



Livrable 5.7.1

Définition des métadonnées pour les services web externes

Projet ANR CHOUCAS (2017-2021)

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

Numéro de livrable	L5.7.1
Partenaire responsable	IGN - Laboratoire en Sciences et Technologies de l'Information Géographique (LaSTIG), Équipe MEIG (Médiation et Enrichissement de l'Information Géographique)
Contributeurs	Marie-Dominique Van Damme, Ana-Maria Raimond
Relecteurs	Eric Gouardères, Meriem Halilali
Description	Méthode d'acquisition et constitution de métadonnée sur les services web de données géographiques (objets de repère et itinéraires) dans le contexte de la recherche de victime en montagne
Type de livrable	Rapport, Fichiers de métadonnées
Date de soumission	15-01-2021





Table des matières

1	Introduction	4
2	Contexte	4
2.1	Architecture SOA	4
2.2	API de données	4
2.3	Contenu du livrable	5
3	Modèles de métadonnées utilisés	6
3.1	Le standard WFS de l'OGC	6
3.2	La norme ISO 19115	6
3.3	Métadonnées de la qualité des données géographiques	7
3.4	Métadonnées de lignage des jeux de données géographiques	10
4	Sources de données, pré-traitements, méthode d'évaluation de la qualité	11
4.1	Présentation des sources de données	11
4.1.1	BDTopo® de l'IGN	13
4.1.2	Refuge.info	13
4.1.3	Camptocamp.org	13
4.1.4	Aires protégées	13
4.1.5	Visorando	14
4.2	Identification des services de données pour Choucas	14
4.3	Proposition d'une méthode d'évaluation de la qualité des sources de données identifiées	15
4.4	Résultats de la construction des itinéraires et de l'appariement multi-critères	17
4.5	Services de construction d'itinéraires	18
4.6	Les propriétés fonctionnelles des services de données	19
5	Les propriétés non fonctionnelles principales de la norme ISO 19115	22
5.1	Métadonnées d'identification	22
5.1.1	La métadonnée <i>fileIdentifier</i>	22
5.1.2	La métadonnée <i>MD_TopicCategory</i>	23
5.1.3	La métadonnée <i>MD_LegalConstraints</i>	23
5.1.4	La métadonnée <i>MD_SecurityConstraints</i>	25
5.2	Métadonnées des systèmes de référence	25
5.2.1	La métadonnée <i>MD_ReferenceSystem</i>	25
5.3	Métadonnées d'information et représentation spatiale et temporelle	26
5.3.1	La métadonnée <i>MD_Resolution</i>	26
5.3.2	La métadonnée <i>MD_SpatialRepresentation</i>	26
5.3.3	La métadonnée <i>MD_TopologyLevelCode</i>	27
5.3.4	La métadonnée <i>MD_GeometricObjectTypeCode</i>	27
5.3.5	La métadonnée <i>EX_Extent</i>	28
6	Les propriétés non fonctionnelles de filiation	30
6.1	Le lignage des objets de repère	30
6.2	Le lignage des itinéraires	31
6.2.1	Méthode d'acquisition des traces	31
6.2.2	Les étapes de transformation	31



7	Les propriétés non fonctionnelles de qualité de la norme ISO 19157	37
7.1	Contexte	37
7.2	Métadonnées des métadonnées de qualité	38
7.2.1	La métadonnée <i>gmd :DQ_MetaqualityConfidence</i>	38
7.3	Les métadonnées d'exactitude de position	40
7.3.1	La mesure MeanAbsolute2D de l'indicateur <i>DQ_AbsoluteExternalPositionalAccuracy</i> pour les objets de repères	40
7.4	Les métadonnées d'exactitude sur la thématique	42
7.4.1	La mesure de similarité de Samal pour l'indicateur <i>DQ_NonQuantitativeAttributeCorrectness</i> (attribut nom des objets de repères)	42
7.4.2	La distance Cosinus pour l'indicateur <i>DQ_NonQuantitativeAttributeCorrectness</i> (attribut titres des itinéraires)	44
7.4.3	La mesure NombreComposantesConnexes pour l'indicateur <i>DQ_NonQuantitativeAttributeAccuracy</i>	45
7.4.4	Les mesures OverallAccuracy et ConfusionMatrix pour l'indicateur <i>DQ_ThematicClassificationCorrectness</i> (attribut nature des objets de repère)	47
7.4.5	Les mesures OverallAccuracy et ConfusionMatrix pour l'indicateur <i>DQ_ThematicClassificationCorrectness</i> (attribut cotation des itinéraires)	52
7.5	Les métadonnées de complétude	54
7.5.1	La mesure <i>MissingItems</i> de l'indicateur <i>DQ_CompletenessOmission</i>	54
7.5.2	La mesure <i>MissingClass</i> de l'indicateur <i>DQ_CompletenessOmission</i> et <i>DQ_CompletenessCommission</i>	55
7.5.3	La mesure <i>Excess</i> de l'indicateur <i>DQ_CompletenessCommission</i>	57
7.5.4	La mesure <i>Duplicate</i> de l'indicateur <i>DQ_CompletenessCommission</i>	59
8	Conclusion	61
9	Bibliographie	62

Table des figures

1	Taxonomie des services défini dans le livrable L5.2.2 ([5] Halilali et Gouardères, 2020)	5
2	Diagramme de classes de la norme ISO 19115-1 :2014 (© norme ISO 19115-1)	7
3	Diagramme de classes de la norme ISO 19157-2013 (© norme ISO 19157-2013)	8
4	Diagramme de classes de l'élément <i>Lineage</i> de la norme ISO de description des métadonnées géographiques (© norme ISO 19115-1)	10
5	Zone d'étude : zone de montagne dans les Alpes françaises	12
6	Trois services web fournissant des objets de repères	14
7	Méthode proposée pour l'appariement des itinéraires et des objets de repère (issue de Van Damme et al., 2019 [16])	15
8	Points GPS (orange) recalés sur le réseau routier de la BDTopo (bleu). Le point P1 est recalé avec le nœud du réseau pour maintenir la consistance de l'itinéraire (issue de Van Damme et al., 2019)	16
9	Trois services web fournissant des traces GPS, trois compositions de services fournissant des itinéraires	18
10	Modèle de la provenance des itinéraires	32
11	Pour chaque service, répartition spatiale des tronçons connectés, le label indique le nombre d'itinéraires	46
12	Matrice de confusion pour comparer la sémantique des objets de Refuges.info avec la BDTopo	48
13	Matrice de confusion pour comparer la sémantique des objets de C2C avec la BDTopo	49
14	Matrice de confusion pour comparer la sémantique des objets des Parc Nationaux avec la BDTopo	50
15	Matrice de confusion pour comparer la sémantique des objets de CamptoCamp avec Aires protégées et Visorando	53
16	Matrice de confusion pour comparer la sémantique des objets des Aires protégées avec Camptocamp et Visorando	53
17	Matrice de confusion pour comparer la sémantique des objets de Visorando avec Camptocamp et les Aires protégées	54
18	Matrice de confusion des noms des toponymes de Refuge.info	56



19 Matrice de confusion des noms des toponymes de Camptocamp 56
 20 Matrice de confusion des noms des toponymes des AP 56

Liste des tableaux

1 URL des sources de données collaboratives des points de repère et des itinéraires 11
 2 URL des sources de données d'autorité des points de repère et des itinéraires 12
 3 Nombre d'instances pour chaque source de données 12
 4 Résultats des appariements des points de repère 17
 5 Résultats d'appariement des itinéraires 18
 6 Description du service "Refuge.info des objets de repère" 19
 7 Description du service "C2C trace GPS" 20
 8 Description du service "WPS Map-matcher" 20
 9 Description du service "WPS RouteBuilding" 21
 10 Liste des métadonnées principales retenues pour évaluer les sources de données 22
 11 Identifiant des fichiers des métadonnées des web services étudiés 22
 12 Catégorie des données des web services étudiés 23
 13 Licence des web services étudiés 24
 14 Les systèmes de référence des web services étudiés 25
 15 Résolution spatiale des web services étudiés 26
 16 «TopologicalLevelCode» des web services étudiés 27
 17 «GeometricObjectTypeCode» des web services étudiés 27
 18 Emprise des API de données étudiées 28
 19 Liste des métadonnées de filiation retenues pour évaluer les services de données 30
 20 Liste des métadonnées de qualité retenues pour évaluer les sources de données 37
 21 Nombre de composantes connexes suivant le regroupement des sources de données 46
 22 «TopicCategoryCode» des web services étudiés 59



1 Introduction

De plus en plus de sites web proposent de partager et de télécharger, entre amateurs de sport de plein air, des traces GPX d'itinéraires ainsi que des points de repères. La connaissance des points de repères le long des itinéraires ou des itinéraires qui passent près de points de repères peut aider les secouristes dans la phase de localisation d'une victime en montagne. Afin de constituer des données métier (i.e. secours en montagne), il est nécessaire d'étudier la pertinence des différentes sources de données et de définir leurs caractéristiques.

Dans ce livrable, nous explorons le potentiel des données issues du web afin de déterminer leurs bien-fondés dans le contexte du sauvetage en montagne. Notre objectif est donc d'étudier le contenu des différentes sources de données afin d'avoir une meilleure connaissance de la description des points de repère et des itinéraires, comment sont-ils saisis, de démontrer la complémentarité des sources de données collaboratives en comparaison avec des données d'autorité et d'étudier la faisabilité de définir des liens entre les itinéraires d'une part et les points de repères d'autre part.

La méthode proposée utilise les techniques d'intégration telles que le recalage de points GPS sur un réseau routier, la reconstruction d'itinéraires et l'appariement de données géographiques. Un premier résultat concerne le grand nombre de données non appariées, ce qui prouve une grande richesse et complémentarité des sources. De plus, les résultats de l'appariement de données font émerger de nouvelles règles sémantiques concernant les points de repère et les géométries des routes.

2 Contexte

2.1 Architecture SOA

Le projet CHOUCAS vise à disposer d'une plateforme d'intégration de sources hétérogènes spatialisables. L'architecture choisie pour répondre à cet objectif repose sur une architecture orientée services (SOA) et mettra à disposition 1/ des sources de données et des services web existants 2/ des services développés dans les autres lots du projet 3/ ainsi que des compositions statiques de services web géospatiaux. Cette architecture a été privilégiée car elle facilite entre autres la fraîcheur des données ([6] Halilali, Gouardères, Devin, & Gaio, 2018).

Un médiateur permet, à l'aide d'un catalogue de services web géospatiaux, de diriger les requêtes sur différentes sources de données indexées par le système au moyen d'un Domain Specific Language (DSL). Il a été défini dans la tâche 5.2.2 du projet « Vers un modèle de classification pour les services géospatiaux ». Afin de retourner un ou plusieurs services pertinents en adéquation avec les besoins de l'utilisateur, les descriptions de services du catalogue devront tenir compte de la sémantique des web services. Le livrable 5.2.2 propose d'utiliser une ontologie pour permettre l'annotation sémantique de la description du service, des entrées-sorties et de la qualité des web services géospatiaux.

Par exemple, au travers du lot 2, ([8] Moncla et al, 2017) a déjà proposé des services web modulaire pour construire de l'information spatialisée à partir de corpus textuels. Ces web services ont été annotés sémantiquement dans le lot 5 ce qui donnera des compositions différentes.

2.2 API de données

Traditionnellement, les objets de repère font partie des données faisant autorité, issues des institutions telles que les agences nationales de cartographie ou les gazetiers (par ex. Wikipedia, GeoNames). La plupart du temps elles sont produites en mode collaboratif ou provenant de sources de données externes, y compris celles faisant autorité (e.g. des organismes public locaux). Au cours de la dernière décennie, de plus en plus de sites web collaboratifs offrent la possibilité de partager et de télécharger des points d'intérêt et des itinéraires. Ces sites fournissent des informations techniques (topoguide, difficulté, etc.), des objets de repère et des itinéraires au format GPX qui servent à différentes fins : utilisation de la trace dans une application mobile, suivi des performances, etc. Cartographiées et saisies par l'utilisateur, ces données collaboratives ont différents degrés de fiabilité et des structures hétérogènes (e.g. catégories d'objets de repère, cotations de difficultés différentes, irrégularités dans la description des itinéraires).



Aux cours de cette analyse, nous avons exploré le potentiel des données collaboratives candidates comme sources d'information dans le contexte du sauvetage en montagne. Aussi, notre objectif est-il d'étudier le contenu des différentes sources d'information afin d'avoir une meilleure connaissance de la description des objets de repère et des itinéraires et d'exprimer leurs caractéristiques sous forme de métadonnées. Dans ce livrable, nous explorons le potentiel des données issues du web, les « Api de données géographiques », afin de déterminer leurs bien-fondés dans le contexte du sauvetage en montagne. Il est donc nécessaire d'identifier les sources de données pertinentes de type d'objets de repère et itinéraires et de décrire leurs caractéristiques pour construire l'indexation de recherche.

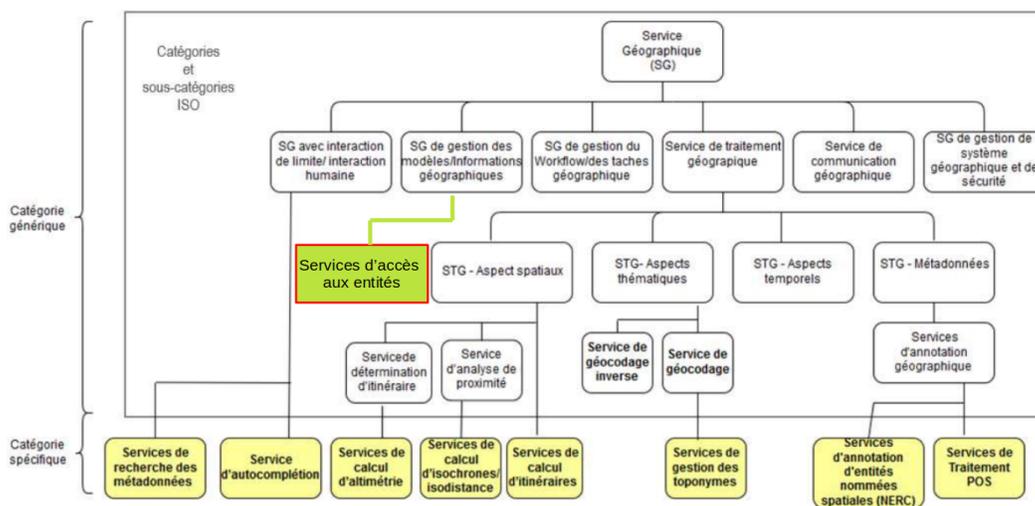


FIGURE 1 – Taxonomie des services défini dans le livrable L5.2.2 ([5] Halilali et Gouardères, 2020)

Dans la taxonomie des services géospatiaux du livrable 5.2.2, issu de l'ISO 19119 et adapté aux concepts spécifiques aux services web du projet, les sources de données des objets de repères et des itinéraires sont dans la catégorie « Information géographique – Accès aux entités simples (ISO 19125-1) » sous-classe de « Services de gestion des modèles/informations géographiques » (Cf. figure 1.).

Pour atteindre cet objectif d'annoter sémantiquement les web services de données géographiques d'objets de repères et d'itinéraires et suite aux recommandations du livrable 5.2.2 (c.f. le modèle générique de description de services web géospatiaux (figure 8 du livrable 5.2.2)), nous utiliserons une combinaison des standards de Web Feature Service (WFS) et de la norme ISO afin de décrire les métadonnées fonctionnelles (i.e. le schéma des données retournées par le service) et non fonctionnelles basées service (i.e. la qualité de service) des services API de données.

2.3 Contenu du livrable

Ce livrable est produit dans le cadre du lot 5 du projet CHOUCAS « Plateforme d'intégration orientée services » au titre de la tâche 5.7 intitulée « Qualifier les données et les web services externes existants ». Il contient :

- ~ le présent rapport qui décrit la méthode d'acquisition et la constitution des métadonnées sur les services API de données géographiques de points de repère et d'itinéraires,
- ~ des fichiers XML du même type que l'opération "DescribeFeatureType" des WFS qui contiendront les métadonnées fonctionnelles,
- ~ des fichiers XML de type "Reporting" qui contiendront les métadonnées non fonctionnelles dont le modèle est recommandé par la directive INSPIRE pour la diffusion de données géographiques.

Il convient de noter que la méthode d'évaluation de la qualité des sources de données identifiées et de la construction des itinéraires a fait l'objet d'une publication de revue (Van Damme et al. 2019 [16])



3 Modèles de métadonnées utilisés

Dans ce paragraphe nous présentons plus en détail les standards des métadonnées utilisés pour documenter les services web de données de notre étude. Parmi ces deux standards :

- ~ un standard de l'Open Geospatial Consortium (OGC)¹ pour décrire les propriétés fonctionnelles des API de données
- ~ une norme de l'ISO recommandée par la directive INSPIRE pour la diffusion de données géographiques et qui décrira les métadonnées non fonctionnelles des API de données.

3.1 Le standard WFS de l'OGC

Le standard WFS de l'OGC fournit un moyen de télécharger des données géographiques vectorielles. La description est faite de manière générique.

La méthode *DescribeFeatureType* est l'opération qui renvoie une description du schéma des données qui sont encodées en sortie lors de l'appel de la méthode de téléchargement. Si la valeur du paramètre *outputFormat* a été fixée à "application/gml+xml; version=3.2", la géométrie des données géographiques est décrite avec le standard GML.

Ce standard WFS sera employé pour décrire les métadonnées fonctionnelles des API de données.

En revanche, les standards des services web de l'OGC (WFS, etc.) ne permettent pas de décrire les métadonnées de description d'un jeu de données ainsi que la qualité des données diffusées. Par conséquent, pour compléter les descriptions WFS, nous nous appuyons sur les standards de l'ISO qui prennent en compte ces métadonnées.

3.2 La norme ISO 19115

La communauté géospatiale utilise traditionnellement la norme ISO 19115 pour encoder les métadonnées utilisées afin de décrire, valider et échanger l'information géographique. La norme ISO 19115 permet de décrire par exemple les métadonnées d'identification, de l'étendue des données, de leur qualité, etc. La figure 2 montre les principales rubriques de la norme ISO 19115-1:2014. Les propriétés entourées en vert sont celles que nous avons retenues d'étudier et de créer dans ce rapport.

1. <https://www.ogc.org/>

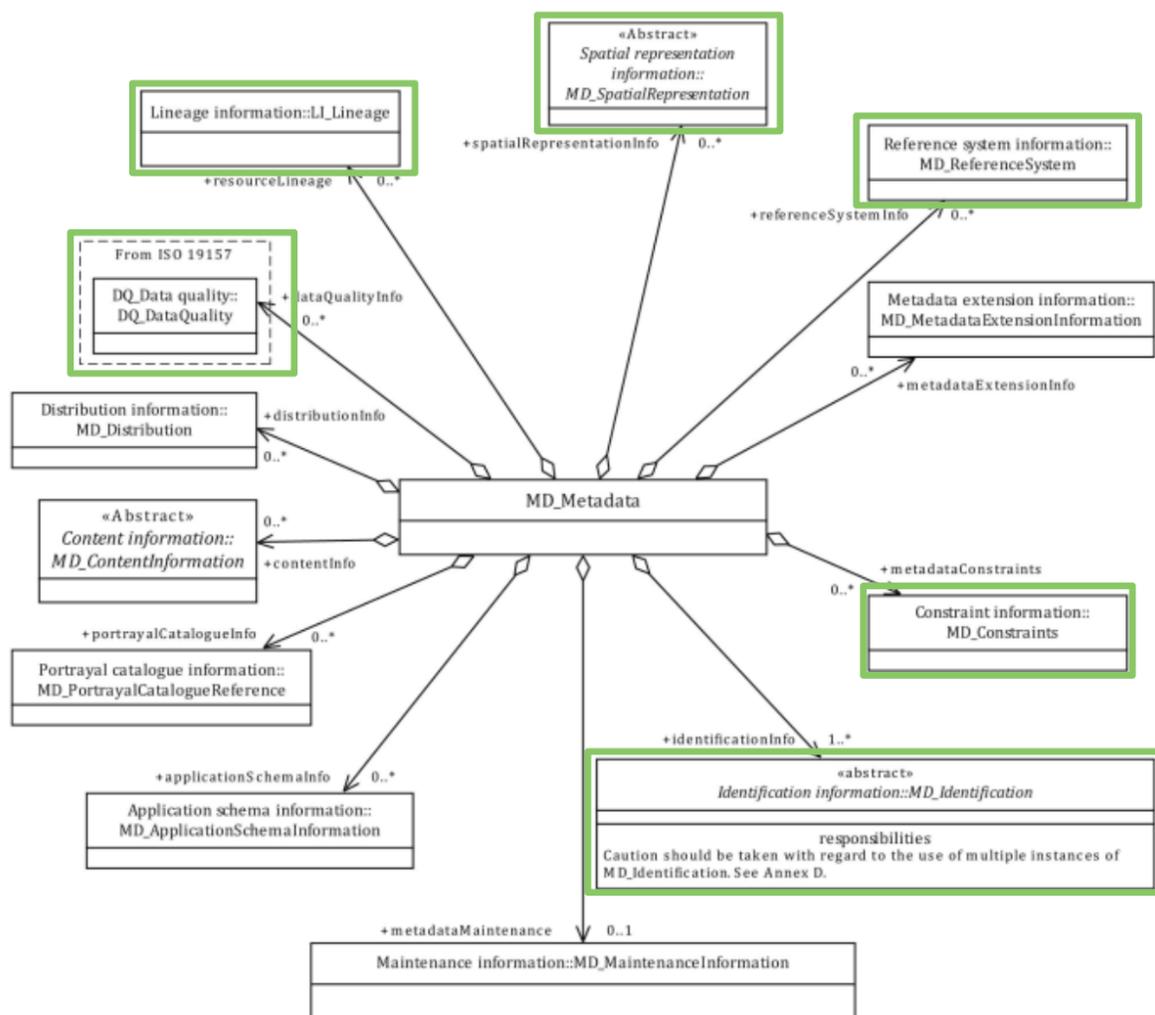


FIGURE 2 – Diagramme de classes de la norme ISO 19115-1 :2014 (© norme ISO 19115-1)

La qualité des données géographiques stockée dans l'élément *DQ_DataQuality* est standardisée via la norme ISO 19157. Certains concepts de cette dernière sont décrits en fonction des concepts fournis dans la norme des métadonnées ISO 19115.

