



Huma-Num
la TGIR des humanités numériques

Huma-Num, l'IR* des humanités numériques

Johann Holland, Chargé de mission

JoSY 2021

09-sept-2021



Aix-Marseille
université

CAMPUS
CONDORCET
Paris-Aubervilliers

Introduction

- Objectifs de la présentation
 - Comprendre ce qu'est Huma-Num, et la culture Huma-Num
 - Comprendre à quoi sert Huma-Num, et dans quelles situations
 - Découvrir l'offre de services et les conditions d'utilisation
 - Se sentir comme partie prenante de la communauté Huma-Num
- Démarche suivie
 - Adopter la position d'un porteur de projet en SHS qui mobilise des données numériques
 - Identifier ses besoins et mieux connaître ses pratiques
 - Fournir quelques exemples concrets d'utilisation des services
 - Montrer les avantages à s'inscrire dans la communauté des utilisateurs

Introduction

- Plan
 1. Les missions d'Huma-Num et la culture de la Science Ouverte
 2. Les situations types pour un porteur de projet et les réponses que peut apporter Huma-Num
 3. L'organisation d'Huma-Num pour construire ces réponses et les modalités d'accès aux services

Huma-Num et la Science Ouverte

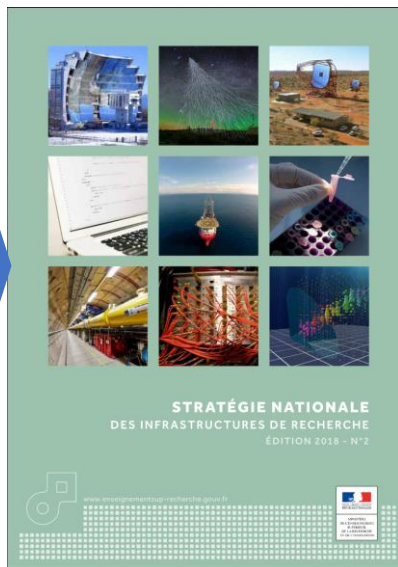
L'IR* des humanités numériques en 2021

**Une TGIR :
Une infrastructure
pour les SHS dans la
science ouverte.**



Une UMS-USR





Des IR françaises, ... dont des infrastructures pour les SHS et l'IST :

CFHT
Canada-France-Hawaii Telescope



Type de l'infrastructure : TGR

ESRF
Source Européenne de Rayonnement Synchrotron/
European Synchrotron Radiation Facility



Type de l'infrastructure : TGR

Huma-Num
la TGIR des humanités numériques

PROGE DO
DATA INFRASTRUCTURE

Réseau national **MSH**

OpenEdition

IR **MÉTOPES**

CollEx-Persée
Développer les services documentaires de demain

CCSD | CENTRE POUR LA COMMUNICATION SCIENTIFIQUE DIRECTE

La Science partagée

DARIAH-EU

CLARIN
Common Language Resources and Technology Infrastructure



« 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 »

EUROPEAN OPEN SCIENCE CLOUD

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.

Science ouverte

« La science ouverte est la diffusion sans entrave des publications et des données de la recherche »



Axe 2
Structurer et ouvrir les données de recherche

Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes.

«Les données de la recherche sont la matière première de la connaissance.»

PARTAGER SUR TWITTER

Mesures concrètes

- 4** Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.
- 5** Créer la fonction d'administrateur des données et le réseau associé au sein des établissements.
- 6** Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles par les chercheurs.

Sur les publications et les données, le cadre législatif est récent : loi pour une République numérique d'octobre 2016.

1. Publications

L'article 30 donne un nouveau droit aux auteurs, celui de publier en accès ouvert leurs articles après un embargo de 6 à 12 mois, pour des recherches financées au moins à 50% sur fonds public.

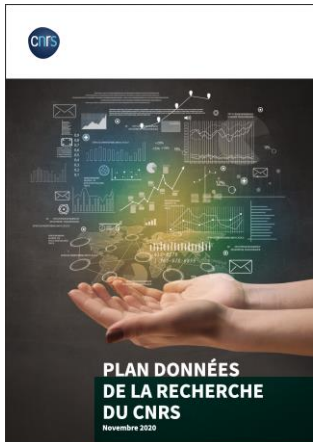
Il crée donc un droit.

2. Données

L'article 6 définit le principe d'ouverture par défaut pour les données administratives : données construites dans le cadre des missions de service public, auxquelles les données de recherche appartiennent.

Il crée donc une obligation.

Le CNRS doit se doter de moyens pour agir de manière coordonnée sur l'ensemble du continuum, qui va du calcul aux publications, en passant par les données massives, la longue traîne de données, les infrastructures matérielles et logicielles, les services aux utilisateurs, le référencement et la documentation.



Situations types et réponses d'Huma-Num

Situations type

- Suivant le degré de maturité de mon projet de recherche, et mes connaissances liées aux « données de la recherche », j'ai besoin :
 - S1 D'obtenir des conseils méthodologiques et techniques pour faire mes choix
 - S2 D'être mis en réseau avec d'autres projets ou acteurs à différentes échelles
 - S3 D'organiser, collecter, traiter, préserver, publier, réutiliser des données
 - S4 De me former à certaines notions et certains services ou outils
 - S5 De faire évoluer des services existants pour mieux les adapter à mes besoins
 - S6 De partager des réflexions et développements autour d'objets de prospective
 - S7 ...

Situations type

S1

Obtenir des conseils méthodologiques et techniques pour faire mes choix : *quelques exemples*

- Mes données sont peu volumineuses mais ont de la valeur pour ma recherche et sont juste dupliquées sur un disque dur : comment les sécuriser dans la durée ?
- J'ai un volumineux corpus d'images à numériser, et je m'interroge sur le meilleur format pour les fichiers de sortie (TIFF ? PNG ? JPEG2000 ?)
- Comment puis-je gérer le respect du RGPD pour mes données d'enquête ? Est-ce lié au plan de gestion de données ?
- Je souhaite rendre mes publications compatibles avec le FAIR : dois-je déposer sur une archive ouverte ?
- Les données que je veux collecter vont servir à alimenter une base de données et une plateforme web : à quel moment dois-je définir mon modèle de données ?

