



La Science Ouverte : une révolution nécessaire

8 oct. 2019 Paris (France)



La TGIR Huma-Num, une infrastructure pour les sciences humaines et sociales à l'ère de la science ouverte

Olivier Baude
www.huma-num.fr

La recherche en SHS

Ouvrages

Revue

Bibliographies

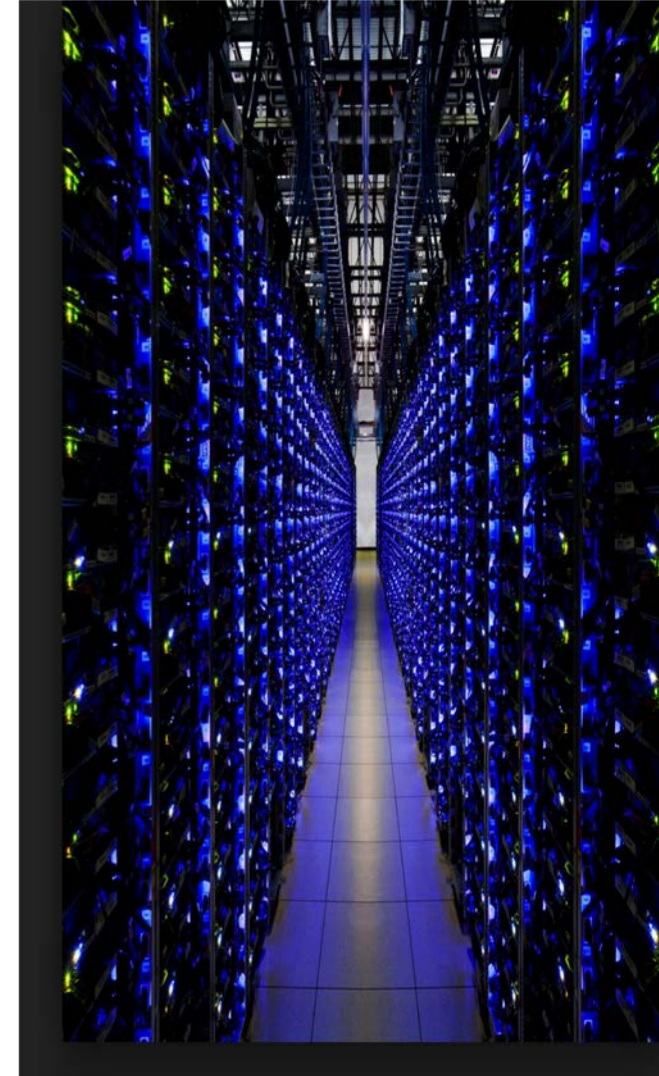
Fonds d'archives

Enquêtes

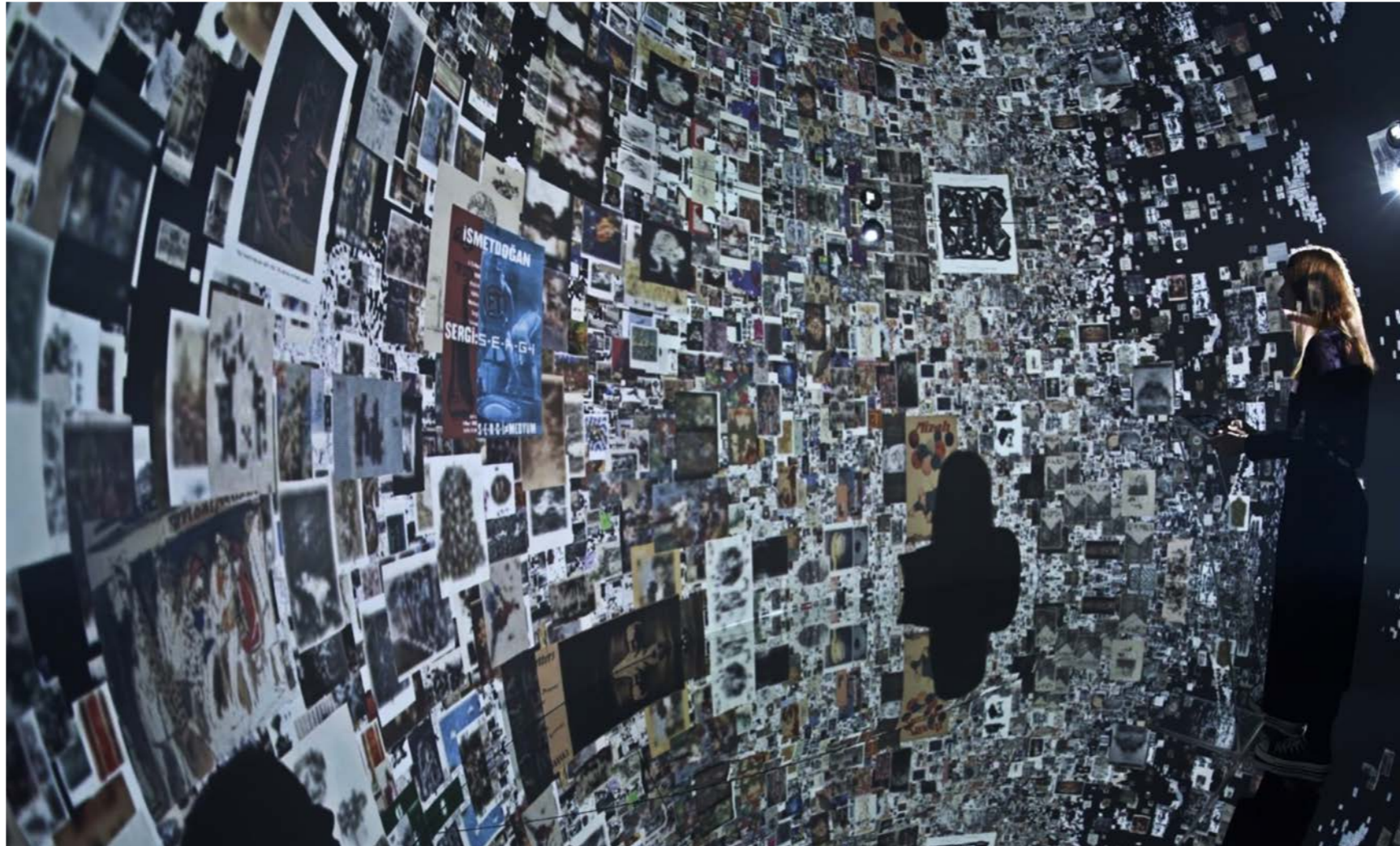
Collections de documents...



