



HAL
open science

Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas

Marie-Dominique van Damme, Ana-Maria Olteanu-Raimond, Yann Méneroux

► To cite this version:

Marie-Dominique van Damme, Ana-Maria Olteanu-Raimond, Yann Méneroux. Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas. *International Journal of Cartography*, 2019, 5 (2-3), pp.195-213. 10.1080/23729333.2019.1615730 . hal-02276065

HAL Id: hal-02276065

<https://hal.science/hal-02276065>

Submitted on 6 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Potential of crowdsourced data for integrating landmarks and routes for rescue in mountain areas

Marie-Dominique Van Damme^{a, *}, Ana-Maria Olteanu-Raimond^a, Yann Méneroux^a

^a Univ. Paris-Est, LASTIG MEIG, IGN, ENSG, F-94160 Saint-Mandé, France; marie-dominique.van-damme@ign.fr, ana-maria.raimond@ign.fr; yann.meneroux@ign.fr

* Corresponding author

Abstract:

Many different websites offer the opportunity to share and download landmarks and routes produced by the crowd. Landmarks near to a route or routes passing near to some landmarks may help in the context of mountain rescue. Therefore, it is necessary to identify relevant data sources and to describe their characteristics. In this paper, we set out to explore the potential of crowdsourced data in order to be considered such as data sources in the context of mountain rescue. Thus, our aim is to study the content of different sources to have a better knowledge on how landmarks and routes are mapped, to demonstrate the complementarity of crowdsourced data with respect with authoritative data, and to study the feasibility of defining links between routes and landmarks. The proposed method used integration techniques such as map matching, route construction and data matching. The results show that the large number of non-matched features proves the richness of crowdsourced data. The matching results generate new semantic rules for both type of landmarks and geometries of route.

Keywords: crowdsourced landmarks, crowdsourced routes, data and map matching, mountain rescue, data integration

1. Introduction

More than 15, 000 SOS messages per year are counted in mountain areas. Despite technological advances in positioning systems and the increasing number of geo-applications, locating victims during the alert is still an issue for rescue teams. In the absence of the exact position, the victims describes the spatial environment by using landmarks ("I'm seeing *Mont Blanc summit*", "I'm near a *shelter*"), or describing their routes ("I'm walking along GR20¹", "I'm walking on the route that runs along a lake"), where routes and landmarks can be associated to proper names (Col des Montets parking, Mont Blanc summit, GR20) or common names (shelter, route that runs along the lake).

These types of descriptions (i.e. locate an object by using spatial relations and reference feature) are qualitative and known in the GIS domain as indirect georeferencing (Hill and Zheng 1999). This is consistent with the concept of Naive Geography coined by Egenhofer and Mark (1995), where the goal is to formalize how people perceive the space in order to propose skills closer to the way of thinking of people. One of the skills necessary in our context is data integration. Qualitative reasoning can be a solution to improve heterogeneous data source integration (Duckham et al., 2006).

To transform the relative position to an absolute position (i.e. geographic coordinates), a first need is to have access to spatial data containing such landmarks and routes. Moreover, knowing the landmarks near to a route or the routes passing near to some landmarks may help the mountain rescuer to validate or invalidate a location hy-

pothesis. Thus, the second need is to define links between landmarks and routes. This situation, where user queries many data sources, is usually part of semantic web (Ferrara et al., 2013). The qualitative reasoning is also explored to integrate heterogeneous spatial data sources from different

Traditionally, landmarks are part of authoritative data produced by institutions such National Mapping Agencies (NMA) or national and global gazetteers (e.g. Wikipedia, GeoNames) mostly produced in a collaborative way or integrating data coming from different sources including authoritative ones. In the last decade, different websites offer the opportunity to share and download landmarks and routes produced by the crowd. These websites provide technical information (e.g. topographic, difficulty), landmarks and GPX routes serving to different purposes: use track in mobile application, following user's performance, etc. Mapped in a collaborative way, crowdsourced data have different degrees of reliability and heterogeneous structures (e.g. types of landmarks, different rating nomenclature, irregularities on route description). Due to the variety of data, many queries need to be executed in heterogeneous data sources on the web to get information related to routes or landmarks. This process is part of the integration data. Moreover,

This research work is part of the project CHOUCAS (choucas.ign.fr) aiming to help the rescue teams to better estimate a probable location area. The architecture of the project is based on Service-Oriented Architecture and Web Services to facilitate the freshness of data (Halilali et al., 2018). The approach is towards process mediation requiring the definition of a catalogue of services characterized by functional and non-functional attributes. Thus, it is necessary to identify relevant sources of information

¹ GR is a well know European abbreviation defining a long distance path used for hiking activities. GR20 is a famous GR path in Corsica having 179 km length.

providing routes and landmarks and to describe their characteristics to facilitate the definition of the mapping rules between schemas sources.

