



LE WEB SÉMANTIQUE AU BRGM

Tendry RANDRIAMALALA
29/11/2022



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité



Geosciences pour une Terre durable

brgm

LES REGISTRES

Un langage commun pour les géosciences



Quelle problématique ?



J'observe sur cet affleurement, la présence potentielle **Mispickel (FeAsS)**, près de Beaulieu



Analysons les métaux lourds dans cet échantillon prélevé sur Beaulieu, alors, **As, Fe, Pb, ...**



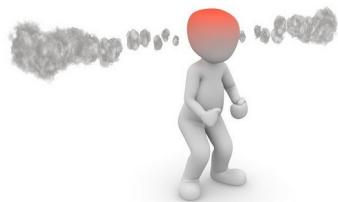
J'observe sur cet affleurement, la présence potentielle **Arsenopyrite**



What a great **dalarnite !**



Tiens, en creusant ici j'ai trouvé du **fer !**

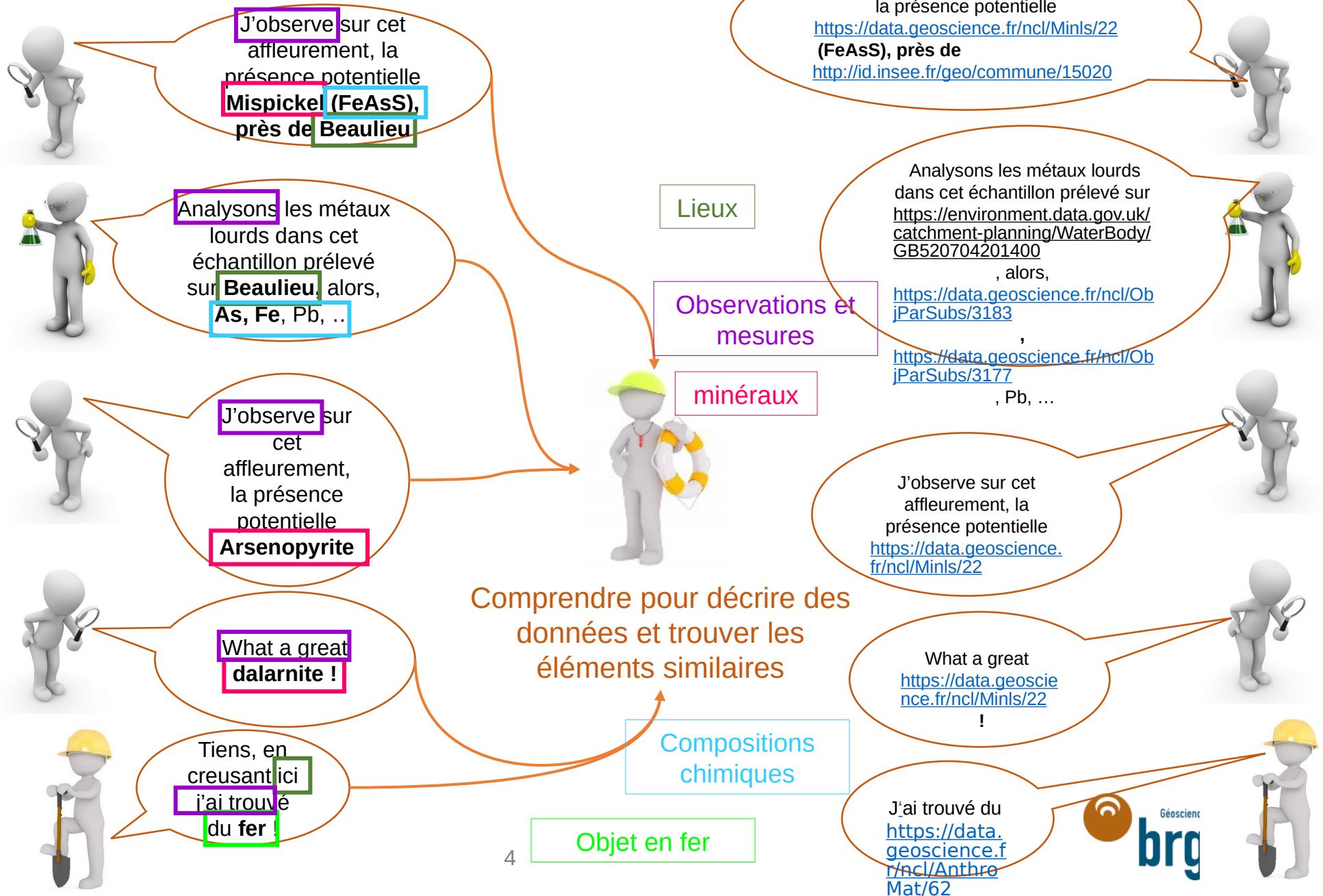


Alors, le Mispickel c'est quoi ? Je le range avec les analyses de Fe et As ? L'Arsenopyrite un autre mineraux ?





Partager le même langage





Partager le même langage

Comprendre pour décrire des données et trouver les éléments similaires

J'observe sur cet affleurement, la présence potentielle
<https://data.geoscience.fr/ncl/Minls/22>
(FeAsS), près de
<http://id.insee.fr/geo/commune/15020>



Analysons les métaux lourds dans cet échantillon prélevé sur
<https://environment.data.gov.uk/catalogue/planning/WaterBody/GB520704201400>, alors,
<https://data.geoscience.fr/ncl/ObjParamSubs/3183>



<https://data.geoscience.fr/ncl/ObjParamSubs/3177>, Pb, ...



J'observe sur cet affleurement, la présence potentielle
<https://data.geoscience.fr/ncl/Minls/22>

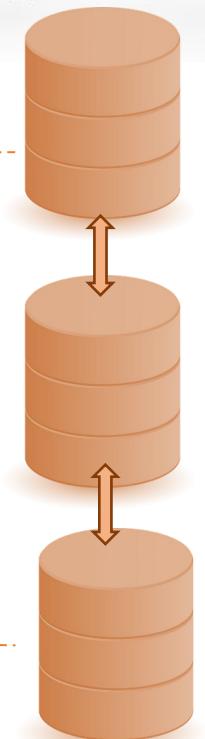
What a great
<https://data.geoscience.fr/ncl/Minls/22> !



J'ai trouvé du
<https://data.geoscience.fr/ncl/AnthroMat/62>



Interopérabilité des données





Un pont entre deux univers

Connaître les systèmes
d'informations et les standards

Comprendre les besoins et les
données thématiques pour les
structurer dans un système
d'information

