Huma-Num
la TGIR des humanités numériques

Les technologies du Web-Semantique

au service de Sciences Humaines

Juin 2015

Nicolas Larrousse

Les données de la recherche en SHS

Une frénésie de productions numériques …
et d'utilisations

Numérisation

Accès

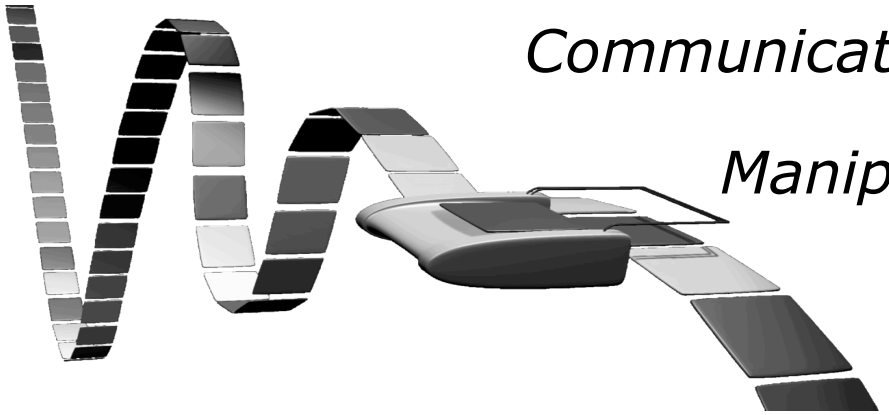
Natif

Communication

Simulation

Manipulation

Analyse



Duplication

Huma-Num

Les données de la recherche en SHS

Une typologie foisonnante

Textes

Images

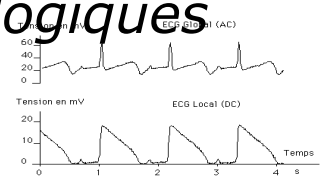
Vidéos

Sons

3D

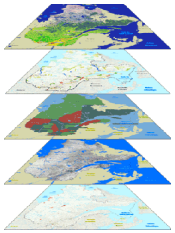
Cartes

Enregistrements physiologiques



Transcriptions

SMS



Géomatique

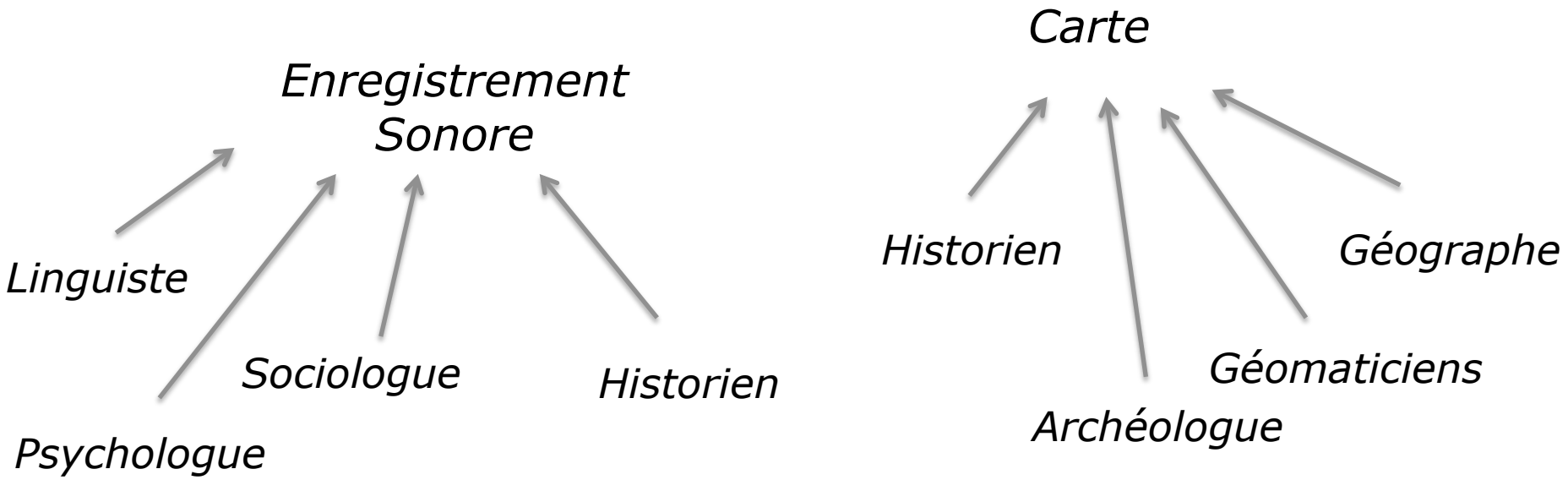
Données SIG

Enrichissements

Huma-Num

Les données de la recherche en SHS

Des usages multiples pour la même donnée



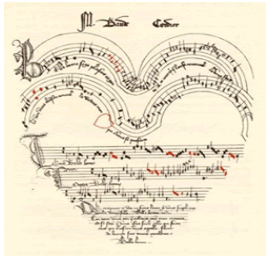
Huma-Num

Les données de la recherche en SHS

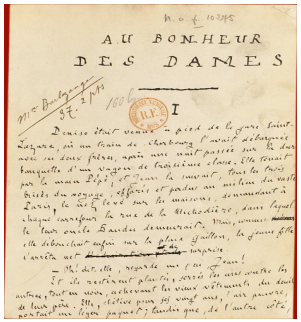
Un intérêt scientifique mais aussi patrimonial

Données de l'archéologie

Enregistrements de langues disparues



Numérisation de manuscrits désormais non consultables



Musique écrite, jouée, encodée

Huma-Num

Les données de la recherche en SHS

*Peu de textes **réglementaires***

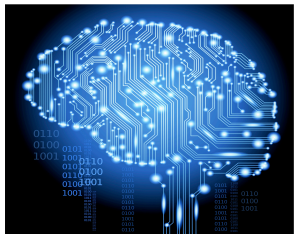
FILATURES H. BERTRAND
Saint-Jean-du-Gard

RÈGLEMENT
pour les fileuses logées

H. BERTRAND

Textes existant plutôt adaptés aux données administratives

La donnée scientifique numérique est un objet nouveau



*Peu **d'intérêt** des tutelles*

Travail non valorisé pour la carrière d'un chercheur

Coûts importants non intégrés dans les projets

Timide prise de conscience dans H2020, mais peu contraignante

Huma-Num

La peste de l'oubli

« 100 ans de solitude » G. G. Marquez
