



# Ontologie d'objets repères CHOUCAS

---

Projet ANR CHOUCAS (2017-2021)

Intégration de données hétérogènes et raisonnement spatial pour l'aide  
à la localisation des victimes en montagne



**Ana-Maria Raimond, Sébastien Mustière, Mattia Bunel, Catherine Domingues, Cécile Duchêne, Laurence Jolivet, Marie-Dominique Van Damme.**

Livrable ANR – IGN / LaSTIG – Projet CHOUCAS

7 février 2020 – v0.3



## 1 Objet du document

Ce document décrit les motivations, le contenu et le processus de création de l'ontologie dite « ontologie des objets de repère » (OOR) réalisée dans le cadre du projet CHOUCAS (ANR-16-CE23-0018). L'ontologie est consultable sur : <http://choucas.ign.fr/doc/ontologies/index-fr.html>

## 2 Contexte

Le projet **CHOUCAS** a pour objectif de proposer des méthodes et des outils permettant de constituer et enrichir des données géographiques issues de sources hétérogènes, et des modèles de raisonnement afin de favoriser le processus de décision dans le contexte de la localisation des victimes en montagne. Plusieurs ontologies sont définies dans ce cadre, sur lesquelles s'appuient les mécanismes de raisonnement et de recherche d'information du projet. L'ontologie décrite ici est l'ontologie des objets de repère (OOR) qui vise à recenser et organiser l'ensemble des types d'objets par rapport auxquels une personne peut se repérer en montagne (lac, rivière, route, forêt, glacier, sommet...).

En raison de ce contexte, on peut noter que l'ontologie est, d'une part, focalisée sur la description du domaine de la montagne en France métropolitaine et, d'autre part, focalisée sur la description de ce qui est visible sur le terrain.

## 3 Motivations

Cette ontologie joue tout d'abord un rôle de pivot ou de vocabulaire contrôlé, utile pour indexer les sources d'information et pour permettre l'interopérabilité des outils développés dans le cadre du projet. Elle a également un rôle d'organisation des concepts, utile en particulier pour évaluer les proximités sémantiques lors de l'appariement de sources de données hétérogènes ou plus généralement de la prise en compte de confusions possibles entre types d'objets.

On peut lister plus précisément un ensemble de cas d'usage typiques envisagés de l'ontologie :

- Indexation d'une base de données de points de repères pour en faciliter l'interrogation, à l'image d'un thésaurus pour organiser une bibliothèque. L'interrogation se fera dans ce cas en identifiant d'abord le concept recherché.
- Interrogation de données sur une base documentaire non maîtrisée, comme le Web. L'interrogation par mot clef s'appuiera dans ce cas sur l'ontologie pour rechercher les différents labels du concept recherché ou de concepts proches, voire de concepts plus éloignés en cas de volonté d'élargir une requête infructueuse.
- Spécification de domaine d'application de méthodes définies dans le cadre du projet. Par exemple, on pourra s'appuyer sur l'ontologie pour spécifier que l'outil « recherche des objets de forme pointue » s'applique aux concepts de pic, aiguille ou sommet mais pas aux concepts de maison ou grotte.



- Servir de référence à d'autres ontologies ou d'autres connaissances en général. Par exemple une ontologie des activités (ex. escalade) pourra faire référence aux types d'objets de repère importants pour cette activité (ex : falaise). Par exemple encore, un score de confusion pourra être défini entre tous les couples possibles de concepts de l'ontologie.
- Guider l'appariement de données de sources hétérogènes. Par exemple, l'ontologie permettra d'apparier des lieux classés comme « sommet » dans une source et comme « pic » dans une autre, car ces deux concepts sont proches dans l'ontologie.

## 4 Principes de modélisation

Owl a été choisi comme langage de représentation de l'ontologie pour des raisons d'interopérabilité, cependant tout son pouvoir d'expressivité n'a pas été exploité.

