

Ecosystème et stratégie de la gestion des échantillons et des données analytiques de l'INSU : Première phase dédiée au domaine Terre-Solide

Christophe Scheffer – Resp. Pilotage & Administration IR RéGEF (christophe.scheffer@univ-lorraine.fr)

Pierre-Yves Arnould – Resp. Service Systèmes d'informations / Gestion des données OTELo (pierre-yves.arnould@univ-lorraine.fr)

Raphaël Pik – chargé de mission INSU – directeur de l'IR RéGEF (raphael.pik@univ-lorraine.fr)

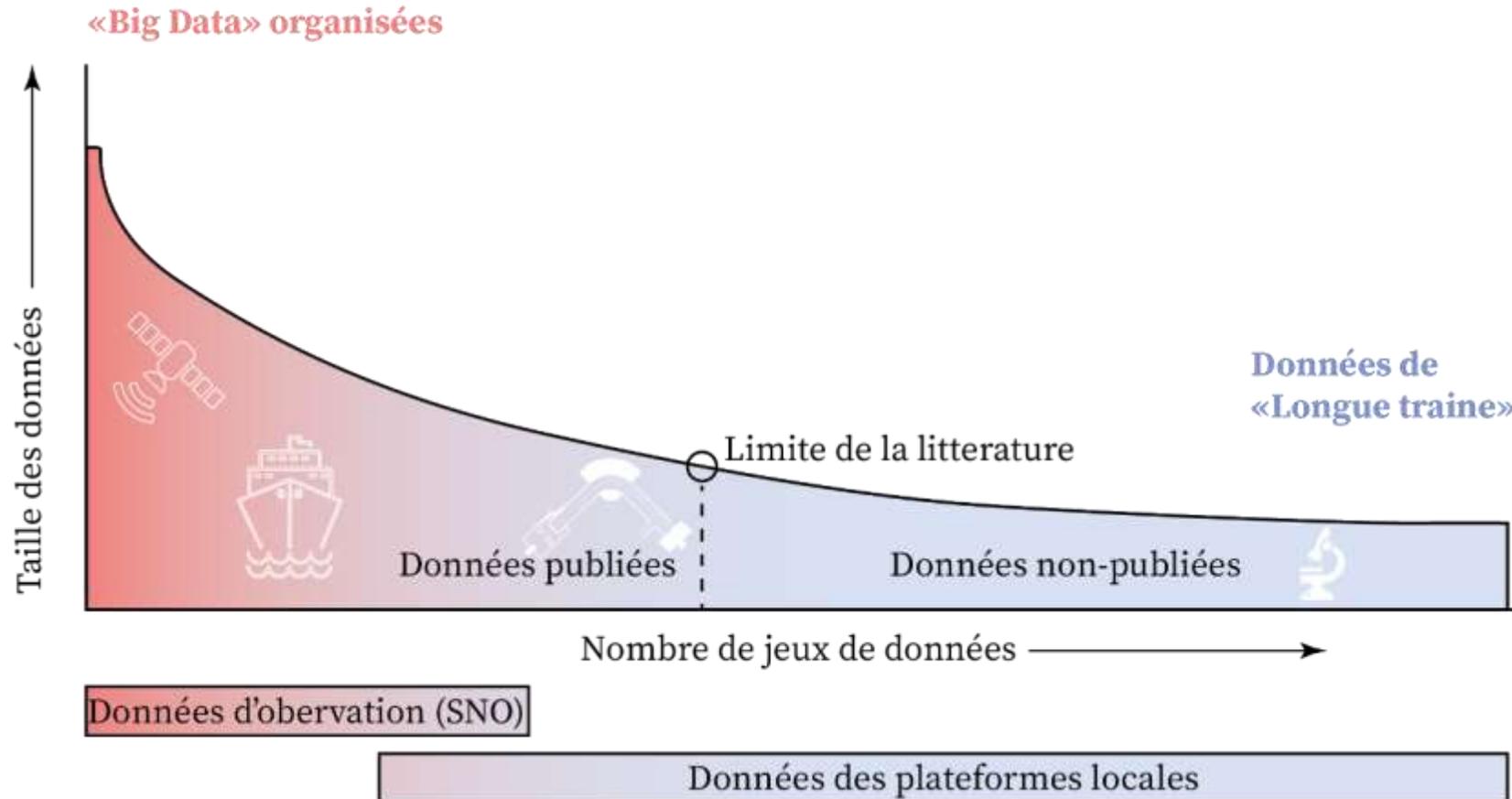
■ Contexte et historique

Cette initiative s'inscrit dans les objectifs de l'IR RéGEF :

- (i) **Maintenir l'innovation et le développement technologique au plus haut niveau** pour faire avancer nos capacités de traçage et de caractérisation et faire reculer les frontières de notre investigation (plus précis & petit, nouveaux traceurs, miniaturisations, etc ...)
=> **Excellence scientifique**