Réponses aux situations type

R1 Réponse d'Huma-Num :

- Si les conseils attendus portent spécifiquement sur les services d'Huma-Num, l'équipe vous apportera une réponse
- Si les conseils attendus portent sur d'autres sujets, l'équipe cherchera à vous rediriger vers les acteurs compétents ou les ressources pertinentes, si elle les connaît
- Si un acteur-relais (ex : MSH) existe à proximité de la personne qui fait la demande, l'équipe vous réorientera vers le service d'appui à la recherche de cette structure
- Si les conseils attendus portent sur des questions juridiques et éthiques, l'équipe pourra vous mettre en lien avec des DPO

Situations type

S2

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

- Je travaille en histoire de l'art sur des objets numérisés en 3D, et je cherche à échanger sur ces problématiques avec d'autres équipes françaises du domaine, ou d'autres disciplines scientifiques mais travaillant sur la 3D
- Je fais actuellement partie d'un petit groupe d'échange sur le TAL au niveau national, et je suis à la recherche d'expériences similaires à l'étranger
- Pour déposer un dossier lors d'un AAP régional, je cherche des partenaires potentiels sur le territoire où je me trouve
- J'organise des écoles d'été avec un partenaire canadien qui sont ouvertes aux étudiants et doctorants de nos deux réseaux, et je cherche à étendre le dispositif à un troisième pays

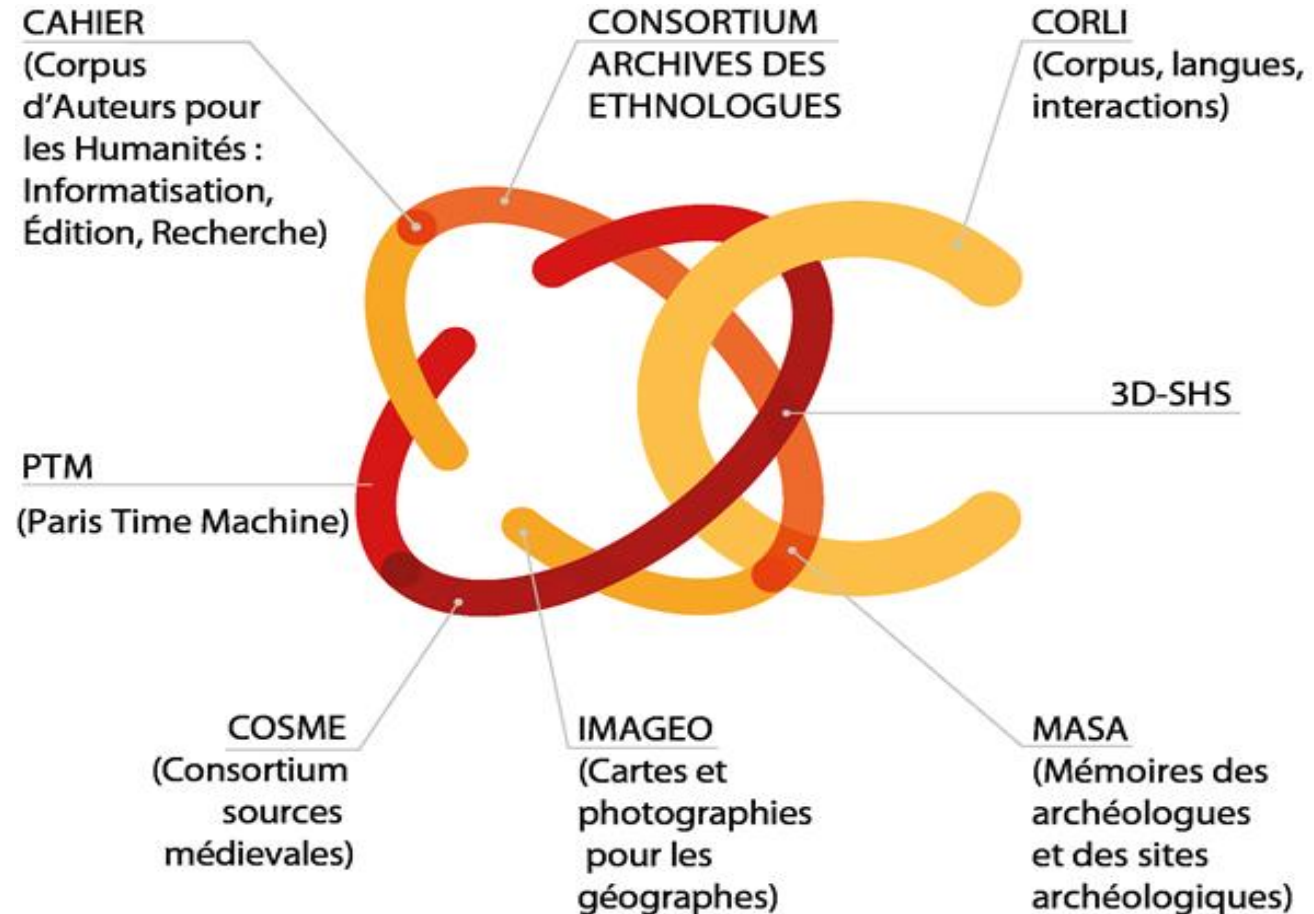
Réponses aux situations type

R2 Réponse d'Huma-Num :

- Mise en lien avec un consortium Huma-Num si l'objet de votre demande correspond à celui du consortium
- Mise en lien avec des infrastructures de recherche européennes, et/ou des projets ou réseaux européens suivant la nature de votre domaine ou objet d'échange scientifique
- Mise en lien avec la communauté des correspondants Huma-Num dans les acteurs-relais (ex : MSH <https://www.huma-num.fr/trouver-son-relais-msh/>)

Les consortiums Huma-Num

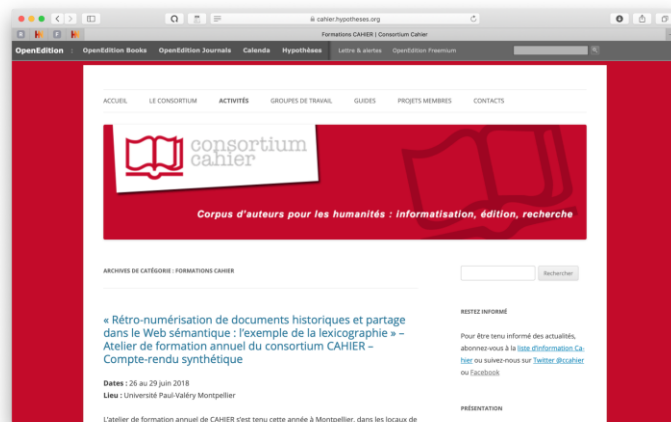
- 8 consortiums, 80-100 équipes de recherches qui couvrent des champs variés des SHS
- Un réseau de chercheurs, ingénieurs, bibliothécaires, archivistes, etc.
- Plateformes, outils, Guides, Formations,...