In this paper, we set out to explore the potential of crowdsourced data in order to be considered such as sources of information in the context of mountain rescue. Thus, our aim is to study the content of different crowdsourced sources to have a better knowledge how landmarks and routes are described, to demonstrate the complementarity of crowdsourced data with respect with authoritative data, and to study the feasibility of defining links between routes and landmarks.

In the remainder of this paper, Section 2 provides an overview of the existing literature. Section 3 describes the sources of information identified as relevant in mountain areas. Sections 4 and 5 expose our approach to evaluate the potential of crowdsourced data in the mountain rescue context and the results. Finally, conclusion and perspectives are presented in Section 6.

2. Related work

Landmarks and routes are essential in different domains such as navigation (Denis, 1997; Zhou et al., 2017), text geocoding (Moncla et al., 2014), geographic information retrieval (Schockaert et al., 2008). The analysis of landmarks and descriptions of routes is also in focus of many research for the purpose of analyzing how people describe their spatial environment, places or routes (Goodchild and Li, 2012; Schwering et al., 2013).

From a cognitive perspective, according to Lynch (1960) landmarks are entities from the real world allowing to a human to better orient himself in a space that is not always familiar to him. In the field of space description, landmarks are considered such as the referent (Zhou et al., 2017), whereas, in the field of text geocoding, landmarks representing place-names are considered as anchor to geocode proper names (Moncla et al., 2014).

In the field of GIS, according to Teulade-Denantes et al. (2015), routes represents paths, journey rides, or trails to follow being carried by a physical network infrastructure composed of roads and pedestrian paths. Routes are characterized by geometry and textual description. In the field of geographic retrieval, route description is studied by using natural language techniques for defining wayfinding instructions where landmarks are identified as crucial elements to confirm the route (Denis, 1997). Thus, landmarks and routes are not separate concepts but have a relation, i.e. a route is composed by a set of landmarks and segments (Laddada and Saux 2017; Denis, 1997).

From the last decade, both descriptions of the route and its GPX trace are available. Many research studies use GPX traces to study leisure activities (Pitman, 2012), to identify categories of hikers itineraries and their stops (Colin et al. 2017), to build pedestrian networks (Kasemsuppakorn, 2013), to highlight updates in authoritative spatial data (Ivanovic et al., 2016) or to enrich thematic information of route networks to propose alternative itineraries, in the context of natural spaces protection (Vias 2014).

Another important aspect is to identify which are the types of landmarks and routes in mountain areas. Although many research is made to study landmarks and routes in urban environment (Timpf, 2005), few research is focus on landmarks and routes in natural and open en-

vironment and even less in mountain areas. Brosset et al. (2008) confirmed the relevant role of landmarks in itinerary description in parks and forest areas and found that features represented by points, lines and polygons are used to describe the itinerary. For coastal navigation, landmarks located on the see or on the coastal are used to plan itinerary for boats directions (Laddada and Saux 2017). Landmarks such as rock, spars, isolated danger buoy, lighthouse are identified from analyzing itinerary descriptions for boats made by experts. Derungs and Purves (2014) are interested on how natural features in mountain environment are described in an annotated corpus on Swiss Alps. They identified 95 natural features types which were compared with other studies from literature discovering that there is an overlap of types for the 10 first ranks and a higher diversity of types du the specificity of the corpus. In the same context of mountain, Egorova et al. (2015) analyze textual description of alpine itineraries coming from a collaborative platform in order to study the influence of landmarks in mountain areas. The different types of identified landmarks are grouped into four categories: landscape features, spatial part, terrain, and constellation.