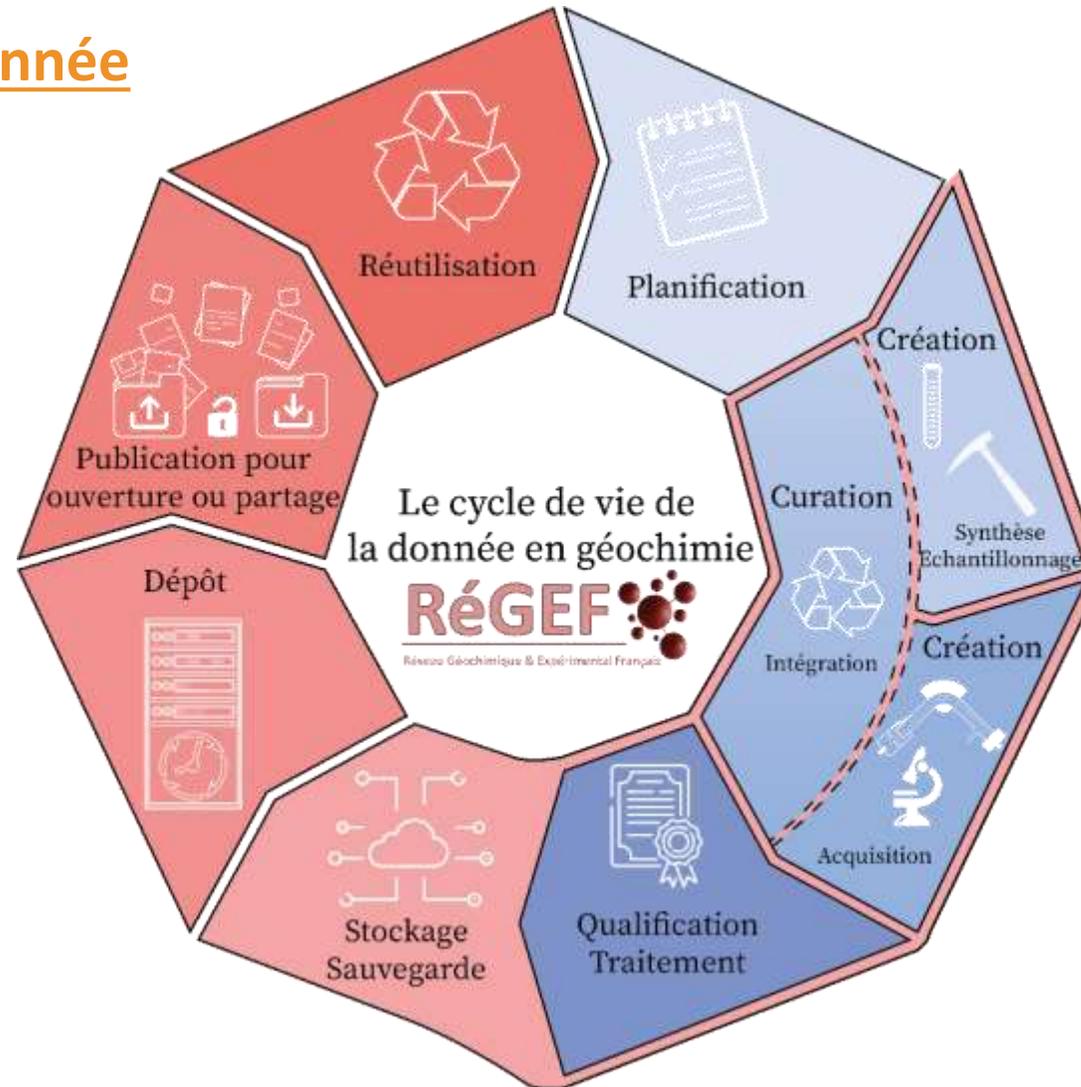
- (ii) **Assurer la production de données et l'accès aux mesures pour l'ensemble de la communauté**, en organisant le transfert des outils et méthodes de géochimie et pétrologie vers les scientifiques travaillant dans le domaine des géosciences et de leurs implications sociétales (environnement, climat, risques, santé, ...)
=> **Production – visibilité – accessibilité – archivage**

Le type de données produites dans le cadre de l'IR RéGEF



Modifié d'après
Ferguson et al. 2014

Le cycle de vie de la donnée



La politique de gestion de donnée

Groupe de Travail : Echantillons & Données

 De la déclaration
de l'échantillon ...

... à l'analyse ...



 ... aux services de la
science ouverte

Métadonnées des Échantillons

Animateur: C Scheffer (CDP RéGEF)

Mission: Définir les bases de métadonnées nécessaires à la déclaration et l'identification des échantillons des géosciences.



Identification des Échantillons

Animateur: PY Arnould (OSU OTELo)

Mission: Assurer la mise en œuvre du service national de numérotation universelle IGSN des échantillons des géosciences.



Qualité des Données

Animateur: C Scheffer (CDP RéGEF)

Mission: Construire et mettre en application les Plan de Gestion des Données des réseaux de l'IR RéGEF.

Valorisation des Données

Animateur: R Pik (DSI, Dir. RéGEF)

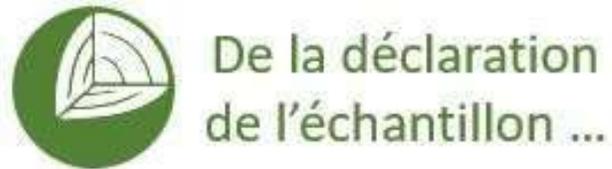
Mission: Accompagner la stratégie de gestion et de pérennisation des données produites sur les échantillons des géosciences.

Réseau de Géothèques

Animateur : M Grégoire (CMI)

Mission: Mettre en place un projet de Géothèque virtuelle pour la pérennisation physique des échantillons.

La politique de gestion de donnée



Métadonnées des Échantillons

Animateur: C Scheffer (CDP RéGEF)

Mission: Définir les bases de métadonnées nécessaires à la déclaration et l'identification des échantillons des géosciences.

Comment déclarer un échantillon des géosciences ?

Quelles sont pour chaque type d'échantillon, **les schémas de métadonnées**, et **le vocabulaire contrôlé** qui vont permettre aux communautés de définir et caractériser efficacement un échantillon collecté ou synthétisé ?