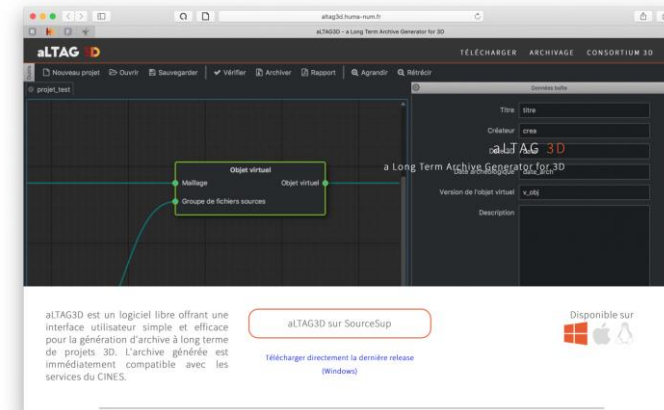
Exemples de réalisations de consortiums



transcrire.huma-num.fr



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



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

Du national à l'international

- **Communautés**
- **Réseaux et relais**
- **Infrastructures**

**Projets de
recherche**



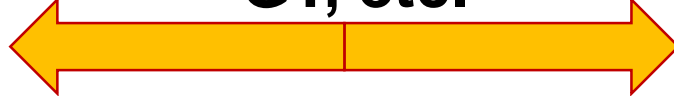
Partenariats



Veille stratégique



GT, etc.



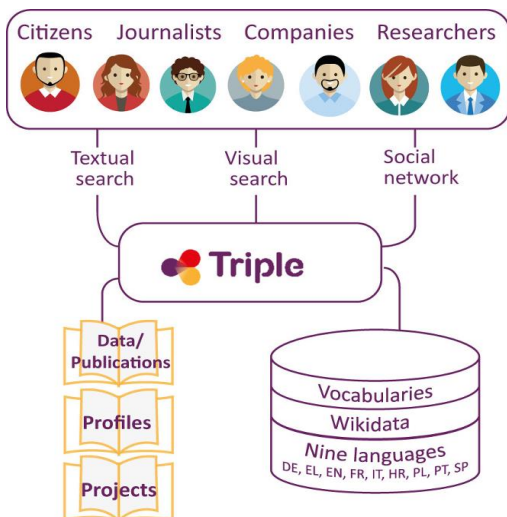
- **Infrastructures
(ESFRI)**
- **Initiatives,
réseaux
internationaux**
- **Réseaux
francophones**

*Construire du lien entre ces différents niveaux
Partager l'expertise
Elaborer des partenariats*

Un exemple: la stratégie de participation à la construction de l'EOSC



Implication de tous les ERICs en SSH / Création du 'Marketplace' pour les SSH



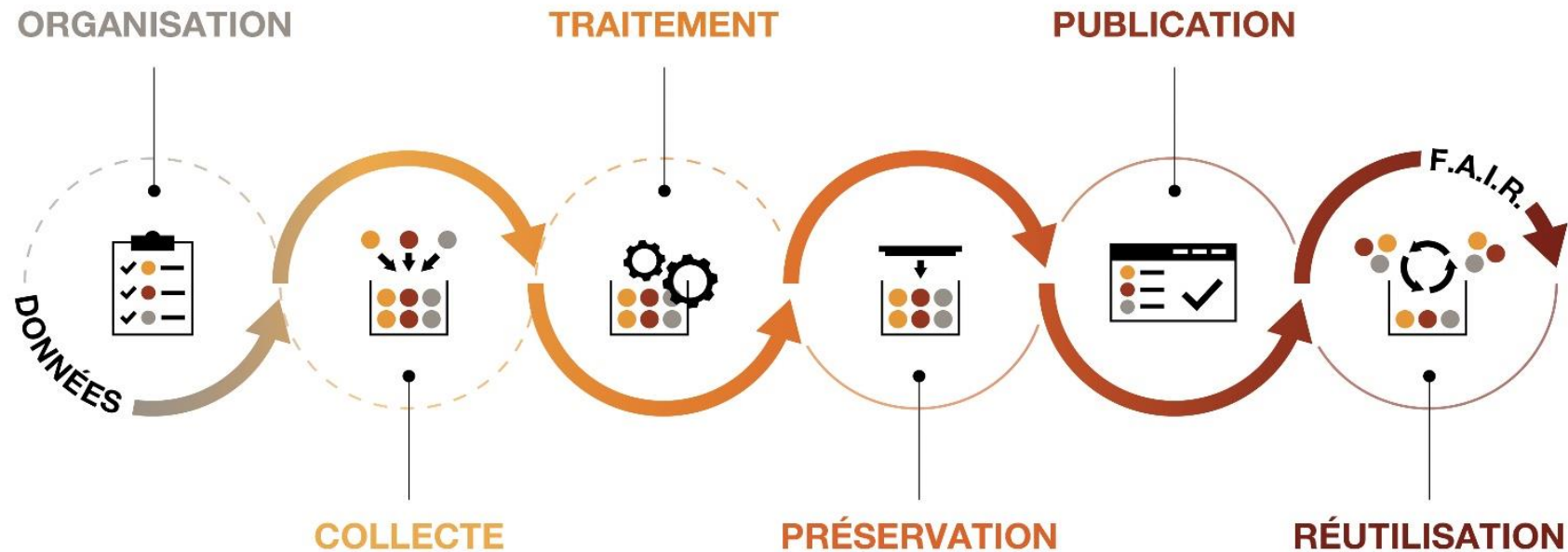
Situations type

S3

Organiser, collecter, traiter, préserver, publier, réutiliser des données :
quelques exemples