**Roberto Busa
(1913 - 2011) :**
Index thomisticus,
lemmatisation
des œuvres de Thomas
d'Aquin sur IBM



Les SHS numériques

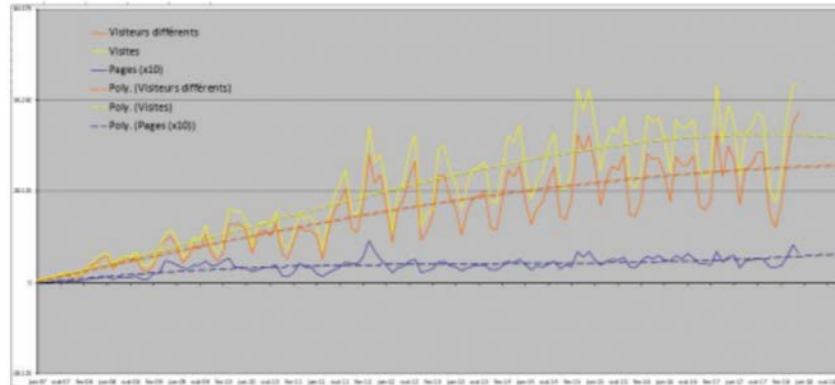


1. Donner accès à de grands corpus d'œuvres, archives multimédias hypertextuelles, outiller l'exploration

« Le caractère ouvert et non clos de l'archive numérique bouleverse les pratiques de monographie et d'éditions savantes, elles mêmes transformées par les mediums » [McGann, 2008]

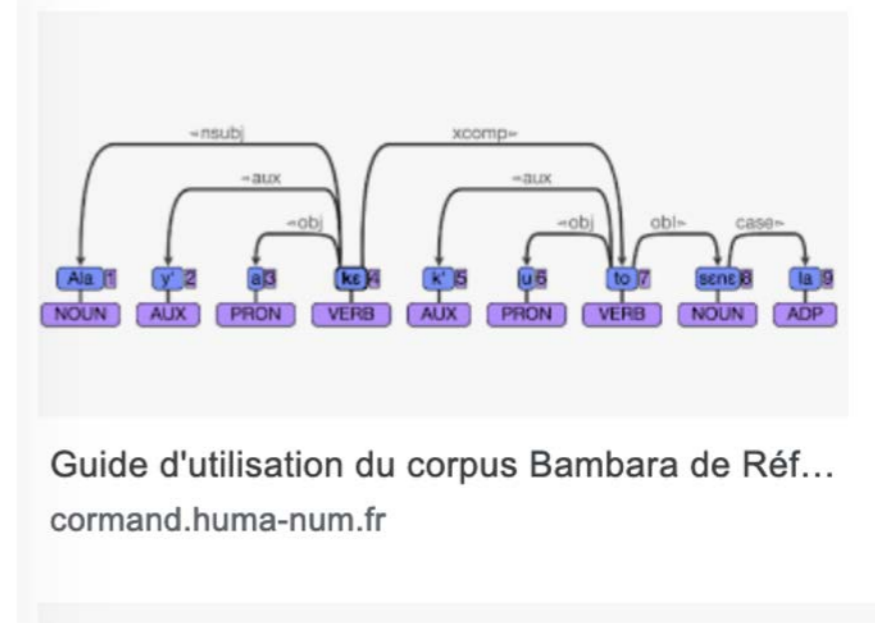
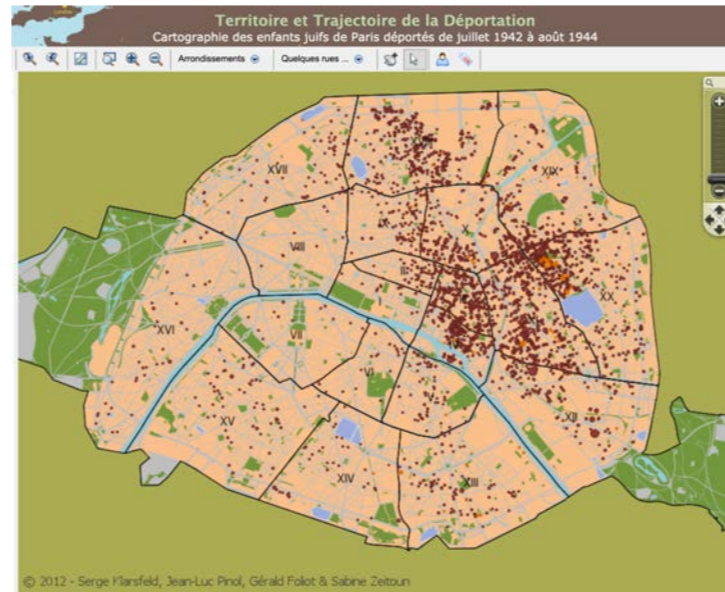
Les SHS numériques

Figure 1 Nombre de visites, de visiteurs différents et de pages vues depuis la création de *Confins*

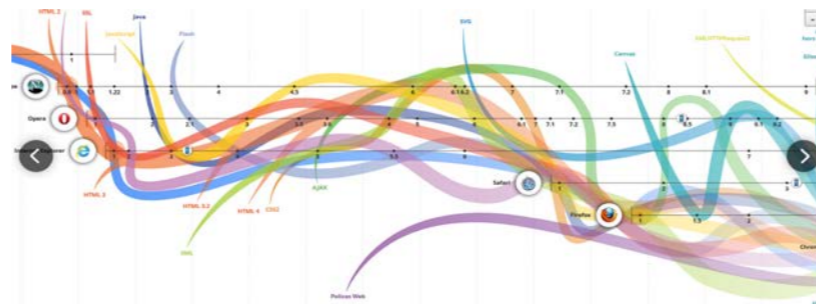


Aumentar Original (, 48k) ↓

Source: Statistiques de: [journals.openedition.org/Confins](https://logs.openedition.org/awstats.pl?config=confins), <https://logs.openedition.org/awstats.pl?config=confins>