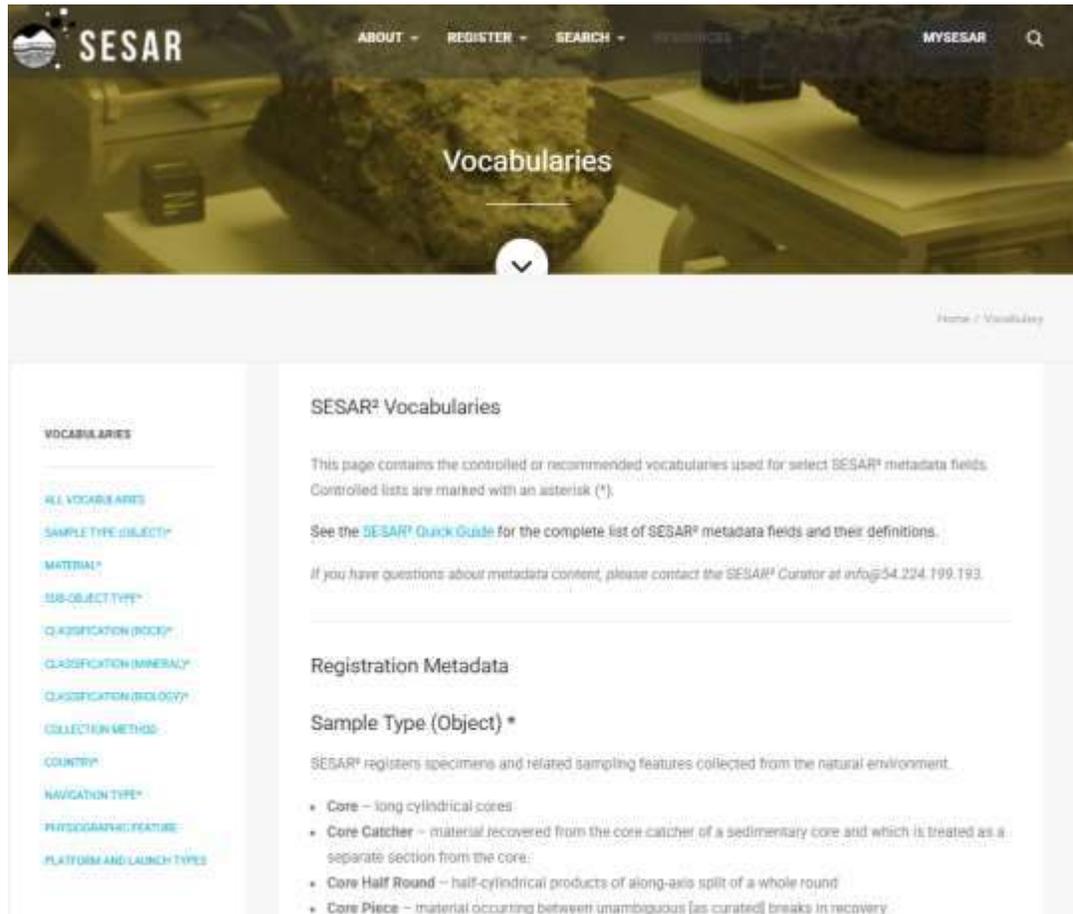


Groupe de travail formé par les représentants des différentes Commissions scientifiques de l'INSU concernés par les échantillons des géosciences (CS-OA, CS-SIC, CS-TS), des acteurs du BRGM, de l'IFREMER, de Reolnat, et de la Cyber-Carothèque Nationale.



Le système de déclaration existant

Système de déclaration IGSN américain



SESAR² Vocabularies

This page contains the controlled or recommended vocabularies used for select SESAR² metadata fields. Controlled lists are marked with an asterisk (*).

See the [SESAR² Quick Guide](#) for the complete list of SESAR² metadata fields and their definitions.

If you have questions about metadata content, please contact the SESAR² Curator at info@54.234.199.193.

Registration Metadata

Sample Type (Object) *

SESAR² registers specimens and related sampling features collected from the natural environment.

- Core – long cylindrical cores
- Core Catcher – material recovered from the core catcher of a sedimentary core and which is treated as a separate section from the core.
- Core Half Round – half-cylindrical products of along-axis split of a whole round
- Core Piece – material occurring between unambiguous (as curated) breaks in recovery

Système actuel de déclaration IGSN du CNRS-INSU

Red fields	Compulsory fields to be declared and obtain IGSN
Blue fields	Fields with vocabulary (pre-defined menus)
Nom du champ	Signification
General	
user_code	Prefix of IGSN code - TOAE and eventually CNRS (not available yet)
name	User name, descriptive phrase or title of the resource
other_names	Other name(s) used for the sample. Provide multiple names delimited by semi-colons
resourceIdentifier	Code IGSN returned - leave empty in table for IGSN request
parent_igsn	Fill in to declare daughters. Blank otherwise (e.g., subsamples of a sample, sampling of cores, etc)
isPublic	The visibility of a resource's metadata. Use 'true' for public and 'false' for private resources
publish_date	A private resource will not be publicly available until the publish date - leave blank if no publication date
sample_type	The type of the sample
Description	
material	From vocabulary, general sample description (e.g., rock, ice, fluid, gas...)
field_name	Taxonomy (field name) Informal classification of sample e.g. basalt, amphibole, sea water, peridotite...
classification(s)	name [rock classification] [mineral classification]
age_min	Numerical value for the minimum age of a sample "4.2"
age_max	Numerical value for the maximum age of a sample "4.5"
age_unit	Unit for the age provided (e.g., "Ma" for million years, "years", etc)
geological_age	Age of a sample as described by the stratigraphic era, period, state, etc. "Cretaceous", "Upper Miocene"
geological_unit	unit for geological_age
collection_method (et descr)	The method/instrument by which the resource was collected/created, e.g., Dredging, Manual Observation
size	Size of the registered object, such as the dimension of a specimen, the length of a core, Or the weight of a dredge, e.g. "2x4" (enter "cm" into 'size unit' - see field below), "45" (enter "kg" into 'size
size_unit	(i.e., cm, in, m, kg) : Unit for the numerical value provided for 'size', e.g. "meters", "kg"
sample_comment	Any free text describing the resource. This can also be used to include additional metadata which cannot be specified with the other fields provided
purpose	Free text to describe purpose for collecting/creating the sample or item
Geolocation	
latitude	Latitude of the location where the sample was collected. ('Start latitude' for linear sampling features such as dredges.) Needs to be entered in decimal degrees. Negative values for South latitudes: "-24.7852" (=24.7852 S); "5.69634" (=5.69634 N)
latitude_unit	Latitude unit, of east or west longitude (e.g., "N" for North, "S" for South)