In contrast with routes (textual descriptions and GPX) which are new data traditionally not produced by NMAs, landmarks can be found in both authoritative and crowdsourced data. Moreover, although landmarks characterized by place-names can be provided by gazetteers, it is not the case for landmarks expressed by vernacular place-names or unnamed places (e.g. shelter). Thus, another important research question is to study the quality of landmarks coming from crowdsourced data. Evaluating the quality of crowdsourced data in general and landmarks in particular has been lately the focus of different research studies. Acheson et al. (2016) found by comparing GeoNames and Getty Thesaurus of Geographic Names gazetteers a thematic heterogeneity inside each gazetteer: populated places are more covered then natural features such as mountains, streams and hills. They suggest “national data should be used preferentially when dealing with natural features”. The same finding of thematic heterogeneity is highlighted in crowdsourced data where contributors prefer mapping certain types of features (Bégin et al., 2013). Concerning the spatial heterogeneities of crowdsourced data, many studies showed that there is a high spatial heterogeneity between urban and less populated places due to density population or places attractiveness (Neis and Zielstra, 2014). Accuracy location of crowdsourced data is also frequently studied in the literature showing that accuracy is less good for natural features (Girres and Touya 2010; Bégin et al., 2013). From a semantic point of view Acheson et al. (2016) noticed a semantic heterogeneity between gazetteers (e.g. mountain represent peaks and ranges in DBPedia and only peaks in GeoNames).

The contributions of this paper are. 1/ The use of landmarks and routes and a combination of them for better localize victims in mountain areas. This approach is slightly different from the approach generally proposed in the navigation context (Denis, 1997), when landmarks are part of the itinerary. 2/ Generate knowledge concerning

the quality of landmarks and routes from crowdsourced data in natural environment. 3/Identify new types of landmarks and routes in mountain areas.

3. Sources of data description

In this study, we are focusing on data coming from active websites dedicated to leisure activities where the community or the authors are doing regular information updates, recommended by the practitioners or by authority, can be partner of NMA, publish enough data and treat a specific thematic source. Four sources of data are considered as the most relevant in mountain area and checked these conditions: RefugeInfo, Camptocamp (C2C), VisoRando. Two authoritative data, BDTPO and Protected areas (PA) are used to complete our analysis. Almost all data are provided by API.

Let us mention that other data portals, like OpenStreetMap (OSM), GeoNames, IGNRando provide landmarks or/and route but data are often in-bulks coming from other sources, where the origin of data is mentioned (e.g. in OSM some landmarks are referenced as C2C or wikidata). In this paper, these sources are not in our focused because the schema is too generalized, specific information being lost (e.g. technical information, GPX, link with waypoints) or is not appropriated (IGNRando use the cirkwi schema dedicated to tourist information).

3.1 Points of interest or activity: BDTopo

Among different themes of BDTPO produced by the French Mapping Agency, there are Points of Interest or Activity (PAI) and Roads (BDTPO, 2018) having a national coverage. The PAI theme groups points of interest and activity classified by category: administrative, relief, transportation, etc. Each feature is represented by its location (x,y), origin of data, type, name and relevance. The Roads layer contains both traffic roads and pedestrian paths being modeled such a road network and described by type, name, relevance, etc.

3.2 Refuge.info

Refuges.info is a crowdsourced website proposing different types of features in mountain areas. Even the website was initially focused on shelters; other types of landmarks can be found such as water point, summit, pass, hotels, camping, parking, bus. Each point is characterized by its location, name, type, and different descriptions such as (how to get access, number of places, origin of data, etc.), and photos. The features are generally coming by digitalization using maps. Regarding the spatial coverage, data are available mainly for France.

3.3 Camptocamp

C2C is a crowdsourced project involving mountain practitioners. The aim of the project is to share technical information: guides, incidents and accidents reports, outings with weather conditions or accessibility, and to contribute to the safety of them. It is an active and sustainable community proposing a broad coverage, both from mountain activities and geographic (France, Germany, Switzerland, Spain) points of view. All data (i.e. routes, waypoints, outings, accidents reports, photos) are referenced to each other's.

Waypoint. Each point is characterized by location, name, type and some thematic information such as altitude, completeness (i.e. qualitative quality showing how much

the contribution heeds the recommendations for sharing), textual description, and photos. Different types such as summit, pass, lake, waterfall, locality, canyon exist.

Itineraries. They represent different types of hikes mountain leisure. Very little itinerary has a GPX trace (25% in our study area). The creation of tracks has done by the contributors with its own data management tools and with its own reference data which can be authoritative or crowdsourced data. A route is characterized by title, type (e.g. loop by foot), rating (i.e. express the global difficulty of an itinerary), different information linked to the altitude (e.g. min, max, difference, slope, etc.), completeness, textual description, and photos.

