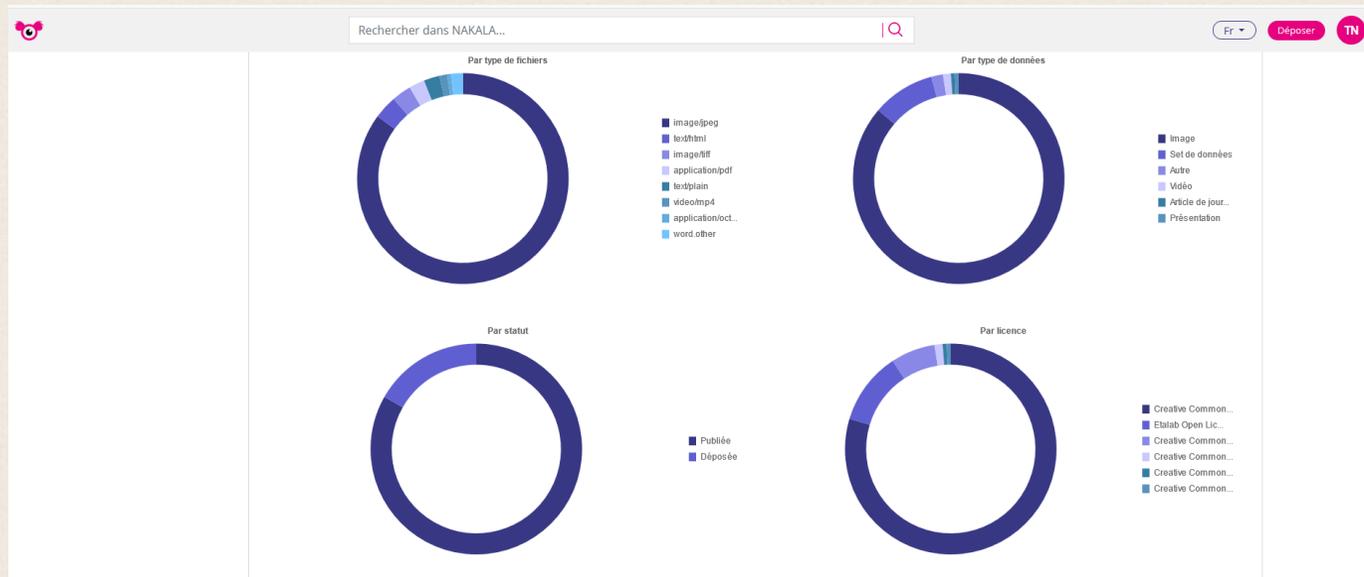
Entrepôt sécurisé pour la publication de données de recherche en accord avec les principes du *FAIR data*

- partager,
- publier
- valoriser tous types de données numériques documentées





FONCTIONNALITÉS



- Tableau de bord
- Données**
- Collections
- Listes
- Sites web

Mes données

Rechercher par titre Filtrer par : Type Statut Année de dépôt Trier par : Date de dépôt (décroissante)

<input type="checkbox"/>		Date de création	Statut	
<input type="checkbox"/>	carnet Mission d'Ethnomusicologie du 5 au 12 Septembre 1969 Charente Maritime - Côte Atlantique - Île d'Oléron	03/05/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	telephone-booth-768610_1920.jpeg	30/04/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	EFEO_BOITE01_102-a_01.jpg	30/04/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	EFEO_BOITE01_100_01.jpg	30/04/2021	Sous embargo	<input type="button" value="⋮"/>
<input type="checkbox"/>	EFEO_BOITE01_102-a_01.jpg	30/04/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	EFEO_BOITE01_100_01.jpg	30/04/2021	Sous embargo	<input type="button" value="⋮"/>
<input type="checkbox"/>	Fiche d'inventaire - Objet n° 102 a	30/04/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	Fiche d'inventaire - Objet n° 101 a	30/04/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	Fiche d'inventaire - Objet n° 100	30/04/2021	Publiée	<input type="button" value="⋮"/>
<input type="checkbox"/>	Fiche d'inventaire - Objet n° 102 a - final	30/04/2021	Publiée	<input type="button" value="⋮"/>