■ Stratégie employée

Thématique choisie :

Terre-Solide

Groupe de travail :

Constitution la plus exhaustive possible (max 15 personnes)

Acteurs & institutions couvrant un maximum de domaines scientifiques et d'institutions (CS, IFREMER, BRGM, Reolnat, CCN, + Experts data)

Premiers échanges en visio

Exposition des objectifs et attendus

Constitution d'une base de travail commune :

Liste IGSN + Modèle SESAR

Première brique de l'outil CNRS de déclaration IGSN

Recensement des vocabulaires existants : SESAR, CCN, IGSN, EPOS, BRGM, Reolnat, USGS, NOAA, PETDB EarthChem, Mindat, GeoCASE...

Collecte et analyse

Recueil de vocabulaires & schémas de métadonnées existants sur un espace de travail partagé : Résana

■ Stratégie employée

Réunion de travail en présentiel à Paris :

10 février 2025

Organisation :

Evaluation et comparaison des vocabulaires existants

Accord sur un vocabulaire de référence, enrichi au besoin

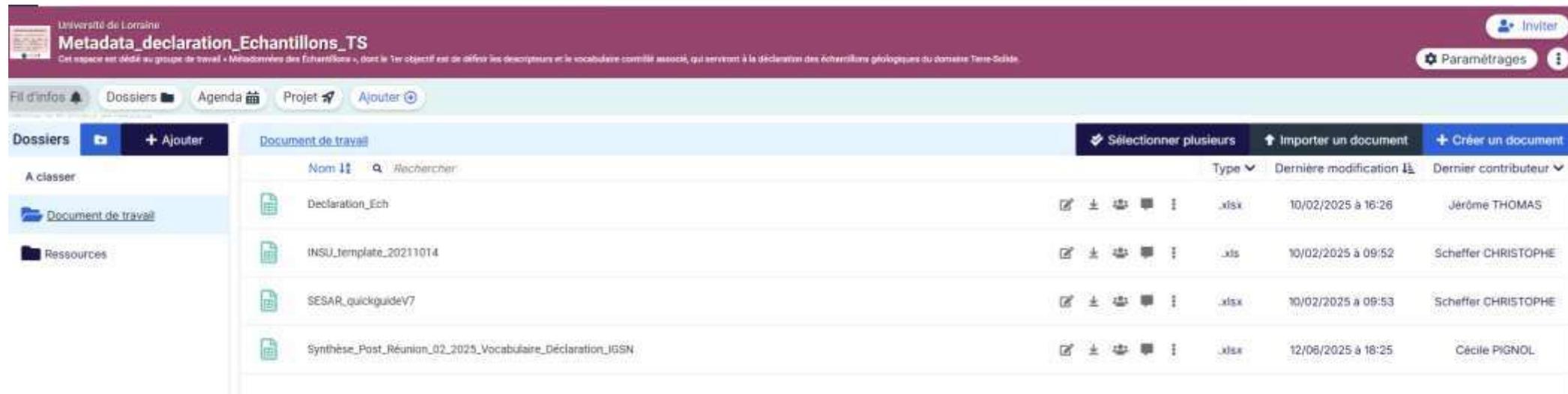
Trouver un équilibre : « *Suffisamment complet pour être d'utilité, Suffisamment épuré pour être utilisé* »

Résultats intermédiaires

Fichier commun : Vocabulaire + Descripteurs (Résana)