3.4 Protected area

Among protected area, the Natural Regional Parks (NRP) and the National Park (NP) are French public institutions having the role to preserve natural environment and to develop economic and touristic activities. In addition, Parks on the French Territory managed information like points of relief, heritage points of interest, services (practical information like water points, awkward crossing) and routes. Most of them use GeoTrek suite softwares and BDTopo data to manage their data and DEM's french NMA to compute 3D attributes. Let us mention that authoritative data (BDTopo and protected area database) are not equivalent in term of content because they have different points of view (topographic versus environment protection and touristic). They are rather complementary with a small redundancy in terms of types of features and thematic information.

The entry for the data provided by the API is the route, the landmarks being linked to it using linear referencing. The landmarks are described by location, type, name, textual description, and altitude. The number of landmarks is set at three minimum and must be well distributed on the route. Types such as viewpoint, waterpoint, shelter, pass, lac, glacier, summit, pastoralism, history are present. Routes are defined by geometry provided in a GPX file (route points), a title, different practices, a rating, information linked to altitude, textual description, etc.

3.5 Visorando

Visorando is a website offering the opportunity to download tracks published online by contributors. Contrary to C2C, 100% of itineraries have been produced from GPS sensors. Besides, trackpoints, attributes such as title, location type (hiking, biking, etc.), textual description, duration, rating and altitude are provided. The website advises contributors to add fewer waypoints (landmarks or a major change direction). Tracks are adjusted and validated by contributors in a cartographic web interface using NMA layers (authoritative data).

4. Methodology

To identify and evaluate the potential of crowdsourced data and to define mapping rules describing semantic and schema relationships between the different data sources, integration techniques such as data matching is used. Our methodology is composed by three steps.

The first step consists in extracting data from different sources of information. To facilitate this task, a QGIS plugin² is developed. The second step is data matching.

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

The goal is to define both homologous routes (e.g. routes following the same paths) and landmarks (e.g. features representing the same object from the real world). The third step consists in evaluating homologous features by defining network connectivity measures such as number of connected components and confusion matrix.

Knowing the heterogeneity of data coming from different sources, data matching is a complex task. Figure 1 illustrates our approach to match landmarks and routes.

Routes are composed from a list of track points or way-points with respect with their original sources. Thus, to facilitate route matching from different routes, there is need to build their linear geometry. Two pre-processing steps are proposed.

First, points are map-matched to a reference road network by applying the Hidden Markov Models algorithm developed by Newson and Krumm (Newson and Krumm, 2009). This method is more suitable in our case because for each point to match it takes into account the points before and after, thus the consistency of the routes is fulfilled.

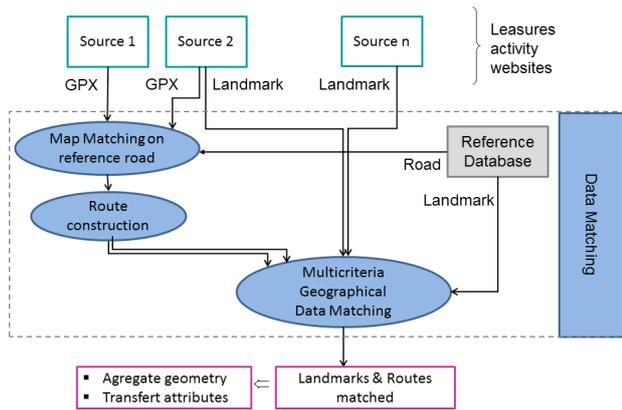


Figure 1. Data matching approach for routes and landmarks.

One of the difficulties of this algorithm is to apply the correct parameters: finding a balance between the location measurements and the tolerance of non-direct routes. For example, in Figure 2, the point P1 should not be map-matched to the closest edge, in order to maintain the consistency of the route and not to create deadlocks.

The second pre-processing step is route construction. Due to original data mapping, some consecutive points are far away one from the other. Thus, once points are map-matched, all edges obtained are not necessarily connected. In our process, two consecutive non-connected edges are connected by computing the shortest path between them. If the ratio between the distance as the crow flies and the distance of the shortest path is below a threshold, the edges of the shortest path are added to the route (green edges in Figure 2). At the end of this stage, the geometry of a route is linear being composed by all edges.