- Je dois concevoir un modèle de données, un plan de stockage, un instrument de recherche...
- Je dois collecter plusieurs To de données issues de numérisations par drones de fouilles archéologiques
- Je dois procéder à des extractions de données par requêtes dans un triple store, et je les visualise à l'aide de graphes
- Je dois préserver à long terme des jeux de données issues de mes enquêtes de terrain
- Je souhaite publier mes jeux de données pour qu'ils puissent être citables par d'autres recherches
- J'ai besoin de consulter des jeux de données et d'en réutiliser une partie pour ma propre recherche

Le cycle de vie des données en SHS



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

Facile à trouver • Accessible • Interopérable • Réutilisable

Réponses aux situations type

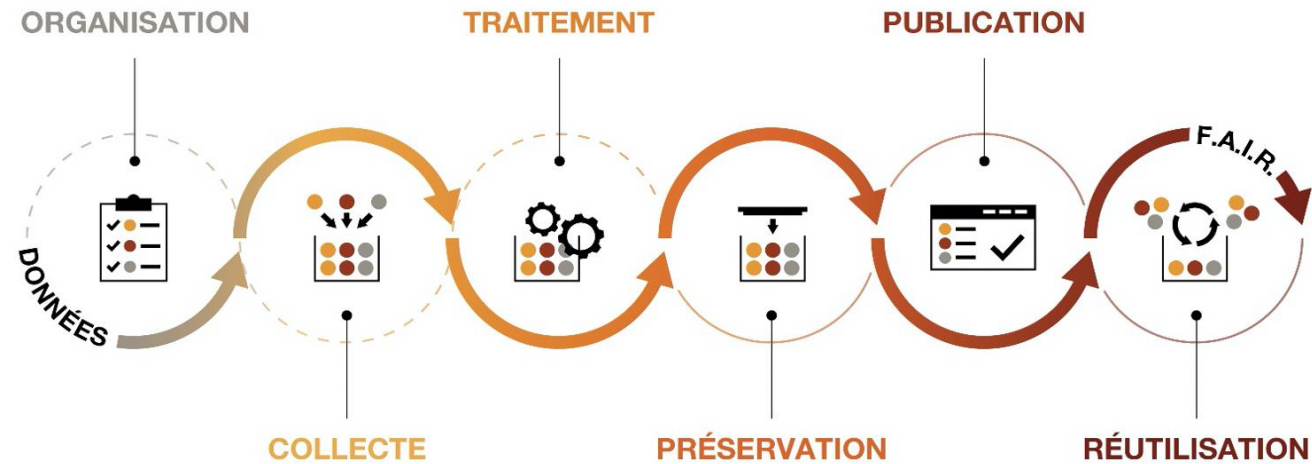
R3 Réponse d'Huma-Num

- Suivant les besoins précis, plusieurs services et outils sont proposés par Huma-Num sur sa grille ; ils sont gratuits à l'usage pour tous les acteurs de l'ESR
- Si les services d'Huma-Num ne sont pas pertinents, l'équipe peut renvoyer vers d'autres solutions ou acteurs qu'elle connaîtrait

Des services pour les données en SHS



Des services classiques ...



Dès le début du projet

- Mettre à l'abri des données, des documents (Huma-Num Box)
- Travailler collectivement sur ces données (Sharedocs)

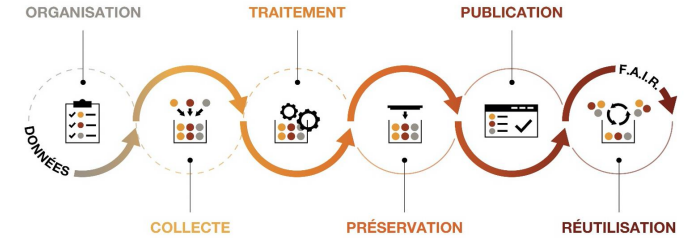
Dès le début du projet et au fil de l'eau

- Travailler les données (formats, organisation, calcul)
- Accès à des logiciels sur HN ou à distance (R Studio Pro, Server FileMaker Pro, SIG...)

A tout moment

- Diffusion rapide mais non pérenne (sites Web)

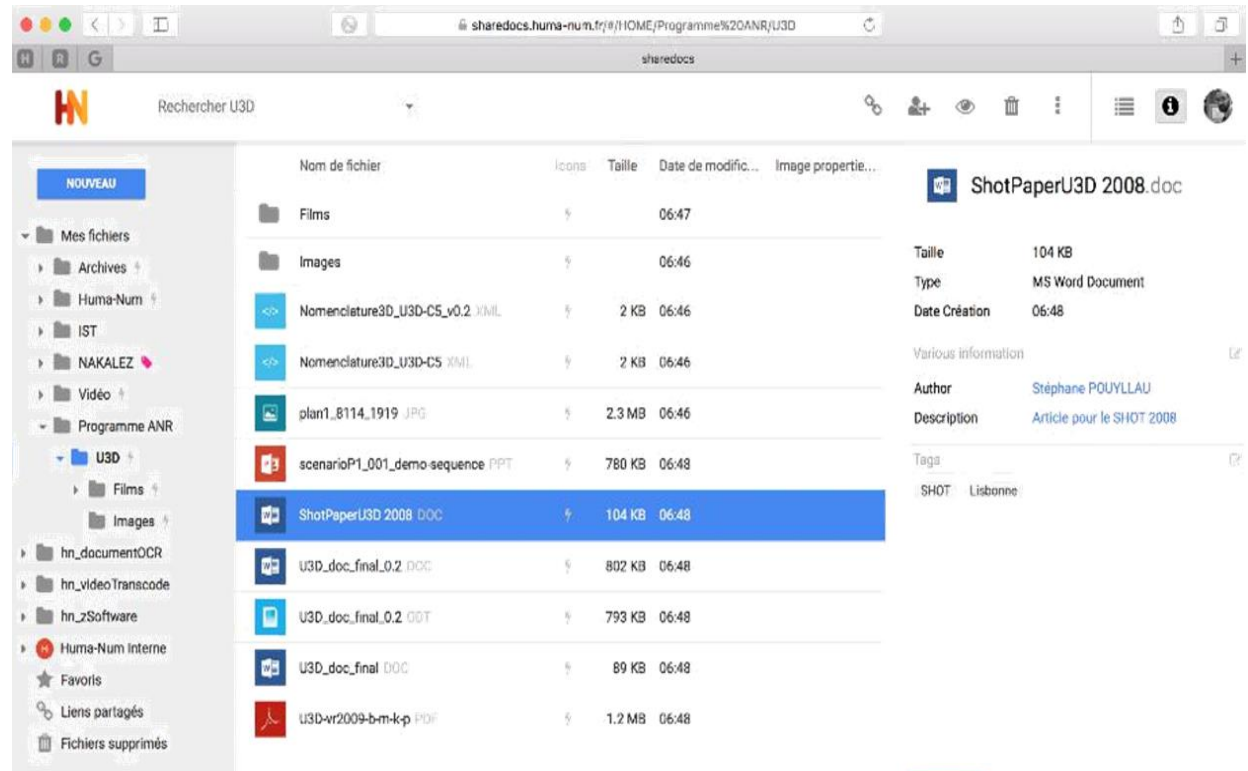
Des services classiques ...