Les habitants de Macondo : une métaphore utile

Mais comment perdre ses données ?

Support

Incompréhension

Format

Ignorance de l'existence

Logiciel

Jeux de caractères

Codage

Fichier endommagé

Pilote

Huma-Num

Huma-Num
la TGIR des humanités numériques

- Très Grande Infrastructure de Recherche pour les SHS

- Financée par le Ministère de l'Enseignement Supérieur et de la Recherche

ACCOMPAGNER LES COMMUNAUTÉS DE RECHERCHE

PROJETS DE RECHERCHE

UMR . ANR . ERC . EA
Labex . Equipex, ...

CONSORTIUMS

Numérisation
Bonnes pratiques
Expertises

OUTILS ET SERVICES

Stocker . Traiter . Diffuser
Exposer . Signaler . Archiver

RESEAU EUROPEEN

Partage . Expertise

Les services d'Huma-Num s'adressent aux porteurs de projets de recherche numériques émanant des communautés de l'ESR.

11 consortiums labellisés par Huma-Num définissent des procédures et standards numériques partagés autour de thématiques et d'objets communs.

Ensemble de services pour le traitement, le stockage, la diffusion et la conservation des données numériques de la recherche en sciences humaines et sociales. A chaque étape du cycle de vie des données correspond un service dédié.

Huma-Num porte la participation française dans le projet européen DARIAH.

Huma-Num

SERVICES POUR LES DONNÉES NUMÉRIQUES

STOCKER
Entreposer . Organiser

TRAITER
Outils . Logiciels

DIFFUSER
Machines virtuelles
Diffusion web

DONNÉES
DE LA RECHERCHE

ARCHIVER
Préservation à long terme

SIGNALER
Enrichissement sémantique
Accès unifié

isidore

EXPOSER
Documenter . Partager

nakala

Partenariat avec le CC-IN2P3, le CINES, et le CCSD

Solutions pour entreposer et organiser les données numériques dans un espace sécurisé professionnel. En partenariat avec le CC-In2p3.