3.3 Métadonnées de la qualité des données géographiques

Le processus d'évaluation de la qualité des données géographiques est de nos jours normalisé par la norme ISO 19157². De même que les métadonnées générales, les propriétés entourées en vert sont celles que nous avons retenues d'étudier et de créer dans ce rapport.

2. <https://www.iso.org/fr/standard/32575.html>

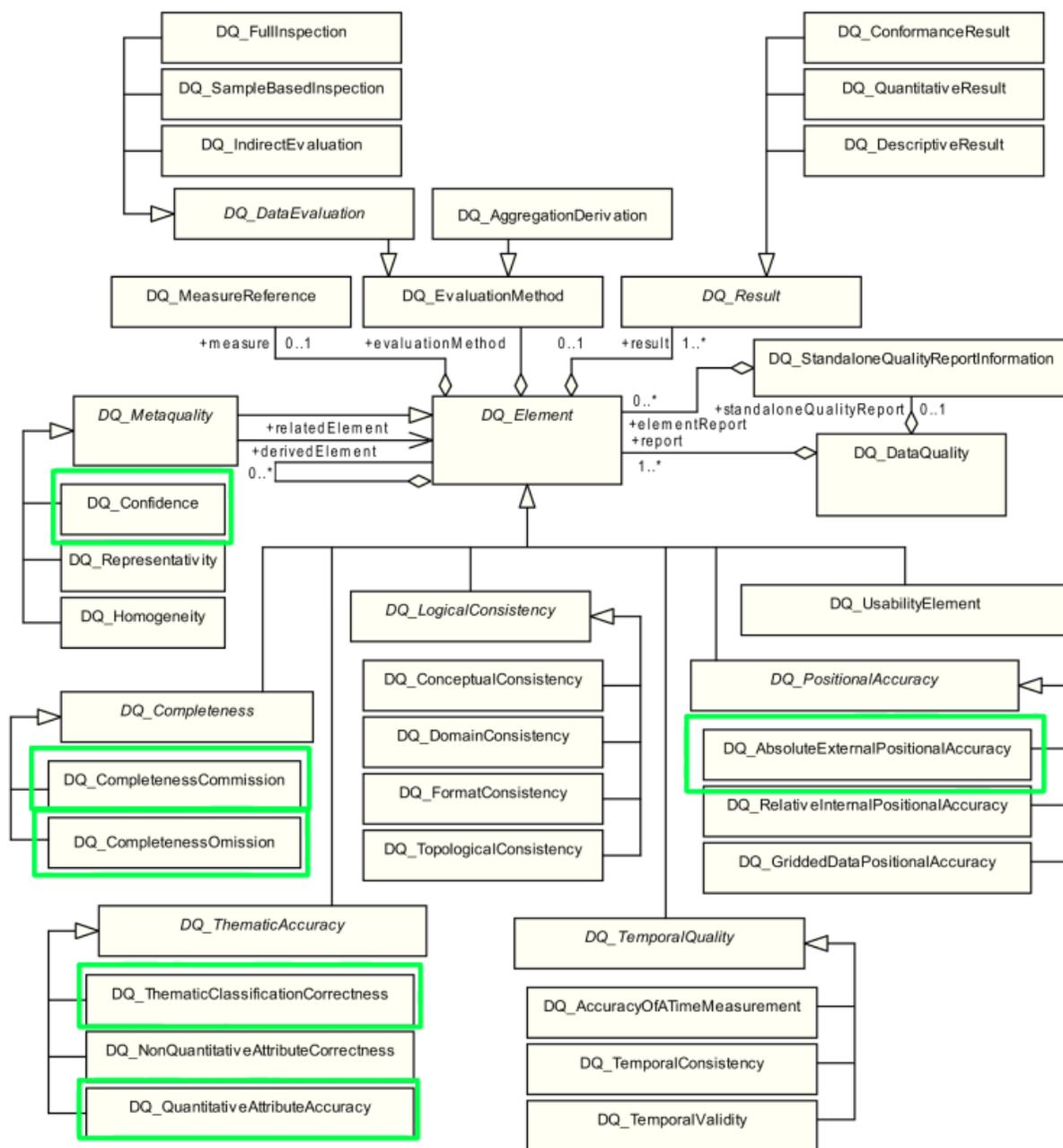


FIGURE 3 – Diagramme de classes de la norme ISO 19157-2013 (© norme ISO 19157-2013)

Six indicateurs de qualité sont définis (Cf. Figure 3) dans la norme ISO 19157-2013 :

1. L'exactitude de position (Positional accuracy) est la proximité des valeurs de coordonnées par rapports aux valeurs vraies (terrain nominal) ou reconnue (données de référence). Elle se compose de trois indicateurs de qualité des données :
 - ~ **absolute ou external accuracy** : la proximité des valeurs de coordonnées déclarées par rapport aux valeurs acceptées comme étant vraies ;
 - ~ relative ou internal accuracy : la proximité des positions relatives des objets d'un jeu de données par rapport à leurs positions relatives respectives acceptées comme vraies ;



- ~ grided data positional accuracy : proximité des valeurs de position spatiale des données maillées par rapport aux valeurs acceptées comme vraies ou étant vraies.
2. L'exactitude thématique (Thematic accuracy) est définie comme étant l'exactitude des attributs quantitatifs ou la justesse des attributs non qualitatifs et l'exactitude des classifications et de leurs relations. Elle se dérive en trois types d'indicateurs :
 - ~ **exactitude de la classification** : comparaison des classes ou des attributs avec les données de référence ou les données réelles ;
 - ~ exactitude des attributs non-quantitatifs : mesure si un attribut est juste ou pas ;
 - ~ **exactitude des attributs quantitatifs** : mesure à quel point les valeurs des attributs quantitatifs sont proches des valeurs réelles ou de celles de données de référence.
 3. La complétude (Completeness) est définie comme étant la présence ou l'absence d'objets géographiques, leurs attributs ou leurs relations. Cet indicateur se divise en deux indicateurs :
 - ~ **commission** : excès des données du jeu de données et redondance (Duplicate) qui définit si dans un jeu de données il existe des doublons (i.e. deux objets géographiques représentent la même réalité de monde réel et qui a les mêmes caractéristiques géométriques, thématiques et sémantiques) ;
 - ~ **omission** : données absentes du jeu de données.
 4. La cohérence logique (Logical consistency) est définie comme le degré d'adhésion aux règles logiques de la structure des données, des attributs et des relations (la structure des données peut être conceptuelle, logique ou physique). Si ces règles logiques sont documentées ailleurs (par exemple, dans une spécification de produit de données), alors la source doit être référencée (par exemple, dans l'évaluation de la qualité des données). Elle se compose de quatre indices de qualité des données :
 - ~ cohérence conceptuelle : respect des règles du schéma conceptuel ;
 - ~ cohérence du domaine : respect des valeurs dans les domaines de valeurs ;
 - ~ cohérence du format : mesure dans laquelle les données sont stockées conformément à la structure physique de l'ensemble de données ;
 - ~ cohérence topologique : exactitude des caractéristiques topologiques explicitement codées d'un ensemble de données.
 5. La qualité temporelle (Temporal quality) définit la qualité des attributs temporels et les relations temporelles entre les objets géographiques. Elle se dérive en trois indicateurs :
 - ~ exactitude de la mesure du temps : la proximité entre les mesures de temps rapportées aux valeurs vraies ou à celles de données de référence ;
 - ~ cohérence temporelle : mesure la justesse de l'ordre des événements ;
 - ~ validité temporelle : validité de données par rapport au temps donné.
 6. L'utilisabilité (Usability) est basée sur les besoins utilisateur (adéquation au besoin). Tous les indicateurs de qualité peuvent être utilisés pour évaluer l'utilisabilité. Les besoins des utilisateurs peuvent être spécifiques et les éléments de qualité insuffisants. Dans ce cas, l'utilisabilité doit décrire les informations de qualité spécifiques au besoin.

Dans la communauté de la géomatique, certains auteurs préfèrent, à juste titre, ajouter l'exactitude sémantique qui se réfère à la nature des objets géographiques afin d'éviter la confusion avec l'exactitude thématique faisant référence aux attributs des objets géographiques ([1]). Un autre indicateur, souvent utilisé en généralisation porte sur la résolution géométrique : c'est-à-dire la distance minimum entre deux points de la géométrie des objets géographiques (Ruas, 2004). Ces deux indicateurs, nous semble pertinents dans notre cas d'étude.

De plus, dans la norme ISO 19157 il existe des éléments de méta-qualité *Metaquality* qui sont un ensemble de déclarations quantitatives et qualitatives concernant une évaluation de la qualité et ses résultats. Dans notre étude, suite à la méthode employée, nous avons défini **la certitude** (Confidence) qui indique la fiabilité d'un résultat de qualité des données.



3.4 Métadonnées de lignage des jeux de données géographiques

Étant donné les spécificités de nos jeux de données (construction des itinéraires à partir de traces GPS), nous détaillons aussi l'élément de lignage des jeux de données géographiques dans la figure 4 que nous allons exploiter :

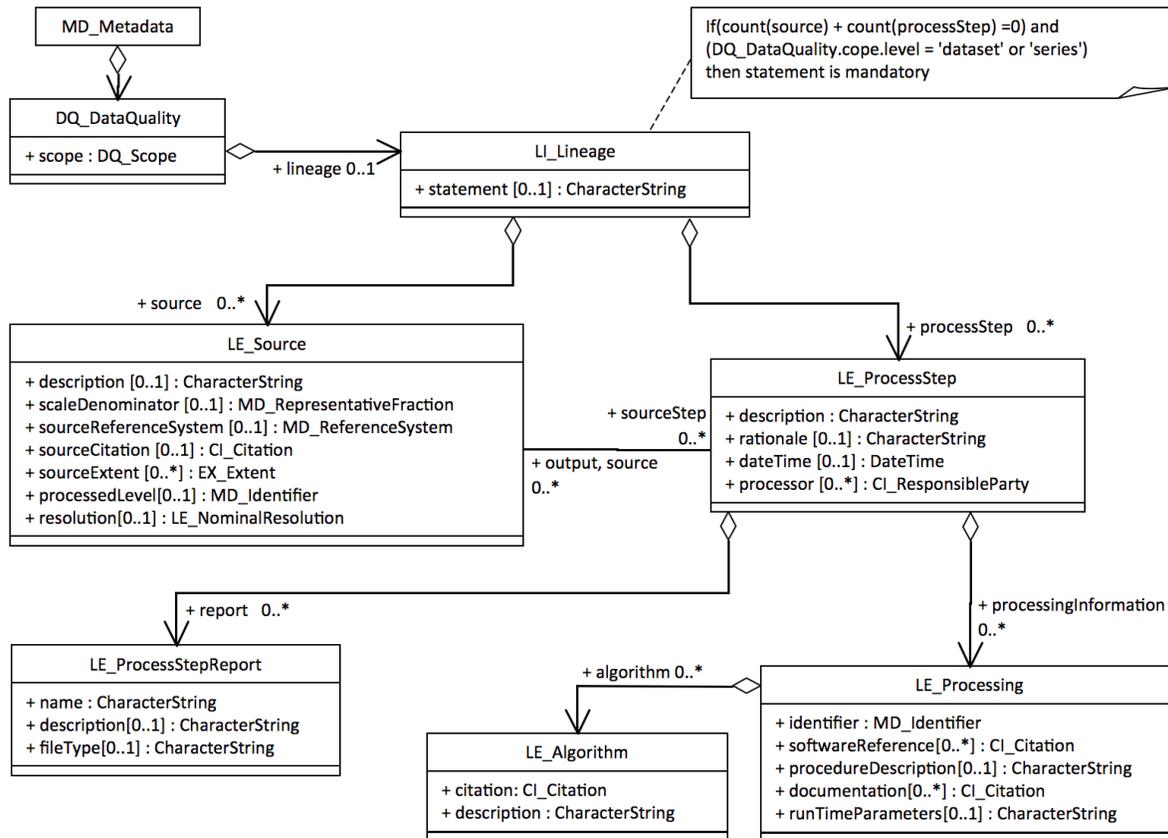


FIGURE 4 – Diagramme de classes de l'élément *Lineage* de la norme ISO de description des métadonnées géographiques (© norme ISO 19115-1)

La provenance des informations, aussi appelée lignage, définit la description des jeux de données d'origine (les sources) et les traitements qui leur ont été appliqués afin de produire le jeu de données du service de données. En spécifiant la série successive des algorithmes qui ont été appliqués, le lignage décrit précisément les entrées-sorties des algorithmes, l'environnement d'exécution, etc. Il joue un rôle important dans l'évaluation de la qualité et de l'usage des données.

Les informations de la provenance sont stockées dans l'élément *LI_Lineage* alors que les différents traitements sont stockés dans les éléments *LI_ProcessStep* (un élément par traitement). Les sources utilisées dans les traitements sont décrites avec *LI_Source*.

Nous pouvons noter que selon Santo (2019) [13], la description de la provenance des jeux de données n'est pas assez satisfaisante avec la norme ISO 19115 car elle ne permet pas de documenter les structures des données d'entrée et de sortie, ni de définir les valeurs utilisées lors de l'exécution des algorithmes.



4 Sources de données, pré-traitements, méthode d'évaluation de la qualité

4.1 Présentation des sources de données

Dans cette étude, nous nous sommes concentrés sur des données provenant de sites web actifs dédiés aux sports de loisirs de montagne, où la communauté et les auteurs des sites font régulièrement des mises à jour, recommandées par les autorités ou par les pratiquants, éventuellement partenaires de l'IGN, publiant suffisamment de données et traitant au moins d'une source thématique spécifique à la montagne.

Parmi les sources de données considérées comme les plus pertinentes, nous avons sélectionné parmi les sites web participatifs (cf. tableau 1.) : Refuge.info, Camptocamp et Visorando

Fournisseur du service	Type du service	Exemple d'URL d'accès aux données
Refuge.info	API REST	https://www.refuges.info/api/bbox?type_points=XXX
Camptocamp.org	API REST	https://api.camptocamp.org/waypoints?limit=30&wtyp=XXX https://api.camptocamp.org/waypoints//xxx https://api.camptocamp.org/outings?limit=30&act=hiking https://api.camptocamp.org/outings/xxx
Visorando	HTTP	https://www.visorando.com/index.php?component=exportData &task=getRandos&duree=0&accessibilite=0&retourDepart=0 &difficulte=0&projection=FXX&idEG=0&zoom=9&bbox=xxx https://www.visorando.com/randonnee-decouverte-des-dolmens-de-martiel /xxx.gpx

TABLE 1 – URL des sources de données collaboratives des points de repère et des itinéraires

Parmi les sources de données d'autorité (cf. tableau 2.), la base de données topographiques *BDTOPO*© de l'IGN ainsi que 2 sites d'Aires Protégées (AP), sont utilisés pour compléter notre analyse.

Presque toutes les données sont fournies via des API. Les données de notre étude ont été extraites au cours du premier trimestre de 2019. L'étude de ces sources de données s'est appliquée sur une partie des Alpes françaises, près de Grenoble (9870 km²), sur un territoire couvrant sept massifs montagneux (voir figure 5.).



Fournisseur du service	Type du service	URL d'accès aux données
BDTopo@ de l'IGN	WFS	<p>https://wxs.ign.fr/cle/geoportail/wfs?service=WFS&request=GetFeature&typeName=BDTOPO_V3 :troncon_de_route&srsName=EPSG :2154&version=2.0.0&outputFormat=json&BBOX=XXX</p> <p>http://wxs.ign.fr/cle/geoportail/wfs?service=WFS&request=GetFeature&typeName=BDTOPO_BDD_WLD_WGS84G :pai_orographie&srsName=EPSG :2154&version=2.0.0&outputFormat=json&BBOX=XXX</p>
Aires protégées (parc du Vercors et parc national des Ecrins)	API REST	<p>http://rando.parc-du-vercors.fr/data/api/fr/pois.geojson</p> <p>http://rando.parc-du-vercors.fr/data/api/fr/treks.geojson</p> <p>http://rando.ecrins-parcnational.fr/fr/files/api/trek/trek.geojson</p> <p>https://rando.ecrins-parcnational.fr/api/fr/pois.geojson</p>

TABLE 2 – URL des sources de données d'autorité des points de repère et des itinéraires

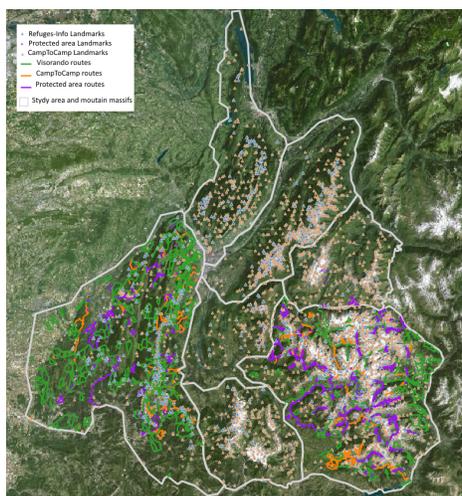


FIGURE 5 – Zone d'étude : zone de montagne dans les Alpes françaises

Le Tableau 3 décrit le nombre d'instance pour chaque source de données dans cette zone d'étude. Les objets de repères de Visorando ont été exclus car leurs catégories n'étaient pas bien identifiées et le nom ne correspondait pas directement à un toponyme. Refuge.info et la BDTOPO sont deux sources de données qui ne contiennent pas d'itinéraires.

	BDTOPO	C2C	Refuge.info	AP	Visorando
Objets de repères	21 097	3226	478	1102	exclus
Itinéraires	–	127	–	181	398

TABLE 3 – Nombre d'instances pour chaque source de données

Mentionnons que d'autres portails de données, comme OpenStreetMap (OSM), GeoNames, IGNRando fournissent des points de repère ou des itinéraires mais ces données sont souvent stockées en masse et proviennent de différentes sources de données externes, la source est alors souvent mentionnée (par exemple, dans OSM, certains points de repère sont référencés C2C ou wikidata). Dans ce livrable, ces sources de données n'ont pas été prises en compte



parce que le schéma de données est trop généralisé, que des informations spécifiques sont perdues (comme des informations techniques, les traces GPX, lien entre les itinéraires et les waypoints) ou ne sont pas appropriées (IGNRando utilise le schéma cirkwi dédié aux informations touristiques).

4.1.1 BDTopo® de l'IGN

Parmi les différents thèmes de la base de données topographiques produite par l'IGN, les données utilisées dans cette étude concernent les Points d'Intérêt ou d'Activités (PAI) et les Routes (BDTOPO, 2018). La couverture des données est nationale. Le thème *PAI* regroupe les points d'intérêt et les activités classées par catégorie : administratif, secours, transport, etc. Chaque élément est représenté par sa localisation (x,y), l'origine des données, son type, son nom et sa pertinence. La couche *Routes* contient à la fois les routes et les chemins ; les données sont modélisées en réseau et sont décrites par type, nom, pertinence, etc.