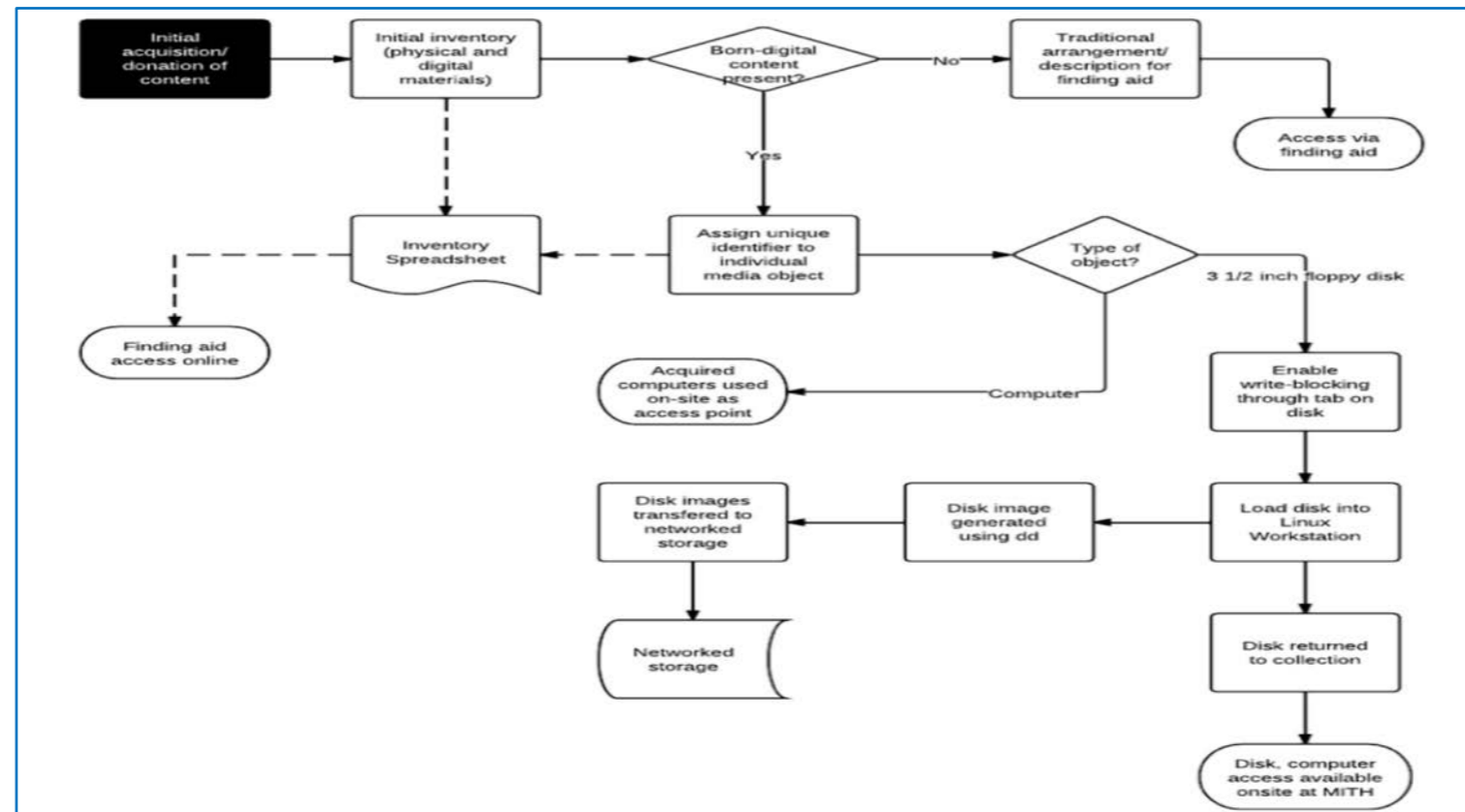
Guide d'utilisation du corpus Bambara de Réf...
cormand.huma-num.fr



2. Visualiser les données, Graphs, Maps, Trees...

« Je dirais simplement que nous avons beaucoup à apprendre des méthodes des sciences sociales et des sciences naturelles » [Moretti, 2003]

Les SHS numériques



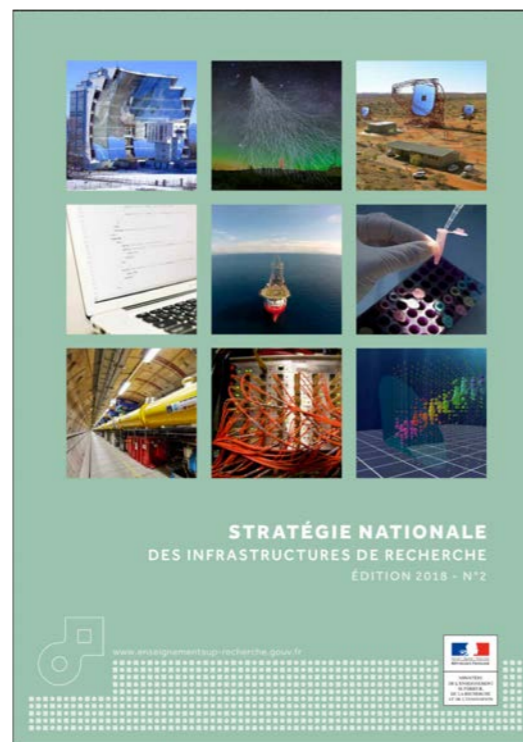
3. Formaliser la matériau (le corpus en data) mais aussi formaliser le raisonnement pour traiter ces données

« L'ordinateur est un télescope pour l'esprit. C'est l'instrument qui permet d'observer d'une autre manière et plus loin ». [McCarty, 2005]

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 le contexte de la science ouverte (dans la logique des *FAIR principes* : *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'EOSC.





CATÉGORIE	NOM	NOM COMPLET	ESFRI
TGIR	Huma-Num ¹	Humanités Numériques	DARIAH (2006) CLARIN (2006)
TGIR	Progedo	PROduction et GEstion de DONnées	ESS (2006) CESSDA (2006) SHARE (2006)
IR	ERIHS-FR ²	European Research Infrastructure for Heritage Science	ERIHS (2016)
IR	MÉTOPES ³	Méthodes et outils pour l'édition structurée	
IR	OpenEdition ⁴	Edition électronique ouverte en Sciences humaines et sociales	
IR	RnMSH ⁵	Réseau national Maison des Sciences de l'Homme	
IR	CollEx-Persée	Collections d'excellence pour la Recherche – Persée	
IR	HAL, CCSD	Archive ouverte HAL (Hyper Article en Ligne), Centre pour la Communication Scientifique Directe	



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

Un écosystème centré sur les projets et données de la recherche...