Solutions adaptées aux besoins de transformations ou d'analyses sur les données : logiciels de traitement, de visualisation, d'encodage, SIG, puissance de calcul, etc.

Différentes solutions pour diffuser les données (packs logiciels et machines virtuelles).

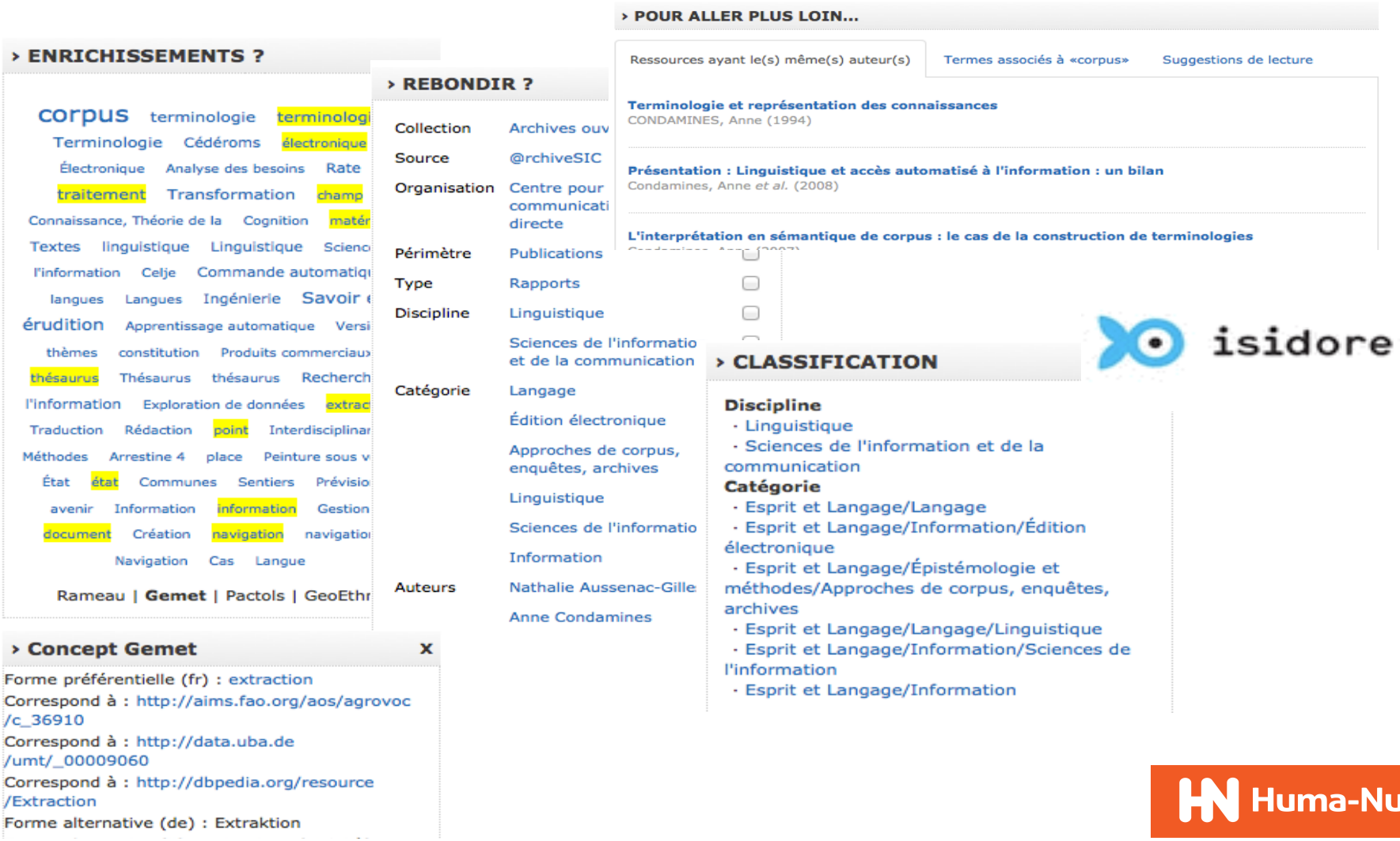
Conseil et accompagnement dans le processus d'archivage à long terme : aide aux choix des formats de données, normalisation des métadonnées, transfert dans le système d'archivage du CINES.

