

## Présentation générale

Huma-Num est une très grande infrastructure de recherche (TGIR) visant à faciliter le tournant numérique de la recherche en sciences humaines et sociales.

Pour remplir cette mission, la TGIR Huma-Num est bâtie sur une organisation originale consistant à mettre en œuvre un dispositif humain (concertation collective) et technologique (services numériques pérennes) à l'échelle nationale et européenne en s'appuyant sur un important réseau de partenaires et d'opérateurs.

La TGIR Huma-Num favorise ainsi, par l'intermédiaire de consortiums regroupant des acteurs des communautés scientifiques, la coordination de la production raisonnée et collective de corpus de sources (recommandations scientifiques, bonnes pratiques technologiques). Elle développe également un dispositif technologique unique permettant le traitement, la conservation, l'accès et l'interopérabilité des données de la recherche. Ce dispositif est composé d'une grille de services dédiés, d'une plateforme d'accès unifié et d'enrichissement des données et d'une procédure d'archivage à long terme dédiée aux données SHS.

La TGIR Huma-Num propose en outre des guides de bonnes pratiques technologiques généralistes à destination des chercheurs. Elle peut mener ponctuellement des actions d'expertise et de formation. Elle porte la participation de la France dans le projet DARIAH-EU en coordonnant les contributions nationales.

## Nakala : un service de stockage pérenne et de signalement des données de la recherche en SHS

Partant du constat que de nombreuses équipes et projets de recherches ne disposent pas de l'infrastructure numérique nécessaire qui permettrait un accès persistant et interopérable à leurs données numériques, la TGIR Huma-Num a mis en œuvre un service d'exposition de données appelé NAKALA ([www.nakala.fr](http://www.nakala.fr)). NAKALA repose sur les principes du web des données (formalise RDF, modèles de données exprimés dans le *linked open data* ou *LOD*, ...) et permet aux utilisateurs d'accéder aux métadonnées et données via un SPARQL endpoint.

NAKALA propose deux grands types de services : des services d'accès aux données elles-mêmes et des services de présentation des métadonnées. Les producteurs de données numériques ainsi soulagés de la gestion purement technique, peuvent ainsi se consacrer à la valorisation scientifique de leurs données et cela dans le cadre des données liées dans le web de données.

Les données hébergées par NAKALA peuvent être éditorialisées à l'aide du Pack OMEKA développé à partir du CMS OMEKA (RRCHNM) et géré par Huma-Num dans le cadre de l'offre NAKALONA (voir NAKALA).

## **ISIDORE : une solution complète d'enrichissement sémantique des données de la recherche et des publications scientifiques en SHS**

ISIDORE est un service qui collecte, enrichit et offre un signalement et un accès unifié aux documents et données numériques des sciences humaines et sociales. Il donne accès plus de 3000 sources de données et contient plus de 4,2 millions de ressources.

ISIDORE « moissonne » - c'est le terme consacré - les notices, les métadonnées et le texte intégral issus des publications électroniques, des corpus, des bases de données et des actualités scientifiques, accessibles sur le web et proposés dans des standards ouverts d'interopérabilité. ISIDORE moissonne principalement des données francophones produites en France ou dans le monde francophone mais intègre aussi des données en anglais et en espagnol. Enfin ISIDORE valorise les données en libre accès (open access).

Une fois moissonnées, ces informations sont enrichies en trois langues (anglais, espagnol et français) par croisement avec des référentiels métiers (listes de vocabulaires, thésaurus, référentiels exprimés en RDF/SKOS) produits soit par la communauté scientifique (GeoEthno, Pactols...), soit par les grandes institutions du domaine de l'enseignement supérieur et de la recherche (Rameau, LCSH, BNE, Gemet, Lexvo, GeoNames...). Les enrichissements multilingues permettent de relier les données entre-elles. Ces informations constituent des points d'entrée vers le texte intégral qui est lui aussi indexé quand cela est possible.

ISIDORE s'inscrit dans la longue tradition des instruments documentaires forgés par les SHS pour le repérage des sources de données quelles soient textuelles, audio-visuelles, sérielles, ou encore – depuis l'arrivée massive des données numériques – complexes et hybrides. L'importance de la citabilité et de la garantie de l'accès dans le temps aux documents, publications, objets de recherche se voit renforcée avec le support qu'est le web, qui évolue tous les jours et mute régulièrement. L'accès multidisciplinaire et les besoins de réutilisation des données de la recherche obligent à proposer des outils puissants pour le signalement et le repérage des données, forgés par les SHS, c'est-à-dire correspondants à leurs paradigmes.

Si les SHS évoluent, le web change aussi : il devient de plus en plus hétérogène et la mise en place du web de données en est la dernière grande évolution. Le Web est non seulement le support des sites mais aussi un espace (au sens d'environnement de stockage, d'édition et de diffusion des données) dans lequel se construisent des territoires pour y stocker à la fois des documents non structurés sémantiquement et des documents contenant une information « structurée », c'est-à-dire une proposition de qualification sémantique de l'information. Dans ce cadre, les outils de recherche doivent s'adapter, changer profondément afin de tirer parti de ces espaces structurés, ouverts et normalisés. Utilisant les méthodes et principes du web des données (modèle RDF) et du linked data (URIs), ISIDORE est le plus gros projet d'open data scientifique, en exploitation en France, reposant sur les principes du *LOD*. Il propose l'accès à plus de trois millions de documents numériques. Il n'est donc pas un moteur de recherche classique : c'est une plateforme de recherche modulable qui a vocation à fédérer l'accès aux données numériques de la recherche en SHS et d'en offrir, pour les enseignants, chercheurs, doctorants et étudiants, un accès unifié tout en reliant les données entre-elles au moyen d'enrichissements multilingues.

ISIDORE associe plus d'une centaine de producteurs de données et moissonne plus de 2000 sources de données : les principales plateformes d'édition électronique, un très grand nombre de bibliothèques (de recherche, universitaires, municipales) mais aussi de nombreuses bases de données des SHS.

Pour les utilisateurs recherchant de l'information et des documents :

- Sur le web, via le portail [www.rechercheisidore.fr](http://www.rechercheisidore.fr)
- Sur les tablettes et téléphones connectés via le site [m.rechercheisidore.fr](http://m.rechercheisidore.fr)
- Au travers d'applications embarquées dans les outils d'édition électronique (blogs et sites web sous WordPress)
- Des widgets pour Windows et Mac.

Pour les développeurs, architectes de données, documentalistes et webmestres :

- Une API du moteur de recherche de [www.rechercheisidore.fr](http://www.rechercheisidore.fr) est disponible
- Une interface d'interrogation SPARQL (endpoint) pour interroger en SPARQL les triplets RDF enrichis et contenus dans le triple-store ISIDORE
  - Des widgets de développement « ISIDORE à la demande » (*ISIDORE on demand*) dans le cadre du programme *ISIDORE Motor Constructor* (ISIDORE-MC)
- Via la technologie OpenSearch

ISIDORE est en exploitation depuis 2011 et fonctionne en mode 24/7.