Figure 2. Map-matched GPS points (orange) to the BDTopo road network (blue). The point P1 is matched with the nodes to maintain consistency of the route.

Once the pre-processed steps finalised, the data matching can be separately carried out for routes and landmarks. An open source³ multi-criteria data-matching algorithm based on Belief Theory is used (Olteanu-Raimond et al., 2015). The data matching algorithm requires defining a reference and a comparison dataset giving in this way the direction of matching (for each feature from the reference dataset, the algorithm looks for homologous features in the comparison dataset). Let us mention that the reference dataset can be either an authoritative or a crowdsourced dataset. Knowing the characteristics of our datasets (see Section 2), two (position and name) and three (position, toponym, and semantic) criteria are defined for routes and landmarks respectively and the dataset coming from BDTPO are used as reference.

Two landmarks are considered as homologous features if they represent the same object in the real world. Knowing that an itinerary is a concept carried by a physical network infrastructure, thus we consider that two routes are homologous features if they are geometrically closed.

The position criterion is based on the distance between the reference feature and a candidate (e.g. Euclidean distance for landmarks and average of minimum of Hausdorff distance between every roads segments for lines). The name criterion compares the name of the reference feature with the name of the candidate. Among, different measures tested to compare strings, Samal distance (Samal et al., 2005) and Cosinus distance is considered as most appropriate for landmarks and routes respectively. Finally, the semantic criterion compares feature types. To compute similarities between features types the Wu-Palmer (Wu-Palmer) measure is applied by using domain ontology (Mustière et al., 2011). All these criteria are merged to take a final decision. Note that, the data matching algorithm do not take any decision (i.e. features are not matched) if the criteria are contradictories (e.g. for two potential homologous features the name is the same, the semantic are different and the distance between features is relevant). These cases are tagged as uncertainty and managed manually.

5. Results

5.1 Description of test area

The methodology is applied on the French Alps, near Grenoble (9,870 km²) on an area covering seven moun-

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

tain massifs (see Figure. 3). Concerning the landmarks, in our test area, the 25,906 features are distributed as follows: BDTPO (21,097 features), C2C (3,229 points), Refuge.info (478 points), PA (1,102 points). For this study the Visorando’s waypoints are excluded because there is no type and the name does not match with a toponym. The routes are localized in two mountain ranges. The 706 routes are distributed as follow: Visorando (398), C2C (127) and PA (181). Our approach is not taking into account existing spatial and topological relations between landmarks and routes in the data matching process. Each data source does not necessarily provide both features. In addition, the types of the relationships between landmark and route are not clearly defined, a link represents proximity relation just as well as intervisibility relation. The results will therefore be analyzed separately.

As seen in Figure. 3, the coverage of landmarks and routes is well dispersed. The accuracy of routes is heterogeneous between sources of data. On average, the distance between two consecutive points is 25m, 50m, 5m for Visorando, C2C, and PA respectively.

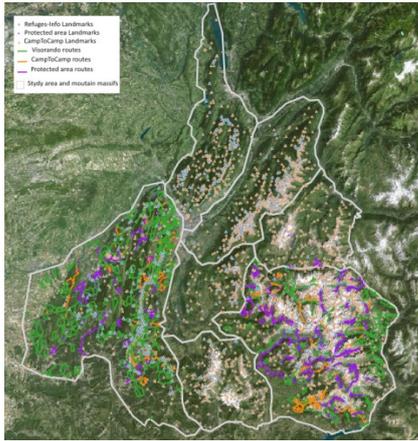


Figure. 3. Test area: mountain area in the French Alps.

5.2 Data matching results

The parameters for the map matching are: 80m to search candidates, and 4m for standard deviation of GPS error measurements. For adding the edges of the shortest path to the route, the maximal ratio between the distance as the crow flies and the distance of the shortest path is empirically estimated to 5.

Overall, 97% of points (298,348 points) are map matched and the root mean square error of displacements induced by map matching is equal to 12.94 m. Note that a few deadlocks are generated after the map matching process.

Regarding the route construction step, the rate between the distance as the crow flies and the distance of the shortest path is empirically estimated to 5. The total length of added missing edges equals to 337 km representing 4% of the total length of the routes. It shows the relevance of route construction process.