NAKALA_PRESS

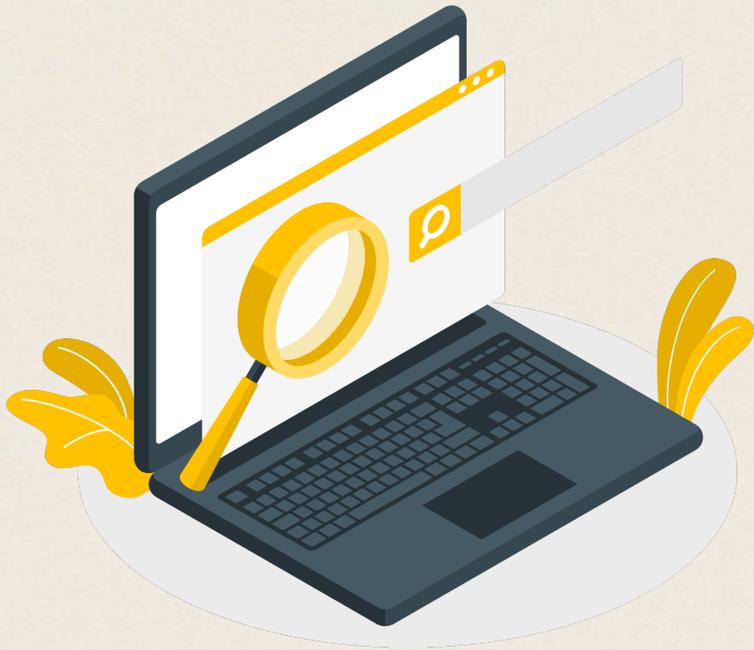


Option de Nakala

Editorialisation des données
d'une collection

- Valoriser
- Diffuser
- Rendre lisible

ISIDORE



Moteur de recherche spécialisé

- Publications,
- Données numériques
- Profils de chercheurs



Votre assistant de recherche
en Sciences Humaines et Sociales

Collections ▾ Rechercher dans les 8 953 collections de ISIDORE...

[i démonstration](#)





CRÉER UN COMPTE ISIDORE – AVOIR UN ASSISTANT PERSONNEL

1

Bienvenue Victoria Le Fournier !

En rejoignant la communauté, vous pouvez garder le fil de vos recherches, suivre vos auteurs et organiser vos documents !

Nous aimerions mieux vous connaître.

Vous pourrez ainsi identifier vos publications et d'autres Isidoriens pourront consulter votre profil.

2

Êtes-vous l'auteur de ces documents ?

Deux siècles de sources disparates sur l'industrie textile en France : comment automatiser les traitements d'un corpus non-uniforme ? | Alix Chagué et al.

Deux siècles de sources disparates sur l'industrie textile en France : comment automatiser les traitements d'un corpus non-uniforme ? | Alix Chagué et al.

3

ID chercheur (Victoria-Le-Fournier par défaut)

titre

décrivez votre activité

Français

UN ASSISTANT DE RECHERCHE PERSONNEL — APPRÉHENDER LES AUTEURS

Possibilité
de faire
un tri

De suivre
un auteur

The screenshot displays a research assistant interface. On the left, a sidebar for author **Manuel Charpy** shows his disciplines and co-authors. The main area lists documents written by or about him, with a filter menu on the right.

Manuel Charpy

Discipline
Histoire (46) - Sociologie (10) - Art et histoire de l'art (10) - Littératures (8) - Anthropologie sociale et ethnologie (6) - Architecture, aménagement de l'espace (3) - Musique, musicologie et arts de la scène (2) - Géographie (2) - Psychologie (1) - Linguistique (1) - Sciences de l'information et de la communication (1) - Education (1) - Droit (1) - Etudes classiques (1) - Archéologie et Préhistoire (1)