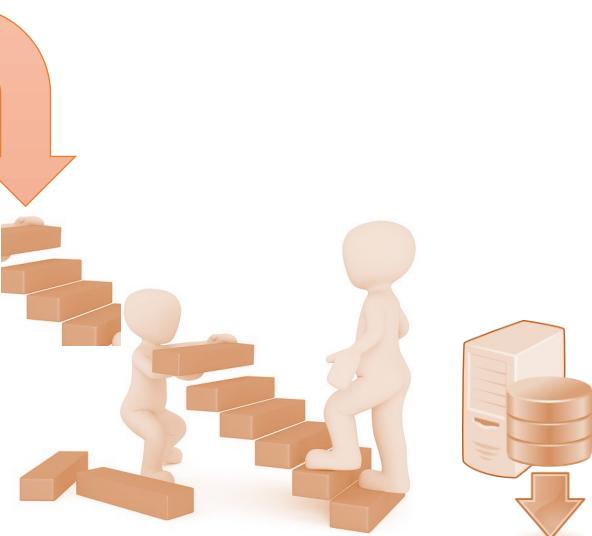


Monde thématique

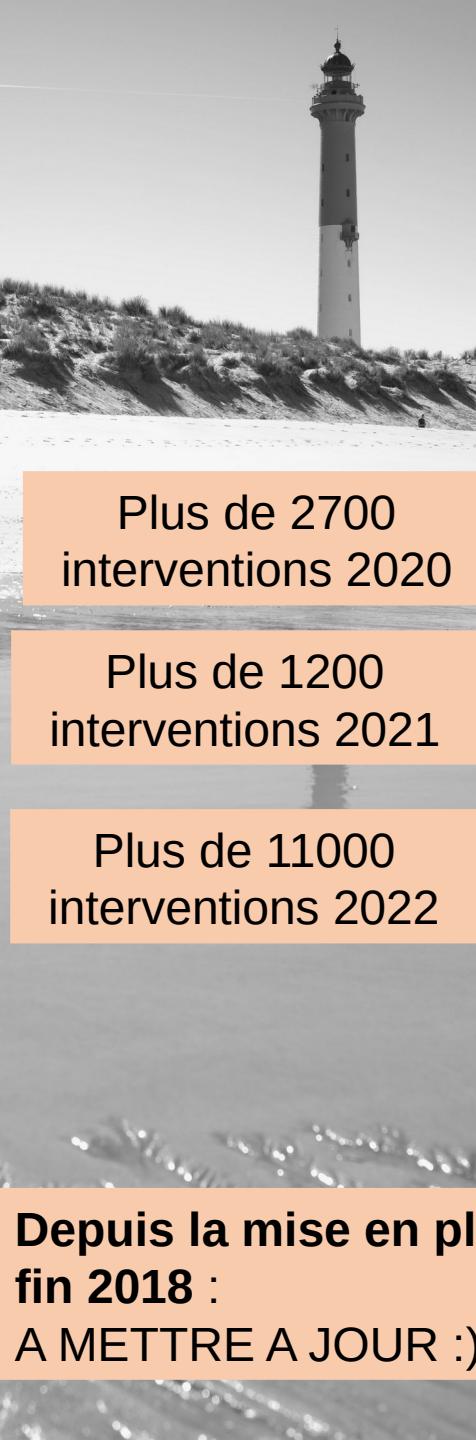
Élabore les listes de termes et leurs
définitions
Précise leurs besoins



Ingénieur de la connaissance, data
qualiticien, architecte sémantique...



Monde IT



Plus de 2700 interventions 2020

Plus de 1200 interventions 2021

Plus de 11000 interventions 2022

Depuis la mise en place fin 2018 :
A METTRE A JOUR :)

En pratique au BRGM



- Définition des données

- Besoin de vocabulaires, registres

- Échanges avec les thématiciens

- Structuration du registre

- Initialisation du registre

- Publication du registre

- Utilisation du registre

Thématiciens



ArchiData



IT



Près de 600 interventions sur les registres dont 330 créations d'éléments en 2020



Registres & Vocabulaires - définitions

Registres

Ensemble de termes homogènes, codifiés et répertoriés (nomenclature, thesaurus, ...) chacun possédant **un identifiant unique externe stable sous la forme d'une HTTP URI**. Ces termes sont cohérents entre eux selon un domaine donné. Un registre peut être matérialisé par différentes représentations informatiques (CSV, JSON-LD, TTL, RDF/XML) selon la requête envoyée à l'outil de gestion de registre.

Outil de gestion des registres du BRGM

Explorer À propos Avancé ▾ Rechercher Soumettre

https://data.geoscience.fr/ncl/_litho

Registre: Lithologie

URI: <https://data.geoscience.fr/ncl/litho>

Ce registre contient la classification lithologique du BRGM. Elle définit l'ensemble des matériaux naturels résultant de la cristallisation de la roche ou de la dissolution des minéraux dans les sols et les eaux. Il comprend les roches magmatiques, métamorphiques et sédimentaires, ainsi que les roches météoriques et les roches artificielles.

Contenu (vue arborescente)

- Roche magmatique Roche résultant de la cristallisation d'un magma
 - + Roche plutonique Roche ignée cristalline phanéritique ayant cristallisées lentement...
 - + Carbonatite Roche ignée aphanitique à cristalline ayant cristallisées à partir de...
 - + Syénitoïde foidique Roche ignée cristalline phanéritique constituée de feldspatho...
 - + Roche volcanique Roche ignée aphanitique, à grain fin et contenant souvent du ...
 - + Gabbroïde foidique Roche ignée cristalline phanéritique constituée de feldspatho...
 - + Dioritoïde foidique Roche ignée cristalline phanéritique constituée de feldspatho...

Géosciences pour une Terre durable

brgm

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix version: <http://purl.org/linked-data/version#> .
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix dgu: <http://reference.data.gov.uk/def/reference#> .
@prefix ui: <http://purl.org/linked-data/registry-ui#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix reg: <http://purl.org/linked-data/registry#> .
@prefix ldp: <http://www.w3.org/ns/ldp#> .
@prefix time: <http://www.w3.org/2006/time#> .
@prefix api: <http://purl.org/linked-data/api/vocab#> .
@prefix vann: <http://purl.org/vocab/vann/> .
@prefix vs: <http://www.w3.org/2003/06/sw-vocab-status/ns#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
```

```
<https://data.geoscience.fr/ncl/litho/189>
a skos:Concept ;
rdfs:label "Clinopyroxénolite"@fr ;
dct:description "Pyroxénite constituée à plus de 90% de clinopyroxène."@fr ;
owl:sameAs <http://data.bgs.ac.uk/id/EarthMaterialClass/RockName/CPXT> ;
skos:broader <https://data.geoscience.fr/ncl/litho/183> ;
skos:inConceptScheme <https://data.geoscience.fr/ncl/litho> ;
skos:prefLabel "Clinopyroxénolite"@fr ;
<http://www.w3.org/ns/adms#status>
<http://inspire.ec.europa.eu/registry/status/valid> ;
<https://data.geoscience.fr/def/litho#mnemo>
"CPYR" ;
<https://data.geoscience.fr/def/litho#nature>
<https://data.geoscience.fr/ncl/natelem/1> ;
<https://data.geoscience.fr/def/litho#rank>
<https://data.geoscience.fr/ncl/lithorank/5> ;
<https://schema.org/sameAs> <http://www.bgs.ac.uk/bgsrscs/rcc\_details.cfm?code=CPXT> .
```