Regarding the parameters for the multi-criteria data matching, the buffer to find candidate is 500m for both routes and landmarks. The thresholds in the matching algorithm are defined with respect to data characteristics obtained by analyzing the distributions of the distances for each criterion. For routes, the thresholds for the position criterion and the name similarity criterion are equal

to 200m and 0.3, respectively. For landmarks, the thresholds are 100m, 0.6 and 0.7 for position, name, and semantic criteria, respectively.

Chosen the appropriate thresholds is a difficult task being always a compromise between matching results quality and number. In this research, thresholds are empirically chosen by computing distributions of distances or by using thresholds already suggested in the literature (Olteanu-Raimond et al., 2015). It was the case for example for thresholds used for name and semantic criteria for landmarks.

5.2.1 Landmarks data matching results

Each data source is separately matched with the authoritative BDTopo data. The results obtained are depicted in Table 1. Link cardinality 1:0 means that a feature from a data source (e.g. C2C) has no homologue feature in the authoritative data (e.g. BDTopo), whereas 1:1 link cardinality means that a homologous feature exists in authoritative data.

	1:0	1:1	Uncertainty
C2C-BDtopo	1167	1938	214
REFuges.info-BDtopo	80	367	34
PA - BDTopo	863	180	61

Table 1. Data matching results with respect with the cardinality.

Concerning the non-matched features (column 2 from Table 1) two patterns can be noticed. First, features are not matched due to the poor accuracy of crowdsourced landmarks (i.e. the landmark is far away from its homologous in BDTopo); secondly, features are not matched because, indeed, there are no homologous features in BDTopo. This last case concerns 70% of the non-matched landmarks. For example, a ruined shelter is tagged as history in PA data but the shelter is not any more in authoritative databases or other data, because the objet has destroyed.

Figure 4 shows the distributions of non-matched landmarks for the most frequent types of landmarks. These results show the complementarity of different sources of data. For example C2C has many types of specific landmarks such as climbing, waterfall, less present in authoritative data.

Concerning the links 1:1, a manually validation is carried out in terms of precision. This is necessary for analyzing the semantic between different sources of data. The precision is equal to 95% for C2C and Refuges.info and 71% for PA. Figure 5, on the left, illustrates typical 1:1 matching links where the object named Lac Gary has homologues in all three data sources.

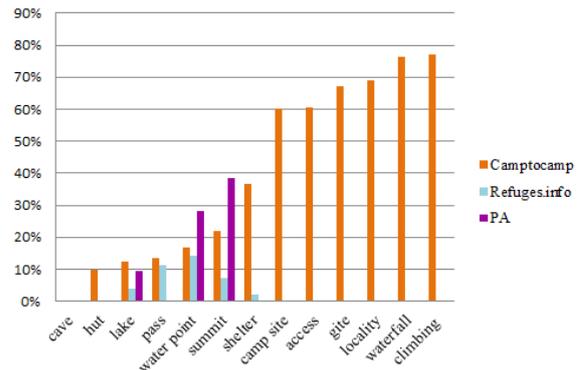


Figure 4. Percentage of non-matched landmarks by type

Finally, uncertainty cases are complex cases where any decision is taken by the data matching algorithm. They are fixed manually. Figure 5, on the right shows an example of such a case. The landmark “Col de l’homme (Sanctuaire Notre Dame de la Salette)” in C2C, having the type equal to access, has three candidates to match in BDTopo: “Col de l’homme” (pass), “Tête de l’homme” (summit) and “Parking Col de l’homme” (parking). In this case, no decision is taken due to the fact that access is not a concept in the domain ontology, and the names of candidates are quite similar.



Figure 5. Matching results: 1:1 cardinality in the left; and uncertainty, in the right.

5.2.2 Routes data matching results

The results obtained for route data matching are depicted in Table 2 having also a good precision (~ 85%).

	1:0	1:1	Uncertainty
C2C- Visorando	76	20	21
Geotrack- Visorando	92	59	14
C2C- Geotrack	77	14	6

Table 2. Data matching results with respect with the cardinality.

In order to build a pedestrian routes dataset, two routes are matched if they are geometrically close. Different specific cases where two routes are homologous are observed: 1/ when it exists a variant of the route and the detour is short; 2/ when two routes are very close and have the same direction (typically when they border a river (which is satisfying in our context when the victim describes the route by “I’m walking along a river”)) (see Figure 6); 3/ when the start points are a few meters one from each other. In this case, a route can do a loop and another route is a round itinerary. In our study, the missing paths on the road network have not been managed (e.g. off road, glacier road, rock road, etc.). These sections are not considered in data matching.