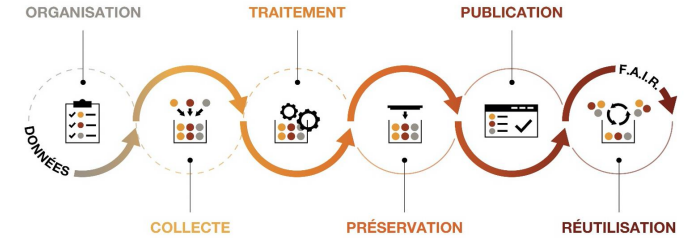
ShareDocs

stockage « outillé » pour les SHS

- 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.

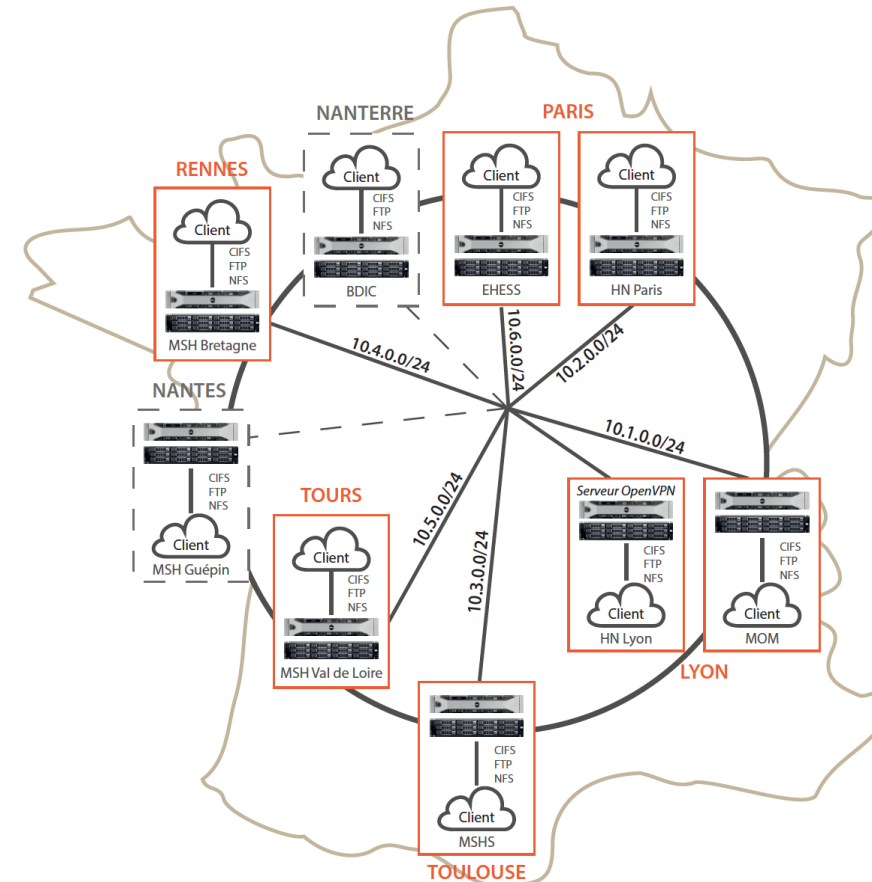


Des services classiques ...

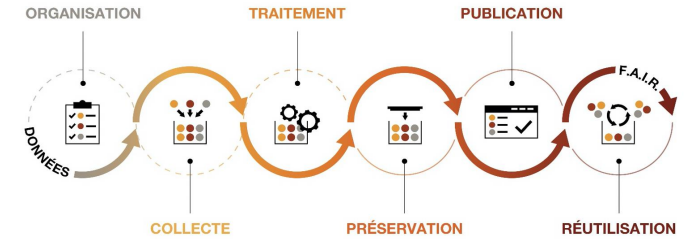


Le réseau des Huma-Num Box

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.



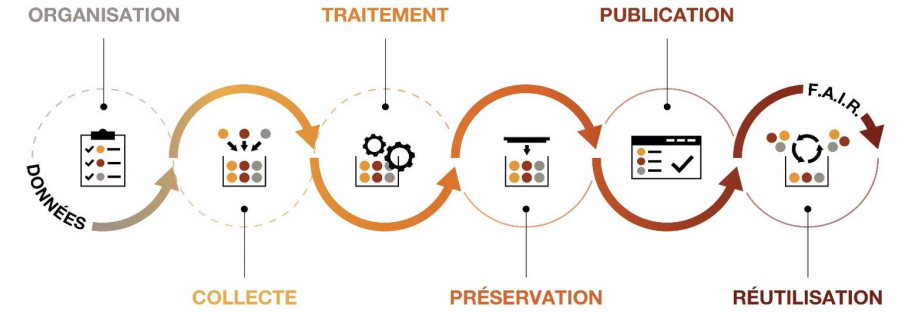
Des services classiques ...