```
<https://data.geoscience.fr/ncl/litho/510>
a skos:Concept ;
rdfs:label "Phyllonite"@fr ;
dct:description "Mylonite riche en phyllosilicates et qui a l'éclat lustré brillant d'une phyllite"@fr ;
skos:broader <https://data.geoscience.fr/ncl/litho/506> ;
skos:exactMatch <http://inspire.ec.europa.eu/codelist/LithologyValue/phyllonite> ;
```



Structuration des registres – mise en place

Les registres répondant à une description thématique

- Registres de la connaissance (registres « métiers »)
- Ils répondent à un besoin de décrire un objet métier et permettent de structurer la connaissance

Lithologie	Minéraux	Matériaux anthropiques	Type plan
Fossiles	Traces fossiles	Types de supports d'observation	Type ligne

Les registres répondant à une organisation des données d'observations

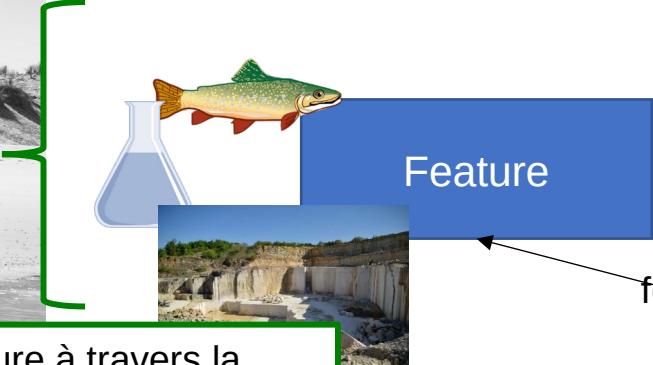
- Registres « O&M »
- Ils répondent à un besoin de structuration des données d'observations et facilitent l'interopérabilité entre les « silos métiers »

Propriétés observées	Objets d'intérêt	Type de capteurs	Méthodes / process	Unités de mesures
----------------------	------------------	------------------	--------------------	-------------------

Structuration des registres – mise en place

Les registres O&M et les registres de la connaissance

Feature / Objets métiers définis par un modèle de données



Description du feature à travers la Connaissance:

<https://data.geoscience.fr/ncl/litho>

<https://data.geoscience.fr/ncl/apt>

<https://data.geoscience.fr/ncl/ObjParSubs>

Méthodes / process :
<https://data.geoscience.fr/ncl/Proc>

Propriétés observées:
<https://data.geoscience.fr/ncl/ObsProp>