4.1.2 Refuge.info

Refuges.info est un site web collaboratif proposant différentes données géographiques ponctuelles en zones de montagne. Même si le site Web était focalisé sur les refuges à l'origine, d'autres types de points d'intérêt peuvent y être trouvés aujourd'hui tels que les points d'eau, sommets, cols, hôtels, camping, parking, bus. Chaque objet est caractérisé par sa localisation, son nom, son type et différentes descriptions, telles que comment y accéder, nombre de places, origine des données, etc. Les localisations sont généralement obtenues par numérisation à l'aide de cartes. Concernant la couverture spatiale, les données sont disponibles principalement en France métropolitaine.

4.1.3 Camptocamp.org

Camptocamp (C2C) est un projet collaboratif auquel participe des pratiquants plutôt avertis des sports de montagne. L'objectif du site web est de partager des informations techniques (guides topo, rapports d'incidents et d'accidents, sorties avec les conditions météorologiques et l'accessibilité) et de contribuer à la sécurité des utilisateurs de l'application. C'est une communauté active qui propose une large couverture, tant du point de vue des activités de montagne que du point de vue géographique (France, Allemagne, Suisse, Espagne). Toutes les données (routes, waypoints, sorties, rapports d'accidents, photos) sont liées les unes par rapport aux autres.

Objet de repère. Chaque objet est caractérisé par sa localisation, son nom, son type et certaines informations thématiques telles que l'altitude, l'exhaustivité (c'est-à-dire la qualité montrant dans quelle mesure la contribution tient compte des recommandations de partage), la description textuelle et les photos. Différentes catégories telles que sommet, col, lac, chute d'eau, localité, canyon existent.

Itinéraire. Les itinéraires représentent différents types de randonnées en montagne, très peu ont une trace GPX (25% dans notre zone d'étude). La création des géométries est réalisée par les contributeurs avec leurs propres outils de gestion de données et leur propre fond cartographique. Un itinéraire se caractérise par un titre, un type (par exemple une boucle, un aller-retour), une cotation (difficulté globale d'un itinéraire), différentes informations liées à l'altitude (altitude minimale, maximale, dénivelé, pente, etc.), une description textuelle et des photos.

4.1.4 Aires protégées

Parmi les aires protégées (AP), les Parcs naturels régionaux (PNR) et les Parcs nationaux (PN) sont des institutions publiques françaises ayant pour mission de préserver l'environnement naturel et de développer des activités économiques et touristiques. Par ailleurs, les Parcs du territoire français gèrent des informations telles que les points de relief, les points d'intérêt patrimoniaux, les services (informations pratiques comme les points d'eau, les passages difficiles) et les itinéraires. La plupart d'entre eux utilisent les logiciels de la suite GeoTrek, les données de la BD-Topo ainsi que le Modèle Numérique de Terrain (MNT) de l'IGN comme référentiel pour administrer leurs données. Mentionnons que ces données d'autorité (BDTopo et base de données sur les aires protégées) ne sont pas équivalentes en termes de contenu parce qu'elles ont des points de vue différents (topographique vs. environnemental et touristique). Elles sont plutôt complémentaires avec une petite redondance en termes de types de fonctionnalités et d'informations thématiques.



Objet de repère. Les objets de repère sont décrits par leur localisation, type, nom, description textuelle et altitude. Le nombre d'objet de repère est fixé à trois au minimum et ces objets doivent être bien répartis sur le parcours. Des types tels que point de vue, point d'eau, abri, col, lac, glacier, sommet, pastoralisme, histoire sont présents.

Itinéraire. les objets de repère sont reliés à l'itinéraire par référencement linéaire. Les itinéraires sont définis par une géométrie fournie dans un fichier GPX, un titre, une activité sportive, une cotation, des informations liées à l'altitude, une description textuelle, etc.

4.1.5 Visorando

Visorando est un site web offrant la possibilité de télécharger des traces publiées en ligne par des contributeurs. Contrairement à C2C, 100% des géométries des itinéraires ont été produites à partir de données issues de capteurs GPS. En plus des traces, des attributs tels que le titre, le type d'activité (randonnée, vélo, etc.), la description textuelle de l'itinéraire, la durée, la difficulté et le dénivelé sont fournis. Le site Web propose aux contributeurs d'ajouter plusieurs points de passages aux traces (points de repère ou changement majeur de direction). Les traces sont ensuite ajustées et validées par les contributeurs dans une interface web cartographique utilisant les fonds cartographiques de l'IGN (données faisant autorité).

4.2 Identification des services de données pour Choucas

Les objets de repère issus des services de données présentés ci-dessus peuvent être utilisés tels quels. En effet les données en sortie sont exploitables directement par les secouristes via le prototype de l'application de géovisualisation et les services développés dans les autres lots du projet. Chacun de ces services propose les deux principaux paramètres en entrée nécessaires : pour filtrer les jeux de données, à savoir l'emprise spatiale de sélection et la catégorie des objets de repère recherchés.

Les trois API de données des sites identifiés des objets de repère vont donc fournir trois services de données (Cf. figure 6.).

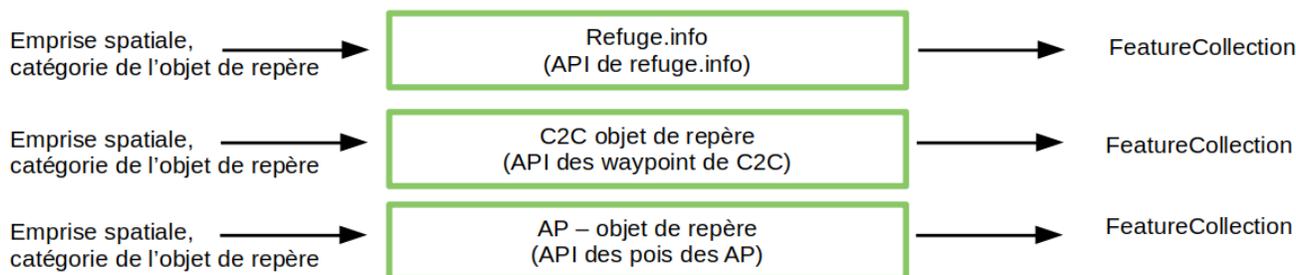


FIGURE 6 – Trois services web fournissant des objets de repères

Concernant les API de données fournissant des traces, l'exploitation est bien plus difficile. Les traces GPS ne fournissent pas directement une structure de données sous forme de graphe afin de pouvoir faire des calculs de plus court chemin par exemple. La détection de la connectivité entre les traces est une opération coûteuse et rend la tâche de recherche d'informations (chercher les itinéraires qui longent des lacs, différentes routes d'un même itinéraire) plus compliquée. Il faut concevoir des pré-traitements.

Dans le prochain paragraphe, nous allons présenter la méthode qui nous a permis d'évaluer la qualité des sources de données et établir une chaîne de pré-traitement destinée à la construction d'un réseau d'itinéraires.



4.3 Proposition d'une méthode d'évaluation de la qualité des sources de données identifiées

La méthode la plus utilisée pour évaluer la qualité d'un jeu de données géographiques est de comparer le jeu de données à évaluer avec un jeu de données dit de référence. Ce dernier peut être issu d'une base de données d'autorité (produite par un organisme habilitée à produire des données géographiques) ou d'une base de données dont la qualité est connue. La démarche pour évaluer la qualité des données géographiques consiste à appairer le jeu de données à évaluer avec un jeu de données (d'autorité ou autre), puis de calculer les indicateurs de la norme ISO 19157.

L'appariement de données consiste à définir des objets homologues de deux jeux de données géographiques (Olteanu, 2008 [10]). Dans le cadre du projet Choucas, on définit :

1. des itinéraires homologues : sachant qu'un itinéraire est un concept porté par une infrastructure de réseau physique, nous considérons donc que deux itinéraires sont homologues s'ils suivent des chemins proches.
2. des objets de repère homologues : deux objets de repères sont homologues s'ils représentent la même entité du monde réel.

Les liens d'appariement de type [n :m] (n objets du jeu de données à évaluer correspondent à m objets du jeu de données de référence) sont exploités pour calculer les indicateurs d'exactitude géométrique, thématique ou sémantique. Les liens de cardinalité [1 :0] et [0 :1] permettent d'estimer l'exhaustivité géométrique. Les matrices de confusion sont également utilisées pour estimer les exactitudes permettant de calculer les taux d'omission et de commission. Nous pouvons citer à titre d'exemples des travaux qui ont suivi cette démarche (Fan et al., 2014 [2] ; Girres and Touya, 2010 [3] ; Haklay, 2010 [4] ; Jokar Arsanjani et al., 2015 [7] ; Neis et al., 2011 [9] ; Touya et al., 2017 [15]).

Notre méthode d'évaluation de la qualité des sources de données est composée de trois étapes. La première étape consiste à extraire les données de différentes sources d'information. Pour faciliter cette tâche, un plug-In QGis³ a été développé. La deuxième étape concerne l'appariement des données. La troisième étape vise à évaluer la qualité des données : d'une part en calculant les indicateurs de qualité, d'autre part en évaluant la redondance d'information en calculant des mesures de connectivité dans le réseau d'itinéraires fusionnés. La Figure 7 illustre notre approche pour l'appariement des objets de repère et des itinéraires.

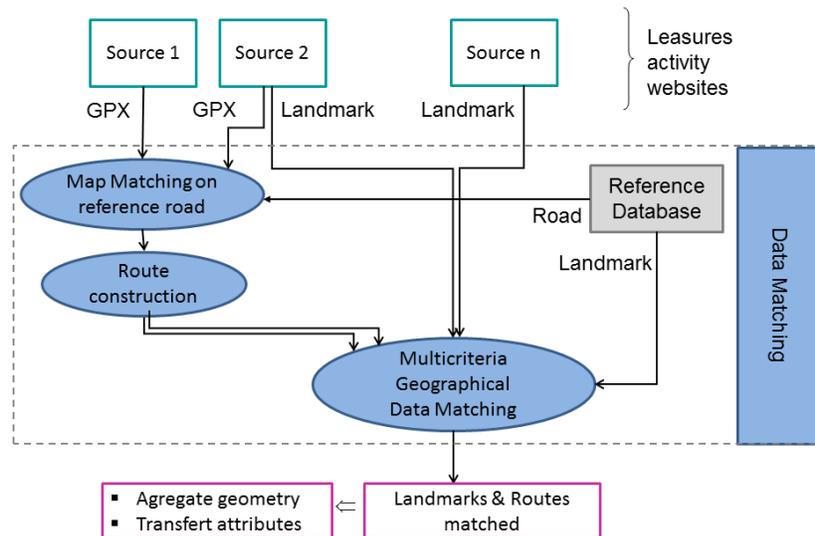


FIGURE 7 – Méthode proposée pour l'appariement des itinéraires et des objets de repère (issue de Van Damme et al., 2019 [16])

Les sites web fournissent, à défaut de la géométrie linéaire des itinéraires, au moins la trace GPX d'origine. Pour faciliter l'appariement des itinéraires, il est donc préférable de reconstruire leur géométrie. Deux étapes de pré-

3. <https://github.com/ANRChoucas/PluginChoucas>

traitement sont proposées. Tout d'abord, les points vont être recalés sur un réseau routier de référence en appliquant l'algorithme proposé par Newson et Krumm (Newson et Krumm, 2009) fondé sur les modèles de Markov cachés. Cette méthode est plus appropriée dans notre cas, car pour chaque point à recalcer sur le réseau, l'algorithme tient compte des points avant et après, la cohérence des itinéraires étant ainsi bien conservée. Une difficulté concernant cet algorithme est de choisir les bons paramètres : trouver le bon rapport entre les erreurs de positionnement et la tolérance à suivre des routes indirectes. Par exemple, dans la Figure 8, le point P1 ne doit pas être mis en correspondance avec le tronçon le plus proche afin de maintenir la cohérence de l'itinéraire et de ne pas créer une impasse.

La deuxième étape du pré-traitement consiste à reconstruire la géométrie linéaire de l'itinéraire. Les points issus de la trace de l'itinéraire sont parfois très éloignés les uns des autres, ce qui crée parfois des cassures entre les tronçons portants les points recalés. Notre processus intègre l'ajout de ces tronçons manquants, en calculant le chemin le plus court sur le réseau routier entre deux tronçons non connectés. Si le rapport entre la distance à vol d'oiseau et la distance du plus court chemin est inférieur à un seuil fixé, les tronçons manquants sont ajoutés à l'itinéraire (tronçons verts sur la Figure 8). A la fin de cette étape, la géométrie d'un itinéraire est composée d'une liste de tronçons du réseau routier. Cet algorithme de construction de routes à partir d'une liste de tronçon a été implémenté dans le plug-In MapMatcher⁴

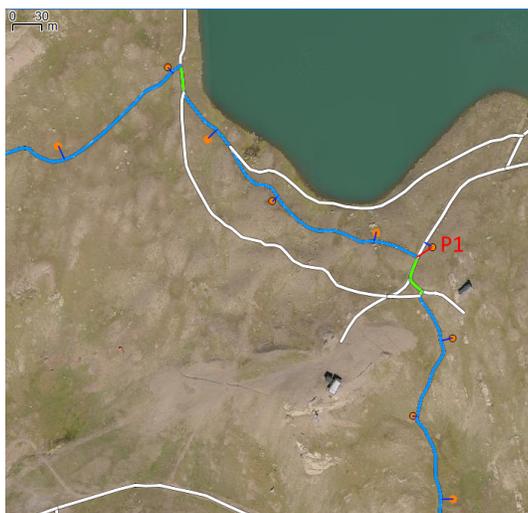


FIGURE 8 – Points GPS (orange) recalés sur le réseau routier de la BDTopo (bleu). Le point P1 est recalé avec le nœud du réseau pour maintenir la consistance de l'itinéraire (issue de Van Damme et al., 2019)

Une fois les étapes de pré-traitement terminées, l'appariement de données peut être effectué séparément pour les routes et les points de repère. L'algorithme d'appariement de données appliqué est un appariement de données multi-critères open-source⁵ s'appuyant sur la théorie des fonctions de croyances (Olteanu-Raimond, Mustière, & Ruas, 2015 [11]). L'algorithme d'appariement s'utilise sur un jeu de données de référence et sur un jeu de données de comparaison donnant ainsi une direction à l'appariement (pour chaque donnée du jeu de référence, l'algorithme recherche les données homologues et candidates dans le jeu de comparaison). Mentionnons que le jeu de données de référence peut être un jeu de données faisant autorité ou un jeu de données collaboratif. Les données de la BDTOPO seront utilisées comme données de référence pour l'appariement des points de repère. Comme il n'existe pas de base de référence pour les itinéraires, les sources de données vont être appariées entre elles et deux à deux. La source de données qui contient le plus grand nombre de données jouera le rôle de jeu de données de référence.

Pour l'étape 2, connaissant les caractéristiques des jeux de données issus des sites web décrits dans la Section 4, nous avons opté pour deux critères (position et nom) pour l'appariement des itinéraires et trois critères (position,

4. <https://github.com/IGNF/mapmatcher>

5. <https://github.com/IGNF/MultiCriteriaMatching>



toponyme et sémantique) pour l'appariement des points de repère.

Le critère de position s'appuie sur la distance géométrique entre l'élément de référence et un candidat (à savoir la distance euclidienne pour les points de repère et la moyenne de la distance moyenne de Hausdorff pour chaque tronçon routier des itinéraires). Le critère toponymique compare le nom de l'élément de référence avec le nom du candidat. Parmi les différentes mesures testées pour comparer les chaînes de caractères, la distance de Samal (Samal et al., 2005 [12]) et la distance Cosinus sont considérées comme les plus appropriées pour les points de repère et les routes respectivement. Enfin, le critère sémantique compare les catégories des points de repère. Afin de calculer les similitudes entre les catégories, la mesure Wu-Palmer (Wu & Palmer, 1994 [17]) est appliquée en utilisant l'ontologie de domaine (Mustière et al., 2011 [14]). Tous ces critères sont enfin fusionnés pour prendre la décision finale.

Nous pouvons noter que l'algorithme d'appariement de données ne prend aucune décision (i.e. les objets géographiques ne sont pas appariés) si les critères sont contradictoires (par exemple dans le cas de deux objets homologues potentiels, si le nom est identique et si la sémantique est différente avec une distance importante). Ces cas sont étiquetés comme étant des cas incertains et seront exprimés à travers l'indicateur Uncertainty.

4.4 Résultats de la construction des itinéraires et de l'appariement multi-critères

Les paramètres concernant le recalage des points des traces sur le réseau routier sont de : 80m pour la recherche de candidats et 4m pour l'écart type des mesures d'erreur GPS. Dans l'ensemble, 97% des points (298348 points) sont recalés et la racine carré de l'erreur quadratique moyenne des déplacements induits par le recalage est égale à 12,94 m. Il est à noter que quelques impasses ont été générées après le processus de recalage des traces. Concernant l'étape de reconstruction de la géométrie linéaire des itinéraires, le rapport maximal entre la distance à vol d'oiseau et la distance du chemin le plus court entre 2 tronçons non connectés est estimé empiriquement à 5. La longueur totale des tronçons du réseau routier manquants ajoutés est égale à 337 km, soit 4% de la longueur totale des itinéraires. Ceci montre la pertinence du processus de reconstruction des routes.

Pour les paramètres de l'appariement de données multi-critères, le buffer choisi pour la recherche des candidats est de 500m et 300m pour les itinéraires et les points de repère respectivement. Choisir les seuils appropriés est une tâche difficile, il faut trouver le bon compromis entre la qualité des résultats et le nombre de liens d'appariement. Les seuils de l'algorithme d'appariement ont été fixés empiriquement en analysant les distributions des distances entre chaque couple d'objets des deux jeux de données et pour chaque critère ou en utilisant des seuils déjà suggérés dans la littérature (Olteanu-Raimond et al., 2015 [11]). C'est le cas, par exemple, des seuils utilisés pour le nom et le critère sémantique pour les points de repère. Ce qui donne pour les itinéraires, des seuils pour le critère de position et le critère de similarité des noms de randonnées, 200m et 0,3, respectivement. Pour les points de repère, les seuils ont été estimés respectivement à 100 m, 0,6 et 0,7 pour la position, le nom et la sémantique.

Les résultats des appariements obtenus sont présentés dans le tableau 4 pour les objets de repère et dans le tableau 5 pour les itinéraires.

Appariement avec BDTopo	1 :0	1 :1	Incertain
C2C (3229)	1167	1938 (précision 93%)	124
RefugeInfo (478)	80	367 (précision 96%)	34
Aires protégées (AP)(1102)	863	180 (précision 71%)	61

TABLE 4 – Résultats des appariements des points de repère

En ce qui concerne les objets non appariés (colonne 2 du tableau 4), deux types de cas peuvent être observés. Premièrement, les objets ne sont pas appariés en raison de la faible précision des points de repère collaboratif (i.e. le point de repère est éloigné de son homologue dans la BDTopo); deuxièmement, les objets ne sont pas appariés parce qu'il n'y a aucun objet homologue dans la BDTopo. Ce dernier cas concerne 70% des points de repère non appariés. Par exemple, un abri en ruine est étiqueté comme historique dans les données des zones protégées et n'existe plus dans les bases de données faisant autorité ou dans les autres sources car l'objet a été détruit.



Appariement	1 :0	1 :1	Incertain	Sans candidat
AP - Visorando	92	59 (précision 87%)	14	16
C2C - Visorando	76	20 (précision 86%)	21	9
C2C - AP	77	14 (précision 83%)	6	29

TABLE 5 – Résultats d'appariement des itinéraires

Les liens 1 :1 des résultats des appariements des itinéraires dans le tableau 5 ont également une bonne précision (85%), précision obtenue manuellement. Afin de construire un réseau d'itinéraires de randonnées, on considère que deux itinéraires sont homologues si leurs géométries sont relativement proches. Différents cas d'itinéraires appariés sont alors observés : 1/ lorsqu'il existe une variante de l'itinéraire et que le détour est court ; 2/ lorsque deux itinéraires sont très proches et ont la même direction (typiquement lorsqu'ils bordent une rivière (ce qui est satisfaisant dans notre contexte puisque la victime décrit l'itinéraire par « je marche le long d'une rivière ») ; 3/ lorsque les points de départ sont à quelques mètres les uns des autres. Dans ce cas, un itinéraire peut faire une boucle et l'autre itinéraire un aller-retour. Précisons que dans cette étude, les tronçons manquants sur le réseau routier n'ont pas été gérés (e.g. hors sentier, chemin sur glacier, voie d'escalade, etc.). Ces tronçons ne sont pas pris en compte dans l'appariement de données. Les itinéraires incertains font partie des trois cas mentionnés ci-dessus. Remarquons que le nombre d'itinéraires appariés est faible. En réalité de nombreux itinéraires n'ont pas le même point de départ ou la même destination, ils n'ont que quelques tronçons en commun.