Pour aider au traitement :

- De la puissance de calcul
- Des services sous licences temporaires
- Un serveur R avec environnement logiciel R Studio Pro
- Un serveur FileMaker Pro
- Du SIG...

...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 :

L'entrepôt



NAKALA est un service 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.

NAKALA sépare le stockage des données (fichiers + métadonnées) des différents modes d'utilisation et éditorialisation.

NAKALA assure à la fois l'accessibilité aux données et aux métadonnées ainsi que leur "citabilité" dans le temps à l'aide d'identifiants stables fournis par Huma-Num.

NAKALA est entièrement basé sur les technologies utilisées pour le Web des données permettant notamment de rendre interopérables les métadonnées, c'est-à-dire la possibilité de pouvoir les connecter à d'autres entrepôts existants suivant ainsi la logique des données ouvertes et liées (LOD) .

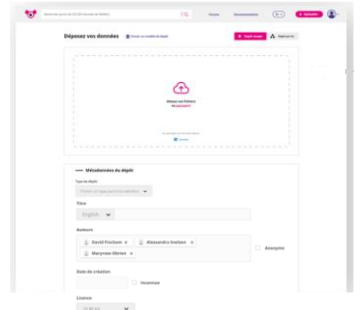
Par ailleurs, NAKALA propose également un dispositif d'exposition des métadonnées qui permet de les référencer par des moteurs de recherche spécialisés comme ISIDORE.

NAKALA s'inscrit dans un dispositif cohérent de services mis en place par Huma-Num pour faciliter l'accès, le signalement, la conservation et l'archivage à long terme des données de la recherche en SHS.

Principes clés

- Dépôt des données dans « tout » Nakala
- 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 étendable)
- 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)

nakala Dépôt



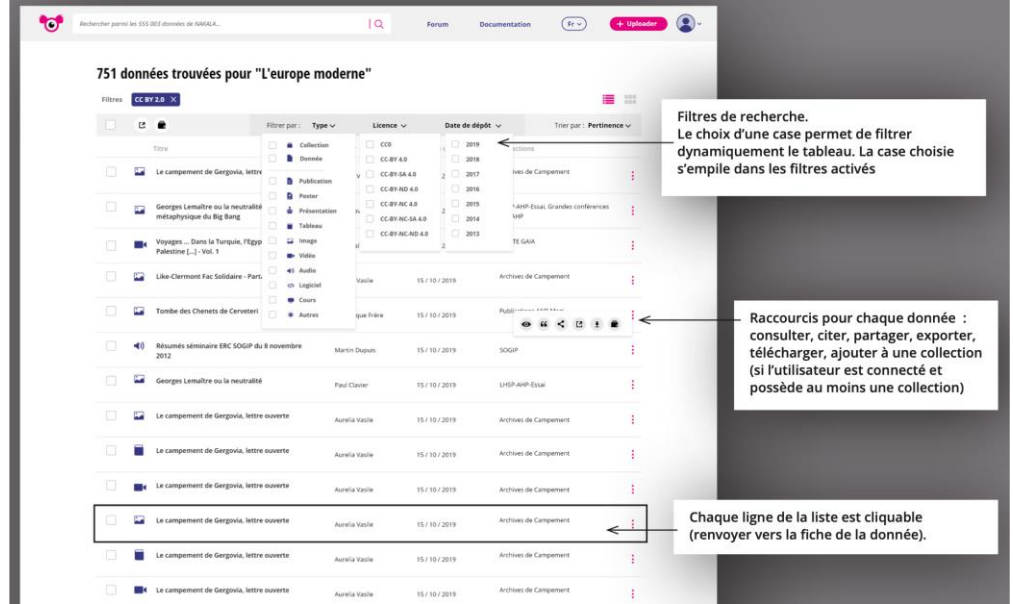
nakala Fiche Donnée



nakala Fiche Collection



nakala Consultation



751 données trouvées pour "L'Europe moderne"

CC BY 4.0

Filtres de recherche. Le choix d'une case permet de filtrer dynamiquement le tableau. La case choisie s'empile dans les filtres activés

Raccourcis pour chaque donnée : consulter, citer, partager, exporter, télécharger, ajouter à une collection (si l'utilisateur est connecté et possède au moins une collection)

Chaque ligne de la liste est cliquable (renvoyer vers la fiche de la donnée).

nakala Consultation



La barre d'action et de filtre est la même pour les deux modes d'affichage

Ce mode permet d'afficher une prévisualisation des fichiers (s'ils sont prévisualisables). Dans le cas échéant, on affiche un message du type : «Aperçu non disponible», «Fichiers sous-embargo», «Fichiers en accès restreint».

Même raccourci pour switcher l'affichage entre liste et galerie

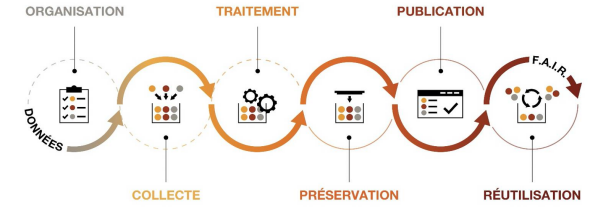
Pour les deux modes d'affichages, si une donnée possède plusieurs auteurs, on affiche un menu déroulant en hover qui liste les auteurs dans l'ordre saisi lors de la création de la donnée.

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



La version 2021 en résumé

- Chaque donnée aura un identifiant pérenne de type DOI
 - Si la donnée est composée de plusieurs fichier(s) chacun aura également un DOI
- La gestion des partages sera beaucoup plus souple (édition, lecture, publique etc.)
- Les métadonnées descriptives sont toujours du dcterms mais pourra être étendu
- Chaque donnée aura une page de présentation (landing page)
- Un moteur de consultation et de recherche sera intégré
- Un module de présentation des collections « éditorialisées » sera présent (NAKALA Press)



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

ISIDORE pour signaler, accéder et enrichir les données

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

ISIDORE permet de rechercher dans des notices et dans le texte intégral dans plusieurs millions de documents en SHS. Ces documents sont des publications (articles, ouvrage, etc.), des événements (annonces de colloques et de séminaires), des données en texte intégral issues de base de données, de corpus de documents et de fonds d'archives scientifiques.