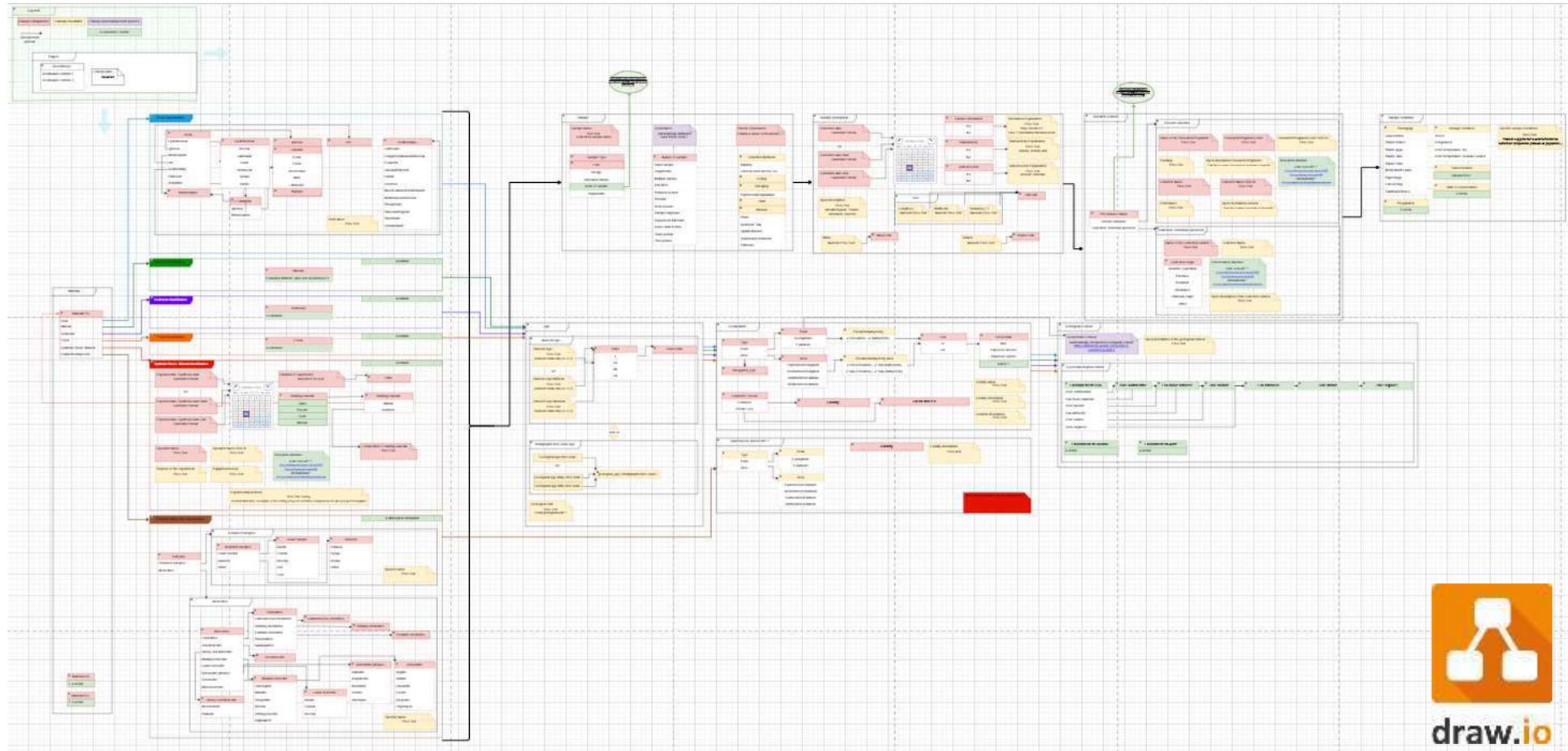
Schéma de métadonnées logique (liens & relations via Draw.io et stockage Google Drive)