4.5 Services de construction d'itinéraires

Afin d'évaluer la qualité des jeux de données des itinéraires issus des sources présentées, deux étapes de pré-traitement sont proposées dans notre méthode :

- ~ tout d'abord, les points vont être recalés sur un réseau routier de référence en appliquant l'algorithme proposé par Newson et Krumm (Newson et Krumm, 2009)
- ~ le second traitement consiste à reconstruire la géométrie linéaire de l'itinéraire. Les points issus de la trace sont parfois très éloignés les uns des autres, ce qui crée parfois des cassures entre les tronçons portant les points recalés. Ce processus intègre l'ajout de ces tronçons manquants.

La mise en service de ces algorithmes, suivant le standard WPS, permet de proposer trois autres services de données fournissant en sortie un réseau d'itinéraires. La figure 9. montre les chaînages à effectuer.

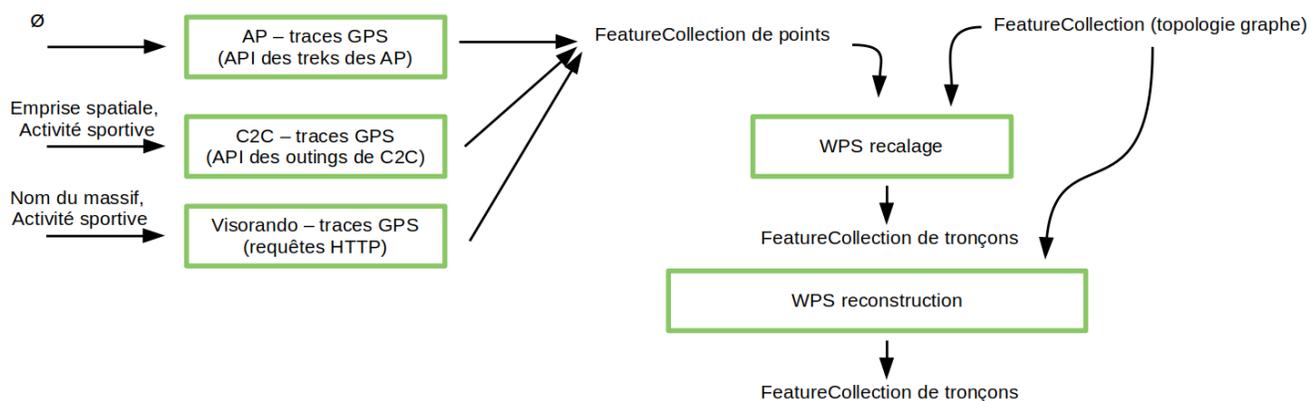


FIGURE 9 – Trois services web fournissant des traces GPS, trois compositions de services fournissant des itinéraires

Comme nous avons décidé de nous concentrer uniquement sur les objets de repère et les itinéraires, nous allons, au



final, étudier les caractéristiques et la qualité et les traduire sous forme de métadonnées, six web services de données (POI : Refuge.info, Camptocamp, Aires protégées, Itinéraires : Camptocamp, Aires protégées et Visorando).

Les informations sémantiques concernant les traces GPS ne seront pas traitées dans ce rapport. Cependant les métadonnées concernant la construction des itinéraires en fonction des traces seront bien définies dans la partie *Lineage* des métadonnées non fonctionnelles du jeu de données.

Mentionnons aussi que nous ne prenons pas dans cette construction de métadonnées les web services de l'IGN (WFS des points d'activité et du réseau routier). Les métadonnées de ces jeux de données sont déjà documentés sur le géocatalogue⁶.

Toutes les traces des API de données n'ont pas forcément un timestamp. Pour harmoniser les traitements, l'heuro-datage n'a pas servi pour le recalage, uniquement l'ordre des points de la trace. C'est pourquoi aucune métadonnée de qualité ne tiendra compte du timestamp des traces.

4.6 Les propriétés fonctionnelles des services de données

Les services maintenant identifiés, il reste à décrire les métadonnées fonctionnelles avec le standard WFS. Elles sont décrites dans les fichiers **XXX_DescribeService.xml** qui font partie du livrable.

Nous allons présenter dans un tableau, de manière informelle, en exemple les entrées-sorties des web services de données, un pour les objets de repère et un pour les itinéraires.

© API Refuge.info des objets de repère

Nom du service	API Refuge.info des objets de repère		
Fonctionnalité	This service returns different types of POI in mountain areas for planning sport activities		
Paramètres d'entrée	Nom	Type	Description
	bbox	LiteralData	Les dimensions de la bbox à exporter
	format	LiteralData	Le format de l'export (geojson, kml, csv)
	type_points	LiteralData	Les types de point à exporter
Paramètres de sortie	Nom	Type	Description
	output	ComplexData	Export de la couche vecteur des points

TABLE 6 – Description du service "Refuge.info des objets de repère"

© WFS du réseau d'itinéraires construits de camptocamp

Description des deux WPS permettant de construire, en chaînant les process, les itinéraires issus de l'API des traces, de la source de données Camptocamp :

6. <http://www.geocatalogue.fr/>



Nom du service	API des sorties GPS de Camptocamp par BBOX		
Fonctionnalité	Ce service retourne les traces GPS des sorties de randonnées sur une emprise en zone de montagne et pour une activité donnée		
Paramètres d'entrée	Nom bbox	Type LiteralData	Description Les dimensions de la bbox à exporter
	activite	LiteralData	Le nom de l'activité sportive (skitouring, snow_ice_mixed, mountain_climbing, rock_climbing, ice_climbing, hingiking, snowshoeing, paragliding, mountain_biking, via_ferrata, slacklining)
Paramètres de sortie	Nom output	Type ComplexData	Description Export de la couche vecteur des points GPS

TABLE 7 – Description du service "C2C trace GPS"

Nom du service	WPS Map-matcher		
Fonctionnalité	Recalage des points sur un réseau routier de référence en appliquant l'algorithme proposé par Newson et Krumm (Newson et Krumm, 2009) fondé sur les modèles de Markov cachés.		
Paramètres d'entrée	Nom sigma	Type LiteralData	Description Ecart type des mesures d'erreur GPS
	radius	LiteralData	distance du buffer pour la recherche de candidat
	reseau_routier_reference	ComplexData	Le reseau routier de reference pour recalcr les points GPS
	trace gps	ComplexData	Les traces GPS à recalcr sur le réseau routier de référence
Paramètres de sortie	Nom output	Type ComplexData	Description Points recalés sur le réseau routier de référence

TABLE 8 – Description du service "WPS Map-matcher"



Nom du service	WPS RouteBuilding		
Fonctionnalité	Reconstruction de la géométrie linéaire de l'itinéraire. Les points issus de la trace de l'itinéraire sont parfois très éloignés les uns des autres, ce qui crée parfois des cassures entre les tronçons portant les points recalés. Le processus intègre l'ajout de ces tronçons manquants, en calculant le chemin le plus court sur le réseau routier entre 2 tronçons non connectés. Si le rapport entre la distance à vol d'oiseau et la distance du plus court chemin est inférieur à un seuil fixé, les tronçons manquants sont ajoutés à l'itinéraire. A la fin de cette étape, la géométrie d'un itinéraire est composée d'une liste de tronçons du réseau routier.		
Paramètres d'entrée	Nom	Type	Description
	ks	LiteralData	Seuil que le rapport maximal entre la distance à vol d'oiseau et la distance du chemin le plus entre 2 tronçons non connectés ne doit pas dépasser.
	reseau_routier _reference	ComplexData	Le reseau routier de reference pour recalcr les points GPS
	tableau	ComplexData	tableau d'index des points recalés avec l'identifiant des tronçons du réseau routier de référence
Paramètres de sortie	Nom	Type	Description
	output	ComplexData	Nom de l'export de la couche vecteur des linéaires des itinéraires

TABLE 9 – Description du service "WPS RouteBuilding"



5 Les propriétés non fonctionnelles principales de la norme ISO 19115

Les propriétés fonctionnelles étant définies, il reste à définir les propriétés relatives à la qualité de service (QoS) des services web de données. Dans un premier temps, nous définirons les métadonnées non fonctionnelles qui peuvent se définir ou se calculer automatiquement (grâce à la documentation) et ne nécessitant pas les outils d'appariement de données géographiques. Elles sont listées dans le tableau 10. Ces métadonnées concernent plutôt les éléments du noyau de la norme ISO 119115.

Catégorie	Indicateur
MD_Metadata	fileIdentifier
MD_Identification	MD_TopicCategoryCode
MD_Constraints	MD_LegalConstraints
MD_Constraints	MD_SecurityConstraints
MD_ReferenceSystem	MD_ReferenceSystem
MD_Identification	spatialRepresentationType
MD_Identification	MD_Resolution.equivalentScale
MD_SpatialRepresentation	MD_TopologyLevelCode
MD_SpatialRepresentation	geometryObjectTypeCode
MD_Identification	EX_Extent.EX_GeographicExtent

TABLE 10 – Liste des métadonnées principales retenues pour évaluer les sources de données

5.1 Métadonnées d'identification

5.1.1 La métadonnée *fileIdentifier*

Le tag du nom de fichier des métadonnées est <gmd:fileIdentifier>. Il faut en définir un pour chaque service identifié. Ce code est choisi de façon arbitraire, les résultats sont donnés dans le tableau 11 :

service	FileIdentifier
RefugeInfo	report_choucas_2019_refuge_infos.xml
C2C - Objets de repères	report_choucas_2019_c2c_objet_de_repere.xml
C2C - Sortie en randonnées	report_choucas_2019_c2c_sortie_pied.xml
PA - Objets de repères	report_choucas_2019_pa_objet_de_repere.xml
PA - Itinéraires	report_choucas_2019_pa_randonnee_pied.xml
Visorando	report_choucas_2019_visorando_trace.xml

TABLE 11 – Identifiant des fichiers des métadonnées des web services étudiés

Exemple en XML :

```
<gmd:fileIdentifier>
  <gco:CharacterString>report_choucas_2019_refuge_infos.xml</gco:CharacterString>
</gmd:fileIdentifier>
```



5.1.2 La métadonnée *MD_TopicCategory*

L'élément «topicCategory» permet de définir la catégorie des données du service. On doit la choisir parmi cette liste :

"farming" | "biota" | "boundaries" | "climatologyMeteorologyAtmosphere" | "economy" | "elevation" | "environment" | "geoscientificInformation" | "health" | "imageryBaseMapsEarthCover" | "intelligenceMilitary" | "inlandWaters" | "location" | "oceans" | "planningCadastre" | "society" | "structure" | "transportation" | "utilitiesCommunication"

Ce qui nous donne dans le tableau 12 :

service	TopicCategoryCode
RefugeInfo	location
C2C - Objets de repères	location
C2C - Traces	location
C2C - Sortie de randonnées	transportation
PA - Objets de repères	location
PA - Traces	location
PA - Itinéraires à pied	transportation
Visorando - Traces	location
Visorando - Itinéraires	transportation

TABLE 12 – Catégorie des données des web services étudiés

Exemple en XML :

```
<gmd:topicCategory>  
  <gmd:MD_TopicCategoryCode>location</gmd:MD_TopicCategoryCode>  
</gmd:topicCategory>
```

5.1.3 La métadonnée *MD_LegalConstraints*

Les licences des services web de données ont été récupérées sur les sites web des sources de données. Elles sont listées dans le tableau 13 :



service	Licence indiquée sur le site internet
RefugeInfo	Les données du site Refuges.info sous sont licence CC-By-Sa 2.0 , cependant l'API permet aussi de retourner des données provenant d'OpenStreetMap via sa fonction de recherche, ces données sont sous licence ODbL. https://opendatacommons.org/licenses/odbl/ , https://creativecommons.org/licenses/by-sa/2.0/deed.fr
C2C - Objets de repères et itinéraires	Le contenu collaboratif coordonnées géographiques est sous double licence CC by-sa et Open Database Licence. https://creativecommons.org/licenses/by-sa/3.0/fr/ , https://opendatacommons.org/licenses/odbl/
C2C - Traces	Les traces sont sous licence CC-by-nc-nd https://creativecommons.org/licenses/by-nc-nd/3.0/fr/
PA - Objets de repères, traces et itinéraires	Opensource, publiée sous licence libre BSD
Visorando - Traces et itinéraire	Licence particulière définie ici : https://www.visorando.com/confidentialite.html

TABLE 13 – Licence des web services étudiés

Exemple en XML :

```

<gmd:resourceConstraints>
  <gmd:MD_LegalConstraints>
    <gmd:useLimitation xsi:type="gmd:PT_FreeText_PropertyType">
      <gco:CharacterString>licence CC-By-Sa 2.0.</gco:CharacterString>
    </gmd:useLimitation>
    <gmd:accessConstraints>
      <gmd:MD_RestrictionCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_RestrictionCode" codeListValue="license"/>
    </gmd:accessConstraints>
    <mcp:licenseLink>
      <gmd:URL>https://creativecommons.org/licenses/by-sa/2.0/deed.fr</gmd:URL>
    </mcp:licenseLink>
    <gmd:useConstraints>
      <gmd:MD_RestrictionCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_RestrictionCode" codeListValue="otherRestrictions"/>
    </gmd:useConstraints>
    <gmd:otherConstraints xsi:type="gmd:PT_FreeText_PropertyType">
      <gco:CharacterString>Cependant l'API permet aussi de retourner des donnees provenant d'OpenStreetMap via sa fonction de recherche, ces donnees sont sous licence ODbL : https://opendatacommons.org/licenses/odbl/. </gco:CharacterString>
    </gmd:otherConstraints>
  </gmd:MD_LegalConstraints>
</gmd:resourceConstraints>

```



5.1.4 La métadonnée *MD_SecurityConstraints*

Concernant les contraintes de sécurité sur les sources de données de notre étude, on n'a aucune restriction. Les limitations sont à choisir parmi : unclassified, restricted, confidential, secret, topSecret.

```
<gmd:metadataConstraints>
  <gmd:MD_SecurityConstraints>
    <gmd:useLimitation>
      <gco:CharacterString>Aucune contrainte</gco:CharacterString>
    </gmd:useLimitation>
    <gmd:classification>
      <gmd:MD_ClassificationCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codelist/ML_gmxCodelists.xml#MD_ClassificationCode" codeListValue="unclassified"/>
    </gmd:classification>
  </gmd:MD_SecurityConstraints>
</gmd:metadataConstraints>
```

5.2 Métadonnées des systèmes de référence

5.2.1 La métadonnée *MD_ReferenceSystem*

Pour les objets de repère, ces métadonnées concernent l'information relative à la description du ou des systèmes de référence spatiale disponible pour le service web.

Pour les itinéraires, les points des traces GPS ont été (re-)projetés à l'entrée du service de recalage en conformité avec le réseau routier de la BDTopo©.

Les informations sont représentées dans le tableau 14

service	EPSG
RefugeInfo	4326
C2C - Objets de repères	3857
PA - Objets de repères	4326
C2C - Itinéraires des sortie de randonnées	2154
PA - Itinéraires à pied	2154
Visorando - Itinéraires	2154

TABLE 14 – Les systèmes de référence des web services étudiés

```
<gmd:referenceSystemInfo>
  <gmd:MD_ReferenceSystem>
    <gmd:referenceSystemIdentifier>
      <gmd:RS_Identifier>
        <gmd:code>
          <gmx:Anchor xlink:href="http://www.opengis.net/def/crs/EPG/0/4326">EPSG:4326</gmx:Anchor>
        </gmd:code>
      </gmd:RS_Identifier>
    </gmd:referenceSystemIdentifier>
  </gmd:MD_ReferenceSystem>
</gmd:referenceSystemInfo>
```



ou

```
<gmx:Anchor xlink:href="http://www.opengis.net/def/crs/EPSSG/0/3857">EPSG:3857</gmx:Anchor>
```

ou

```
<gmx:Anchor xlink:href="http://www.opengis.net/def/crs/EPSSG/0/2154">EPSG:2154</gmx:Anchor>
```

5.3 Métadonnées d’information et représentation spatiale et temporelle

L’information et la représentation spatiale et temporelle se trouve dans 2 éléments de la classe MD_Metadata : “MD_Identification” et “MD_SpatialRepresentation”. Ce dernier élément apporte une description détaillée des représentations vectorielles ou matricielles et la nature des objets géométriques utilisés.

L’élément “SpatialRepresentation” décrit le type de représentation de la source de données

5.3.1 La métadonnée MD_Resolution

La résolution spatiale décrit le niveau de détail des données. Elle s’exprime soit avec une échelle (échelle optimale d’utilisation des données) soit avec une distance de résolution. Pour les objets de repère et les itinéraires (traces recalées sur le réseau routier de l’IGN) nous utiliserons l’échelle.

La valeur correspond à la valeur de l’échelle, nous avons pris la même valeur que celle de la BDTopo (Cf. tableau 15).

Service	Dénominateur de l’échelle
RefugeInfo	10000
C2C - Objets de repères	10000
C2C - Itinéraires de randonnées	10000
PA - Objets de repères	10000
PA - Itinéraires à pied	10000
Visorando - Itinéraires	10000

TABLE 15 – Résolution spatiale des web services étudiés

Ce qui donne par exemple en XML :

```
<gmd:spatialResolution>
  <gmd:MD_Resolution>
    <gmd:equivalentScale>
      <gmd:MD_RepresentativeFraction id="Refuges.info2019.spatialResolution">
        <gmd:denominator>
          <gco:Integer>10000</gco:Integer>
        </gmd:denominator>
      </gmd:MD_RepresentativeFraction>
    </gmd:equivalentScale>
  </gmd:MD_Resolution>
</gmd:spatialResolution>
```

5.3.2 La métadonnée MD_SpatialRepresentation

L’élément “SpatialRepresentationType” décrit le type d’objet pour représenter l’information géographique des web services de données. Il prend une valeur parmi : vector, grid, textTable, tin, stereoModel, video



Dans notre étude, toutes les sources sont classées «vector».

```
<gmd:spatialRepresentationType>
  <gmd:MD_SpatialRepresentationTypeCode codeList="http://standards.iso.org/iso/19139/resources/gmxCodeList
</gmd:spatialRepresentationType>
```

5.3.3 La métadonnée *MD_TopologyLevelCode*

“MD_TopologyLevelCode” représente le complexe topologique bidimensionnel qui est isomorphe à un sous-ensemble d’une surface. Les valeurs possibles : geometryOnly, topology1D, planarGraph, fullPlanarGraph, surfaceGraph, fullSurfaceGraph, topology3D, fullTopology3D, abstract

Le réseau routier de la BDTopo© n’est pas un graphe planaire car l’attribut *Position par rapport au sol* donne le niveau de l’objet par rapport à la surface du sol (valeur négative pour un objet souterrain, nulle pour un objet au sol et positive pour un objet en sursol). Si l’objet en sursol passe au dessus d’autres objets en sursol, sa valeur ‘Position par rapport au sol’ est égale à "1 + le nombre d’objets intercalés". De la même façon, un souterrain peut prendre une valeur ‘Position par rapport au sol’ égale à "-1 – le nombre d’objets souterrains intercalés". Cependant en zone de montagne, les itinéraires passent quasi jamais au dessus d’un autre itinéraire. Le réseau construit à l’aide de la chaîne de composition de notre étude est donc bien planaire.

Service	TopologicalLevelCode
RefugeInfo	geometryOnly
C2C - Objets de repères	geometryOnly
C2C - Sortie de randonnées	planarGraph
PA - Objets de repères	geometryOnly
PA - Itinéraires à pied	planarGraph
Visorando - Itinéraires	planarGraph

TABLE 16 – «TopologicalLevelCode» des web services étudiés

5.3.4 La métadonnée *MD_GeometricObjectTypeCode*

Le type de géométrie utilisé pour représenter les objets géographiques de la source de données est défini par l’élément : “MD_GeometricObjectTypeCode”. Les valeurs possibles : complex, composite, curve, point, solid, surface.

Ce qui donne pour nos services de données le tableau 17 :

Service	GeometricObjectTypeCode
RefugeInfo	point
C2C - Objets de repères	point
C2C - Sortie de randonnées	curve
PA - Objets de repères	point
PA - Itinéraires à pied	curve
Visorando - Itinéraires	curve

TABLE 17 – «GeometricObjectTypeCode» des web services étudiés



5.3.5 La métadonnée *EX_Extent*

L'information de la couverture spatiale et temporelle de la ressource est dans décrite dans l'élément "gmd:EX_Extent". La couverture spatiale du jeu de données dans son ensemble est stockée dans l'identification des métadonnées : MD_Metadata.MD_DataIdentification.extent.