The screenshot shows the ISIDORE search interface. At the top is the ISIDORE logo, a blue stylized 'X' with a black dot in the center, followed by the word 'isidore' in a bold, lowercase sans-serif font. Below the logo is the tagline: 'Votre assistant de recherche en Sciences Humaines et Sociales'. A search bar is positioned below the tagline, containing a dropdown menu with 'Documents' and a search input field with the text 'Rechercher dans les 6 220 016 documents de ISIDORE...'. To the right of the search bar is a magnifying glass icon. Below the search bar is a link labeled 'Recherche avancée'.

+ de 6 millions de données signalées et reliées entre elles avec des référentiels scientifiques

+ d'1,3 million d'utilisateurs/an

+ de 7000 « sources » de données collectées dans le monde

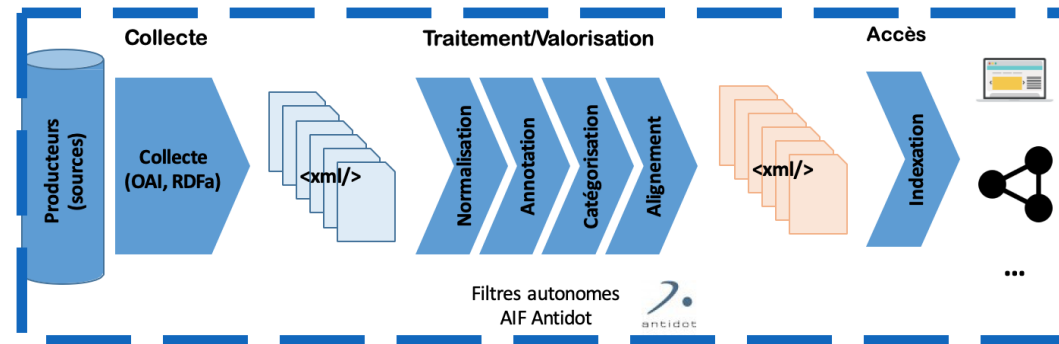
Technologies du web sémantique



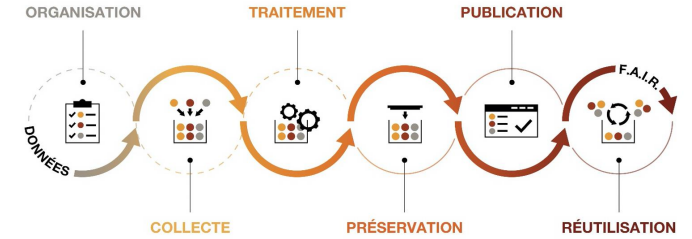
Collecter
Traiter

Enrichir
Classer
Catégoriser

Accéder



Préserver les données sur le long terme



Fragilité des objets numériques

Instabilité des formats

Description & Métadonnées

Supports abîmés

Obsolescence des applications

Obsolescence technique

Huma-Num accompagne les projets de préservation à long terme

Liens entre les producteurs de données et le CINES

Suggestion de nouveaux formats et prise en compte (e.g. formats pour la 3D)

Liens avec la communautés des archivistes

PoC Vitam



Une infrastructure pour la préservation, le calcul et la sécurité



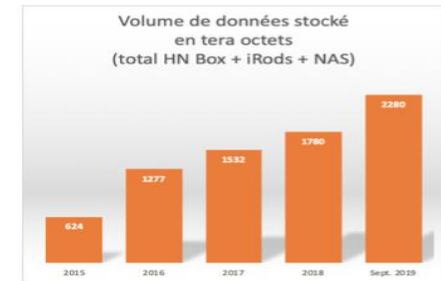
Huma-Num Box
Un stockage distribué
pour les données « tièdes »



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 :



Projet sur la sécurité des données sensibles

Situations type

S4 Se former à certaines notions et certains services ou outils : *quelques exemples*

- J'ai besoin de mieux comprendre les notions de « FAIRisation des données » incluses dans la dynamique de la Science Ouverte
- Je cherche une vidéo de présentation générale de certains outils souvent cités par les collègues comme Nakala, Isidore, Huma-Num Box
- J'ai besoin d'une formation de prise en main sur ces outils pour juger de leur pertinence au vu de mon projet et de mes compétences
- J'ai besoin d'une formation avancée pour tirer le meilleur parti de ces outils dans le cadre de mon projet

Réponses aux situations type

R4 Réponse d'Huma-Num

- Je m'adresse à l'acteur-relais de la zone géographique ou disciplinaire où je me trouve pour voir s'il propose des formations adaptées à mon besoins, et en particulier à la culture et aux services / outils d'Huma-Num
- Sinon je m'adresse directement à Huma-Num pour profiter de l'offre de formations avancée dédiée aux correspondant.e.s Huma-Num (*en cours de production*)

Situations type

S5 De faire évoluer des services existants pour mieux les adapter à mes besoins : *quelques exemples*

- Je constate que le modèle de description d'un jeu de données dans Nakala ne me permet pas de fournir des métadonnées critiques pour moi
- Pour organiser des formations techniques ponctuelles j'aurais besoin de plusieurs licences Oxygen sur une courte période mais en nombre suffisant
- J'aurais besoin d'interface AtoM avec Nakala pour visualiser et manipuler des instruments de recherche associés à des jeux de données
- J'aimerais faire développer de nouvelles IHM dans Heurist pour simplifier une partie de la gestion d'une base de données

Réponses aux situations type

R5 Réponse d'Huma-Num

- Suivant les services ou les outils concernés par la proposition, l'équipe cherchera à évaluer si le besoin s'avère suffisamment partagé au niveau national, ainsi que les moyens à réunir pour faire évoluer la grille
- Si le besoin est confirmé et les moyens réunis, l'équipe travaillera en mode projet avec la communauté qui aura fait la proposition

Situations type

S6 Partager des réflexions et développements autour d'objets de prospective : *quelques exemples*