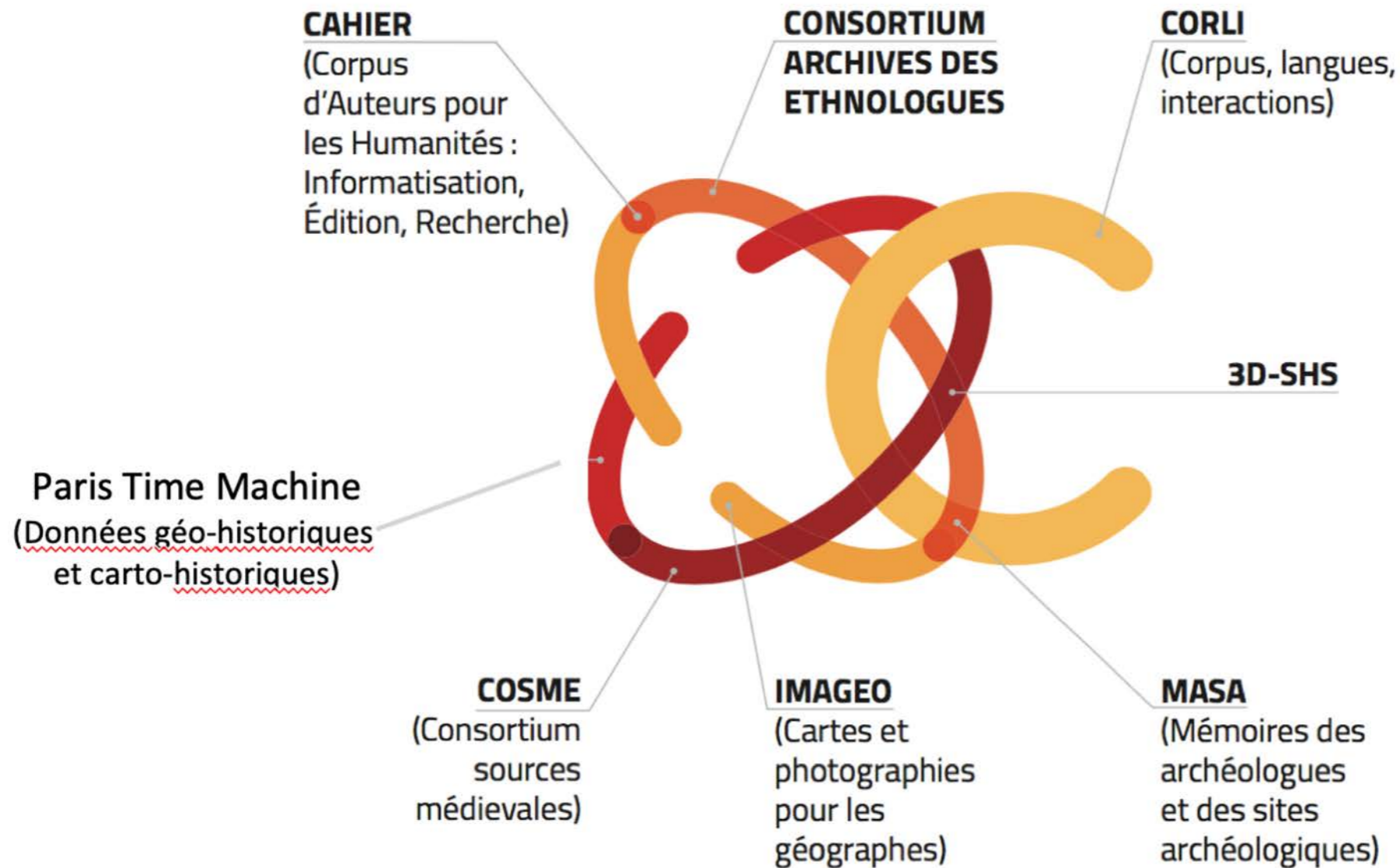


pour la



ISATION

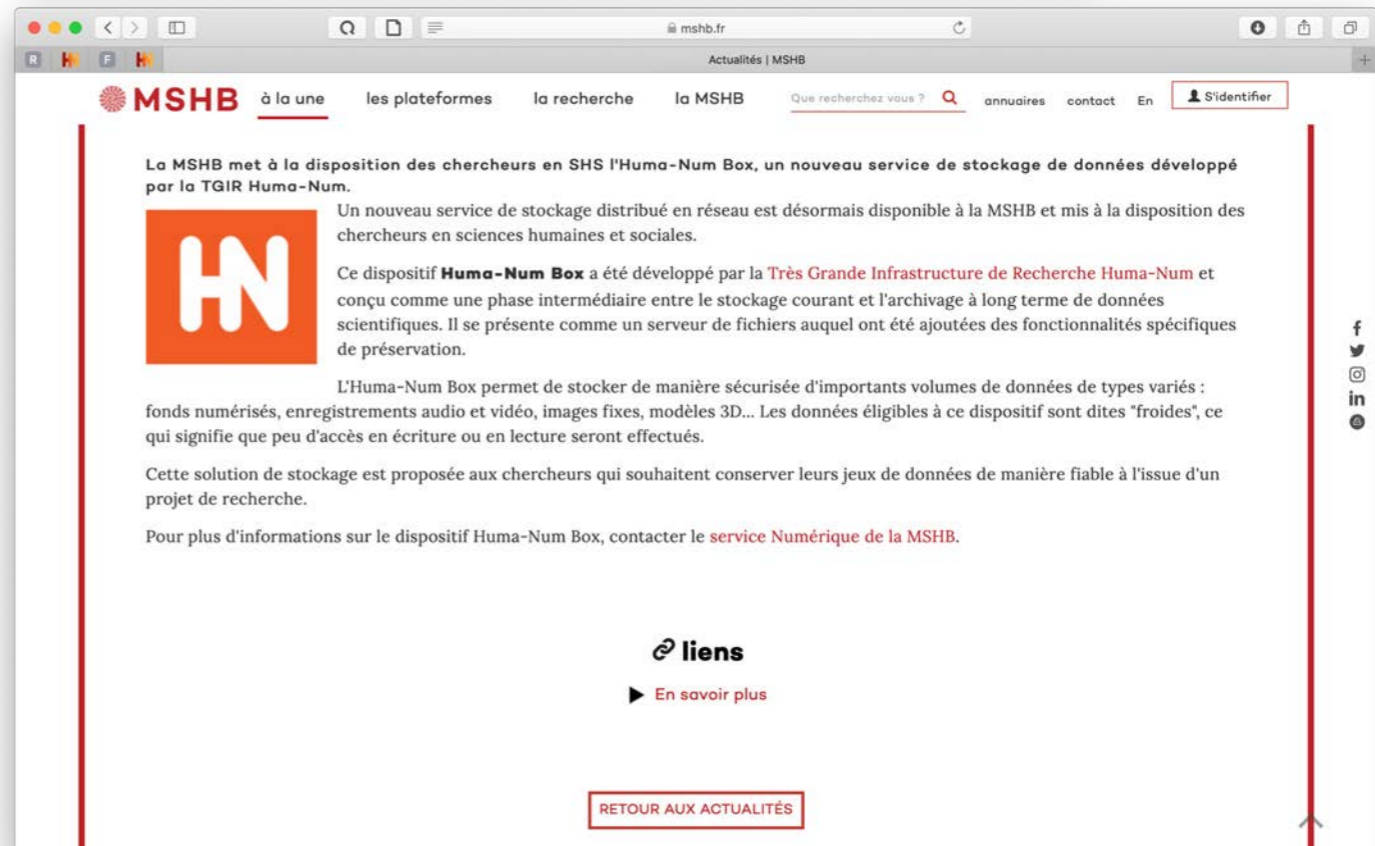
Consortiums



- + Environ 80 équipes de recherche en France
- + de 40 actions (formations, ateliers, séminaires) en 2019
- Labellisation pour 4 ans par le CP d'Huma-Num, évaluation par le CS tous les ans
- 45-50% du budget de la TGIR