Les usages envisagés de l'ontologie sont proches de l'usage habituel d'un thésaurus ou d'une taxonomie. Celle-ci sera plus utilisée pour naviguer de concepts proches en concepts proches, que pour réaliser un raisonnement en logique formelle. Pour ces raisons les choix de modélisation suivants ont été effectués :

- Une organisation principalement sous la forme d'une hiérarchie de types d'objets, chaque type étant une classe au sens du langage owl.
  - A noter : cette hiérarchie n'est pas stricte, des classes pouvant avoir plusieurs parents. L'organisation principalement hiérarchique est un donc principe général et non une contrainte de modélisation.
- Chaque classe de l'ontologie est associée à des labels, potentiellement multiples. Chaque classe possède au minima un label français et un label anglais. Chaque classe peut également posséder plusieurs labels, en cas de termes spécifiques (spécificités locales, spécificités de domaines spécialisés comme la spéléologie par exemple). Sur le modèle du thésaurus RAMEAU de la BNF, un unique label est qualifié de 'préféré' par langue (annotation 'prelabel'), et les éventuelles autres labels sont qualifiés d'alternatifs (annotation 'altlabel')
  - A noter : Un choix délicat a été de choisir entre spécialiser une classe en sous classes d'une part, et associer plusieurs labels à une même classe d'autre part (par exemple, un « gave » doit-il être vu comme un label de rivière lié à un régionalisme ou une sous-classe de torrent ?). Dans ce cas, la règle retenue, malgré son imprécision, a été d'arrêter la spécialisation en sous-classe aux termes estimés connus d'un public de randonneur averti, mais non nécessairement spécialiste de la montagne.
- Une définition a été fournie pour certaines classes, sans que cela soit systématique
- Des propriétés d'objets (ObjectProperties) ont été définies pour modéliser les spécificités des objets ou celles du secours en montagne. Parmi ces propriétés on peut noter : `estComposéDe` ; `peutEtreColocaliséAvec` (pour modéliser des relations particulières entre des objets liés sur le terrain comme les névés et les glaciers) ; `peutEtreConfonduAvec` (pour modéliser les objets qui peuvent être confondu de par leur forme, comme les lignes électriques et les câbles transporteurs) ; `aPourForme` ; `aPourCouleur`
  - A noter : Cette partie de la modélisation doit être vue comme en cours, aucune systématisme dans l'instanciation n'ayant été recherchée, les propriétés instanciées l'ayant été en raison de besoins particuliers ou pour l'exemple.



## 5 Méthode de création de l'ontologie

L'ontologie a été créée en partant d'une autre ontologie réalisée dans le cadre du projet ANR GEONTO (ANR-07-MDCO-005). Cette ontologie, que l'on peut qualifier de taxonomie, avait elle-même été réalisée pour organiser l'ensemble des termes rencontrés pour décrire le territoire dans les spécifications de deux bases de données de l'IGN [Mustière et al 2009], ensemble complété par des termes rencontrés dans des récits de voyage extraits par analyse automatique de texte [Bessagnet et al. 2010].

Cette ontologie GEONTO a été ensuite complétée par des termes issus de 1/ dires d'expert (termes proposés par le PGHM), 2/ analyse de sources de données géolocalisées spécialisées montagne (camp2camp, refuges.info, Parcs Nationaux Régionaux, OSM), 3/ analyse des termes utilisés lors de traitement de recherches réelles réalisées par le PGHM (suite à la retranscription de 45 bandes audio), et enfin 4/ une ontologie spécialisée sur les formes du relief LandFormREference Ontology [Sinha et al. 2018]

L'ontologie résultante a ensuite été filtrée puis réorganisée dans le cadre d'un travail collaboratif, pour suivre au mieux les principes de modélisation listés ci-dessus.