- J'ai besoin de cadrer juridiquement l'utilisation d'une plateforme web de gestion de données de santé issues d'enquêtes sociologiques
- Je souhaite éprouver le modèle des datapapers suivant les domaines scientifiques mobilisés dans les publications et jeux de données
- Je cherche à vérifier la pertinence des réseaux de neurones (AI) dans l'apprentissage profond en sciences des textes
- Je m'interroge sur le développement des épi-revues et leur articulation avec les jeux de données bien exposés

Réponses aux situations type

R6 Réponse d'Huma-Num

- Suivant l'objet prospectif du projet proposé, l'HNLab pourra devenir partenaire et aider à son développement technique et à sa pertinence scientifique
- L'HNLab pourra partager ses résultats de veille sur l'objet prospectif qui intéresse un porteur de projet ou une communauté
- L'HNLab pourra accueillir des chercheurs et des ingénieurs en résidence

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 (Projet en collaboration avec le GED et avec les membres de l'EPCC Condorcet)
- Labex PP



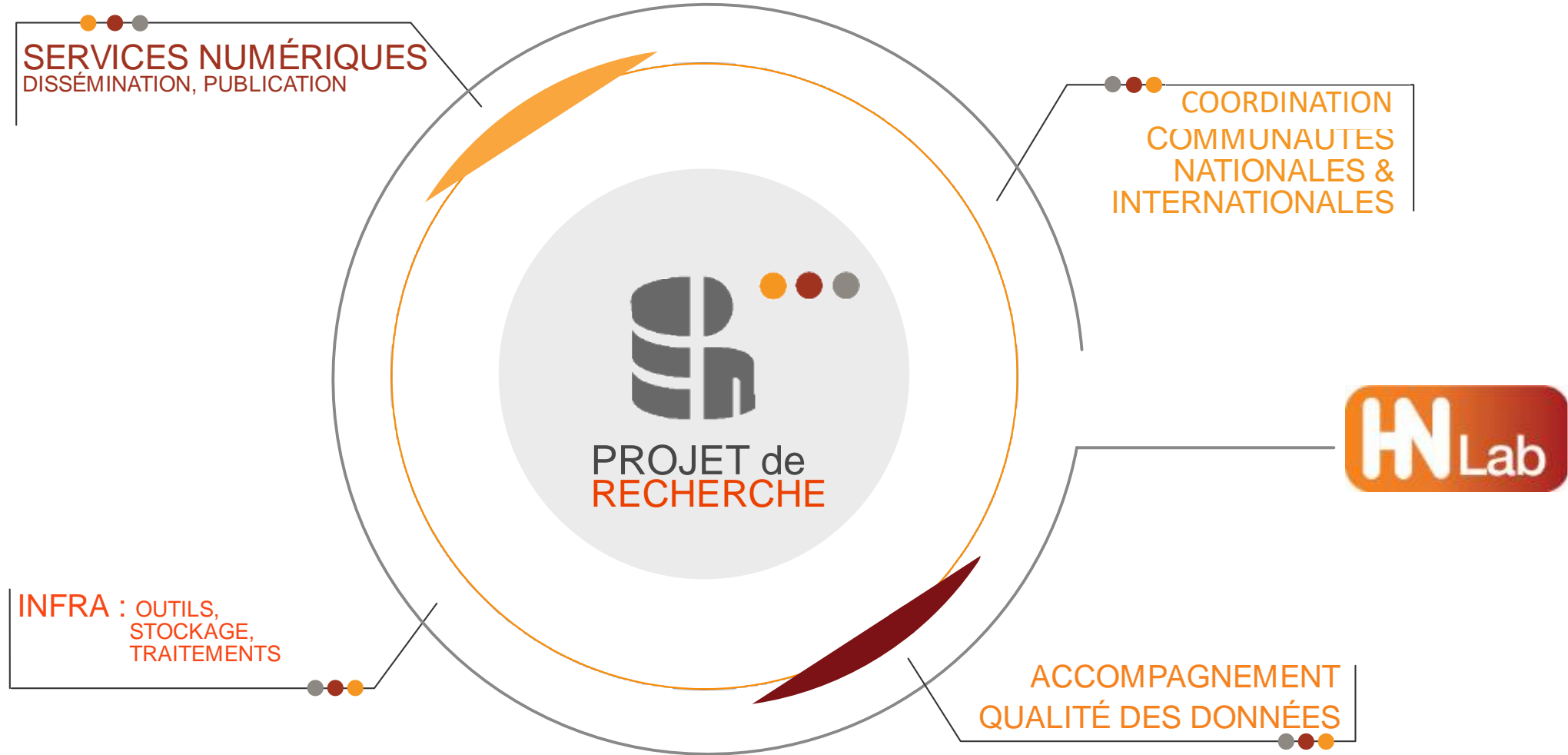
Organisation d'Huma-Num et modalités d'accès

Problématique générale

Structurer les usages à l'écoute des pratiques

- Structurer les usages par l'**offre de services**
 - en vue d'aider à la FAIRisation des données et au développement de la Science ouverte
 - avec pour objectif d'augmenter l'appropriation des services d'Huma-Num par les communautés
- Écouter les pratiques à l'aide de "**lieux-relais**"
 - dans la diversité des "zones" disciplinaires et géographiques existantes, avec comme idéal un maillage le plus fin possible
 - avoir une présence attentive où la recherche se fait afin de mieux rencontrer les utilisateurs des services

Huma-Num



Les canaux et bases de connaissances

- Sur le web

- Le site web : www.huma-num.fr
- Le carnet de recherche : <https://humanum.hypotheses.org/>
- Le site de documentation : <https://documentation.huma-num.fr/>

- Listes et réseaux sociaux

- Liste de diffusion : humanum-diffusion@listes.huma-num.fr
- Twitter : [@Huma Num](https://twitter.com/Huma_Num)

Les points de contact

- Via les acteurs-relais
 - Contacter la MSH la plus proche
 - Annuaire public des correspondant.e.s HN (<https://www.huma-num.fr/trouver-son-relais-msh/>)
- Contact direct
 - Une adresse mail : cogrid@huma-num.fr
 - Soit y formuler une demande précise
 - Soit y demander de l'aide pour préciser votre besoin
 - Pôle qualité des données et accompagnement des demandes