Property

Observation

Values

Process

featureOfInterest

observedProperty

result

Unités de mesure :
<https://data.geoscience.fr/ncl/uom>

Salmo trutta

8,2 mg/L

Connaissances:

<https://data.geoscience.fr/ncl/litho>

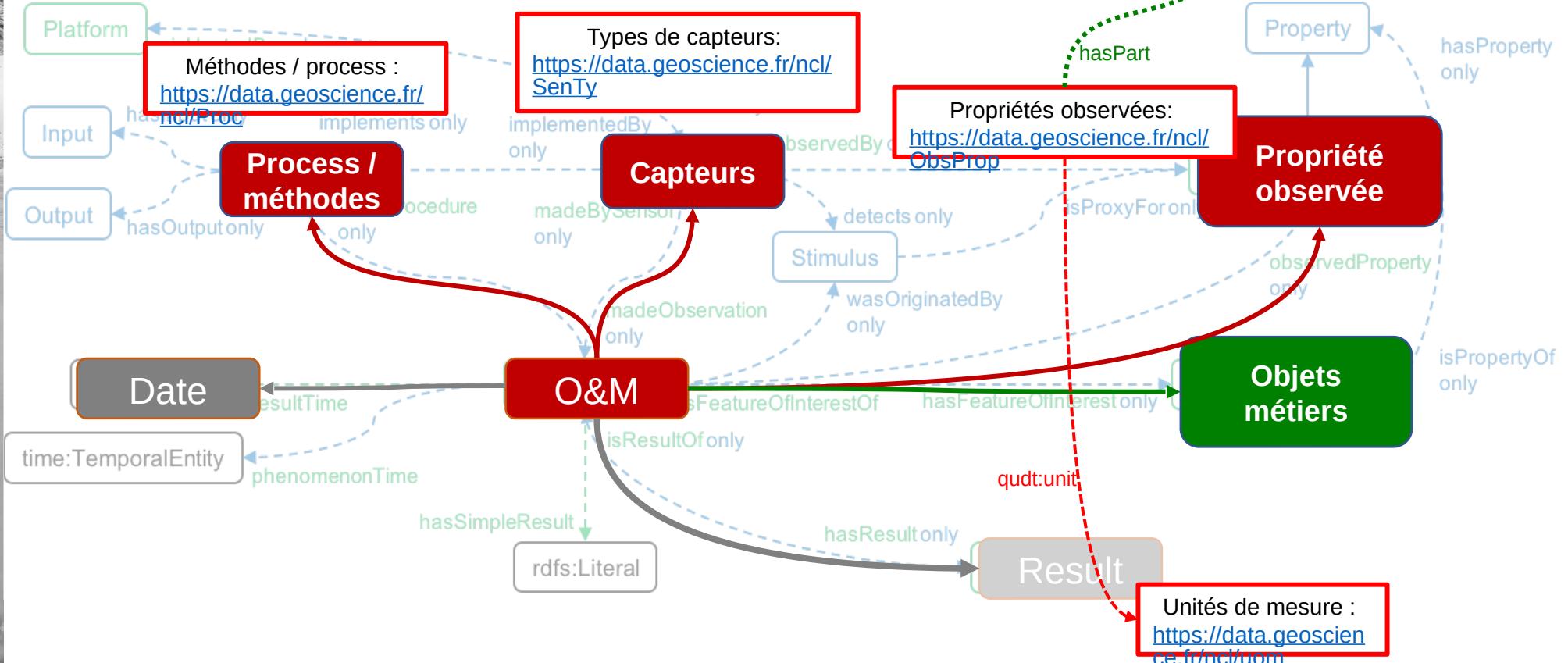
<https://data.geoscience.fr/ncl/apt>

<https://data.geoscience.fr/ncl/ObjParSubs>

Structuration des registres – les standards internationaux

Les registres O&M et les registres de la connaissance

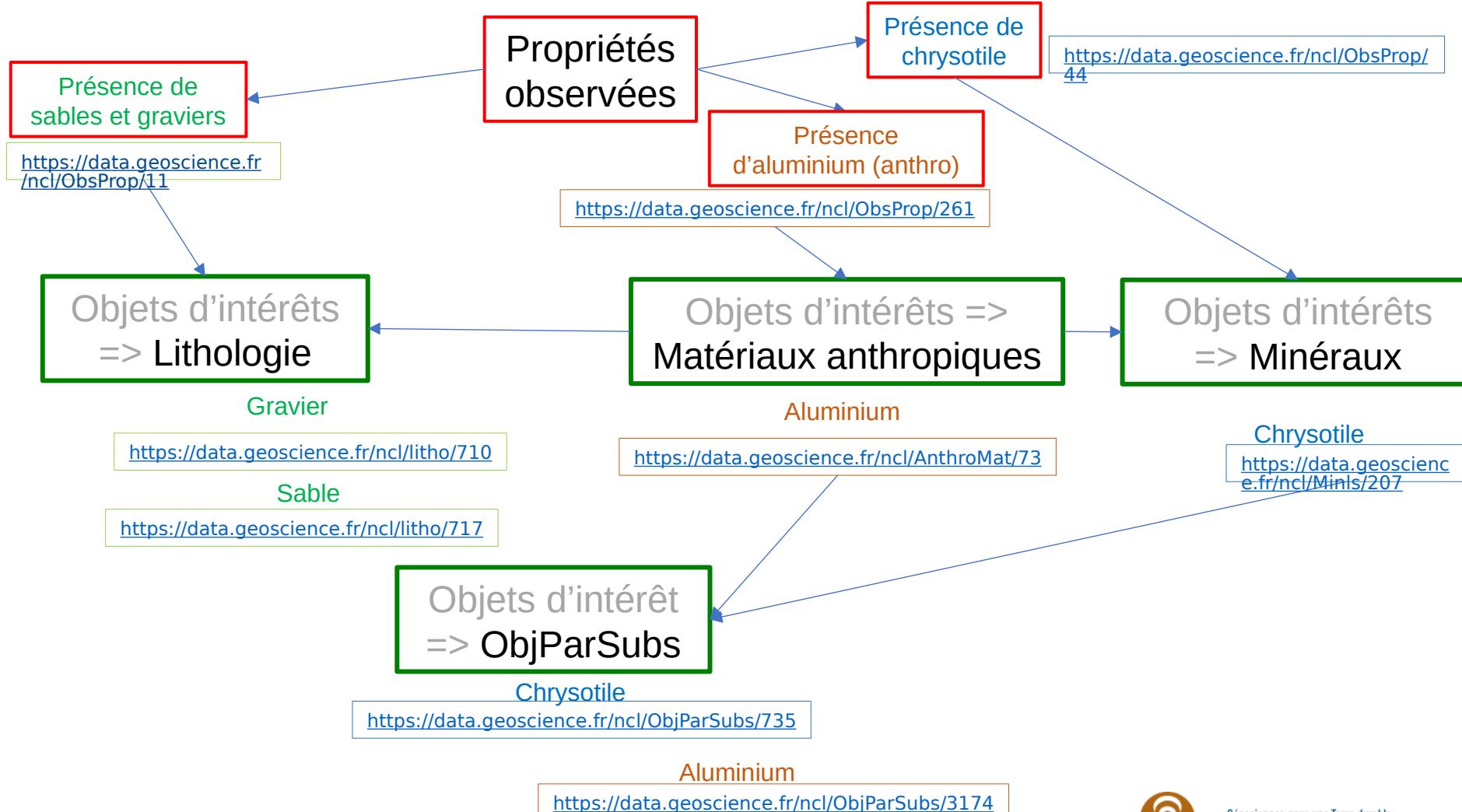
Connaissances:
<https://data.geoscience.fr/ncl/litho>
<https://data.geoscience.fr/ncl/apt>
<https://data.geoscience.fr/ncl/ObjParSubs>





Structuration des registres

Des liens entre les registres « O&M » et les registres de la connaissance



Registres O&M et registres de la connaissance

Propriétés observées	Résultats observation	Unité d'expression du résultat	Méthode d'observation	Sur quoi ?
registre ObsProp https://data.geoscience.fr/nd/ObsProp	xxx	registre uom https://data.geoscience.fr/nd/uom	registre Proc https://data.geoscience.fr/nd/Proc	feature =Objet métier URI Objet métier
Présence de calcaire	Oui	-	CIEL du géologue	"Formation des calcaires du Maastrichtien"
https://data.geoscience.fr/nd/ObsProp/91	Oui	https://data.geoscience.fr/ncl/uom/su	https://data.geoscience.fr/ncl/Proc/xx	Unité géologique de code yyyy
Température de l'eau https://data.geoscience.fr/nd/ObsProp/187	8	8 °C https://data.geoscience.fr/ncl/uom/607	Sonde https://data.geoscience.fr/ncl/Proc/yy	La Dordogne à Lalinde http://id.eaufrance.fr/CEA/P--0000
Canette de soda https://data.geoscience.fr/nd/O	Oui		CIEL du thématicien https://data.geoscience.fr/ncl/Proc/xx	Prélèvement effectué sur les berges de la Dordogne Prélèvement de code XXXX