Type de document
Articles (40) - Billets de blog (37) - Livres et chapitres d'ouvrages (18) - Périodiques (14) - Autres (9) - Photos et images (9) - Documents audiovisuels (3) - Textes imprimés (2) - Recensions (2) - Page Web (1) - Mémoires, Thèses et HDR (1)

Co-auteur

François Jarrige	8
Judith Lyon-Caen	6
Raymond Huard	6
Quentin Deluermoz	5
Caron (Jean-Claude)	5
Gil Bartholeyns	5
Tangi Villerbu	4

Documents écrit par Manuel Charpy (125) | parle de Manuel Charpy (376)

	Mode d'emploi. Sur quelques feuillets dans une machine à coudre Éric Perrin et al. (2021)	Articles		
	Moi & Arnys. Entretien avec Philippe Trétiack manuel charpy (14 déc. 2020)	Articles		
	« Mon chemin dans la sape ». Une vie de vêtements au Congo Brazzaville, 1971-2019 manuel charpy (14 déc. 2020)	Articles		
	"Arbres à loques. Senarpont, juillet 2020." manuel charpy (14 déc. 2020)	Articles		
	Ma vie dans la sape par Séverin Mouyengo (préface seule) manuel charpy (11 déc. 2020)	Livres et chapitres d'o...		
	Dans l'intimité de l'exil Sylvie Aprile et al. (1 déc. 2020)	Périodiques		
	"Objets" manuel charpy (1 déc. 2020)	Livres et chapitres d'o...		
	"Vêtement" manuel charpy (1 déc. 2020)	Livres et chapitres d'o...		
	Mode d'emploi. Sur quelques feuillets dans une machine à coudre Éric Perrin et al. (1 déc. 2020)	Articles		

PROPOSITIONS EN LIEN

Par les mêmes auteurs

Le XIXe siècle au prisme des Visual Studies,
Entretien de Quentin Delu...

Manuel Charpy et al.

2014

Articles

Art et histoire de l'art



La vente de journaux dans une ville 1900. Détail
d'un diorama réalisé...

manuel charpy

21 oct. 2018

Autres

MédHAL



Diorama intitulé « Le Tour du Monde » réalisé par
Dory77 et Ti'poussin...

manuel charpy

21 oct. 2018

Autres

MédHAL



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.

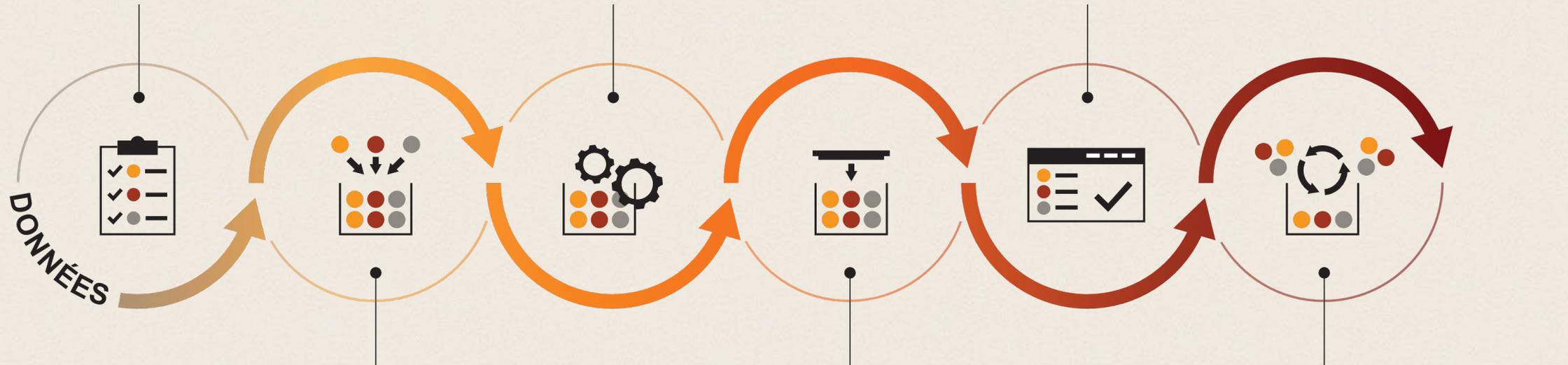
- Calcul statistique et environnements R
- Logiciels d'enquête et d'analyse de données
- Reconnaissance de caractères
- Puissance de calcul (+ CC-IN2P3)

...