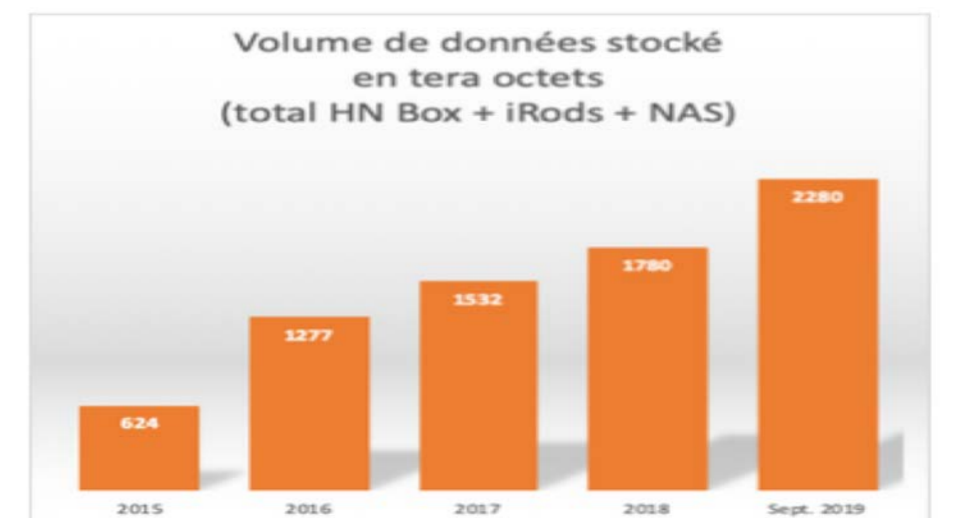
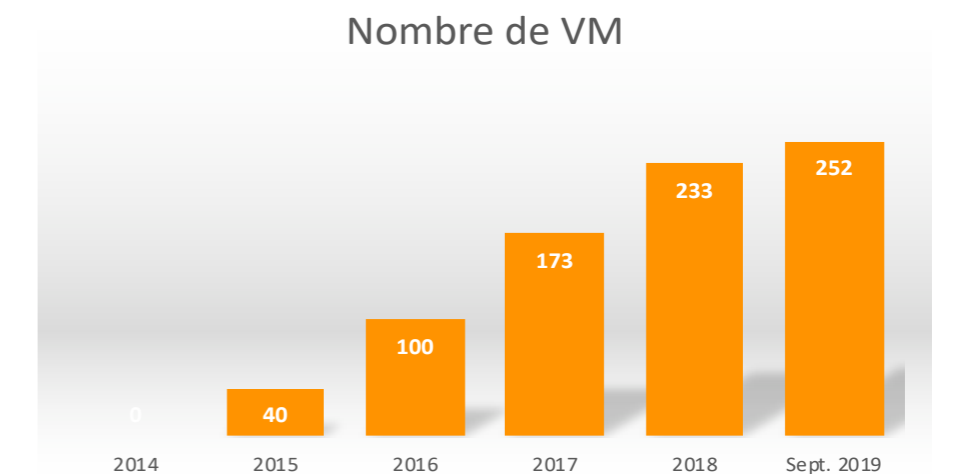
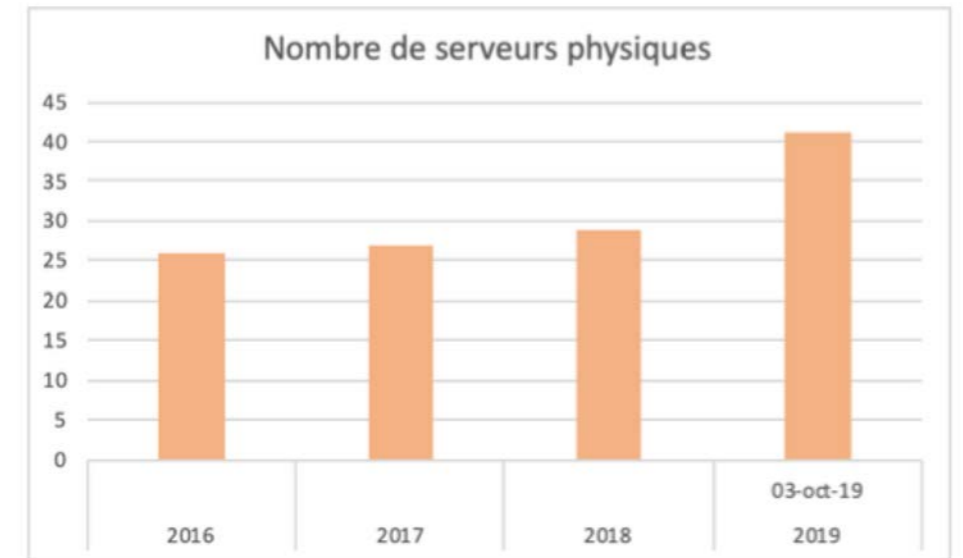
- Des relais dans les maisons des sciences de l'Homme
- Un ancrage des consortiums dans les MSH
- + de 600 projets de recherche en France utilisent Huma-Num
- Un double réseau :
 - Disciplinaire, les consortiums d'HN
 - Géographique, les relais dans les maisons des sciences de l'Homme (MSH)
- Des collaborations



Une infrastructure



- Paris, Lyon
- Deux partenaires historiques :
Centre de calcul de l'IN2P3-CNRS et le CINES
- Budget : 2,3 M€/an
- Une UMS support composée de 15 personnes



SERVICES POUR LES DONNÉES NUMÉRIQUES

HN



STOCKER

Entreposer . Organiser



TRAITER

Outils . Logiciels



DIFFUSER

Machines virtuelles
Diffusion web



DONNÉES DE LA RECHERCHE

Partenariat avec le CC-IN2P3
et le CINES

ARCHIVER

Préservation à long terme



SIGNALER

Enrichissement sémantique
Accès unifié



isidore

EXPOSER

Documenter . Partager



nakala

nakal(©)na

F
indable



A
ccessible



I
nteroperable



R
eusable



Des services de base...

 **STOCKER**

 **TRAITER**

 **DIFFUSER**

... aux services à valeur ajoutée

 **EXPOSER**

 **SIGNALER**

 **ARCHIVER**

F_{indable}


A_{ccessible}


I_{nteroperable}


R_{eusable}




- Stockage et pérennisation des données



- Plusieurs points de présence assurés en France permettant de rapprocher la données du traitement (documentaire, transformation, etc.)
- Un stockage illimité
- Un cout assuré par Huma-Num
- 400 To de données

Huma-Num est une très grande infrastructure de recherche (TGR) visant à faciliter le tournant numérique de la recherche en Sciences Humaines et Sociales. Elle propose une grille de services numériques à destination de l'ensemble de la communauté SHS.