Les registres O&M permettent de structurer une observation de manière similaire

Les registres de la connaissance permettent de structurer les spécificités métiers et thématiques



RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité



Geosciences pour une Terre durable

brgm

ANR APRIL

Besoins et solutions pour une capitalisation structurée du RETEX

Auteur(s)

Date



DE QUELLES DONNÉES DISPOSE T'ON?

EDUCATION - JUSTICE

- LOCAL EDUCATION AUTHORITY
- COURTHOUSE

LOGISTICS - TOWN PLANNING

- HARBOR
- DIRECTION SEA
- DEAL⁽¹⁾
- URBAN PLANNING POLICY

HEALTH - SOCIAL

- FAMILY SOLIDARITY CENTRE
- REGIONAL HEALTH OFFICE
- FRENCH CROIX-ROUGE
- EMERGENCY MEDICINE DEPARTMENT

SECURITY - EMERGENCY SERVICES

- FIRE RESCUE CENTRE
- POLICE FORCE
- BORDER POLICE
- CUSTOMS

NETWORK OPERATORS

- EDF
- ORANGE
- DAUPHIN TELECOM

ADMINISTRATIVE OFFICES

- COMMUNICATION
- OFFICE MANAGEMENT
- SIDPC⁽³⁾

CIC - COGIC⁽⁴⁾

EMBASSY NETHERLANDS

COZ⁽⁵⁾

Légende :

- Homeland
- Martinique
- Guadeloupe
- Saint-Martin

⁽¹⁾ Environment, Planning, and Accommodation Direction

⁽²⁾ Geology and Mining Research Institute

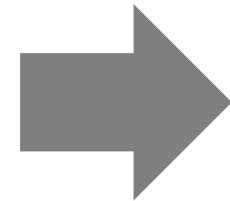
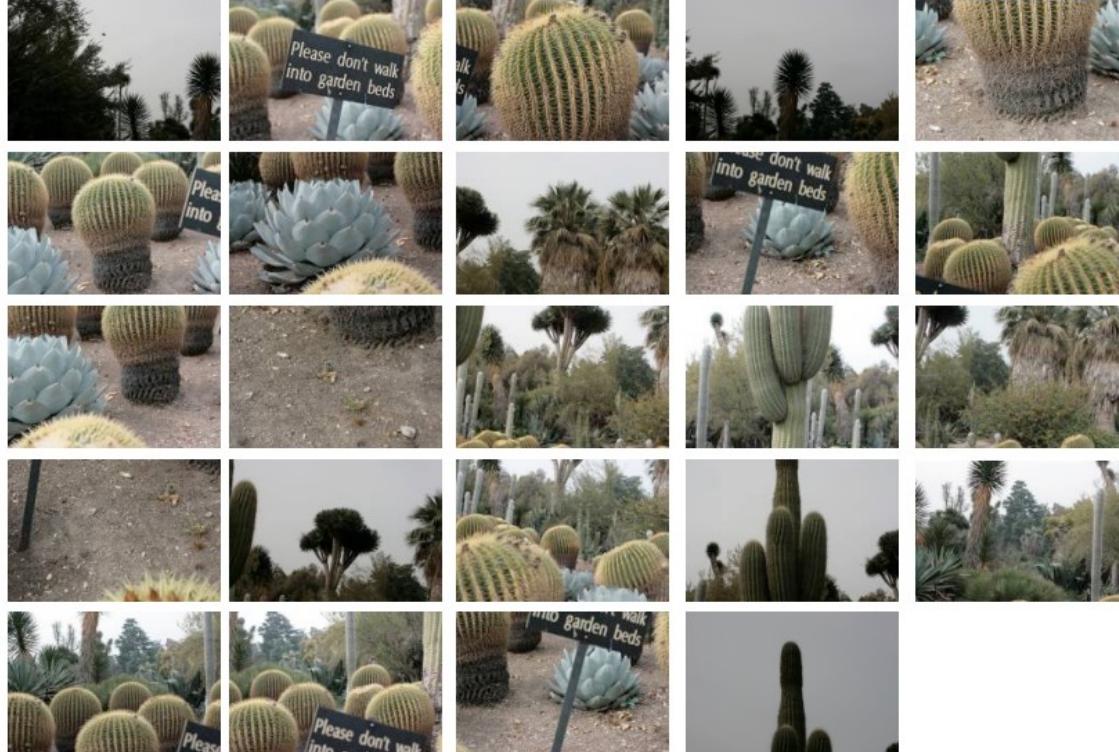
⁽³⁾ Interministerial Civil Defense and Protection Office

⁽⁴⁾ Interministerial Crisis Unit - Interministerial Operational and Crisis Management Centre

⁽⁵⁾ Operational Area Centre

46 interviews -Average interview time: 1h45 (30 minutes to more than 3 hours)

Les besoins et les défis



Restitutions fragmentaires de la même
expérience...¹⁶