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.

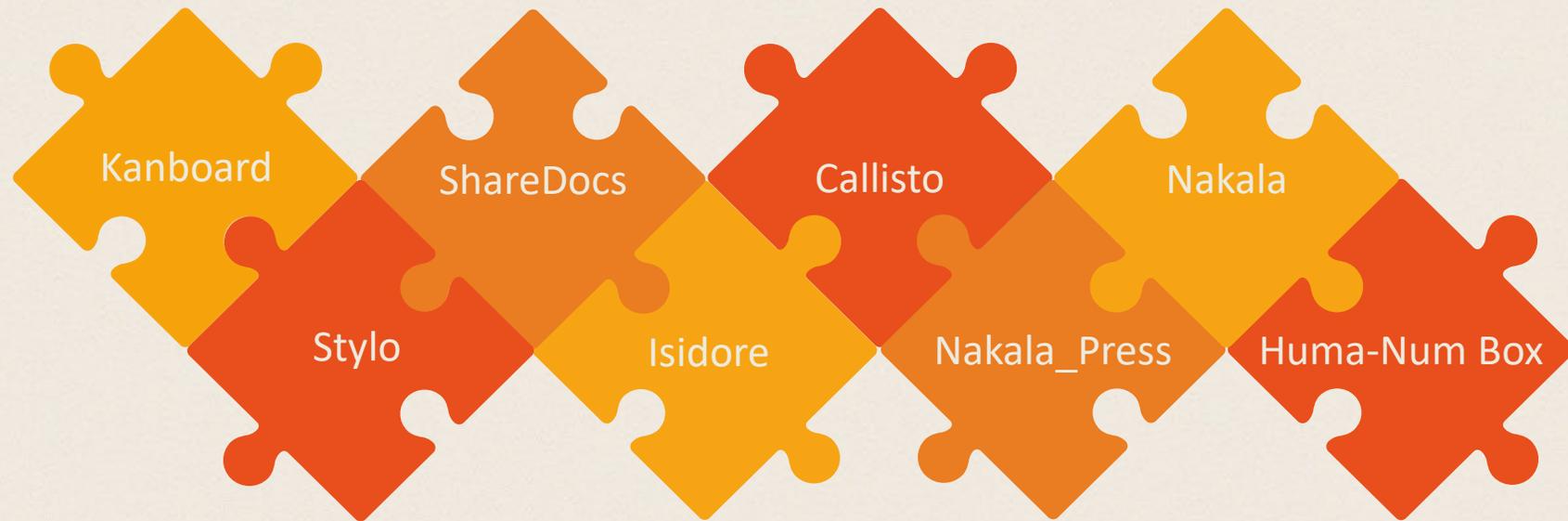
- Nakala
- Huma-Num Box
- Préservation à long terme (+ CINES)

RÉUTILISATION

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

- Portail web
- API
- Triplestore
- OAI-PMH

DES BRIQUES EN FONCTION DES PROJETS





POUR ALLER PLUS LOIN...



RESSOURCES

- Documentation : <https://documentation.huma-num.fr/>
- Carnet de recherche : <https://humanum.hypotheses.org/>
- Twitter : https://twitter.com/Huma_Num
- Canal-U : <https://www.huma-num.fr/chaine-huma-canal-u/>
- Site : <https://www.huma-num.fr/>

CONTACTS

- Relais MSH en région Hauts-de-France : victoria.lefourner@univ-lille.fr
- Questions générales sur Huma-Num : contact@huma-num.fr
- Demandes concernant un projet ou ouverture d'un service non accessible via HumanID : cogrid@huma-num.fr
- Besoin d'assistance sur un service ouvert : assistance@huma-num.fr



DES QUESTIONS ?

MERCI DE VOTRE ATTENTION !

Crédit images : [Storyset](#)