HUMA-NUM BOX : STOCKAGE DISTRIBUÉ SÉCURISÉ POUR LES SCIENCES HUMAINES ET SOCIALES

Objectif : 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. Service simple à utiliser par des chercheurs en SHS par présentation d'un service NAS sur le réseau local.

Données FROIDES ou TIEDES

- peu d'accès par unité de temps,
- pas d'accès simultanés

• issues de campagnes de numérisation de fonds anciens, photos, enregistrements audio, cartes, vidéos, voire modèles 3D
 • fichiers, souvent de grande taille, parfois accompagnés de fichiers de métadonnées
 • documents à valeur patrimoniale

Solution technique
 Ensemble de nœuds au matériel banalisé (serveurs R730 + baies MD1400 avec disques SATA) et avec le logiciel commercial Active-Circle assurant la virtualisation du stockage matériel, les fonctions de stockage, réplication, historisation et archivage des données, ainsi que les services NAS (CIFS, NFS, FTP).

Intégration

- interconnexion des nœuds via OpenVPN, pour communication de chaque nœud avec tous les autres nœuds,
- annuaire LDAP avec gestion déléguée des utilisateurs et des groupes par site via le logiciel libre FusionDirectory,
- travail important par des correspondants techniques et fonctionnels au sein des Maisons des Sciences de l'Homme (MSH) partenaires, en direction des chercheurs.

À date :

- 7 nœuds
- 50 partages, dont 30 actifs
- 200 To de données

Partenaires :

- EHES (Paris)
- MDM (Lyon)
- MSHB (Rennes)
- MSHST (Toulouse)
- MSHVDL (Tours)

Développements

- 2 autres nœuds à venir
- passage en cours sur service L3VPN de RENATER
- ajout d'un stockage sur bandes magnétiques LTO-7 (apport de sécurité et économie financière)
- développement des usages (service de préarchivage avant le CINES, vérification des checksums, applications Web au-dessus des données)

jres
Nantes 2017

- Stockage « outillé » pour les SHS



Stockage des données documentées



- Identifiants handle
- Préparation archivage (CINES)
- Vocabulaire des métadonnées : DCTerms
- Vocabulaire des métadonnées : DCTerms
- Interopérabilité OAI-PMH
- Conversion RDF / triple-store (réservoir)
- Outil de chargement des lots
- API (pour les webmasters)
- Fédération d'identité / LDAP





Exposer et diffuser des données



Screenshots of the nakalona website interface:

- Left screenshot:** Shows the 'Rumble in the Jungle À Kinshasa' resource page with detailed metadata (dc:creator, dc:subject, etc.) and a video player.
- Right screenshot:** Shows the 'isidore' search results page for the same resource, displaying a video thumbnail and metadata.

films-sciences-sociales.nakalona.fr



Votre assistant de recherche
en Sciences Humaines et Sociales

Documents ▾

Rechercher dans les 6 220 016 documents de ISIDORE...



[Recherche avancée](#)

- isidore.science : un moteur de recherche, un réseau social de scientifiques SHS, un outil de veille scientifique et de diffusion dans le cadre de la science ouverte
- + 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