A close-up photograph of a man's face. He has dark hair, wears round-rimmed glasses, and has a beard. He is looking slightly to his left with a contemplative expression, resting his chin on his hand. The background is dark and out of focus.

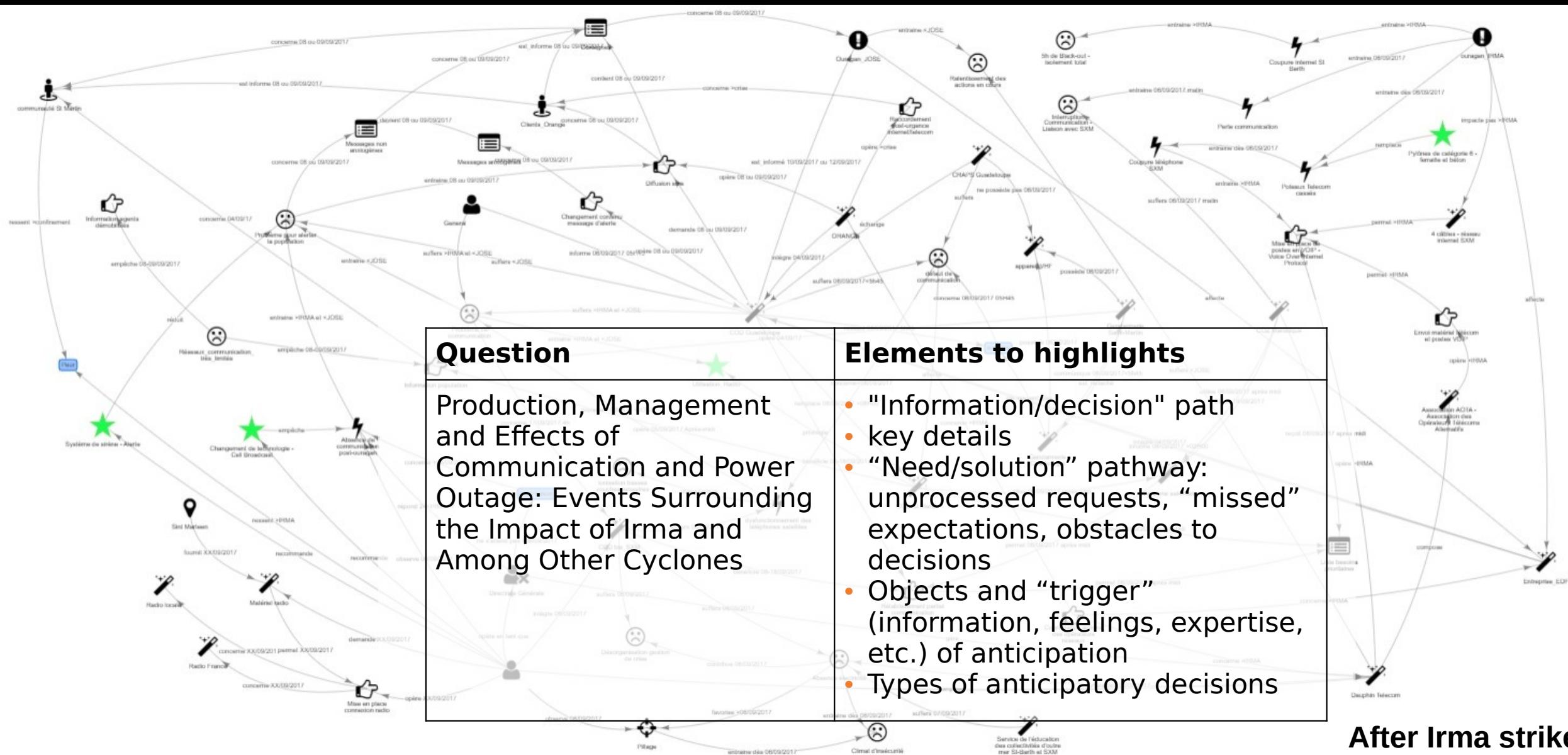
... multi-
perspective

Défi

- Transformer une narration en données en conservant

la richesse
des
informations





Traduire l'expérience en « graphe des connaissances

Finalités

- Capitaliser sur les expériences
- Créer des outils d'aide à la décision

raphe des connaissances

Théorie des Catégories et ses applications concernent l'intérêt et dont les liens représentent des relations entre ces entités.

Un topo élémentaire est une catégorie avec des limites finites,

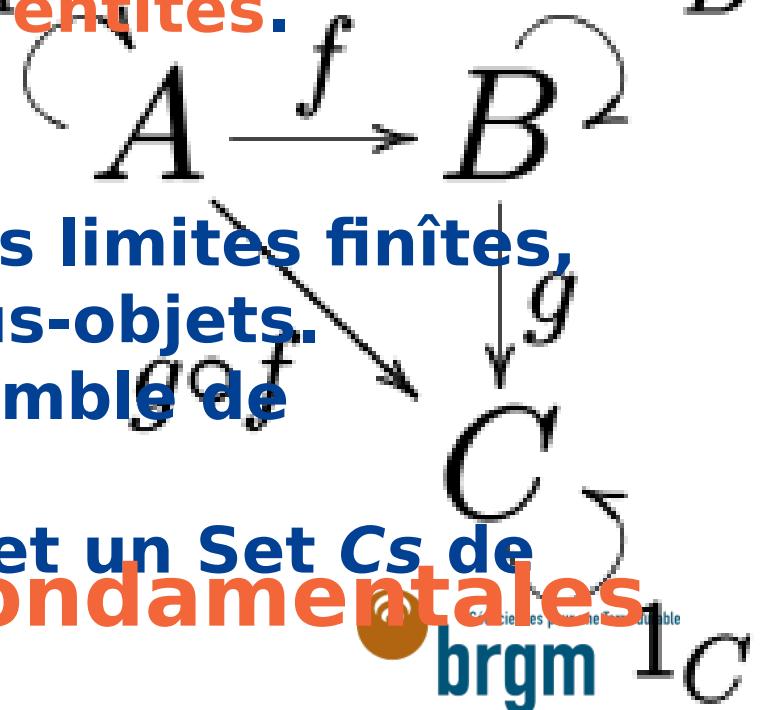
est fermé cartésien et a un classificateur de sous-objets.

Un sketch consiste en un graphique G_s , un ensemble de diagrammes

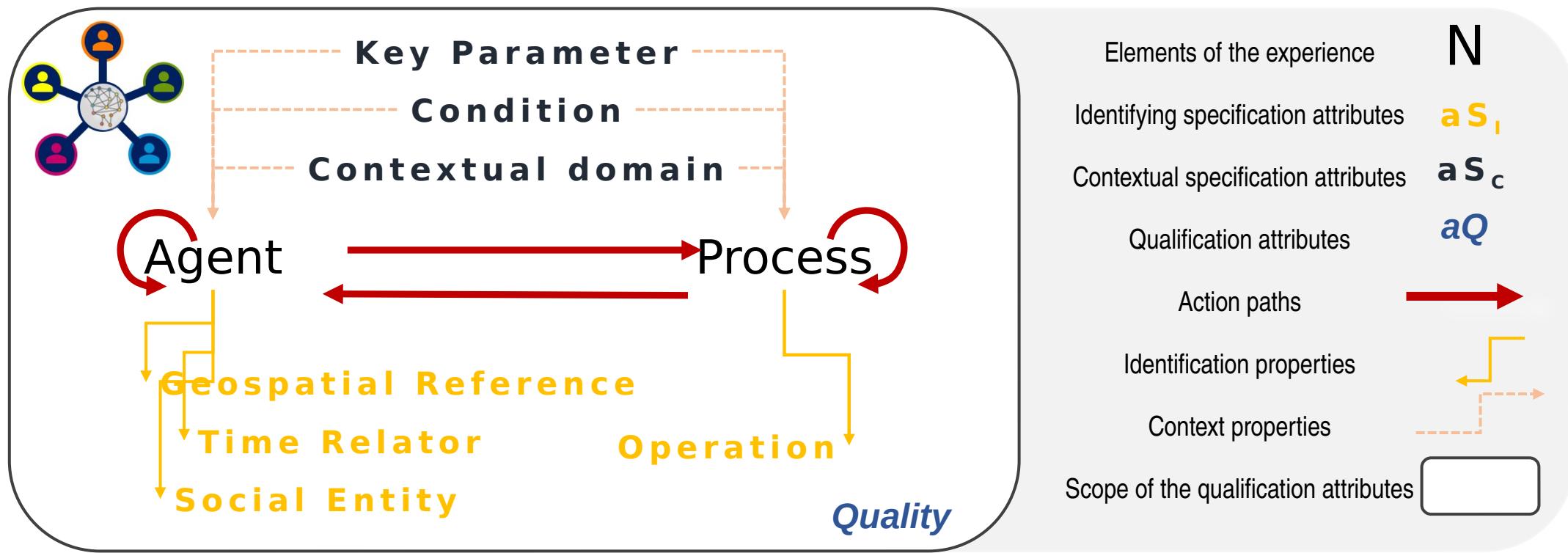
D_s en G_s , un ensemble (Set) L_s de cônes en GS et un Set C_s de co-cônes en GS .

un graphique de données destiné à accumuler et à transmettre la connaissance du monde réel, dont les nœuds représentent des entités

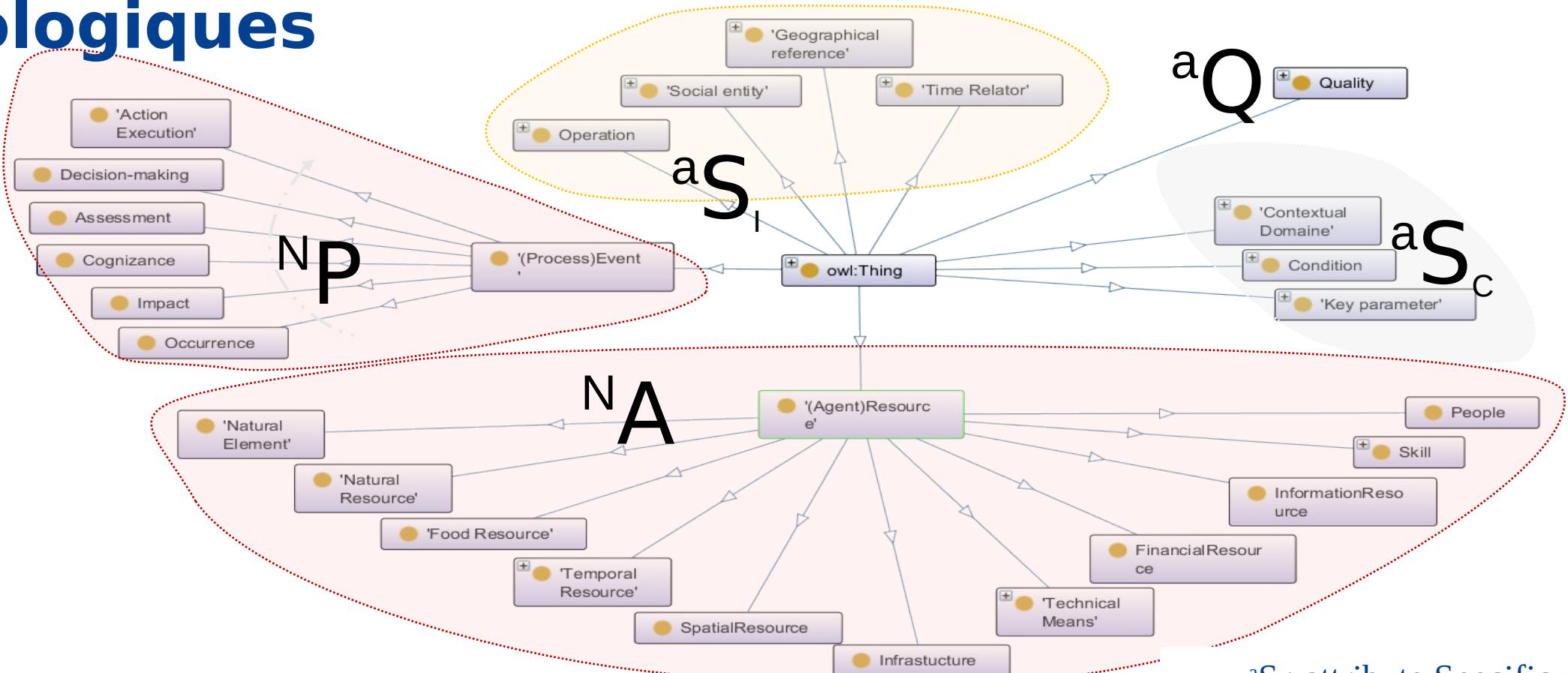
des relations entre ces entités.



Conceptualisation: modèle sémantique de données



Conceptualization: concepts ontologiques



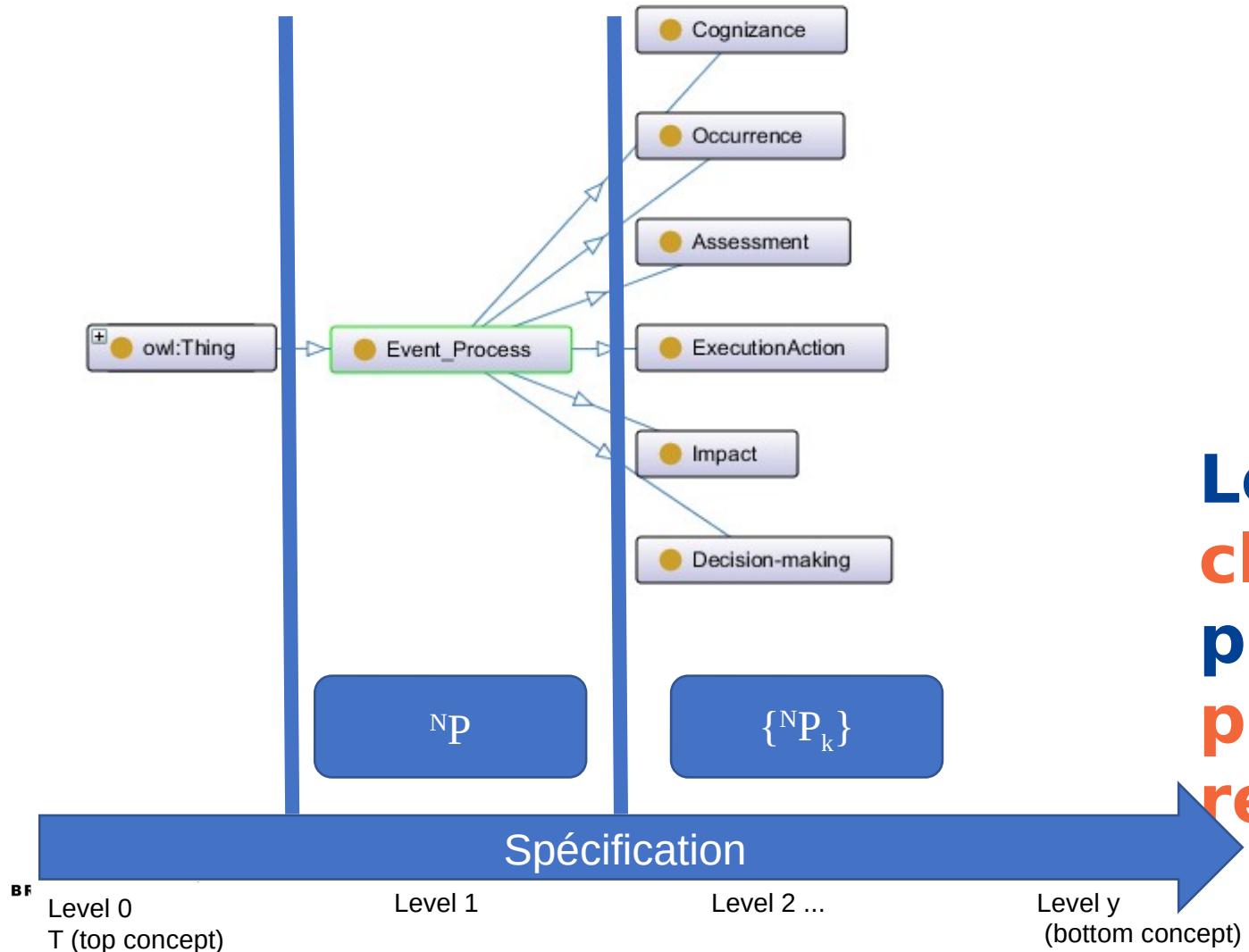
$^N A$: élément (ou Nom) Agent ou ressource

$^N P$ = élément (ou Nom) Processus, évènement, ou action

$^a S_I$: attribute Specific d'Identité

$^a S_C$: attribute Specific de Contexte

$^a Q$: attribute Qualification



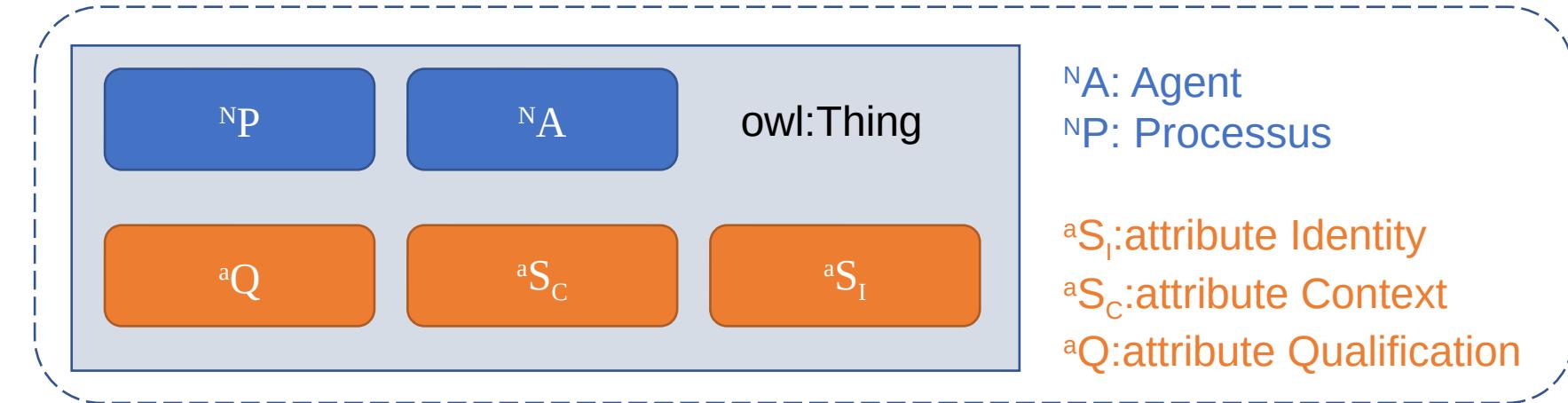