L'extension spatiale des services de données fournissant les objets de repère correspond à l'extension spatiale décrite dans la documentation des API de données. C'est une extension décrite par une chaîne de caractères.

Concernant l'extension spatiale des services de données fournissant les itinéraires, l'extension spatiale a été réduite à la zone d'étude du projet CHOUCAS. Seules les traces GPS de cette zone ont été téléchargées.

Nous verrons plus tard que les indicateurs de qualité ont été calculés sur cette même emprise de la zone d'étude et elle sera décrite dans l'élément MD_Metadata.MD_Quality.DQ_Scope

Service	Emprise
RefugeInfo	France métropolitaine
C2C	France, Allemagne, Suisse, Espagne
PA	Parc régional du Vercors + Parc National des Écrins
Visorando	Monde

TABLE 18 – Emprise des API de données étudiées

Exemple XML pour l'emprise d'un service objet de repères :

```
<gmd:extent>
  <gmd:EX_Extent>
    <gmd:geographicElement>
      <gmd:EX_GeographicDescription>
        <gmd:geographicIdentifier>
          <gmd:MD_Identifier>
            <gmd:code>
              <gco:CharacterString>France metropolitaine</gco:CharacterString>
            </gmd:code>
          </gmd:MD_Identifier>
        </gmd:geographicIdentifier>
      </gmd:EX_GeographicDescription>
    </gmd:geographicElement>
  </gmd:EX_Extent>
</gmd:extent>
```

Exemple XML pour l'emprise d'un service itinéraires :

```
<gmd:extent>
  <gmd:EX_Extent>
    <gmd:geographicElement>
      <gmd:EX_GeographicBoundingBox>
        <gmd:westBoundLongitude>
          <gco:Decimal>5.003586</gco:Decimal>
        </gmd:westBoundLongitude>
        <gmd:eastBoundLongitude>
```



```
    <gco:Decimal>6.63400719025316</gco:Decimal>
  </gmd:eastBoundLongitude>
  <gmd:southBoundLatitude>
    <gco:Decimal>44.5175629999999</gco:Decimal>
  </gmd:southBoundLatitude>
  <gmd:northBoundLatitude>
    <gco:Decimal>45.8508029999999</gco:Decimal>
  </gmd:northBoundLatitude>
</gmd:EX_GeographicBoundingBox>
</gmd:geographicElement>
</gmd:EX_Extent>
</gmd:extent>
```



6 Les propriétés non fonctionnelles de filiation

L'élément "gmd:LI_Lineage" permet de définir la filiation de construction des données de la donnée brute à la donnée calculée, en précisant les protocoles de saisie, les différents traitements opérés dans un certain contexte.

Catégorie	Indicateur	Type de données
LI_Lineage	Protocole de saisie des objets de repère	Points de repère
LI_Lineage	Méthode d'acquisition et intégration des traces GPS	Itinéraires
LI_ProcessStep	Algorithme de map matching	Itinéraires
LI_ProcessStep	Algorithme de construction d'itinéraires	Itinéraires
LI_Source	Sources des traces GPS, réseau routier de référence	Itinéraires

TABLE 19 – Liste des métadonnées de filiation retenues pour évaluer les services de données

6.1 Le lignage des objets de repère

Pour définir le protocole de saisie des données issues des services des objets de repère, nous avons défini une description de cette filiation à partir de la documentation des sites internet et des guides d'utilisation des interfaces utilisateurs. Nous avons surtout pris en considération les types d'outils utilisés pour la saisie (bureautique, capteur, terrain) et la nature du contributeur (citoyen, personne d'autorité)

Ce qui donne pour les objets de repères :

- ~ **Refuges.info** RefugeInfo : RefugeInfo est un site web collaboratif proposant différentes données géographiques ponctuelles en zones de montagne. Les localisations sont généralement obtenues par numérisation à l'aide de cartes. L'API permet aussi de retourner des données provenant d'OpenStreetMap.
- ~ **Objets de repère de CampToCamp** Camptocamp (C2C) est un projet collaboratif auquel participe des pratiquants plutôt avertis des sports de montagne. La saisie des objets de repère se fait via une application web par les contributeurs avec leurs propres outils de gestion de données et leur propre fond cartographique. L'altitude est saisie par les utilisateurs.
- ~ **Objets de repère des Aires Protégées** Parmi les aires protégées (AP), les Parcs naturels régionaux (PNR) et les Parcs nationaux (PN) sont des institutions publiques françaises ayant pour mission de préserver l'environnement naturel et de développer des activités économiques et touristiques. Par ailleurs, les Parcs du territoire français gèrent des informations telles que les points de relief, les points d'intérêt patrimoniaux, les services (informations pratiques comme les points d'eau, les passages difficiles) et les itinéraires. La plupart d'entre eux utilisent les logiciels de la suite GeoTrek, les données de la BDTopo ainsi que le Modèle Numérique de Terrain (MNT) de l'IGN comme référentiel pour administrer leurs données. Mentionnons que ces données d'autorité (BDTopo et base de données sur les aires protégées) ne sont pas équivalentes en termes de contenu parce qu'elles ont des points de vue différents (topographique vs. environnemental et touristique). Elles sont plutôt complémentaires avec une petite redondance en termes de types de fonctionnalités et d'informations thématiques. L'entrée des données fournies par l'API est l'itinéraire ; les objets de repère y sont reliés par référencement linéaire. Le nombre d'objet de repère est fixé à trois au minimum et ces objets doivent être bien répartis sur le parcours.

Ce qui va donner en XML :

```
<gmd:lineage>
  <gmd:LI_Lineage>
    <gmd:statement>
      <gco:CharacterString>A COMPLETER par les textes au-dessus</gco:CharacterString>
    </gmd:statement>
  </gmd:LI_Lineage>
</gmd:lineage>
```



```
</gmd:LI_Lineage>  
</gmd:lineage>
```

6.2 Le lignage des itinéraires

Le lignage des itinéraires est un peu plus complexe car il prend en compte deux traitements sur les données sources des traces GPS.

6.2.1 Méthode d'acquisition des traces

Comme pour les objets de repère, nous avons défini une description de la filiation des données sources GPS à partir de la documentation des sites internet et des guides d'utilisation des interfaces utilisateurs. Nous avons pris en considération les types d'outils utilisés pour la saisie (bureautique, capteur, terrain), la nature du contributeur (citoyen, personne d'autorité) et les fonds de carte accessibles dans les outils cartographiques. Ce qui donne pour les traces GPS :

- ~ **Réseau d'itinéraire de CampToCamp** Camptocamp (C2C) est un projet collaboratif auquel participe des pratiquants plutôt avertis des sports de montagne. La création des géométries est réalisée par les contributeurs avec leurs propres outils de gestion de données et leur propre fond cartographique. L'altitude est saisie par les utilisateurs. Très peu de sorties de randonnées pédestres ont une trace GPX (25% dans la zone d'étude du projet CHOUCAS).
- ~ **Réseau d'itinéraires des aires protégées** Parmi les aires protégées (AP), les Parcs naturels régionaux (PNR) et les Parcs nationaux (PN) sont des institutions publiques françaises ayant pour mission de préserver l'environnement naturel et de développer des activités économiques et touristiques. Par ailleurs, les Parcs du territoire français gèrent des informations telles que les points de relief, les points d'intérêt patrimoniaux, les services (informations pratiques comme les points d'eau, les passages difficiles) et les itinéraires. La plupart d'entre eux utilisent les logiciels de la suite GeoTrek, les données de la BDTopo ainsi que le Modèle Numérique de Terrain (MNT) de l'IGN comme référentiel pour administrer leurs données. Mentionnons que ces données d'autorité (BDTopo et base de données sur les aires protégées) ne sont pas équivalentes en termes de contenu parce qu'elles ont des points de vue différents (topographique vs. environnemental et touristique). Elles sont plutôt complémentaires avec une petite redondance en termes de types de fonctionnalités et d'informations thématiques. L'entrée des données fournies par l'API est l'itinéraire; les objets de repère y sont reliés par référencement linéaire. Les itinéraires sont définis par une géométrie fournie dans un fichier GPX.
- ~ **Réseau d'itinéraires de Visorando**
Visorando est un site web offrant la possibilité de télécharger des traces publiées en ligne par des contributeurs. 100% des géométries des itinéraires ont été produites à partir de données issues de capteurs GPS. En plus des traces, des attributs tels que le titre, le type d'activité (randonnée, vélo, etc.), la description textuelle de l'itinéraire, la durée, la difficulté et le dénivelé sont fournis. Le site Web propose aux contributeurs d'ajouter plusieurs points de passages aux traces (points de repère ou changement majeur de direction). Les traces sont ensuite ajustées et validées par les contributeurs dans une interface web cartographique utilisant les fonds cartographiques de l'IGN (données faisant autorité).

6.2.2 Les étapes de transformation

On définit le modèle de données des étapes de transformation des itinéraires dans la figure 10 suivante. Le traitement comprend deux étapes (le recalage des données GPS et la construction de l'itinéraire), deux sources externes (les traces GPS et un réseau routier) et un produit intermédiaire (les points recalés sur le réseau routier). Notons qu'il faut le même réseau routier en entrée des deux étapes du traitement.

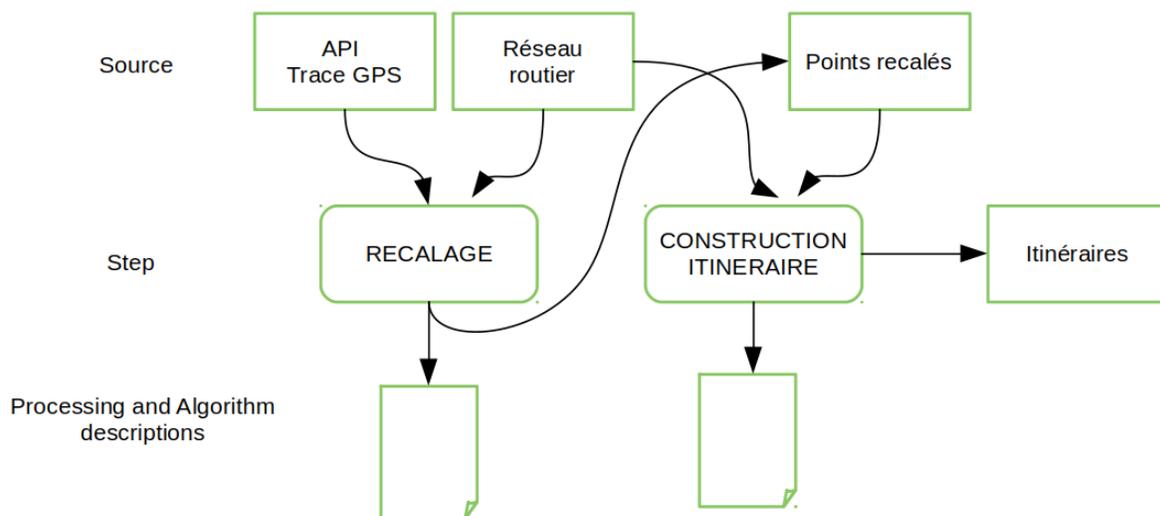


FIGURE 10 – Modèle de la provenance des itinéraires

Ce qui va donner en XML :

```
<gmd:LI_Lineage xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:srv="http://www.isotc211.org/2005/srv" xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:gsrc="http://www.isotc211.org/2005/gsrc" xmlns:gss="http://www.isotc211.org/2005/gss"
xmlns:gts="http://www.isotc211.org/2005/gts" xmlns:gmx="http://www.isotc211.org/2005/gmx"
xsi:schemaLocation="http://www.isotc211.org/2005/gmi
http://www.ngdc.noaa.gov/metadata/published/xsd/schema.xsd">

<gmd:statement>
  <gco:CharacterString>Camptocamp (C2C) est un projet collaboratif auquel participe des pratiquants
  plutot avertis des sports de montagne. La creation des geometries est realisee par les contributeurs
  avec leurs propres outils de gestion de donnees et leur propre fond cartographique. L'altitude
  est saisie par les utilisateurs. Tres peu de sorties de randonnees pedestres ont une trace GPX
  (25% dans la zone d'etude du projet CHOUCAS).</gco:CharacterString>
</gmd:statement>

<gmd:processStep>
  <gmi:LE_ProcessStep id="ps_MAPMATCHING">
    <gmd:description>
      <gco:CharacterString>Recalage des points sur un reseau routier de reference en appliquant
      l'algorithmme propose par Newson et Krumm (Newson et Krumm, 2009) fonde sur les modeles
      de Markov caches.</gco:CharacterString>
    </gmd:description>
    <gmd:dateTime gco:nilReason="2018/12/05" />
    <gmd:processor>
      <gmd:CI_ResponsibleParty>
        <gmd:organisationName>
          <gco:CharacterString>LASTIG, Univ Gustave Eiffel, ENSG, IGN,
          F-94160 Saint-Mande, France</gco:CharacterString>
        </gmd:organisationName>
      </gmd:CI_ResponsibleParty>
    </gmd:processor>
  </gmi:LE_ProcessStep>
</gmd:processStep>
```



```

    </gmd:organisationName>
  </gmd:CI_ResponsibleParty>
</gmd:processor>
<gmd:source xlink:href="#src_API_C2C_GPX" />
<gmd:source xlink:href="#src_BDTopo_V2" />
<gmi:output xlink:href="#POINTS_RECALES" />
<gmi:processingInformation>
  <gmi:LE_Processing>
    <gmi:identifiant>
      <gmd:MD_Identifier>
        <gmd:code>
          <gco:CharacterString>Map-matcher</gco:CharacterString>
        </gmd:code>
      </gmd:MD_Identifier>
    </gmi:identifiant>
    <gmi:algorithm>
      <gmi:description>algorithme de recalage de traces de Newson et Krumm (2009)</gmi:description>
      <gmi:citation>https://dl.acm.org/doi/10.1145/1653771.1653818</gmi:citation>
    </gmi:algorithm>
    <gmi:softwareReference>
      <gmd:CI_Citation>https://github.com/IGNF/mapmatcher,
        https://tel.archives-ouvertes.fr/tel-02493936</gmd:CI_Citation>
    </gmi:softwareReference>
  </gmi:LE_Processing>
</gmi:processingInformation>
<gmi:report>
  <gmi:LE_ProcessStepReport>
    <gmi:name>
      <gco:CharacterString>C2C_MAPMATCHING_REPORT</gco:CharacterString>
    </gmi:name>
    <gmi:description>
      <gco:CharacterString>Les parametres concernant le recalage des points des traces
        sur le reseau routier sont de : 80m pour la recherche de candidats et 4m pour
        l'ecart type des mesures d'erreur GPS. Dans l'ensemble, 97% des points (298348 points)
        sont recales et la racine carre de l'erreur quadratique moyenne des déplacements induits
        par le recalage est egale a 12,94 m. Il est a noter que quelques impasses ont ete generees
        apres le processus de recalage des traces.</gco:CharacterString>
    </gmi:description>
  </gmi:LE_ProcessStepReport>
</gmi:report>

</gmi:LE_ProcessStep>
</gmd:processStep>

<gmd:processStep>
  <gmi:LE_ProcessStep id="ps_ROUTEBUILDING">
    <gmd:description>
      <gco:CharacterString>Reconstruction de la geometrie lineaire de l'itineraire. Les points issus
        de la trace de l'itineraire sont parfois tres eloignes les uns des autres, ce qui cree parfois
        des cassures entre les troncons portant les points recales. Le processus integre l'ajout
        de ces troncons manquants, en calculant le chemin le plus court sur le reseau routier entre
        2 troncons non connectes. Si le rapport entre la distance a vol d'oiseau et la distance
        du plus court chemin est inferieur a un seuil fixe, les troncons manquants sont ajoutees
        a l'itineraire. A la fin de cette etape, la geometrie d'un itineraire
    </gco:CharacterString>
  </gmd:description>

```



```

    est composee d'une liste de troncons du reseau routier.</gco:CharacterString>
</gmd:description>
<gmd:dateTime gco:nilReason="2018/11/29"/>
<gmd:processor>
  <gmd:CI_ResponsibleParty>
    <gmd:organisationName>
      <gco:CharacterString>LASTIG, Univ Gustave Eiffel, ENSG, IGN,
        F-94160 Saint-Mande, France</gco:CharacterString>
    </gmd:organisationName>
  </gmd:CI_ResponsibleParty>
</gmd:processor>
<gmd:source xlink:href="#src_BDTopo_V2" />
<gmi:source xlink:href="#POINTS_RECALES" />
<gmi:output xlink:href="#C2C_ITINERAIRE" />
<gmi:processingInformation>
  <gmi:LE_Processing>
    <gmi:identifiser>
      <gmd:MD_Identifier>
        <gmd:code>
          <gco:CharacterString>Route_Building</gco:CharacterString>
        </gmd:code>
      </gmd:MD_Identifier>
    </gmi:identifiser>
    <gmi:softwareReference>
      <gmd:CI_Citation>https://github.com/IGNF/mapmatcher</gmd:CI_Citation>
    </gmi:softwareReference>
  </gmi:LE_Processing>
</gmi:processingInformation>
<gmi:report>
  <gmi:LE_ProcessStepReport>
    <gmi:name>
      <gco:CharacterString>C2C_ROUTEBUILDING_REPORT</gco:CharacterString>
    </gmi:name>
    <gmi:description>
      <gco:CharacterString>L'etape de reconstruction de la geometrie lineaire des itineraires,
        le rapport maximal entre la distance a vol d'oiseau et la distance du chemin le plus court entre
        2 troncons non connectes est estime empiriquement a 5. La longueur totale des troncons du reseau
        routier manquants ajoutes est egale a 337 km, soit 4% de la longueur totale des itineraires.
      </gco:CharacterString>
    </gmi:description>
  </gmi:LE_ProcessStepReport>
</gmi:report>
</gmi:LE_ProcessStep>
</gmd:processStep>

<gmd:source>
  <gmd:LI_Source id="src_API_C2C_GPX">
    <gmd:description>
      <gco:CharacterString>API des sorties de C2C par BBOX sous format GPX</gco:CharacterString>
    </gmd:description>
    <gmd:sourceCitation>
      <gmd:CI_Citation>
        <gmd:date>
          <gmd:CI_Date>

```



```

        <gmd:date>
            <gco:Date>2018-10-26</gco:Date>
        </gmd:date>
    </gmd:CI_Date>
</gmd:date>
<gmd:identifiant>
    <gmd:MD_Identifier>
        <gmd:code>
            <gco:CharacterString>src_API_C2C_GPX</gco:CharacterString>
        </gmd:code>
    </gmd:MD_Identifier>
</gmd:identifiant>
<gmd:citedResponsibleParty>
    <gmd:CI_ResponsibleParty>
        <gmd:organisationName>
            <gco:CharacterString>Camptocamp.org.</gco:CharacterString>
        </gmd:organisationName>
    </gmd:CI_ResponsibleParty>
</gmd:citedResponsibleParty>
</gmd:CI_Citation>
</gmd:sourceCitation>
<gmd:sourceStep xlink:href="#ps_MAPMATCHING"/>
</gmd:LI_Source>
</gmd:source>

<gmd:source>
    <gmd:LI_Source id="src_BDTopo_V2">
        <gmd:description>
            <gco:CharacterString>WFS Geoportail. Ce service permet la visualisation de données vecteur IGN
            au travers d'un flux WFS</gco:CharacterString>
        </gmd:description>
        <gmd:sourceCitation>
            <gmd:CI_Citation>
                <gmd:title>
                    <gco:CharacterString>BDTopo_V2:troncon_de_route</gco:CharacterString>
                </gmd:title>
                <gmd:date>
                    <gmd:CI_Date>
                        <gmd:date>
                            <gco:Date>2018-06</gco:Date>
                        </gmd:date>
                    </gmd:CI_Date>
                </gmd:date>
            </gmd:CI_Citation>
            <gmd:identifiant>
                <gmd:MD_Identifier>
                    <gmd:code>
                        <gco:CharacterString>IGN</gco:CharacterString>
                    </gmd:code>
                </gmd:MD_Identifier>
            </gmd:identifiant>
        </gmd:sourceCitation>

        <gmd:sourceStep xlink:href="#ps_MAPMATCHING"/>
    </gmd:LI_Source>
</gmd:source>

```



```

    <gmd:sourceStep xlink:href="#ps_ROUTEBUILDING"/>
  </gmd:LI_Source>
</gmd:source>

<gmd:source>
  <gmd:LI_Source id="POINTS_RECALES">
    <gmd:description>
      <gco:CharacterString>Points GPS recales sur un reseau routier de reference,
      la BDTopo V2</gco:CharacterString>
    </gmd:description>
    <gmd:sourceCitation>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>Points recales sur BDTopo</gco:CharacterString>
        </gmd:title>
        <gmd:identifiant>
          <gmd:MD_Identifier>
            <gmd:code>
              <gco:CharacterString>POINTS_RECALES</gco:CharacterString>
            </gmd:code>
            </gmd:MD_Identifier>
          </gmd:identifiant>
          <gmd:citedResponsibleParty>
            <gmd:CI_ResponsibleParty>
              <gmd:organisationName>
                <gco:CharacterString>LASTIG, Univ Gustave Eiffel, ENSG, IGN,
                F-94160 Saint-Mande, France</gco:CharacterString>
              </gmd:organisationName>
            </gmd:CI_ResponsibleParty>
          </gmd:citedResponsibleParty>
        </gmd:CI_Citation>
      </gmd:sourceCitation>

      <gmd:sourceStep xlink:href="#ps_MAPMATCHING"/>
      <gmd:sourceStep xlink:href="#ps_ROUTEBUILDING"/>
    </gmd:LI_Source>
  </gmd:source>
</gmd:LI_Lineage>

```



7 Les propriétés non fonctionnelles de qualité de la norme ISO 19157

Dans la norme ISO 19115, l'élément `DQ_DataQuality` permet de formaliser les informations de qualité. Comme cela a été décrit dans le chapitre 2 cet élément est implémenté suivant le standard 19157. Elles sont décrites dans les fichiers `XXX_Report.xml` qui font partie du livrable.