## 6 Résultat

L'ontologie contient au final 458 classes, associées à 1641 labels, en anglais et français, et possède une profondeur maximale de 6 niveaux. Le premier niveau d'organisation de la hiérarchie est constitué des 5 classes principales suivantes :

- 'Construction', qui regroupe toutes les constructions humaines dont la fonction apparente ne permet pas le rattachement aux classes hydrographie, végétation et relief.
- 'Hydrographie', qui regroupe tous les éléments hydrographiques et les constructions dont la fonction apparente est liée à l'hydrographie
- 'Relief', qui regroupe tous les éléments particuliers du relief
- 'Végétation', qui regroupe la végétation, qu'elle soit "naturelle" ou artificialisée
- 'Zonage', qui regroupe les délimitations immatérielles du territoire, qui peuvent-être de fait (agglomération) ou de jure (commune, zone industrielle). Les marqueurs matériels de ces délimitations sont dans la catégorie "construction".

Les deux premiers niveaux d'organisation sont listés dans la figure ci-dessous :

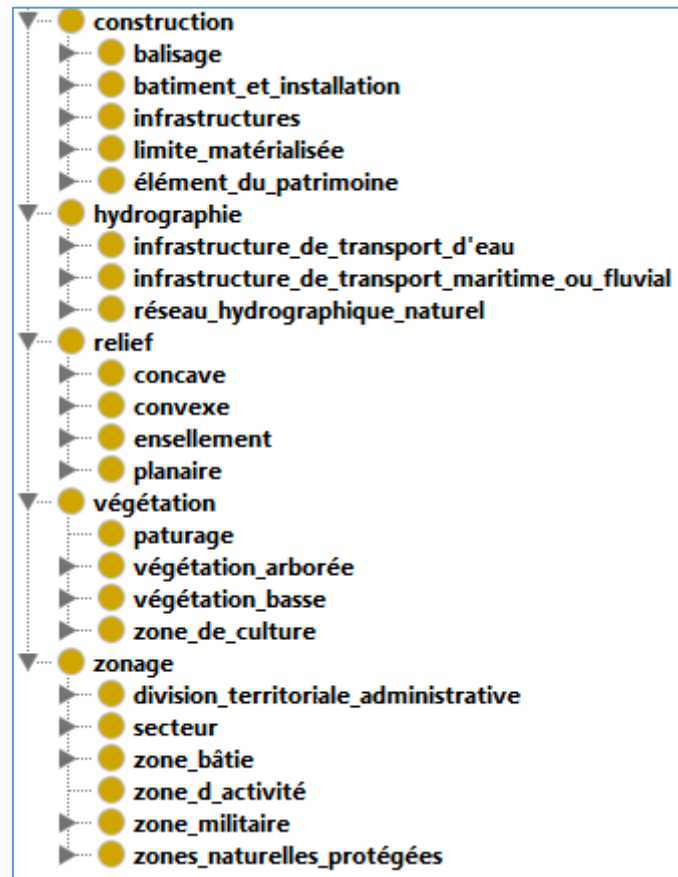


Figure 1. Premiers niveaux d'organisation de l'ontologie OOR

## 7 Références

- Bessagnet Marie-Noëlle, Gaio Mauro, Kergosien Eric, Sallaberry Christian, 2010. Extraction automatique d'un lexique à connotation géographique à des fins ontologiques dans un corpus de récits de voyage. Actes de TALN 2010 – conférence sur le traitement automatique des langues naturelles, Montréal, 19-23 juillet 2010.
- Mustière S., Abadie N., Aussenac-Gilles N., Bessagnet M.;-N., Kamel M., Kergosien E., Reynaud C., Safar B. 2009. GéOnto : Enrichissement d'une taxonomie de concepts topographiques. Actes de la conférence SAGEO, Paris, Novembre 2009.
- Sinha Gaurav, Arundel Samantha T., Hahmann Torsten, Uesry E. Lynn, Stewart Kathleen., Mark David M, 2018. The Landform Reference Ontology (LFRO):A Foundation for Exploring Linguistic andGeospatial Conceptualization of Landforms. Proceedings 10th International Conference on Geographic Information Science (GIScience 2018, 28-31 aout 2018, Melbourne, Australie).