Les informations doivent être :

structurées 

catégorisées 

Les entonnoirs de classification choisis permettent de changer de perspective et de résolution.

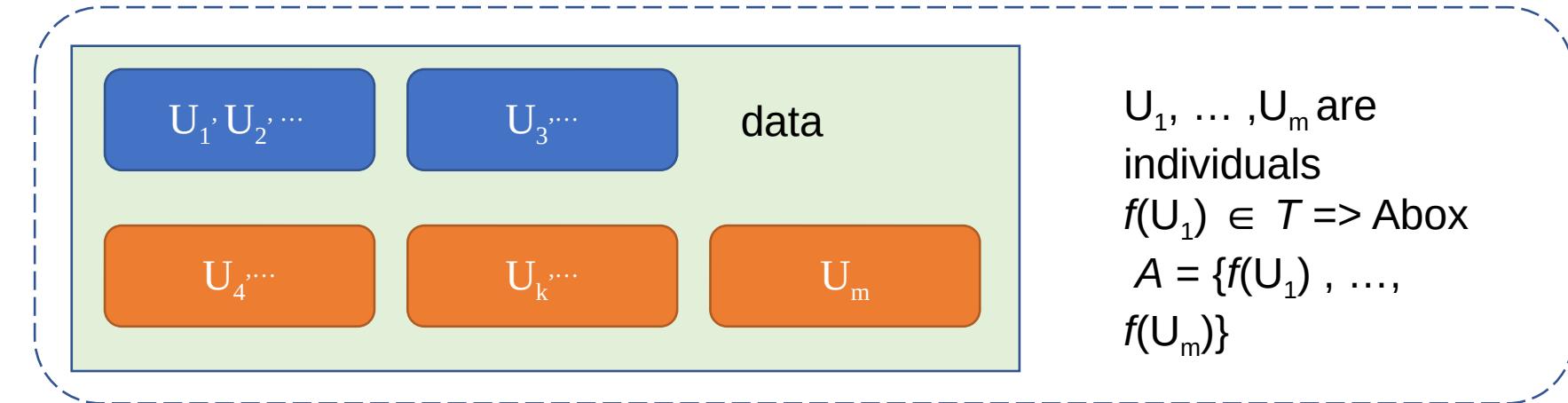
Let $T = \{^N A, ^N P, ^A S_I, ^A S_C, ^A Q\}$, it represents the TBox component, a set of ontology classes at the highest level.
 The dataset is represented by a set of instances, $U = \{u_1, \dots, u_m\}$



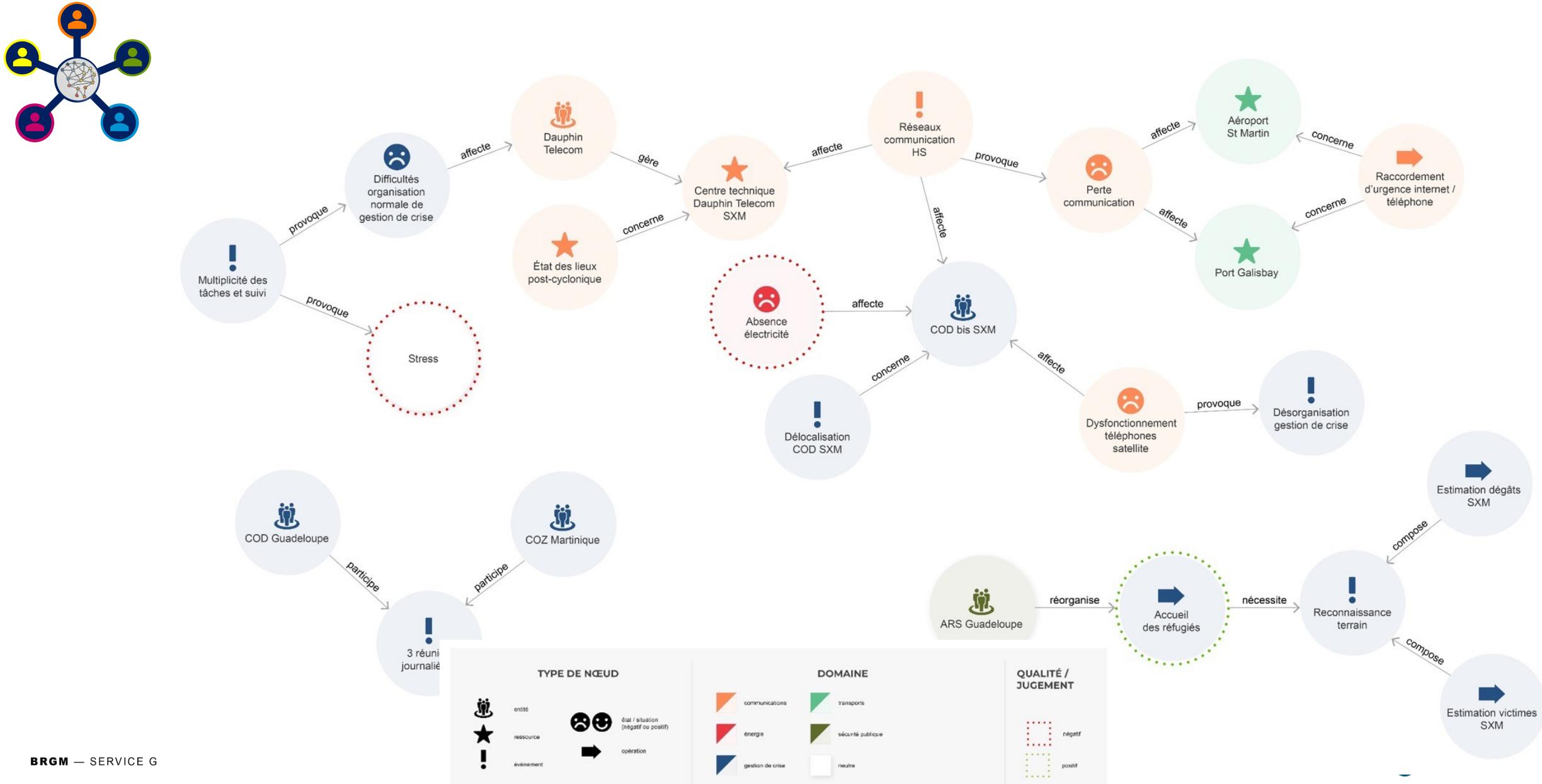
$^N A$: Agent
 $^N P$: Processus

$^A S_I$: attribute Identity
 $^A S_C$: attribute Context
 $^A Q$: attribute Qualification

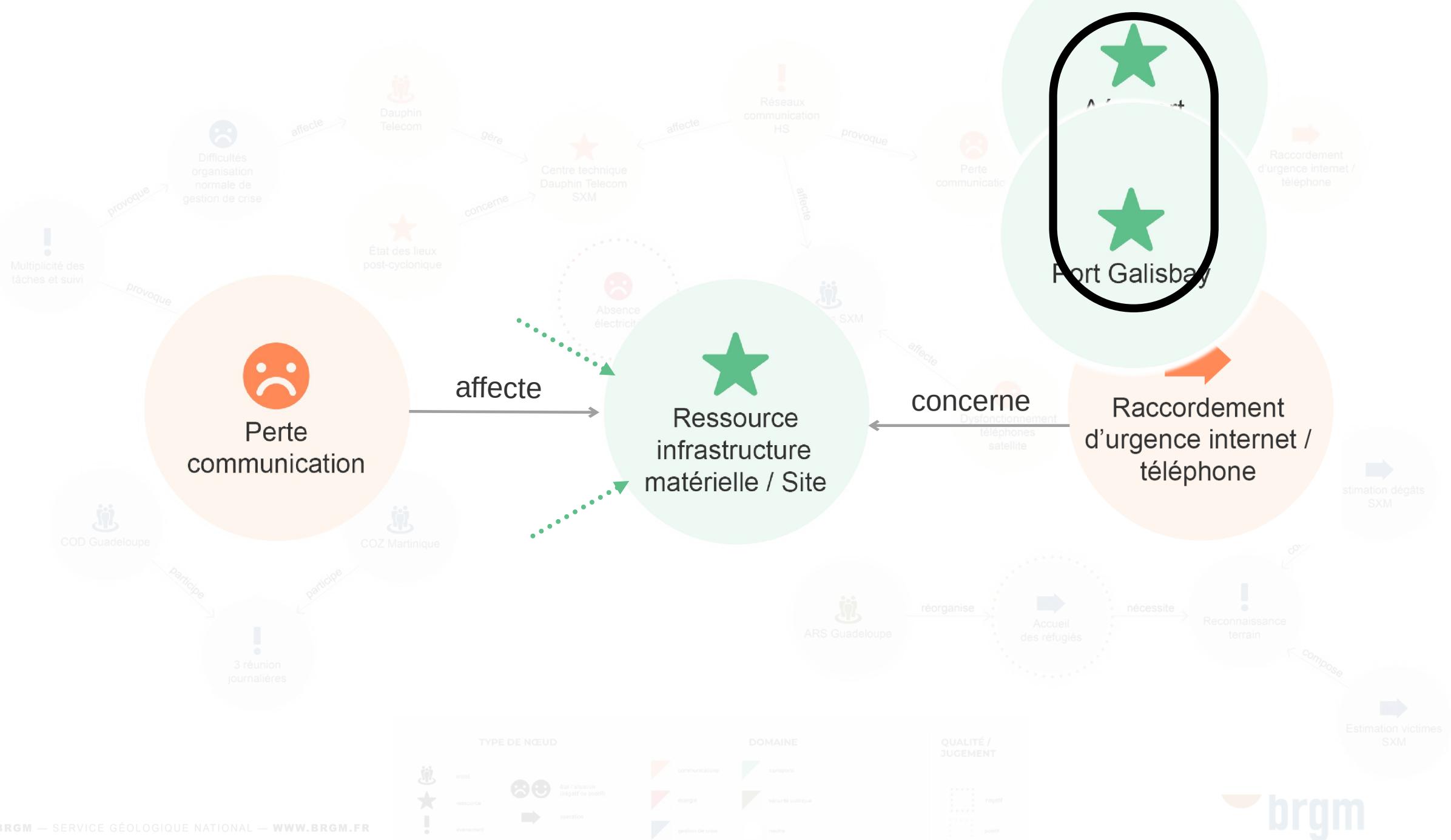
Let $f: A \rightarrow T$ be the mapping from real world instances (individuals) to abstraction (Tbox)
 The knowledge base K is defined by $K = (T, f(U))$



U_1, \dots, U_m are individuals
 $f(U_1) \in T \Rightarrow Abox$
 $A = \{f(U_1), \dots, f(U_m)\}$











RÉPUBLIQUE
FRANÇAISE

Liberté
Égalité
Fraternité

FAIRISATION DE DONNÉES AU BRGM

Auteur(s)

Date



Geosciences pour une Terre durable

brgm

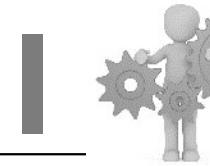


Mise en place d'un catalogue entrepôt de données

CUPIDON - Respect des Principes FAIR

Données

Métadonnées

FINDABLE 	F1 : Données et métadonnées ont un identifiant unique et pérenne	<input checked="" type="checkbox"/>	
	F2 : Données décrites par des métadonnées riches	<input checked="" type="checkbox"/>	
	F3 : Données et métadonnées enregistrées ou indexées dans un dispositif permettant de les rechercher		
	F4 : Métadonnées contiennent l'identifiant unique et pérenne des données	<input checked="" type="checkbox"/>	