Toutes les informations ont été calculées à partir des résultats issus d'un appariement multi-critères décrit précédemment.

Notons que l'algorithme d'appariement s'utilise sur un jeu de données de référence et sur un jeu de données de comparaison donnant ainsi une direction à l'appariement (pour chaque donnée du jeu de référence, l'algorithme recherche les données homologues et candidates dans le jeu de comparaison). Mentionnons que le jeu de données de référence peut être un jeu de données faisant autorité ou un jeu de données collaboratif. Dans notre étude, les sources de données des objets de repères ont été appariés au jeu de données de référence des points de repère de la *BDTopo*© de l'IGN alors que les sources de données des itinéraires, toutes collaboratives, ont été appariées entre elles afin de constituer les métadonnées de qualité. C'est pourquoi dans les fichiers XML de rapportage des métadonnées de qualité, on trouve un indicateur de qualité par source de données comparées.

Catégorie	Indicateur	Mesure	Type de données
Metaquality	Confidence	Confidence	Objets de repère, itinéraires
PositionalAccuracy	AbsoluteExternalPositionalAccuracy	MeanAbsolute2D, Root-MeanSquareError, Agreement Rate	Objets de repères, itinéraires
ThematicAccuracy	NonQuantitativeAttributeAccuracy	Similarité de Samal	Objets de repère (sur l'attribut nom)
ThematicAccuracy	NonQuantitativeAttributeAccuracy	Distance cosinus	Itinéraires (sur l'attribut titre)
ThematicAccuracy	NonQuantitativeAttributeAccuracy	Connectivity	Itinéraires (sur l'attribut géométrie)
ThematicAccuracy	ThematicClassificationCorrectness	OverAllAccuracy, ConfusionMatrix	Objets de repère (sur l'attribut nature), itinéraires (sur l'attribut cotation)
Completeness	CompletenessOmission	MissingItems	Objets de repère
Completeness	CompletenessOmission	MissingClass	Objets de repère (sur l'attribut nom)
Completeness	CompletenessComission	Excess	Objets de repère
Completeness	CompletenessComission	Duplicate	Objets de repère, itinéraires

TABLE 20 – Liste des métadonnées de qualité retenues pour évaluer les sources de données

7.1 Contexte

Avant de commencer la description des métadonnées de qualité, commençons par décrire le contexte. L'objet `DQ_Scope` permet de définir le contexte spatio-temporel de l'étude de la qualité.

L'analyse de la qualité des différentes sources de données s'est réalisée sur l'emprise géographique de la zone d'étude du projet CHOUCAS. Les données ont été extraites à une date fixée, ce qui permet de définir l'emprise temporelle.

Ce qui donne en XML pour le web service des objets de repère de la source `Refuge.info` :



```

<gmd:EX_Extent>
  <gmd:description>
    <gco:CharacterString>Actualite des objets de reperes de la source Refuges.info
    et la zone d'etude</gco:CharacterString>
  </gmd:description>
  <gmd:temporalElement>
    <gmd:EX_TemporalExtent>
      <gmd:extent>
        <gml:TimePeriod gml:id="Refuges.info.temporalExtent">
          <gml:beginPosition>2008-10-20</gml:beginPosition>
          <gml:endPosition>2018-07-22</gml:endPosition>
        </gml:TimePeriod>
      </gmd:extent>
    </gmd:EX_TemporalExtent>
  </gmd:temporalElement>
  <gmd:geographicElement>
    <gmd:EX_GeographicBoundingBox>
      <gmd:westBoundLongitude>
        <gco:Decimal>5.003586</gco:Decimal>
      </gmd:westBoundLongitude>
      <gmd:eastBoundLongitude>
        <gco:Decimal>6.63400719025316</gco:Decimal>
      </gmd:eastBoundLongitude>
      <gmd:southBoundLatitude>
        <gco:Decimal>44.5175629999999</gco:Decimal>
      </gmd:southBoundLatitude>
      <gmd:northBoundLatitude>
        <gco:Decimal>45.8508029999999</gco:Decimal>
      </gmd:northBoundLatitude>
    </gmd:EX_GeographicBoundingBox>
  </gmd:geographicElement>
</gmd:EX_Extent>

```

7.2 Métadonnées des métadonnées de qualité

7.2.1 La métadonnée *gmd :DQ_MetaqualityConfidence*

Nous avons calculé la mesure de confiance pour les objets de repère et les itinéraires.

© Méthode

Cet indicateur mesure la confiance dans les mesures de qualité obtenues à partir des résultats d'appariement. Nous exprimons cette confiance en fonction de l'incertitude liée à l'appariement de données. Elle est définie comme étant le taux d'objets pour lequel l'algorithme n'a pas pu prendre une décision (e.g. les cas indécis) et ceux appariés à tort (e.g. nous savons que l'appariement est faux mais nous ne savons pas avec certitude quels sont les objets homologues dans la source de comparaison) par rapport au nombre total des objets à appairer.

Confiance = 1 – *Incertainde*,

où *Incertainde* = $\frac{NbObjetsSansDecision+NbObjetsAppariésATort}{nombretotald'objetsdelasource}$

© Résultats



Ce qui donne pour les métadonnées des objets de repère :

- ~ **Refuges.info**
DQ_Metaquality_Confidence = 0.84
- ~ **Camptocamp - objets de repère**
DQ_Metaquality_Confidence = 0.93
- ~ **Aires protégées - objets de repère**
DQ_Metaquality_Confidence = 0.9

Et pour les métadonnées des itinéraires :

- ~ **Camptocamp - itinéraires**
DQ_Metaquality_Confidence_AP = 0.72
DQ_Metaquality_Confidence_Visorando = 0.76
- ~ **Aires protégées - itinéraires**
DQ_Metaquality_Confidence_C2C = 0.81
DQ_Metaquality_Confidence_Visorando = 0.83
- ~ **Visorando - itinéraires**
DQ_Metaquality_Confidence_C2C = 0.92
DQ_Metaquality_Confidence_AP = 0.92

Nous observons que les valeurs de certitude liées à l'appariement de données sont élevées pour toutes les sources. Ceci signifie que les valeurs des indicateurs de qualité calculées en se basant sur les liens d'appariement sont fiables.

© Exemple XML

```
<!-- ===== -->
<!-- The trustworthiness of a data quality result.-->
<gmd:DQ_MetaqualityConfidence>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Confidence</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>Data_quality_confidence_Refuges_info</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString>Expresses the trustworthiness of a data quality result.
    It is computed by 1-(uncertainty of the matching results), where uncertainty
    of the matching is the rate of features not matched by the algorithm because of conflict
    and features which were wrongly matched by the algorithm with respect to the total number
    of the refuges_info features.
    </gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unit"/>
```



```

<gmd:value>
  <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xsi:type="xs:double">0.84</gco:Record>
</gmd:value>
</gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_MetaqualityConfidence>
    
```

7.3 Les métadonnées d'exactitude de position

Dans cette partie nous étudierons l'exactitude de position, c'est à dire celles qui définissent la proximité des données du service par rapport à celles du terrain nominal ou celles reconnues (données de référence).

7.3.1 La mesure MeanAbsolute2D de l'indicateur *DQ_AbsoluteExternalPositionalAccuracy* pour les objets de repères

© Méthode

Pour mesurer l'indicateur *DQ_AbsoluteExternalPositionalAccuracy*, plusieurs métriques complémentaires peuvent être utilisées :

- **Erreur planimétrique (MeanAbsolute2D)** est défini comme la moyenne des écarts en x et en y :

$$m_{XY} = \frac{1}{n} \sum_{i=1}^n dXY_i$$

$$dXY_i = \sqrt{(X_{i,TN} - X_{i,JD})^2 + (Y_{i,TN} - Y_{i,JD})^2}$$

où dXY_i représente l'écart de distance en 2D entre l'objet i du jeu de données à évaluer (JD) et son objet homologue j du jeu de données de référence (TN)

- **Écart quadratique moyen (RootMeanSquareError)** et l'estimation de l'incertitude : mesure les fluctuations des valeurs d'une série de mesures autour de la valeur nominale (l'erreur moyenne quadratique). Cette mesure doit être accompagné par l'estimation de l'incertitude (écart-type planimétrique)

$$EMQ_{XY} = \sqrt{\frac{1}{n} \sum_{i=1}^n dXY_i^2}$$

Et :

$$\sigma_{XY} = \frac{EMQ_{XY}}{2\sqrt{n}}$$

où dXY_i représente l'écart de distance en 2D entre l'objet i du jeu de données à évaluer (JD) et son objet homologue j du jeu de données de référence (TN)

- **Taux d'accord**, c'est la proportion du nombre de positions dont l'erreur planimétrique est inférieure à un seuil divisé par le nombre total des positions mesurées :

$$taux = \frac{\sum_{i=1}^n dXY_i \leqslant seuil}{n}$$

$$dXY_i \leqslant seuil = \begin{cases} 1 & \text{si vrai} \\ 0 & \text{si faux} \end{cases}$$



Afin de calculer ces trois mesures, nous exploitons les liens d'appariement de cardinalité 1 :1 (un objet du jeu de données à évaluer a comme homologue un objet du jeu de données de référence). Pour les objets de repères nous utilisons la distance euclidienne.

© **Résultats**

~ **Refuges.info**

Nous avons utilisé les liens 1 :1 pour calculer les trois mesures décrites ci-dessous. Pour les objets de repères nous avons obtenu 353 liens d'appariement justes de cardinalité 1 :1. Ceci représente 73% de la totalité des objets de la source Refuges.info.

- erreur planimétrique calculée : 35 m
- écart quadratique moyen : 52 m, incertitude : 0.6
- taux d'accord : 61% avec un seuil de 30 m ;

~ **Camptocamp**

Nous avons utilisé les liens 1 :1 pour calculer les trois mesures décrites ci-dessous. Pour les objets de repères nous avons obtenu 1790 liens d'appariement justes de cardinalité 1 :1. Ceci représente 61% de la totalité des objets de la source C2C.

- erreur planimétrique calculée : 48.7 m
- écart quadratique moyen : 75.9 m, incertitude : 0.9
- taux d'accord : 53% avec un seuil de 30 m ;

~ **Aires protégées**

Comme pour les autres sources, nous avons utilisé les liens 1 :1 pour calculer les trois mesures décrites ci-dessous. Pour les objets de repères nous avons obtenu 133 liens d'appariement justes de cardinalité 1 :1. Ceci représente uniquement 12% de la totalité des objets de la source PNR.

- erreur planimétrique calculée : 59 m
- écart quadratique moyen : 82 m, incertitude : 3.59
- taux d'accord : 42% avec un seuil de 30 m ;

© **Exemple XML**

Nous avons renseigné dans le rapport des métadonnées de qualité les trois métriques. Exemple d'une mesure pour un service de données d'objets de repère :

```
<!-- ===== -->
<!-- Absolute Positional Accuracy ; Metric Mean Absolute 2D -->
<gmd:DQ_AbsoluteExternalPositionalAccuracy>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Mean Absolute 2D</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString> http://www.qualityml.org/1.0/
          metrics/MeanAbsolute2D</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString>Mean value of the uncertainties in x and y
      of a number of features (N) with respect to a reference dataset.
      The value is computed by taking into account the homologous features in
```



```

    Refuges.info and the reference dataset.</gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/metre" />
      <gmd:value>
        <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema"
          xsi:type="xs:double">35</gco:Record>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_AbsoluteExternalPositionalAccuracy>

```

7.4 Les métadonnées d'exactitude sur la thématique

Dans cette partie nous étudierons l'exactitude des attributs quantitatifs ou la justesse des attributs qualitatifs ainsi que l'exactitude des classifications et de leurs relations dans les conditions décrites dans le contexte.

7.4.1 La mesure de similarité de Samal pour l'indicateur *DQ_NonQuantitativeAttributeCorrectness* (attribut nom des objets de repères)

Pour les objets de repères, nous nous intéressons aux noms des objets (toponymes).

© Méthode

Afin d'évaluer la justesse des noms des objets, nous calculons la distance de Samal (Samal et al., 2005) pour chaque couple d'objet homologue. La distance de Samal est comprise entre 0 et 1, où 0 signifie que les deux objets ont exactement le même nom et 1 signifie que les deux objets ont des noms très différents.

Nous identifions trois métriques : moyenne, écart-type et taux d'accord. Le taux d'accord représente le nombre d'objets dont la distance de Samal est inférieure à un seuil divisé par le nombre total des objets correctement appariés.

© Résultats

~ **Refuges.info** Sur l'ensemble de liens d'appariement justes, nous obtenons les résultats suivants :

- moyenne = 0.19
- variance = 0.25
- taux d'accord = 0.64 (64% des objets ont le même noms que les objets homologues dans le jeu de données de référence).

Les métadonnées renseignées dans le report sont : le couple moyenne-variance (0.19-0.25) et le taux d'accord (0.64).

~ **Camptocamp**

Sur l'ensemble de liens d'appariement justes, nous obtenons les résultats suivants :

- moyenne = 0.02
- variance = 0.13
- taux d'accord = 0.83 (83% des objets ont le même noms que les objets homologues dans le jeu de données de référence)

Les métadonnées renseignées dans le report sont : le couple moyenne-variance (0.02-0.13) et le taux d'accord (0.83).

~ **Aires protégées** Sur l'ensemble de liens d'appariement justes, nous obtenons les résultats suivants :



- moyenne = 0.25
- variance = 0.33
- taux d'accord = 0.65 (65% des objets ont le même noms que les objets homologues dans le jeu de données de référence).

Les métadonnées renseignées dans le report sont : le couple moyenne-variance (0.25-0.33) et le taux d'accord (0.65).

© Exemple XML

```

<gmd:DQ_NonQuantitativeAttributeAccuracy>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Value domain</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>http://www.qualityml.org/
          1.0/measure/ValueDomain</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString>Measures the correctness of the names. It is computed by estimating
      the mean and the variance of the distances between two names (two strings) belonging to
      two homologous features. Only the matched features (1:1) are taken into account for
      the computation. The string metrics is based on Samal distance : "Samal, A., Seth, S.
      and Cueto, K. (2004). A feature-based approach to conflation of geospatial sources.
      IJGIS, vol. 18, no5, p. 459-489".</gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueType>
        <gco:RecordType xlink:href="http://www.uncertml.org/
          distributions/normal">Value for string distances</gco:RecordType>
      </gmd:valueType>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unit"/>
      <gmd:value>
        <gco:Record>
          <un:NormalDistribution>
            <un:mean>0.19</un:mean>
            <un:variance>0.25</un:variance>
          </un:NormalDistribution>
        </gco:Record>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_NonQuantitativeAttributeAccuracy>

<!-- ===== -->
<!-- DQ_NonQuantitativeAttributeAccuracy ; Agreement Rate -->
<gmd:DQ_NonQuantitativeAttributeAccuracy>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Agreement Name Rate</gco:CharacterString>

```



```

</gmd:nameOfMeasure>
<gmd:measureIdentification>
  <gmd:MD_Identifier>
    <gmd:code>
      <gco:CharacterString>Agreement_Name_Rate</gco:CharacterString>
    </gmd:code>
  </gmd:MD_Identifier>
</gmd:measureIdentification>
<gmd:measureDescription>
  <gco:CharacterString>The proportion of features having the same
  name with the homologous features from the reference
  divided by the total number of measured positions. </gco:CharacterString>
</gmd:measureDescription>
<gmd:result>
  <gmd:DQ_QuantitativeResult>
    <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unity" />
    <gmd:value>
      <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema"
        xsi:type="xs:double">0.64</gco:Record>
    </gmd:value>
  </gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_NonQuantitativeAttributeAccuracy>

```

7.4.2 La distance Cosinus pour l'indicateur *DQ_NonQuantitativeAttributeCorrectness* (attribut titres des itinéraires)

Pour les itinéraires, nous nous intéressons aux titres des randonnées et nous vérifions si les sources sont d'accord entre elles.

© Méthode

Afin d'évaluer la justesse des titres des randonnées, nous allons évaluer leur proximité avec la distance cosinus pour chaque couple d'objet homologue (i.e. avec des liens d'appariement 1 :1). Nous prendrons le taux d'accord comme métrique c'est à dire le nombre d'objets dont la distance est inférieure à un seuil divisé par le nombre total des objets correctement appariés. Pour le seuil, nous prenons le même que celui qui a servi à l'appariement multi-critères : 0.3.

© Résultats

Si le taux d'accord est de x , cela signifie que $x \cdot 100\%$ des objets ont un titre quasi semblable aux objets homologues dans le jeu de données de comparaison.

~ CampToCamp

Taux d'accord avec Visorando = 0.25
Taux d'accord avec Aires protégées = 0.36

~ Aires protégées

Taux d'accord avec Camptocamp = 0.36
Taux d'accord avec Visorando = 0.49

~ Visorando

Taux d'accord avec AP = 0.49
Taux d'accord avec C2C = 0.25



© Exemple XML

```

<!-- ===== -->
<!-- DQ_NonQuantitativeAttributeAccuracy ; Agreement Rate -->
<gmd:DQ_NonQuantitativeAttributeAccuracy>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Agreement Name Rate with Cosinus's similarity.
    Aires protegees web sources</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>http://www.qualityml.org/1.0/metrics/items/</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString>Measures the correctness of the names. It is
    computed by estimating the mean and the variance of the
    distances between two names (two strings) belonging to two
    homologous features. Only the matched features (1:1) are taken
    into account for the computation. The string metrics is based on
    Cosinus distance. The proportion of features having the same
    name with the homologous features from the reference divided
    by the total number of measured positions.
    </gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unity" />
      <gmd:value>
        <qml:Items>
          <qml:rate max="1">0.36</qml:rate>
        </qml:Items>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_NonQuantitativeAttributeAccuracy>

```

7.4.3 La mesure NombreComposantesConnexes pour l'indicateur *DQ_NonQuantitativeAttributeAccuracy*

L'indicateur *DQ_NonQuantitativeAttributeAccuracy* permet de calculer la précision des attributs non quantitatifs. La connectivité d'un graphe d'itinéraire est un attribut calculé (à partir de la géométrie des tronçons) intéressant pour observer la couverture et la présence d'itinéraires. En examinant séparément les différents réseaux des itinéraires pour les trois services de données, on obtient dans la figure 11 la répartition spatiale des composantes connexes (aires) par sources de données (couleurs) et labellisés par le nombre d'itinéraires (Tableau 21) : 58 routes connectées pour Visorando, 28 pour C2C et 28 pour AP.