Stratégie employée

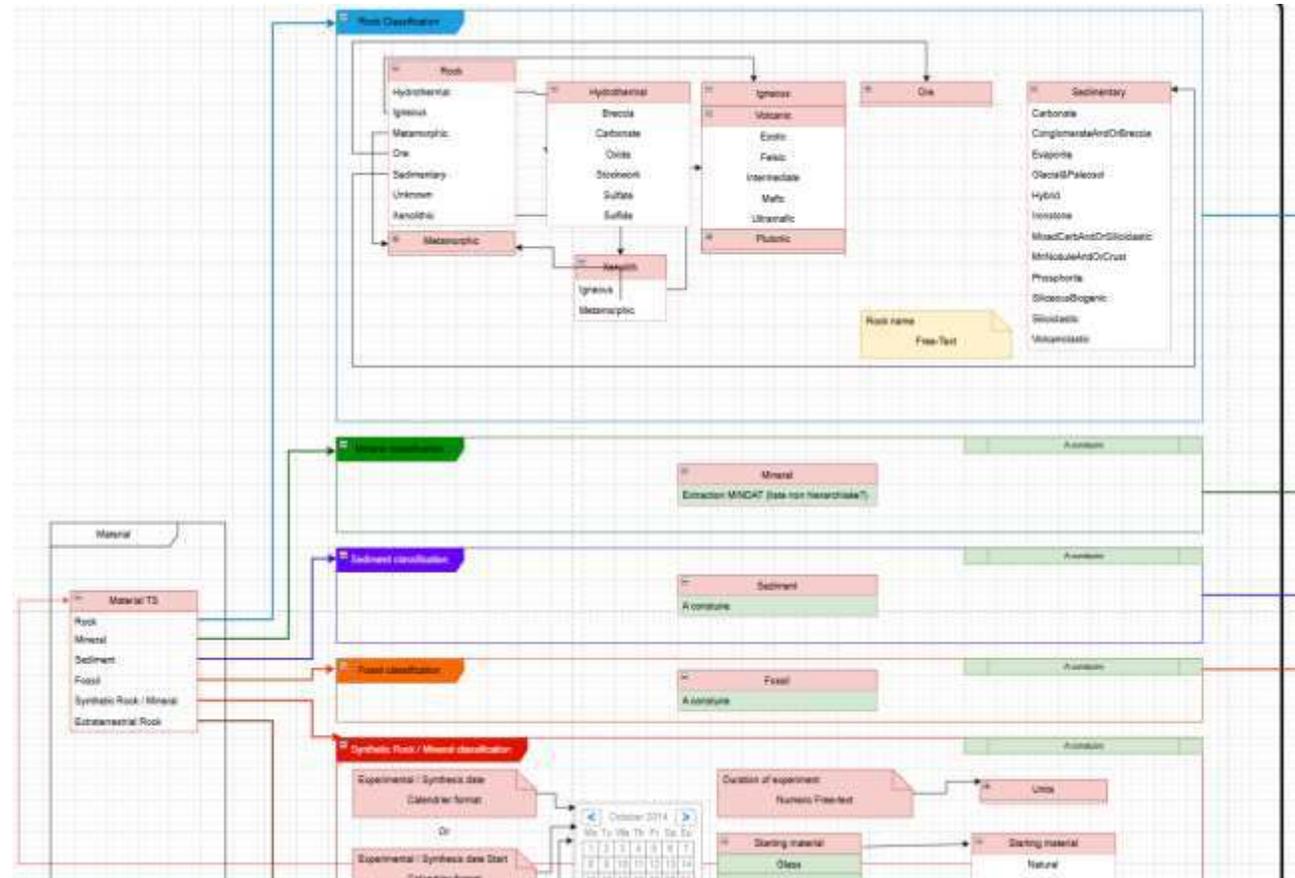



Nom	Type	Dernière modification	Dernier contributeur
Declaration_Ech	.xlsx	10/02/2025 à 16:26	Jérôme THOMAS
INSU_template_20211014	.xls	10/02/2025 à 09:52	Scheffer CHRISTOPHE
SESAR_quickguideV7	.xlsx	10/02/2025 à 09:53	Scheffer CHRISTOPHE
Synthèse_Post_Réunion_02_2025_Vocabulaire_Declaration_IGSN	.xlsx	12/06/2025 à 18:25	Cécile PIGNOL

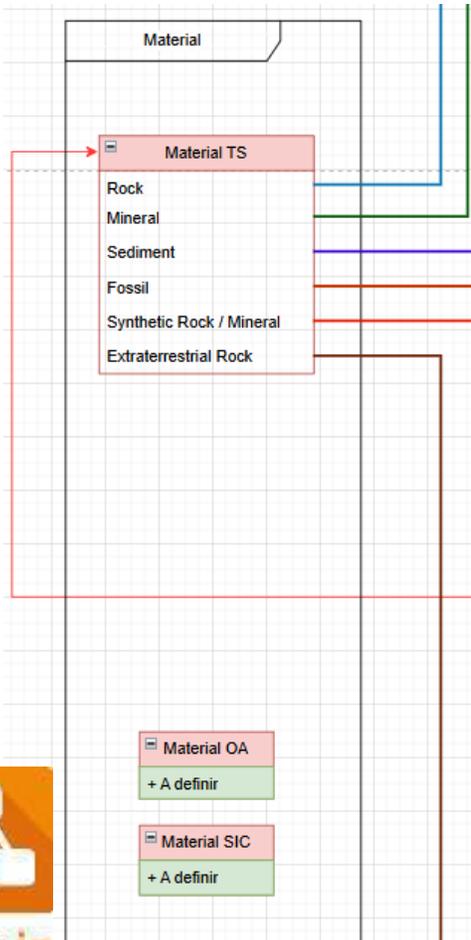
■ Stratégie employée



Stratégie employée



■ Stratégie employée



Pour continuer...

Finalisation des vocabulaires et lien entre descripteurs

Quelques manques en TS : Sédiments, fossiles, environnements magmatique...

Quid des autres communautés (OA, SIC) ?

Intégration dans le groupe de travail international IGSN

Suivi de l'évolution des fonctionnalités et des normes déclaratives du modèle IGSN

Participation active à l'évolution des vocabulaires

Alignement avec les modèles internationaux

Adaptation et cohérence avec le modèle CNRS

Développements de l'outil d'attribution IGSN ...

■ Développement de l'outil



De la déclaration
de l'échantillon ...

Identification des Échantillons

Animateur: PY Arnould (OSU OTELo)

Mission: Assurer la mise en œuvre du service national de numérotation universelle IGSN des échantillons des géosciences.



Mettre en place d'un **service national de numérotation universelle des échantillons**, basés sur l'IGSN (International Generic Sample Number)



Application et dispositif **déjà existant** mais devant **évoluer** afin de couvrir la **diversité et l'hétérogénéité des échantillons et de leurs métadonnées associées**

Il est nécessaire de penser un **outil numérique, optimisé, versatile, interopérable**