ACCESSIBLE 	A1 : Données et métadonnées accessibles par leur identifiants via un protocole standard		
	A1.1 : Le protocole utilisé est ouvert, libre et peut-être implémenté de manière universelle	<input checked="" type="checkbox"/>	
	A1.2 : Le protocole peut gérer des procédures d'authentification et d'autorisation si nécessaire		
	A2 : Métadonnées accessibles même si les données ne sont plus disponibles	<input checked="" type="checkbox"/>	
INTEROPERABLE 	I1 : Données et métadonnées utilisant un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances		
	I2 : Données et métadonnées utilisant des vocabulaires suivant eux-mêmes les principes FAIR	<input checked="" type="checkbox"/>	
	I3 : Données et métadonnées contenant des liens vers d'autres (méta)données	<input checked="" type="checkbox"/>	
REUSABLE 	R1 : Données et métadonnées richement décrites, par une pluralité d'attributs précis	<input checked="" type="checkbox"/>	
	R1.1 : Fournies avec une licence d'utilisation claire	<input checked="" type="checkbox"/>	
	R1.2: Associées à leur provenance	<input checked="" type="checkbox"/>	
	R1.3 : Suivent les standards des communautés		



CUPIDON : un périmètre large

CUPIDON « interne BRGM »

- « Grandes données »
 - ⇒ Données d'observation du sous-sol (BSS)
 - ⇒ Données d'observation des eaux souterraines (ADES)

- « Données externes »
 - ⇒ Données nécessaires à la réalisation des projets (IGN, données du SHOM, Corin Land Cover, ...)

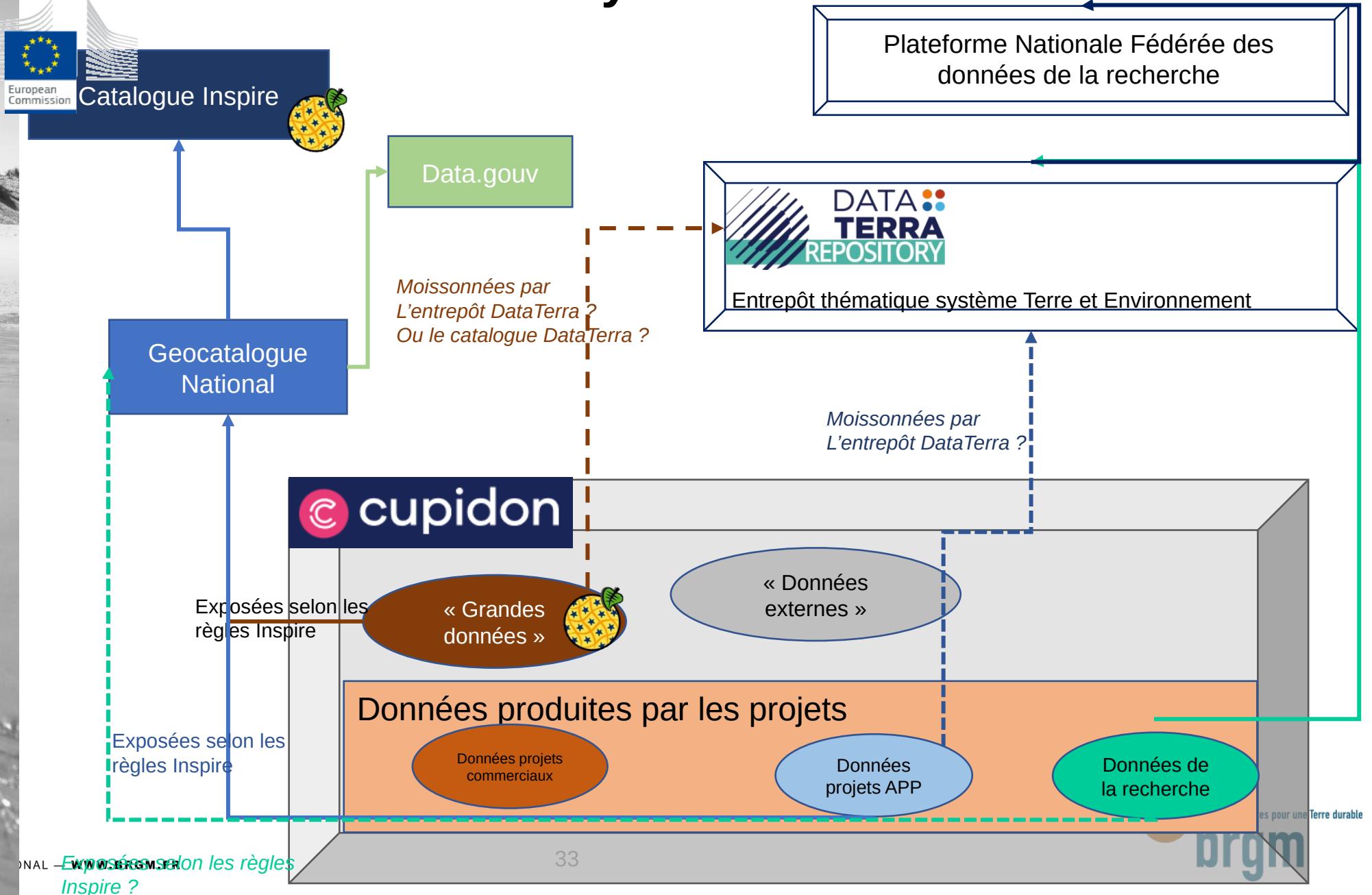
Données produites par les projets

Données projets commerciaux

Données projets APP

Données de la recherche (projets ANR, H2020, ...)

CUPIDON : dans l'écosystème Science Ouverte



Merci