Service multilingue de référence pour le signalement, l'enrichissement, et l'accès aux données en libre accès de la recherche en sciences humaines et sociales (partenariat CCSD)

Entrepôt sécurisé pour déposer et partager tout type de données : fichiers texte, son, image, vidéo.

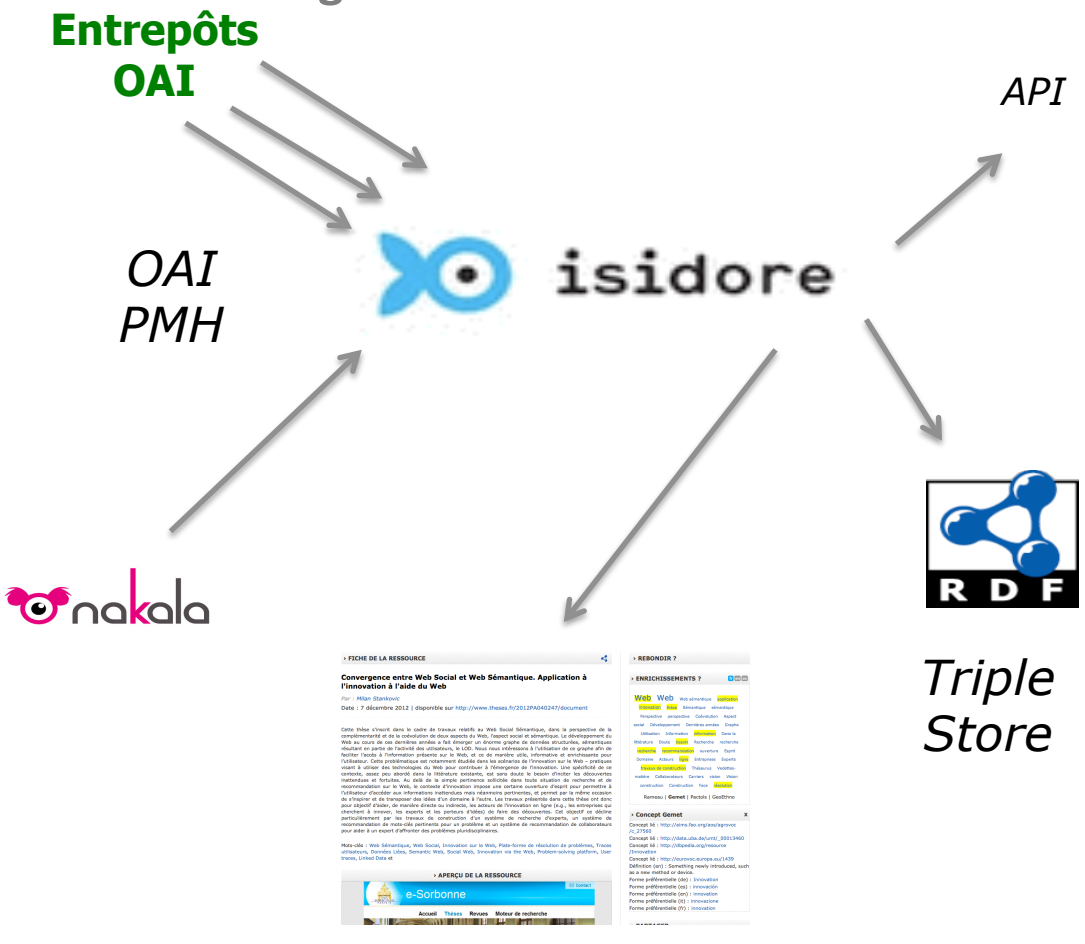
Huma-Num

Signaler les données … ISIDORE



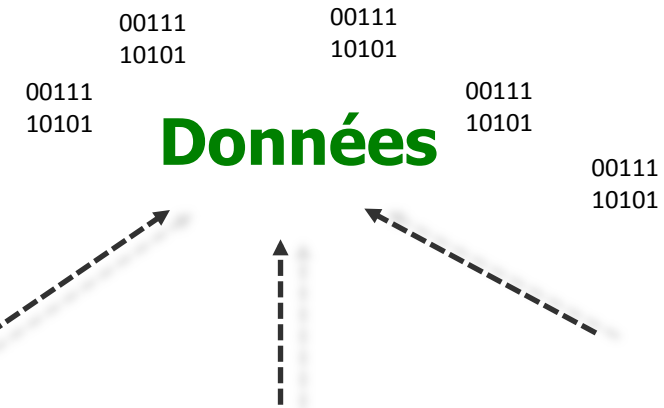
Huma-Num

Signaler les données … ISIDORE



Interface Web et moteur de recherche