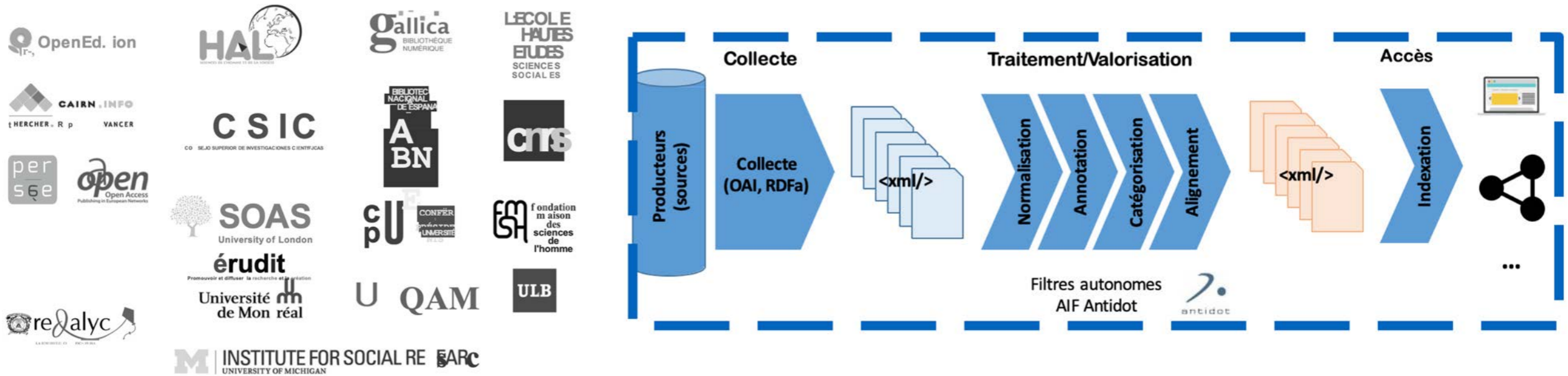


Technologies du web sémantique

Collecte
Traiter

Enrichir
Classer
Catégoriser

Accéder



Accéder

pour les humains

pour les machines

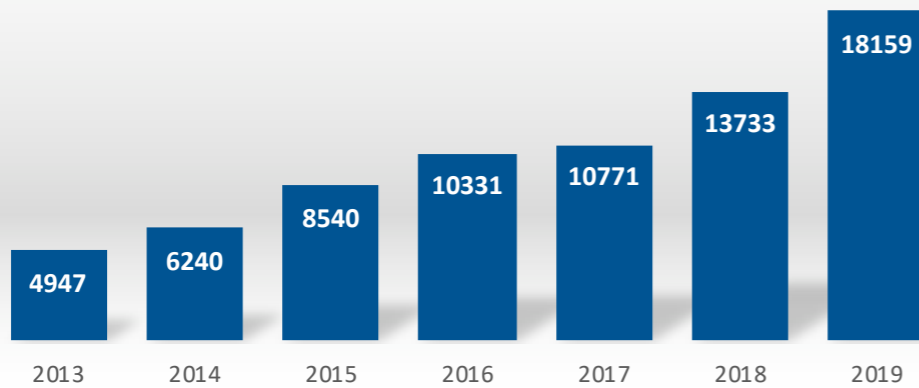


ARCHIVER

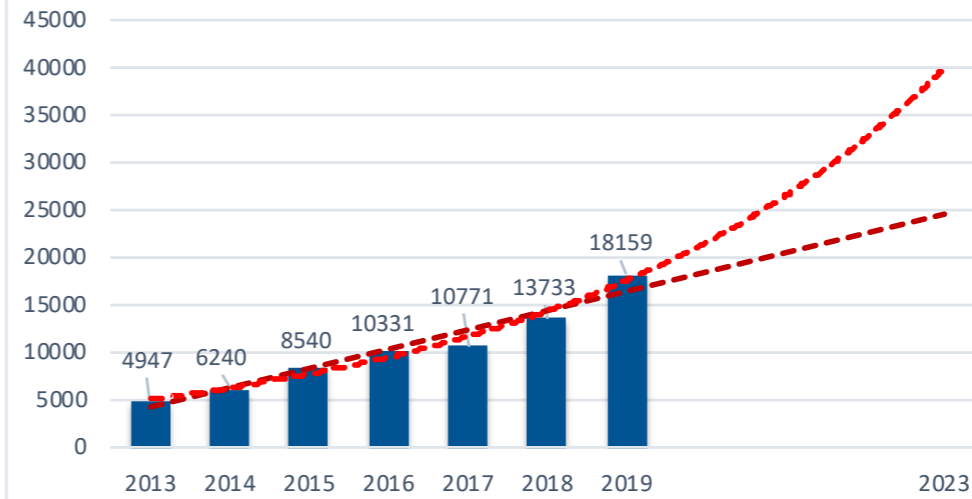
DIFFUSER



Nb de paquets archivés au CINES (cumul)



Nb de paquets archivés au CINES (cumul)



La Science Ouverte en marche au niveau international



ERIC et projets H2020 :



**EUROPEAN OPEN
SCIENCE CLOUD**

TRIPLE

EOSC-PILLAR

SSHOC

PARTHENOS

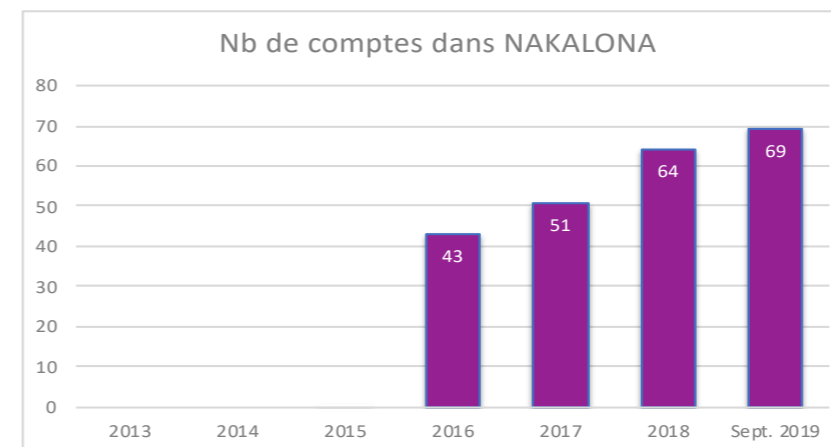
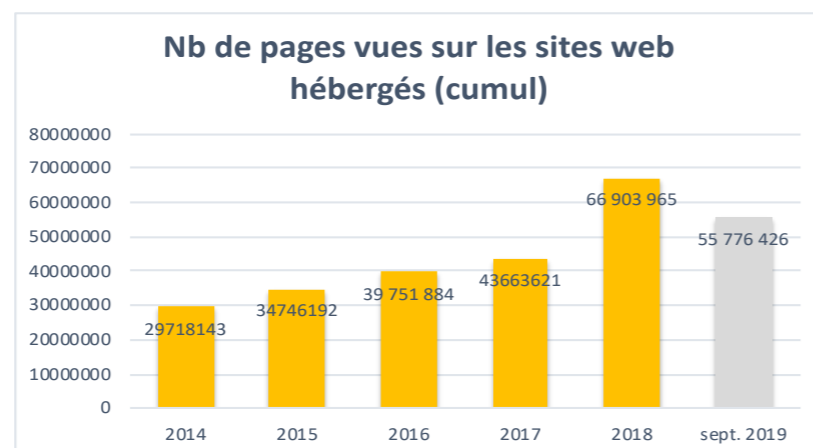
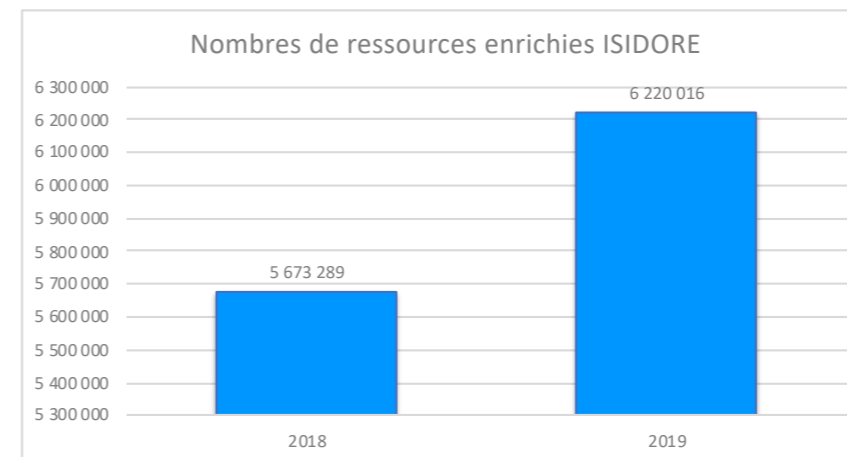
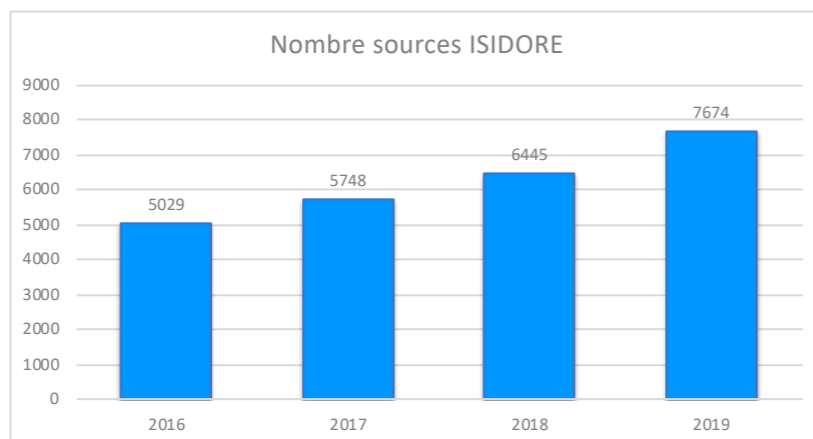
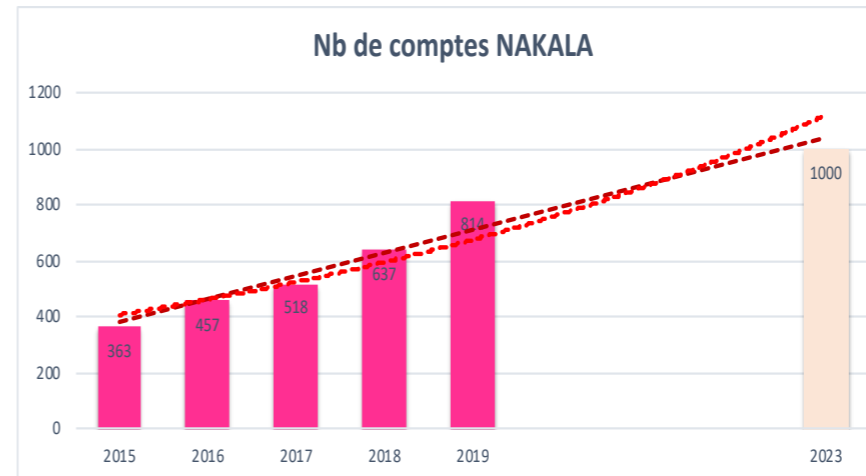
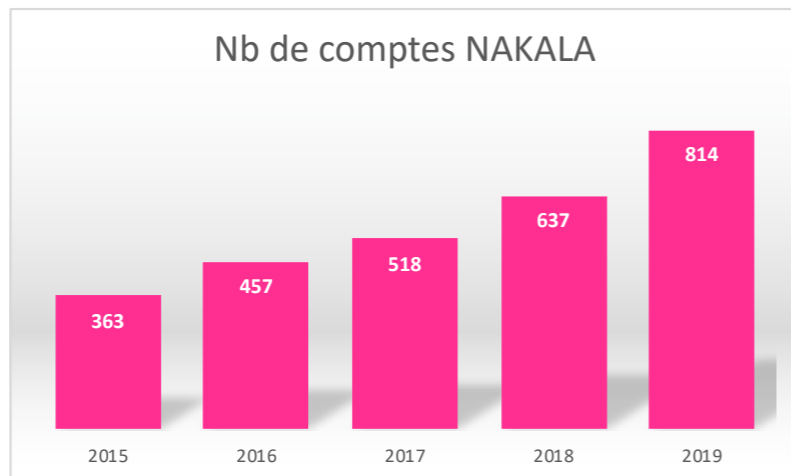