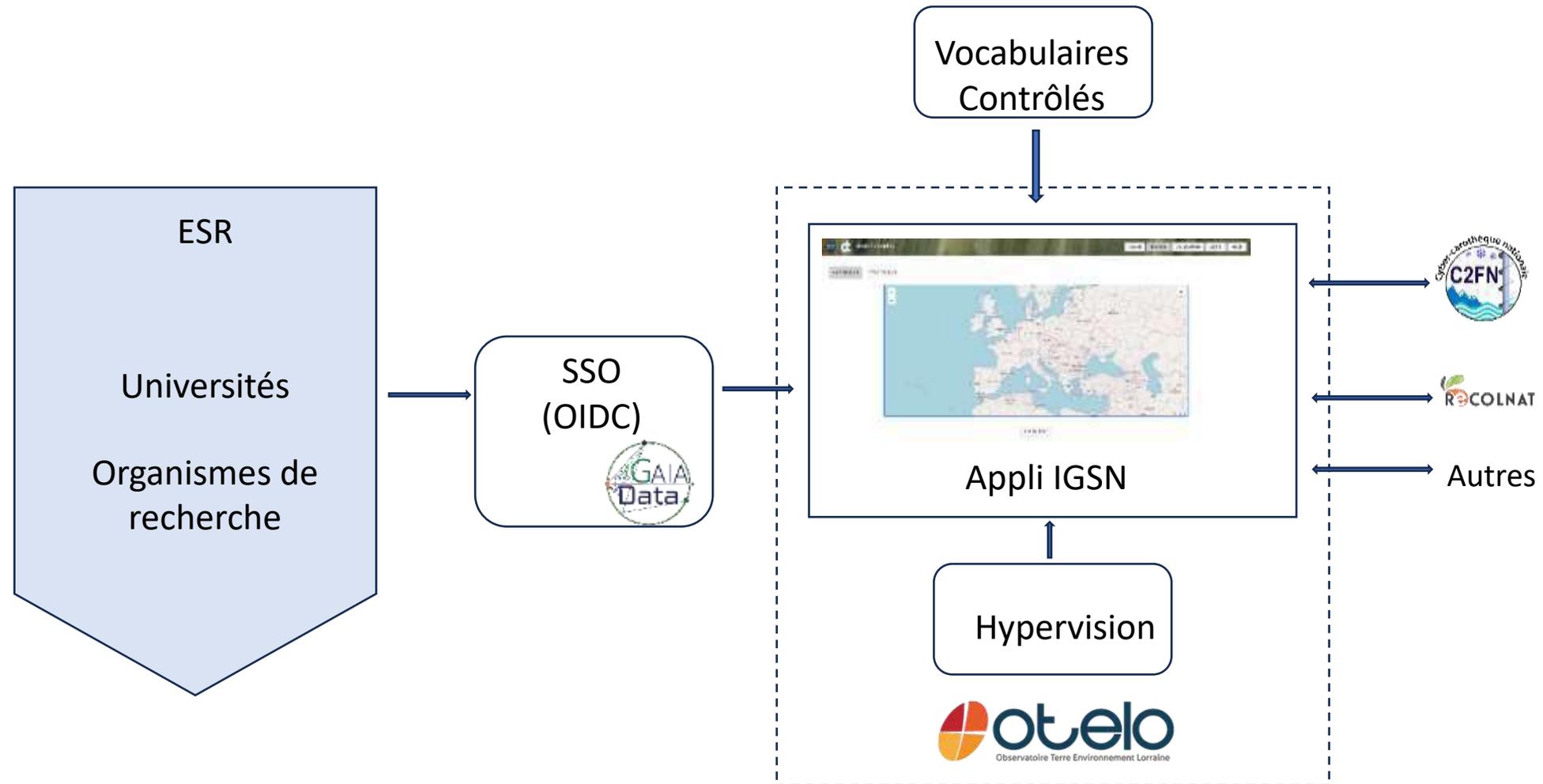
■ Développement de l'outil : Un cahier des charges précis

- Une authentification basée sur le SSO de Gaia Data (OIDC)
- La gestion des accès, rôles et droits des utilisateurs ;
- La gestion de groupes d'utilisateurs ;
- La gestion de déclaration « brouillon » par profil utilisateur ;
- L'accès aux déclarations d'échantillons en cours d'élaboration aux seules personnes accréditées et authentifiées ;
- La réservation de plages de code IGSN en vue de campagne
- La déclaration, modification, suppression (deprecated) d'identifiant échantillons par saisie manuelle ou assistée (masque de saisie, listes de choix), contrôlée et autorisée ;

■ Développement de l'outil : Un cahier des charges précis

- La déclaration, modification, suppression (deprecated) d'identifiant échantillons par utilisation d'API, contrôlée et autorisée ;
- La déclaration, modification, suppression (deprecated) d'identifiant de sous-échantillons.
- La recherche d'échantillon avec filtrage (type, lieu, responsable technique, scientifique...) par interface et API.
- Permettre le moissonnage des échantillons via API (Protocole standardisé, OAI-PMH ...)
- L'extraction des contenus afin de faciliter la constitution de différents types de documents (par exemple: fiche de suivi de déclaration échantillons) ;
- L'export des données au format JSON dans le respect du maintien de l'authentification (respectant des schémas de métadonnées Internationaux comme Datacite, ...);

■ Développement de l'outil : Un cahier des charges précis



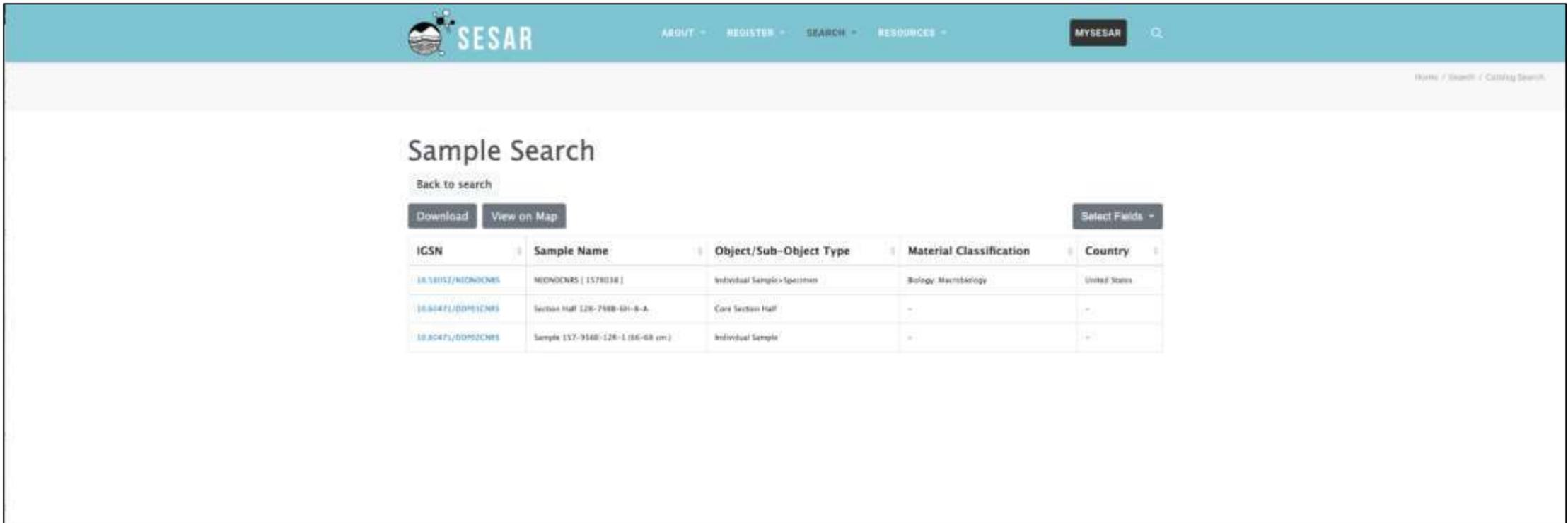
■ Développement de l'outil : Un cahier des charges précis