Exposer les données …



Identifiant

Affectation d'un **handle** utilisé pour la **citation** et **l'accès**

Accès permanent à la ressource

Triple Store

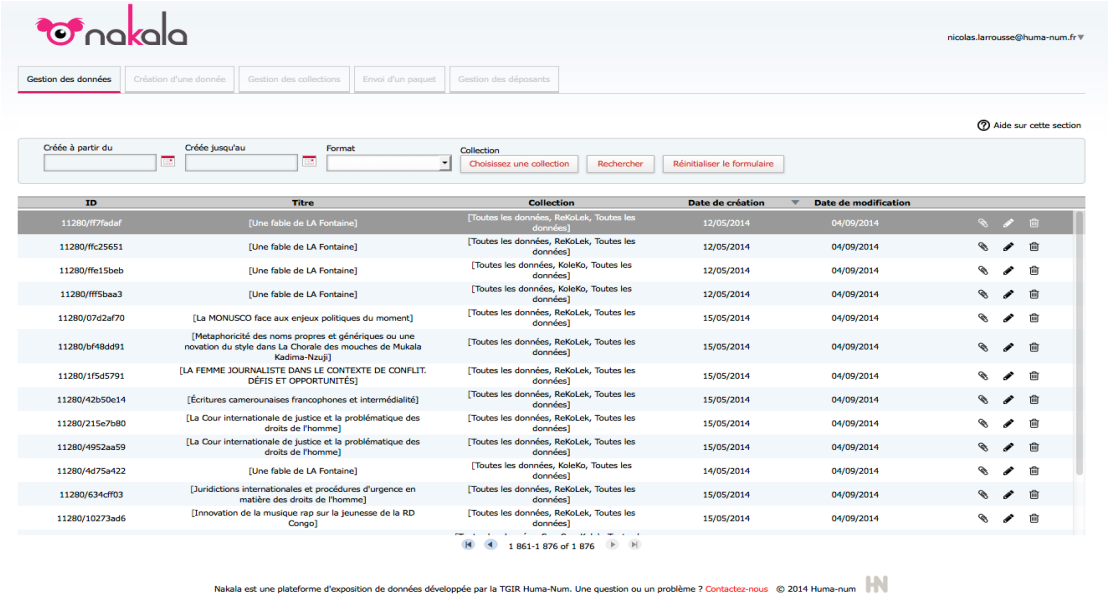
Entrepôt OAI

Accès interopérable à la description

Documentation par des **métadonnées** standardisées associées à des **protocoles** d'échange

NAKALA en pratique …

Une interface web



Dépôt des données

Gestion des métadonnées

Gestion des collections

Gestion des droits

Huma-Num

NAKALA en pratique ...