**ARIADNE
PLUS**

OPERAS-P

HAS

La Science Ouverte en marche



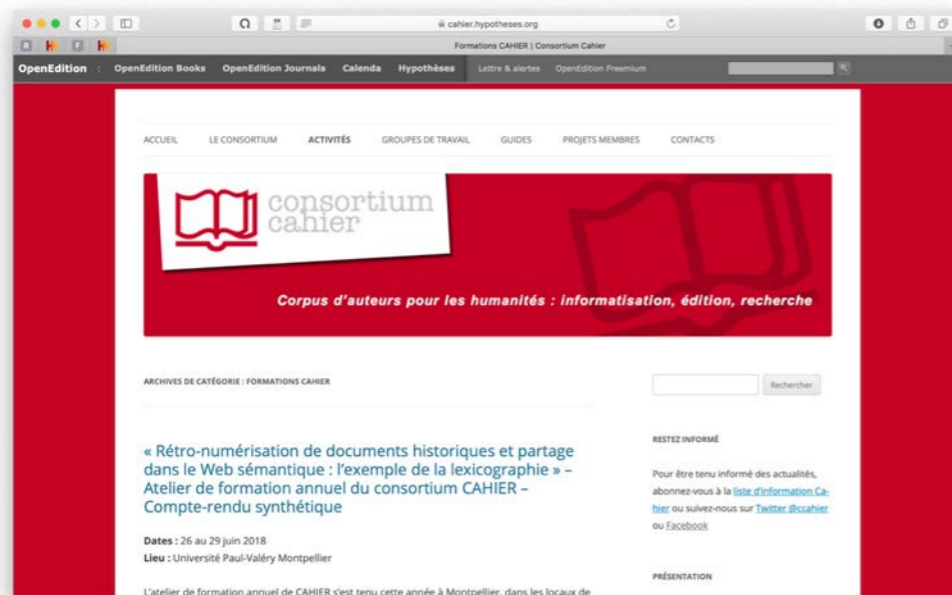
La Science Ouverte en marche



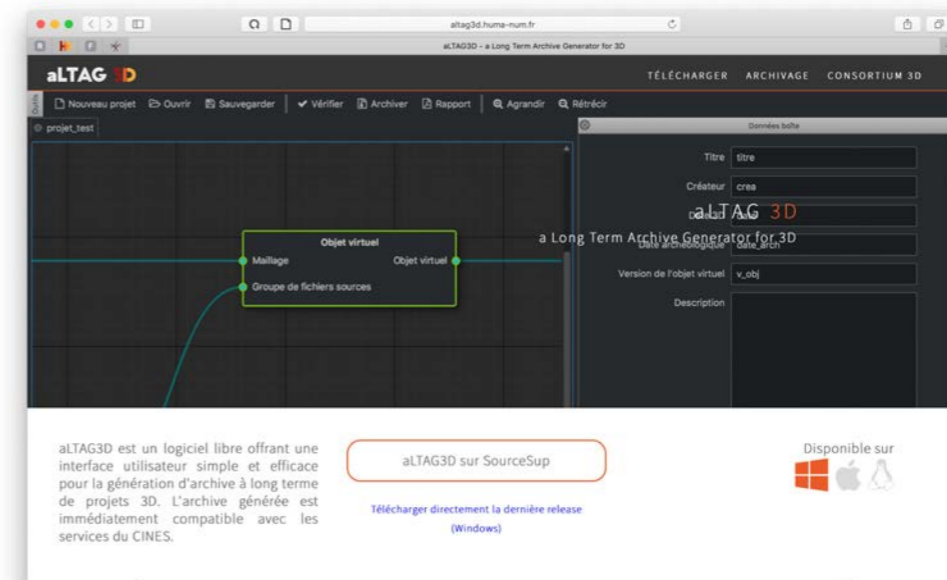
Guide juridique et éthique - consortiums, MSH, communautés...
[voir ethiquedroit.hypotheses.org/1937](http://ethiquedroit.hypotheses.org/1937)



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





SERVICES POUR LES DONNÉES NUMÉRIQUES

STOCKER
Entreposer . Organiser

ARCHIVER
Préservation à long terme

TRAITER
Outils . Logiciels

SIGNALER
Enrichissement sémantique
Accès unifié
isidore

DIFFUSER
Machines virtuelles
Diffusion web

EXPOSER
Documenter . Partager
nakala
nakal@na



DONNÉES
DE LA RECHERCHE

Partenariat avec le CC-IN2P3
et le CINES



- www.huma-num.fr
- humanum.hypotheses.org
- twitter @huma_num
- contact@huma-num.fr