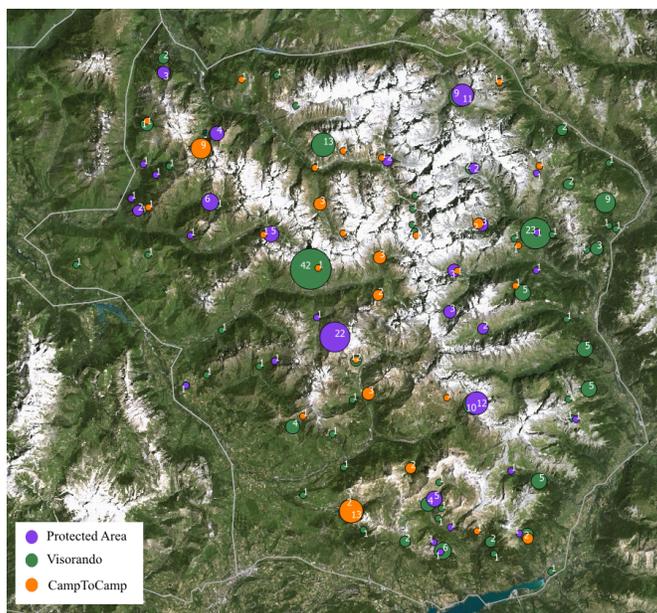


FIGURE 11 – Pour chaque service, répartition spatiale des tronçons connectés, le label indique le nombre d’itinéraires

Cette répartition spatiale signifie que les sources de données couvrent une grande superficie. En calculant ce même indicateur, mais avec les graphes fusionnés deux par deux (Tableau 21), le nombre de composantes connexes reste stable (59, et 58, pour la fusion de Visorando et C2C, respectivement pour la fusion de Visorando et AP). Avec la fusion des sources, le nombre de tronçons augmente par composant connecté, mais les itinéraires des différentes sources ne se sont pas beaucoup plus connectés avec la fusion. Les différentes sources de données sont donc non redondantes.

Nombre de cluster	AP	C2C	Viso	C2C & AP	C2C & Viso	AP & Viso	Massif
Massif Ecrins	28	28	58	39	59	58	58
Massif Vercors	32	33	57	41	60	56	59

TABLE 21 – Nombre de composantes connexes suivant le regroupement des sources de données



Ce qui donne en XML pour le service Camptocamp :

```
<!-- ===== -->
<gmd:DQ_NonQuantitativeAttributeAccuracy>
  <gmd:nameOfMeasure>
    <gco:CharacterString>CorrespondenceMatrix</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString> http://www.qualityml.org/1.0/metrics/CorrespondenceMatrix/</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>

  <gmd:measureDescription>
    <gco:CharacterString>Cette matrice indique le nombre de clusters de randonnees connectees avec les graphes de randonnees fusionnees ou non</gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/units" />
      <gmd:value>
        <qml:ConfusionMatrix max="100">
          <qml:actualCategories>C2C C2C-AP C2C-Visorando Massif</qml:actualCategories>
          <qml:predictedCategories>Massif Ecrins </qml:predictedCategories>
          <qml:values>
            28 39 59 58
            33 41 60 59
          </qml:values>
        </qml:ConfusionMatrix>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_ThematicClassificationCorrectness>
```

7.4.4 Les mesures OverallAccuracy et ConfusionMatrix pour l'indicateur *DQ_ThematicClassificationCorrectness* (attribut nature des objets de repère)

Cette méthode a été appliquée au trois sources de données : Refuges.info, C2C et Aires protégées.

© Méthode

L'exactitude sémantique est calculé grâce à la matrice de confusion qui compare les classifications des objets homologues. Les lignes représentent les types d'objets du jeu de données à comparer tandis que les colonnes représentent les types d'objets du jeu de données de référence. Deux types d'exactitude sémantique peuvent être calculés :

Exactitude sémantique de la classe A : la proportion de nombre d'objets correctement classifiés dans la classe A par rapport au nombre total d'objets appariés de classe A.

Exactitude globale : le nombre d'objets du jeu de données à évaluer correctement classés divisé par le nombre total d'objets appariés.

© Résultats



~ **Refuges.info** Afin de calculer la mesure d'exactitude sémantique nous utilisons les résultats d'appariement entre Refuges.info et la BDTOPO et nous nous appuyons sur le calcul de la matrice de confusion (voir la figure ci-après). Les 5 premières lignes de la matrices représentent les types d'objets de la source Refuges.info et les 13 premières colonnes représentent les types d'objets de la BDTopo. La colonne Non-appariés Refuges.info représente le nombre d'objets de Refuges.info qui n'ont pas d'homologues dans la BDTopo (il s'agit des liens de cardinalité 1 :0), tandis que la ligne Non-appariés BDTOPO représente le nombre d'objets de la BDTopo qui n'ont pas d'objets homologues dans Refuges.info (il s'agit des liens de cardinalité 1 :0). Par exemple, nous pouvons remarquer que parmi les 44 objets de type Lac de Refuge.info appariés, 43 ont été appariés avec des objets de la BDTopo de même type, 1 objet a été apparié avec un objet de la BDTOPO de type Parc de loisirs et 2 objets de Refuges.info n'ont pas d'homologues dans la BDTOPO.

La colonne Semantic Accuracy exprime l'exactitude sémantique pour chaque type d'objet. Par exemple, nous observons que le type Refuge a la meilleure exactitude sémantique (0.98) et que la plus faible est celle pour le type sommet (0.73). Dans ce dernier cas, ceci s'explique par la proximité sémantique entre les deux concepts pic et sommet et sur le fait que dans Refuges.info il n'y a pas cette finesse sémantique. Par exemple, des objets du monde réel de type pic sont classés en sommet dans Refuges.info ; il s'agit sur notre cas d'étude de 16 objets. Si la plupart de différences peuvent être expliquées par les différences de classifications entre les deux modèles de Refuges.info et de BDTopo, seulement pour trois d'entre elles il s'agit d'une erreur de classification dans Refuge.info : il s'agit de passage-cascade ; passage-gorge et point d'eau-col.

Pour la zone étudiée, **l'exactitude sémantique globale est égale à 0.92**, qui montre une bonne classification des objets de Refuges.info. Cette exactitude sémantique sera ajoutée dans le fichier de métadonnées (e.g. DQ_ThematicClassificationCorrectness = 0.92). Il convient de noter qu'il est aussi possible de renseigner les exactitudes sémantique par type d'objet ou de renseigner la matrice de confusion. Nous avons renseigné dans la figure 12 à titre d'exemple la matrice de confusion pour la source Refuge.info.

	Cascade	Col	Crête	Escarpe	Gorge	Lac	Montagn	Parc de loisirs	Pic	Point d'eau	Refuge	Rochers	Sommet	Total appariés (1:1)	Non-appariés Refuges.info (1:0)	Total Refuges_info	Semantic Accuracy	Commission
Lac						43		1						44	2	46	0,96	0,04
Passage	1	52			1									54	8	62	0,87	0,13
Point d'eau	2	1								108				111	27	138	0,80	0,20
Refuge											41			41	1	42	0,98	0,02
Sommet			1	1			2		16			1	82	103	9	112	0,92	0,08
Non-appariés BDTOPO (0:1)	172	1301	843	44	96	289	84	53	423	6236	306	781	1547			400		
Total BDTOPO	175	1354	844	45	97	332	86	54	439	6344	347	782	1629		12528			
Omission	0,98	0,96	1,00	0,98	0,99	0,87	0,98	0,98	0,96	0,98	0,88	1,00	0,95					

FIGURE 12 – Matrice de confusion pour comparer la sémantique des objets de Refuges.info avec la BDTopo

Nous pouvons observer que cette matrice de confusion nous permet également de calculer deux paramètres de qualité : la commission (DQ_CompletenessCommission) et l'Omission (DQ_CompletenessOmission) (voir description Section 6.5).

~ **Camptocamp**

De la même manière que pour la source Refuges.info, les résultats d'appariement obtenus en appliquant l'algorithme d'appariement pour la source C2C et BDTopo, sont utilisés pour évaluer l'exactitude sémantique par type et globale.

La figure 13 illustre la matrice de confusion générée à partir des résultats d'appariement. Cette matrice a la même structure que celle présentée en Figure 12. Les types Grottes, Lac, Refuges et Col ont les meilleures exactitudes sémantiques. Nous remarquons une grande confusion pour le type Sommet (0.61)



qui sont classés selon la référence (BDTopo) de types crêtes, pics ou rochers. Une vérification manuelle a montré qu'il s'agirait plutôt des objets mal classés (e.g. un objet de C2C nommé "crête des charmilles" est classé de type sommet). L'exactitude sémantique n'est pas calculée lorsqu'il n'y a pas de type d'équivalent dans la BDTopo tels que Accès, Site d'escalade, etc.

	Barra	Camp	Casca	Cirque	Col	Construc	Crête	Escar	Espace	Gare	Gare	Gare	Glaci	Gorg	Grott	Lac	Mais	Mont	Musé	Parki	Pic	Plain	Point	Pont	Refu	Roch	Som	Vallé	Versa	Villag	Total	Non-	Total	Semantic	Comm		
	ge	ing	de	e		on		em	publi	routi	voya	voya	er	e	e		on du	agne	e	ng		e	d'eau		ge	ers	met	e	de	vacan	appariés	appariés	C2C	Accuracy	ission		
Accès						33	1				7	1	1	3																		97	415	578	-	0,72	
Cabane																									23							23	12	41	0	0,29	
Camping		1																								23						1	7	5	1	1,40	
Cascade			9																																		
Col					466									5										6								15	100	122	0,6	0,82	
Gîte																											5	4					480	81	573	0,97	0,14
Grotte																	1		1								2						5	48	58	-	0,83
Lac															10																	10	0	12	1	0,00	
Lieu-dit	1					1				1				12	1																	118	18	139	1	0,13	
Plan d'eau																								2	1							20	51	71	0	0,72	
Refuge																								3								3	2	5	1	0,40	
Site d'escalade			4	2	5		1																			32						93	17	110	0,99	0,15	
Sommet							29	4																1			12	6				31	250	304	-	0,82	
Non-appariés	20	214	159	5	796	15	813	40	1005	8	7	48	181	90	702	171	15	76	107	1152	165	211	6251	380	189	712	990	1610	1121	108			3228				
Total BDTOPO	21	215	172	7	1301	16	843	44	1013	9	8	51	193	96	712	289	17	84	108	1192	424	211	6263	392	306	781	1547	1611	1122	109	19157						
Omission	0,95	1,00	0,92	0,71	0,61	0,94	0,96	0,91	0,99	0,89	0,88	0,94	0,94	0,94	0,99	0,59	0,88	0,90	0,99	0,97	0,39	1,00	1,00	0,97	0,62	0,91	0,64	1,00	1,00	0,99							

FIGURE 13 – Matrice de confusion pour comparer la sémantique des objets de C2C avec la BDTopo

Pour la zone étudiée, l'exactitude sémantique globale est égale à 0.69, qui montre une moins bonne classification des objets de C2C. Cette exactitude sémantique sera ajoutée dans le fichier de métadonnées (e.g. DQ_ThematicClassificationCorrectness = 0.69). Il convient de noter qu'il est aussi possible de renseigner les exactitudes sémantique par type d'objet. Voir refuges.info.

~ Aires protégées

Afin de calculer la mesure d'exactitude sémantique nous utilisons les résultats d'appariement entre PNR et la BDTOPO et nous nous appuyons sur le calcul de la matrice de confusion (voir la figure ci-après). La matrice se lit de la même manière que celle présentée dans la Figure 12. Il convient de noter que certains types d'objets de la source PNR n'ont trouvé aucun homologue dans la BDTOPO. Il s'agit des types Faune, Flore, Architecture, Savoir-faire, Glacier, Petit Patrimoine. C'est pour cette raison que ses types n'apparaissent pas dans la matrice de confusion (lignes). La matrice de confusion illustrée en figure Figure 14, montre surtout que ces deux sources de données ne sont pas comparables concernant les objets de repères car elles sont des points de vue très différents.

Parmi les objets appariés, nous remarquons par exemple que les objets de repères ayant une catégorie telle que « géologie », « histoire » sont associés à des cols, des grottes. De plus, les points de vue sont souvent associés à des sommets et des cols. Notre recommandation pour notre cas d'application est de ne pas utiliser cette sources de données pour chercher des types d'objets qui sont présents dans la BDTOPO. Cependant, il sera utile d'utiliser cette source de données pour les types d'objets spécifiques tels que le pastoralisme, archéologie, flore faune, etc.. Si par exemple, la victime dit "je suis dans un champ où il y a des gentianes ponctuées", les objets de PNR de type flore peuvent être utiles pour montrer les lieux où il peut y avoir de la gentienne.



	Casca de	Col	Cons truct ion	Crête	Dépre ssion	Espace public	Gorge	Grotte	Lac	Maison du parc	Mon umen t	Musé e	NR	Point d'eau	Pont	Refuge	Roche rs	Sommet	Versant	Vestiges archéolo giques	Total appariés (1:1)	Non- appariés PNR (1:0)	Total PNR	Semantic Accuracy	Comm ission
Abri																11					11	0	11	0	0
Archéologie et préhistoire								1						1							2	1	3	-	0,33
Col		2																			2	1	3	1,00	0,33
Eau	8								2					1							11	11	22	0,00	0,5
Géologie			1		1	1		1	7					1			1				13	30	43	-	0,70
Histoire												1	2			1					4	62	66	-	0,94
Histoire et patrimoine		8	1			1		2		1	1		3	7	3	1				1	29	0	29	-	0
Lac									14												14	7	21	1,00	0,33
Pastoralisme																2					2	36	38	-	0,95
Petit patrimoine														1							1	12	13	-	0,92
Point de Vue	1	9									1					1	1	9	1		23	8	31	-	0,26
Refuge																14					14	5	19	1,00	0,26
Résistance											2	1									3	0	3	-	0
Sommet																		4			4	5	9	1,00	0,56
Non-appariés BDTOPO (0:1)	163	1281	15	833	49	1012	95	702	273	16	323	106	378	6225	389	276	779	1534	1121	20			311		
Total BDTOPO	172	1301	16	834	50	1013	96	712	289	17	327	108	383	6236	392	306	781	1547	1122	21			15723		
Omission	0,95	0,98	0,94	1,00	0,98	1,00	0,99	0,99	0,94	0,94	0,99	0,98	0,99	1,00	0,99	0,90	1,00	0,99	1,00	0,95					

FIGURE 14 – Matrice de confusion pour comparer la sémantique des objets des Parc Nationaux avec la BDTopo

Pour la zone étudiée, l'**exactitude sémantique globale** (DQ_ThematicClassificationCorrectness) est égale à **0.26**, qui montre une très grande hétérogénéité des classifications des objets de PNR.

© Exemple XML

```

<!-- ===== -->
<!-- DQ_ThematicClassificationCorrectness using the confusion Matrix-->
<gmd:DQ_ThematicClassificationCorrectness>
  <gmd:nameOfMeasure>
    <gco:CharacterString>ConfusionMatrix</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString> http://www.qualityml.org/1.0/
          metrics/ConfusionMatrix</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString>Matrix that indicates the number of features of class
      (i; the reference) classified as class (j; refuges.info). The confusion matrix is a NxN matrix
      with N columns and N rows. N denotes the number of classes under consideration. MCM (i,j) =
      ( # features of class (i) classified as class (j) ) The diagonal elements of the confusion
      matrix contain the correctly classified features, and the off diagonal elements contain the
      number of misclassification errors. In the matrix NA means that the class does not exist
      in Refuges.info. It exists only in BDTopo. UserAccuracy : quantifies the proportion of
      correct classifications of the class based on the reference data and indicates accuracy
      from the map user's point of view. ProducerAccuracy: quantifies the proportion of correct
      classifications for the class based on the Refuges_info, and it indicates the accuracy of
      the classification from the map maker's point of view. </gco:CharacterString>
  </gmd:measureDescription>
</gmd:DQ_ThematicClassificationCorrectness>

```



```

</gmd:measureDescription>
<gmd:result>
  <gmd:DQ_QuantitativeResult>
    <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/units" />
    <gmd:value>
      <qml:ConfusionMatrix max="100">
        <qml:actualCategories>Cascade Col Crete Escarpement Gorge Lac Montagne Parc de loisirs
        Pic Point d'eau Refuge Rocher Sommet UserAccuracy</qml:actualCategories>
        <qml:predictedCategories>Cascade Col Crete Escarpement Gorge Lac Montagne
        Parc de loisirs Pic Point d'eau Refuge Rocher Sommet
        ProducerAccuracy</qml:predictedCategories>
        <qml:values>
          NA NA
          1 52 0 0 1 0 0 0 0 0 0 0 0 0 0.96
          NA NA
          NA NA NA NA NA NA NA NA NA NA NA NA NA NA
          NA NA NA NA NA NA NA NA NA NA NA NA NA NA
          0 0 0 0 0 43 0 1 0 0 0 0 0 0.87
          NA NA
          NA NA NA NA NA NA NA NA NA NA NA NA NA NA
          NA NA NA NA NA NA NA NA NA NA NA NA NA NA
          2 1 0 0 0 0 0 0 0 108 0 0 0 0.8
          0 0 0 0 0 0 0 0 0 41 0 0 0 0.98
          NA NA
          0 0 1 1 0 0 2 0 16 0 0 1 82 0.92
          0 0.04 0 0 0 0.13 0 0 0 0.02 0.12 0 0.05
        </qml:values>
      </qml:ConfusionMatrix>
    </gmd:value>
  </gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_ThematicClassificationCorrectness>

<!-- ===== -->
<!-- DQ_ThematicClassificationCorrectness using Overall Accuracy-->
<gmd:DQ_ThematicClassificationCorrectness>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Overall accuracy</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString> http://www.qualityml.org/1.0/
        metrics/OverallAccuracy</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString>The overall accuracy is computed as the total number of correctly
    classified features divided by the total number of features. Only matched features (1:1)
    are taken into account for the computation.</gco:CharacterString>
  </gmd:measureDescription>
</gmd:result>
  <gmd:DQ_QuantitativeResult>

```



```

<gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unit" />
<gmd:value>
  <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xsi:type="xs:double">0.92</gco:Record>
</gmd:value>
</gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_ThematicClassificationCorrectness>

```

7.4.5 Les mesures OverallAccuracy et ConfusionMatrix pour l'indicateur *DQ_ThematicClassificationCorrectness* (attribut cotation des itinéraires)

Étant donné qu'il n'existe pas une source de données de référence pour les itinéraires, nous allons comparer la cotation (niveau de difficulté) des itinéraires avec la cotation des autres sources pour évaluer l'exactitude sémantique de cet attribut.

Cette méthode a été appliquée aux trois sources de données d'itinéraires : Camptocamp, Aires protégées et Visorando.

© Méthode

L'exactitude sémantique est calculé grâce à la matrice de confusion qui compare les classifications des objets homologues. Les lignes représentent les types d'objets du jeu de données à comparer tandis que les colonnes représentent les types d'objets du jeu de données de référence. Pour calculer l'exactitude sémantique du niveau de difficulté d'un itinéraire nous utilisons les résultats d'appariement et la matrice de confusion qui en résulte pour les liens 1 :1. L'exactitude sémantique globale est alors calculée en évaluant le nombre d'objets du jeu de données à évaluer correctement classés divisé par le nombre total d'objets appariés.

© Résultats

Dans l'ensemble, les sources sont en désaccord, malgré le peu d'itinéraires appariés trouvés. Par exemple, le niveau moyen est réparti sur les niveaux facile, moyen et difficile. Plus précisément, le nombre d'itinéraires de niveau facile peut prendre toutes les valeurs des niveaux de l'autre source. Un parcours de niveau difficile est classé dans le niveau facile et vice-versa.

~ Camptocamp

Pour la zone étudiée, l'exactitude sémantique globale est égale à 0.17 pour la source "Aires protégées" et 0.32 pour la source de données "Visorando" ce qui confirme le désaccord exprimé plus haut. Cette exactitude sémantique sera ajoutée dans le fichier de métadonnées (e.g. *DQ_ThematicClassificationCorrectness*).

	C2C						
	T1	T2	T3	T4	T5	NULL	Total apparié 1:1
Très facile							0
Facile	2	1					3
Moyen	1		1			1	3
Difficile		4	2		1	1	8
Très difficile							0
Aires protégées							



C2C							
	T1	T2	T3	T4	T5	NULL	Total apparié 1:1
Facile							0
Moyenne	7	2					9
Difficile		2	4		1	1	8
Très difficile		2	1				3
Extrêmement difficile							0
Visorando							

FIGURE 15 – Matrice de confusion pour comparer la sémantique des objets de CamptoCamp avec Aires protégées et Visorando

~ Aires protégées

Pour la zone étudiée, l'exactitude sémantique globale est égale à 0.17 pour la source "Camptocamp" et 0.29 pour la source de données "Visorando" ce qui confirme le désaccord exprimé plus haut. Cette exactitude sémantique sera ajoutée dans le fichier de métadonnées (e.g. DQ_ThematicClassificationCorrectness).