Une interface pour traiter les lots de données (batch)

`<?xml?>`

Description des métadonnées en XML

Envoi sous forme de paquets

Script multi-plate-forme en Java

```
<nkl:Data xmlns:nkl="http://nakala.fr/schema#"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xmlns:dcterms="http://purl.org/dc/terms/"
        xsi:schemaLocation="http://purl.org/dc/terms/ http://dublincore.org/schemas/
xmls/qdc/2008/02/11/dcterms.xsd">
   <dcterms:title>Venezuela : Estado Portuguesa : Guanare : Esquina Calle 13 et
Carrera 5 : Vue d'une rue bordée d'habitations collectives et d'une friche urbaine</
dcterms:title>
   <dcterms:creator>Pouyllau, Michel</dcterms:creator>
   <dcterms:type>Image</dcterms:type>
   <dcterms:created>1983</dcterms:created>
<dcterms:subject>[SHS:GEO] Humanities and Social Sciences/Geography</
dcterms:subject>
<nkl:inCollection>11280/a19a5e26</nkl:inCollection>
   <nkl:relation type="http://purl.org/dc/terms/isVersionOf">11280/bf86bec2</
nkl:relation>
   <nkl:relation type="http://purl.org/dc/terms/isrequiredby">11280/bf86bec2</
nkl:relation>
</nkl:Data>
```

Huma-Num

NAKALA en pratique …

Utilisation de la fédération d'identité RENATER

Permettre de se connecter avec les identifiants de son établissement

Protéger l'accès aux données

Validation des formats avec le service FACILE du CINES

Intégration de la validation de format dans l'interface web et dans le batch

FACILE - validation du Format d'Archivage du CInes par anaLyse et Expertise

Vérifier l'éligibilité de vos documents à un archivage sur la plate-forme PAC du CINES, c'est FACILE. Vous pouvez analyser leur degré de conformité à un format grâce à cet assistant. Attention : La taille de votre fichier ne doit pas dépasser 900 Mo.

Cliquez ici pour demander l'aide d'un expert du CINES.

Sélectionnez le format que vous désirez contrôler puis sélectionnez un fichier sur votre système à l'aide du bouton [Parcourir], et enfin cliquez sur [Analyser votre fichier] pour voir le résultat.

choisissez un format : Liste des formats (encodage)

choisissez un fichier sur votre système : Parcourir... Aucun fichier sélectionné. Analyser votre fichier

format	nom	PRONOM PSUD	Type MIME	commentaire
AAC AAC	Advanced Audio Codings	[fmt/199]		format Mpeg-4 contenant uniquement un flux audio au format AAC
AIFF PCM	Audio Interchange File Format	[fmt/414]	[audio/x-aif, audio/x-aiff]	format audio contenant uniquement un flux PCM
FLAC FLAC 1.2.1	Free Lossless Audio Codec	[fmt/279]	[audio/ogg, audio/x-flac]	format audio compressé sans perte
GIF 87a	Graphics Interchange Format	[fmt/3]	[image/gif]	format image pouvant contenir également des animations
GIF 89a	Graphics Interchange Format	[fmt/4]	[image/gif]	format image pouvant contenir également des animations
GeoTIFF	Geographic Tagged Image File Format	[fmt/155]	[image/tiff]	format TIFF contenant des informations de géoréférencement et de géolocalisation
HDF5 1.0	Hierarchical Data Format	[fmt/286]		format de données à caractère scientifique
HDF5 2.0	Hierarchical Data Format	[fmt/287]		format de données à caractère scientifique
JPEG	Joint Photographic Experts Group	[fmt/41, fmt/44]	[image/jpeg]	format de représentation compressée d'une image fixe.
JPEG2000	JPEG 2000	[fmt/151, x-fmt/392]		extension du format JPEG standard
JPEG 1.00	Joint Photographic Experts Group	[fmt/42]	[image/jpeg]	format de représentation compressée d'une image fixe.
JPEG 1.01	Joint Photographic Experts Group	[fmt/43]	[image/jpeg]	format de représentation compressée d'une image fixe.
MKV AVC/FLAC	Matroska	[fmt/569]	[video/x-matroska, audio/x-matroska]	conteneur matroska contenant un flux video h.264 (ou appelé également avc) et un flux audio au format flac uniquement

Utiliser les données de NAKALA

Un exemple d'utilisation par le CFEETK

Huma-Num

Utiliser les données de NAKALA

Application(s) spécifique(s)
Editorialisation

Huma-Num
la TGIR des humanités numériques
OMEKA
Huma-Num

omeka

Autre(s)
application(s)