The screenshot displays the SESAR Catalog Search interface. At the top, there is a navigation bar with 'ABOUT', 'REGISTER', 'RESOURCES', and 'MYSESAR' links, along with a search icon and the text '1/14'. Below the navigation bar, the text 'Catalog Search' is centered. The main content area is titled 'Sample Search' and contains a form with the following fields and options:

- Sample Name: equals [input field]
- ICIN: equals [input field]
- Search by multiple ICINs: (Copy/Paste your sample ICINs here. The ICINs need to be comma-separated.) [input field]
- Object/Task-Object Type: [dropdown menu]
- Material: [dropdown menu]
- Classification: Select a material first [dropdown menu]
- Field name (external classification): [input field]
- Geographic Region: [input field]
- Country: [dropdown menu]
- Register: [input field]
- Collection: [input field]
- Archives: [input field]
- Field Program / Cruise: [input field]
- Platform Name: [input field]
- Collection Method: [input field]
- Registration date between: YYYY-MM-DD to YYYY-MM-DD [input fields]
- Exclude samples without latitude/longitude information [checkbox]
- Exclude child samples [checkbox]
- Only include records with images [checkbox]
- Only include records with linked datasets, paper DOI, or URL [checkbox]

Below the form, there is a map showing a geographical area with a 'Change Projection' button. At the bottom of the form, there are 'Search' and 'Clear' buttons.

■ Développement de l'outil : Un cahier des charges précis



The screenshot shows the SESAR Sample Search interface. The header includes the SESAR logo, navigation links (ABOUT, REGISTER, SEARCH, RESOURCES), and a MYSESAR button. The main content area is titled "Sample Search" and includes a "Back to search" link, "Download" and "View on Map" buttons, and a "Select Fields" dropdown menu. A table displays search results with columns for IGSN, Sample Name, Object/Sub-Object Type, Material Classification, and Country.

IGSN	Sample Name	Object/Sub-Object Type	Material Classification	Country
18.58152/NEON/CNRS	NEON/CNRS (1578138)	Individual Sample/Specimen	Biology: Microbiology	United States
18.60472/DDPS/CNRS	Section Hall 126-7988-01-B-A	Core Section Hall	-	-
18.60472/DDPS/CNRS	Sample 127-9348-126-1.186-68 cm	Individual Sample	-	-

■ Développement de l'outil : Un cahier des charges précis

IGSN: 10.58052/NEON0CNRS   



IGSN:	10.58052/NEON0CNRS
Sample Name:	NEON0CNRS [1578038]
Other Name(s):	621ec9ef-a648-406b-a5b4-405c4946033a, C00000092457
Sample Type:	Individual Sample>Specimen
Parent IGSN:	Not Provided

Description

Material:	Biology
Classification:	Macrobiology
Field Name:	Soil Microbe
Description:	Bulk soil subsample collected from up to 30 cm depth as part of NEON terrestrial soil sampling protocols conducted 3 times annually.
Age (min):	Not Provided
Age (max):	Not Provided
Collection Method:	Manual
Collection Method Description:	PreservedSpecimen
Size:	Not Provided
Geological Age:	Not Provided
Geological Unit:	Not Provided
Comment:	Not Provided
Purpose:	Not Provided

Geolocation

Latitude (WGS84):	40.460611
Longitude (WGS84):	-103.030496
Northing (m) (UTM NAD83):	Not Provided
Easting (m) (UTM NAD83):	Not Provided
Zone:	Not Provided
Vertical Datum:	Not Provided
Elevation Start:	1363 meters
Elevation End:	Not Provided
Nav Type:	Not Provided
Physiographic Feature:	Not Provided
Name Of Physiographic Feature:	Not Provided
Location Description:	Not Provided
Locality:	Central Plains (D10), North Sterling NEON (STER), Plot STER_010 (plot dimensions: 40m x 40m)
Locality Description:	Not Provided
Country:	United States
State/Province:	Colorado
County:	Logan
City:	Not Provided

Collection

Field Program/Cruise:	Not Provided
Platform Type:	Not Provided
Platform Name:	Not Provided

■ Développement de l'outil : Un cahier des charges précis

MySESAR

My Home My Samples Sample Groups Teams Register/Update Samples Transfer Ownership Search My Account Help Logout

Individual Sample Registration

Sample

Select Type of Object: *

Sample Name: *

Does the object have a parent?

Enter Parent IGSN (if in SESAR)

The parent is not in the system?

Enter the object type of the external parent:

Enter the name of the external parent:

* required field



Merci !

Christophe Scheffer – Resp. Pilotage & Administration IR RéGEF (christophe.scheffer@univ-lorraine.fr)

Pierre-Yves Arnould – Ingénieur Systèmes d'informations / Gestion des données OTELO (pierre-yves.arnould@univ-lorraine.fr)

Raphaël Pik – chargé de mission INSU – directeur de l'IR RéGEF (raphael.pik@univ-lorraine.fr)