Aires protégées						
	Très facile	Facile	Moyen	Difficile	Très difficile	Total apparié 1:1
T1		2	1			3
T2		1		4		5
T3			1	2		3
T4						0
T5				1		1
NULL			1	1		2
Total C2C	0	3				

Aires protégées						
	Très facile	Facile	Moyen	Difficile	Très difficile	Total apparié 1:1
Facile	2	4	5	1		12
Moyenne		6	15	7		28
Difficile	1		7	9		17
Très difficile				2		2
Extrêmement difficile						0
Total visorando						

FIGURE 16 – Matrice de confusion pour comparer la sémantique des objets des Aires protégées avec Camptocamp et Visorando

~ Visorando

Pour la zone étudiée, l'exactitude sémantique globale est égale à 0.32 pour la source "Camptocamp" et 0.29 pour la source de données "Aires protégées" ce qui confirme le désaccord exprimé plus haut. Cette exactitude sémantique sera ajoutée dans le fichier de métadonnées (e.g. DQ_ThematicClassificationCorrectness).

Visorando						
	Facile	Moyenne	Difficile	Très difficile	Extrêmement difficile	Total apparié 1:1
T1		7				7
T2		2		2		6
T3			4	1		5
T4						0
T5			1			1
NULL			1			1
Total C2C						



Visorando						
	Facile	Moyenne	Difficile	Très difficile	Extrêmement difficile	Total apparié 1:1
Très facile	2		1			3
facile	4	6				10
Moyen	5	15	7			27
Difficile	1	7	9	2		19
Très difficile						0
Total Aires protégées						

FIGURE 17 – Matrice de confusion pour comparer la sémantique des objets de Visorando avec Camptocamp et les Aires protégées

⊙ **Exemple XML**

Voir dans le fichier

7.5 Les métadonnées de complétude

Dans cette partie nous étudierons, comme décrit dans le contexte, la complétude de la source de données, plus précisément la présence (excès et doublon)(comission) ou l’absence (omission) des objets géographiques, de leurs attributs ou de leurs relations.

7.5.1 La mesure *MissingItems* de l’indicateur *DQ_CompletenessOmission*

Cette méthode à été appliquée au trois services de données des points de repère : Refuges.info, C2C et AP.

⊙ **Méthode**

Nous distinguons deux types d’omission :

- Omission de la classe A : la proportion de nombre d’objets de la classe A absents par rapport au nombre total d’objets de classe A dans le jeu de données à évaluer.
- Omission globale : le nombre d’objets du jeu de données à évaluer absents par rapports au nombre total d’objets du jeu de données à évaluer.

Les deux types d’omissions sont calculées en exploitant les liens d’appariement 0 :1 (1 objet du jeu de données de référence n’a pas d’homologues dans le jeu de données de comparaison).

⊙ **Résultats**

- ~ **Refuges.info** La Figure 12 illustre l’indicateur d’omission par type d’objet pour la source Refuges.info. Par exemple, nous remarquons que l’omission a des valeurs élevées pour la totalité des types d’objets. Le type d’objet ayant la plus faible omission est le Refuge 0.88, qui reste néanmoins élevés par rapport aux besoins. Pour la zone étudiée, l’omission globale est égale à 0.92, qui montre que la source Refuges.info est une source de données non-exhaustive. Cette omission globale sera ajoutée dans le fichier de métadonnées (e.g. *DQ_CompletenessOmission* = 0.92).
- ~ **Camptocamp** La Figure 13 illustre l’indicateur d’omission par type d’objet pour la source C2C. Par exemple, nous remarquons que l’omission a des valeurs élevées pour la plupart de types d’objet. Les types d’objets Lac, Col, Refuge, Sommet font partie des types dont l’omission est moins élevées (e.g. 0.6). Pour la zone étudiée, l’omission globale est égale à 0.91, qui montre que la source Refuges.info est une source de données non-exhaustive. Cette omission globale sera ajoutée dans le fichier de métadonnées (e.g. *DQ_CompletenessOmission* = 0.91). Il convient de noter qu’il est aussi possible de renseigner les omissions par type d’objet. Il convient de noter qu’il est aussi possible de renseigner dans un autre report les omissions par type d’objet.



~ **Aires protégées** La Figure 14 illustre l'indicateur d'omission par type d'objet pour la source PNR. Par exemple, nous remarquons que l'omission a des valeurs extrêmement élevées pour la totalité des types d'objets.

Pour la zone étudiée, l'omission globale est égale à 0.99, qui montre que la source AP est une source de données non-exhaustive pour les types d'objets présents dans la BDTOPO. De la même manière que pour l'exactitude sémantique nous recommandons de ne pas renseigner cette mesure dans les métadonnées à l'échelle globale, mais uniquement pour les types d'objets de la BDTOPO. Il convient de noter qu'il est aussi possible de renseigner dans un autre report les omissions par type d'objet.

© Exemple XML

```
<!-- ===== -->
<!-- Rate of features from Refuges.info missing from the dataset -->
<!-- with respect with a reference dataset -->
<!-- ===== -->
<gmd:DQ_CompletenessOmission>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Omission : features non matched with a reference dataset</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>http://www.qualityml.org/1.0/measure/MissingItems</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString> Rate of missing Refuges.info features. Features absent from the dataset with respect to a reference dataset.
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unity"/>
      <gmd:value>
        <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type="xs:double">0.92</gco:Record>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_CompletenessOmission>
```

7.5.2 La mesure *MissingClass* de l'indicateur *DQ_CompletenessOmission* et *DQ_CompletenessCommission*

Ces métadonnées permettent de mesurer l'exhaustivité des attributs. Nous nous intéressons aux attributs dont la valeur peut être non déterminé (s'il est interdit qu'une valeur soit déterminée il s'agit d'une incohérence logique et non pas d'exhaustivité). Dans notre cas, il s'agit de l'attribut qui porte sur le *nom des objets*.

© Méthode

Les taux de déficit et d'excédent de remplissage se mesurent à travers la matrice de confusion. Les taux sont exprimés par rapport au nombre d'objets. Comme pour les autres mesures de qualité, la matrice de qualité est calculée en se basant sur les objets correctement appariés.

© Résultats



Les trois matrices de confusion représentant les taux de remplissages sont présentées ci-après. Nous remarquons que contrairement aux objets de la BDTOPO, pour tous les objets des sources étudiées, l'attribut *nom* est renseigné. Ceci montre la complémentarité des sources d'informations pour les objets de repères.

- ~ **Refuges.info** Les deux mesures de qualité qui seront renseignées seront :
 DQ_CompletenessOmission, métrique : NameAttributeOmission= 0
 DQ_CompletenessCommission, métrique : NameAttributeCommission= 1
 Ces résultats montrent que la totalité des objets ont l'attribut *nom* rempli.

		BDTOPO	
		Nom determ	Déterminé
Refuges.info	Non determ	0,0	0,0
	Déterminé	0,14	0,86

FIGURE 18 – Matrice de confusion des noms des toponymes de Refuge.info

- ~ **Camptocamp** Les deux mesures de qualité qui seront renseignées seront :
 DQ_CompletenessOmission, métrique : NameAttributeOmission= 0
 DQ_CompletenessCommission, métrique : NameAttributeCommission= 1

		BDTOPO	
		Nom determ	Déterminé
C2C	Non determ	0,0	0,0
	Déterminé	0,09	0,91

FIGURE 19 – Matrice de confusion des noms des toponymes de Camptocamp

- ~ **Aires protégées** Les deux mesures de qualité qui seront renseignées seront :
 DQ_CompletenessOmission, métrique : NameAttributeOmission= 0
 DQ_CompletenessCommission, métrique : NameAttributeCommission= 1

		BDTOPO	
		Nom determ	Déterminé
PA	Non determ	0,0	0,0
	Déterminé	0,21	0,79

FIGURE 20 – Matrice de confusion des noms des toponymes des AP

© Exemple XML

```
<!-- ===== -->
<!-- DQ_CompletenessCommission : Rate of features having the attribute name filled in -->
<!-- ===== -->
<gmd:DQ_CompletenessOmission>
  <gmd:nameOfMeasure>
    <gco:CharacterString>MissingClass "nom"</gco:CharacterString>
  </gmd:nameOfMeasure>
</gmd:measureIdentification>
```



```

    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>http://www.qualityml.org/1.0/measure/MissingClass/</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString> Rate of features having the attribute not filled in</gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unity"/>
      <gmd:value>
        <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type="xs:double">0</gco:Record>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_CompletenessOmission>

<!-- ===== -->
<!-- DQ_CompletenessCommission : Rate of features having the attribute name filled in-->
<gmd:DQ_CompletenessCommission>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Name Attribute Commission</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>Name_Attribute_Comission_Refuges_info</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
  <gmd:measureDescription>
    <gco:CharacterString> Rate of features having the attribute name filled in</gco:CharacterString>
  </gmd:measureDescription>
  <gmd:result>
    <gmd:DQ_QuantitativeResult>
      <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unity" />
      <gmd:value>
        <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type="xs:double">1</gco:Record>
      </gmd:value>
    </gmd:DQ_QuantitativeResult>
  </gmd:result>
</gmd:DQ_CompletenessCommission>

```

7.5.3 La mesure *Excess* de l'indicateur *DQ_CompletenessCommission*

Cette méthode à été appliquée au trois services de données des points de repère : Refuges.info, C2C et AP.

© Méthode

Nous distinguons deux types de commission :

- Commission de la classe A : la proportion de nombre d'objets de la classe A absents présents dans le jeu de données à évaluer et absents du jeu de données de référence.



- Commission globale : le nombre d'objets du jeu de données à évaluer présents dans le jeu de données à évaluer et absents du jeu de données de référence.

Les deux types de commissions sont calculées en exploitant les liens d'appariement 1 :0 (1 objet du jeu de données à évaluer n'a pas d'homologues dans le jeu de données de référence).

© Résultats

- ~ **Refuges.info** La Figure 12 illustre l'indicateur de commission par type d'objet pour la source Refuges.info. Par exemple, nous remarquons que la commission a des valeurs très faibles pour la totalité des types d'objets. Les types d'objets ayant les valeurs de commissions les plus élevées sont les objets de type *Passage* (0.13) et *Point d'eau* (0.2).
Pour la zone étudiée, la commission globale est égale à 0.12, qui montre que la source Refuges.info peut compléter, même si faiblement, la source de données de référence, la BDTopo. Ainsi, la source Refuges.info peut être utilisée pour notre objectif de localisation de victimes en montagne, mais pas comme une source prioritaire. Cette commission globale sera ajoutée dans le fichier de métadonnées (e.g. DQ_CompletenessCommission = 0.12).
- ~ **Camptocamp** La Figure 13 illustre l'indicateur de commission par type d'objet pour la source C2C. Par exemple, nous remarquons que la commission a des valeurs assez élevée pour la plupart des types d'objets de C2C. Ceci montre donc une complémentarité de cette source par rapport à la BDTopo. Les types d'objets ayant les valeurs de commissions les plus élevées sont *Camping* (0.88), *Gite* (0.83), *Site d'escalade* (0.82) qui montre les caractéristiques métier de cette source.
Pour la zone étudiée, la commission globale est égale à 0.4, qui montre que la source C2C peut compléter, la source de données de référence, la BDTopo. Ainsi, la source C2C peut être utilisée pour notre objectif de localisation de victimes en montagne ; elle peut-être considérée une source prioritaire. Cette commission globale sera ajoutée dans le fichier de métadonnées (e.g. DQ_CompletenessCommission = 0.4). Il convient de noter qu'il est aussi possible de renseigner les omissions par type d'objet.
- ~ **Aires protégées** La Figure 14 illustre l'indicateur de commission par type d'objet pour la source AP. Par exemple, nous remarquons que la commission a des valeurs très hétérogènes d'un type à un autre (e.g. 0.2 pour les *points de vue* et 0.92 pour le type *patrimoine*).
Pour la zone étudiée, la commission globale est égale à 0.57, qui montre que la source PNR peut compléter, la source de données de référence, la BDTopo, mais uniquement pour les types d'objets not-présents dans la BDTOPO. Elle peut donc être considérée comme une source complémentaire. Cette commission globale sera ajoutée dans le fichier de métadonnées (e.g. DQ_CompletenessCommission = 0.57). Il convient de noter qu'il est fortement recommandé de renseigner les omissions par type d'objet.

Il convient de noter qu'il est aussi possible de renseigner les omissions par type d'objet en utilisant la matrice de confusion.

© Exemple XML

```
<!-- ===== -->
<!-- Rate of features from Refuges.info having no homologous features in the reference dataset-->
<!-- ===== -->
<gmd:DQ_CompletenessCommission>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Excess</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString> http://www.qualityml.org/1.0/measure/Excess</gco:CharacterString>
      </gmd:code>
    </gmd:MD_Identifier>
  </gmd:measureIdentification>
</gmd:DQ_CompletenessCommission>
```



```

<gmd:measureDescription>
  <gco:CharacterString> Rate of excess Refuges.info features. Features present in the Refuges.info
</gmd:measureDescription>
<gmd:result>
  <gmd:DQ_QuantitativeResult>
    <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unity" />
    <gmd:value>
      <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type="xs:double">80</gco:Record>
    </gmd:value>
  </gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_CompletenessCommission>

```

7.5.4 La mesure *Duplicate* de l'indicateur *DQ_CompletenessCommission*

© Méthode

Il s'agit d'une métrique de l'indicateur *DQ_CompletenessCommission* qui mesure les doublons dans le jeu de données. Pour renseigner cette mesure nous avons calculé :

- pour les objets de repères : 1/le nombre d'objets géographiques ayant exactement la même géométrie ou 2/ le même nom et sont localisés à une distance inférieure à 5m.
- pour les itinéraires : nombre d'itinéraire qui suivent un chemin proche avec la même distance que celle qui a servi dans l'appariement multi-critère : distance moyenne de Hausdorff pour chaque tronçon routier des itinéraires.

© Résultats

service	Nombre doublon
RefugeInfo	0
C2C - Objets de repères	14
C2C - Sortie de randonnées	1
PA - Objets de repères	0
PA - Itinéraires à pied	0
Visorando - Itinéraires	4

TABLE 22 – «TopicCategoryCode» des web services étudiés

© Exemple XML

```

<!-- =====>
<!-- Number of features that are duplicated in the dataset -->
<!-- =====>
<gmd:DQ_CompletenessCommission>
  <gmd:nameOfMeasure>
    <gco:CharacterString>Duplicate</gco:CharacterString>
  </gmd:nameOfMeasure>
  <gmd:measureIdentification>
    <gmd:MD_Identifier>
      <gmd:code>
        <gco:CharacterString>http://www.qualityml.org/1.0/measure/Duplicate</gco:CharacterString>

```



```

        </gmd:code>
    </gmd:MD_Identifier>
</gmd:measureIdentification>
<gmd:measureDescription>
    <gco:CharacterString> Number of features that are duplicated in the database.
Features duplicated are features having the same geometry.
    </gco:CharacterString>
</gmd:measureDescription>
<gmd:result>
    <gmd:DQ_QuantitativeResult>
        <gmd:valueUnit xlink:href="http://www.opengis.net/def/uom/OGC/1.0/unit" />
        <gmd:value>
            <gco:Record xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:type="xs:int">0</gco:Record>
        </gmd:value>
    </gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_CompletenessCommission>

```



8 Conclusion

Dans cette étude nous avons étudié le potentiel des sources de données hétérogènes spatiales participatives et issues du web pour l'intégration de données pour le secours en montagne. Nous nous sommes concentrés sur les itinéraires et les points de repère. Pour atteindre cet objectif, nous avons proposé une méthode fondée sur les techniques d'appariement afin de comparer et d'évaluer ces sources. Cette approche nous a permis d'ajouter aux métadonnées classiques des jeux de données géographiques, des métadonnées de qualité.

Afin de produire les métadonnées des sources de données des objets de repères et des itinéraires, nous avons utilisé une combinaison des standards de Web Feature Service (WFS) et de la norme ISO (19115 et 19157) afin de décrire les métadonnées fonctionnelles (description des données en sortie des web services) et non fonctionnelles basées service (qualité de service). Ce qui a produit au final deux fichiers de métadonnées par sources de données et type de données.

Ces résultats ont mis en valeur la richesse et la complémentarité des différentes sources de données et montrent une bonne couverture de la zone d'étude. Suite à cette analyse, de nouvelles connaissances concernant la sémantique des données (classification des points de repères), la géométrie et le niveau de difficulté des routes sont apparues.

Dans ce travail, nous nous sommes concentrés sur les points de repères et les itinéraires. Cependant nous pensons que notre méthodologie peut être appliquée à d'autres sources de données (gazetier, objets du relief, etc.).

Dans le cadre de recherches futures, nous prévoyons plusieurs pistes :

- ~ de travailler pour concevoir un algorithme d'appariement multi-sources. En effet, l'algorithme multi-critères compare un ensemble de données à une référence ou au plus grand ensemble de données (Visorando pour les routes). Les cardinalités des liens appariés correspondent à 1 :1 ou 1 :n (par post-traitement). Pour n sources de données et sans ensemble de données de référence, nous devons exécuter l'algorithme $n(n-1)$ fois, ce qui n'est pas une solution optimale,
- ~ dans le cadre de travail de thèse dans le lot 5 sur la découverte et l'exploitation de services géospatiaux, ces métadonnées permettront d'alimenter la plateforme logicielle pour l'intégration et la composition de services géospatiaux,
- ~ alimenter l'ontologie des points de repères,
- ~ définir les relations spatiales entre les objets de repères et les itinéraires



9 Bibliographie

Références

- [1] R. Devillers and R. Jeansoulin. Qualité de l'information géographique. [http://lst-iiiep.iiep-unesco.org/cgi-bin/wwwi32.exe/\[in=epidoc1.in\]/?t2000=021596/\(100\)](http://lst-iiiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=021596/(100)), 10 2005.
- [2] H. Fan, A. Zipf, Q. Fu, and P. Neis. Quality assessment for building footprints data on openstreetmap. *International Journal of Geographical Information Science*, 28(4) :700–719, 2014.
- [3] J.-F. Girres and G. Touya. Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 14(4) :435–459, 2010.
- [4] M. Haklay. How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and Planning B : Planning and Design*, 37(4) :682–703, 2010.
- [5] M. Halilali and E. Gouardères. Vers un modèle de classification pour les services géospatiaux. Research report, LMAP, Feb. 2020.
- [6] M. S. Halilali, E. Gouardères, F. Devin, and M. Gaio. Software platform for the integration and composition of geospatial services. In *Sageo 2018*, Montpellier, France, Nov. 2018.
- [7] J. Jokar Arsanjani, P. Mooney, A. Zipf, and A. Schauss. *Quality Assessment of the Contributed Land Use Information from OpenStreetMap Versus Authoritative Datasets*. 03 2015.
- [8] L. Moncla and M. Gaio. Services Web pour l'annotation sémantique d'information spatiale à partir de corpus textuels. *Revue Internationale de Géomatique*, 28(4) :439–459, 2018.
- [9] P. Neis, D. Zielstra, and A. Zipf. The street network evolution of crowdsourced maps : Openstreetmap in germany 2007–2011. *Future Internet*, 4 :1–21, 01 2012.
- [10] A.-M. Olteanu. *Fusion de connaissances imparfaites pour l'appariement de données géographiques : proposition d'une approche s'appuyant sur la théorie des fonctions de croyance*. Theses, Université Paris-Est, Oct. 2008.
- [11] A.-M. Olteanu-Raimond, S. MUSTIERE, and A. Ruas. Knowledge formalization for vector data matching using belief theory. *Journal of Spatial Information Science*, (10) :pp 21–46, Jan. 2015.
- [12] A. Samal, S. Seth, and K. Cueto1. A feature-based approach to conflation of geospatial sources. *International Journal of Geographical Information Science*, 18(5) :459–489, 2004.
- [13] G. Santos, J. Masó, A. Zabala Torres, L. Pesquer, and X. Pons. A provenance metadata model integrating iso geospatial lineage and the ogc wps : Conceptual model and implementation. *Transactions in GIS*, 07 2019.
- [14] M. Sébastien, N. Abadie, N. Aussenac-Gilles, M.-N. Bessagnet, M. Kamel, E. Kergosien, C. Reynaud, B. Safar, and C. Sallaberry. Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique. *Revue Internationale de Géomatique*, 21(2) :155–180, 2011.
- [15] G. Touya, V. Antoniou, A.-M. Olteanu-Raimond, and M.-D. Van Damme. Assessing crowdsourced poi quality : Combining methods based on reference data, history, and spatial relations. *ISPRS International Journal of Geo-Information*, 6(3), 2017.
- [16] M.-D. Van Damme, A.-M. Olteanu-Raimond, and Y. Méneroux. Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas. *International Journal of Cartography*, 5(2-3) :195–213, May 2019.
- [17] Z. Wu and M. Palmer. Verbs semantics and lexical selection. pages 133–138, 01 1994.