WordPress

hypotheses

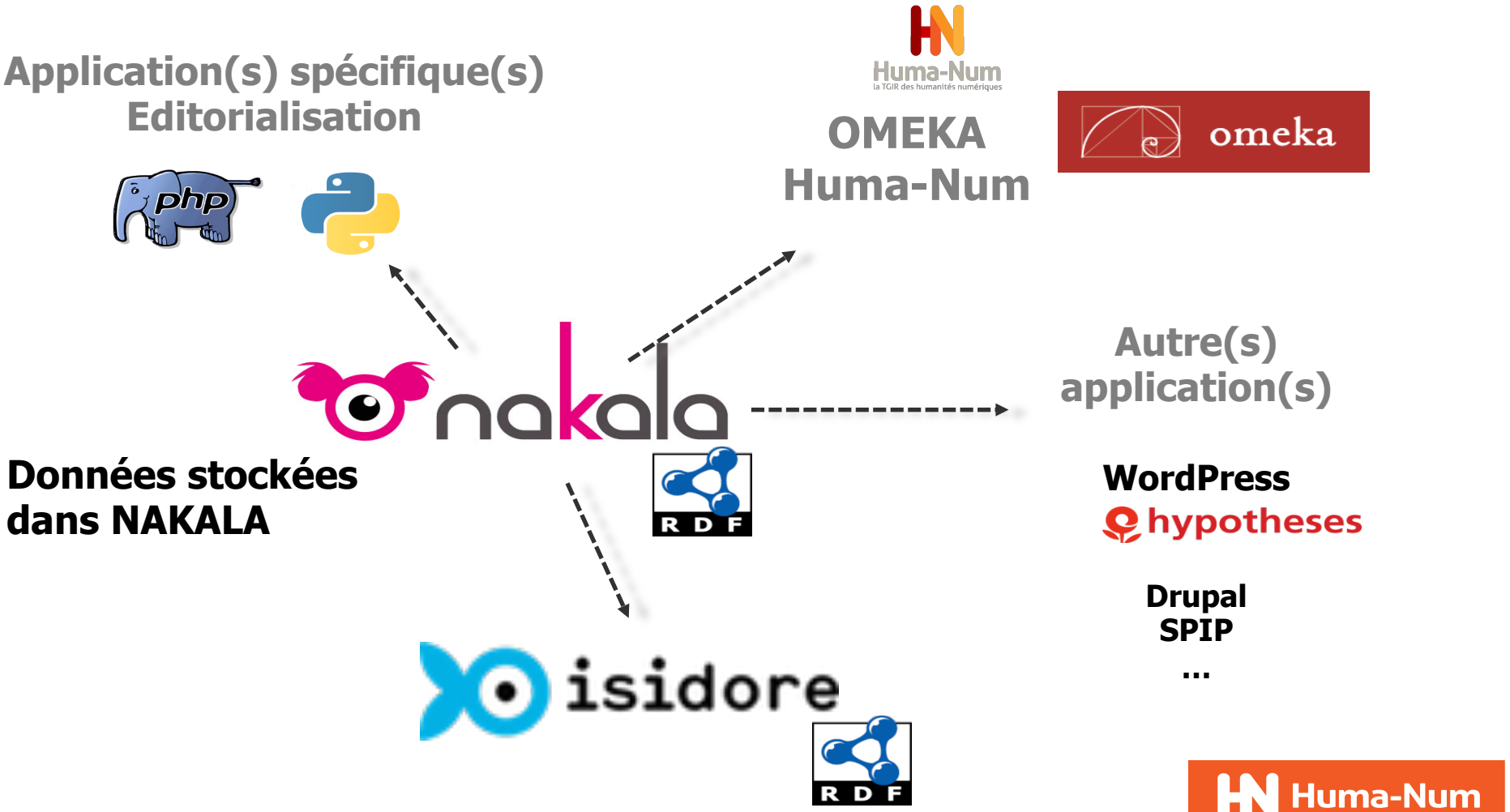
Drupal
SPIP
...

nakala

Données stockées dans NAKALA

Huma-Num

Utiliser les données de NAKALA



Effets attendus …

→ Ne pas perdre les données

→ Promouvoir de bonnes pratiques

→ Produire de « meilleures » données documentées

→ Prendre en compte ces aspects au début du projet

Huma-Num