Figure 6. Route data matching result.

The non-matched and the uncertainty cases represent most of the two cases mentioned above where the distance is more or less relevant. The number of matched routes matching is little. Many routes have not the same starting point or destination, they only show some sections sharing.

In conclusion, landmarks and routes matching give good results. The large number of non-matched features proves the richness of crowdsourced data. In both situations, uncertainty represents in most of the cases, non-matched features, which corresponds to a good behavior for our need. Concerning the landmarks, we have noticed that the matching failed when concepts do not exist in the used ontology and when the name is not filled in authoritative data. In this case, the relevant criterion is the distance. In route matching, the effect of the name criterion is very difficult to analyze. The location criterion plays an important role in the final decision for both cases, matched and not matched routes.

5.3 Data coverage and redundancy

Among the network connectivity measures, the number of connected components of one or two sources can indicate the evolution of sub-networks. In Figure 7, all connected components for each source, represented by bubbles, indicate the number of routes: 58 connected routes for Visorando, 28 for C2C and 28 for PA. This means that data sources cover a large area. In our experiments, with over two data sources, the number of connected components approximates the maxima of the numbers of connected components (59, and 58, for the merging of Visorando and C2C, respectively for the merging of Visorando and PA). The connectivity between the two networks improves (the sub-networks contains more edges), admittedly, but the cover of hiking spots does not change (all sub-networks are distant enough in order not to be merged).

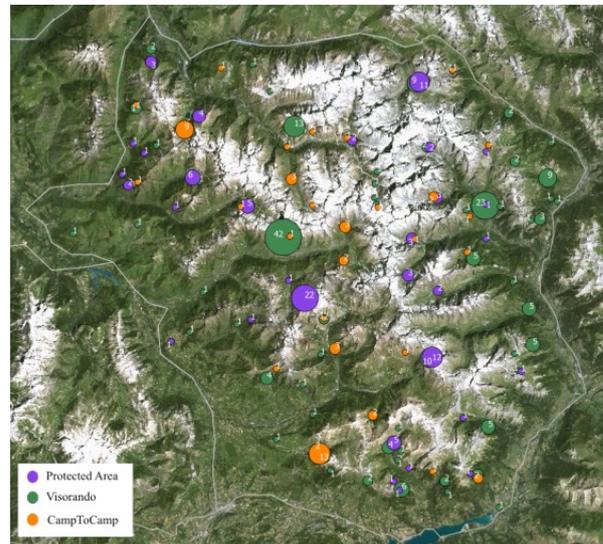


Figure 7. Connected components in each source, the label indicate the number of routes.

Concerning the redundancy of landmarks between different data sources, only 37 homologous landmarks are present in all data sources. Table 3 shows the redundancy of landmarks with respect to data sources. The main diagonal represents the total number of landmark for each source. It can be noticed that C2C and BDTopo are more comparable in terms of data: 1,790 landmarks from C2C have a homologous landmark in BDTopo.

	C2C	Refuge. info	PA	BDTopo
C2C	3,009	217	81	1,790
Refuge.info		478	41	352

PA			1102	127
BDTopo				21,097

Table 3. Redundancy of homologous landmarks in data sources

5.4 Analysis of semantic mapping rules

As mentioned in Section 1, one important need is to query landmarks, routes, and a combination of them. For this purpose, it is necessary to define semantic mapping rules between different concepts of data sources (e.g. pass in the real world is mapped as an access in C2C and as a pass in Refuges.info). To identify the semantic mapping rules, we compute confusion matrices by using homologous landmarks. For lack of space, they are not all depicted; only the main resultants are described.

An interested mapping rule to define queries for routes on the web concerns the difficulty level of the route (see Table 4). People can describe their route such as “we take an easy way to go the shelter Mont Blanc”, “we are on the difficult stretch of the route”. In spite of the little route matched, we analyze the confusion matrix between the Visorando routes and PA routes (59 routes matched).

Each source contains five levels, the names of the levels are different, but the parameters considered are similar (e.g. effort, physical mental and altitude difference). Overall, the two sources do not agree. For example, the medium level is spread over the easy, medium and difficult levels in both cases. More precisely, the number of route with easy level in Visorando takes all values in PA. One route with difficult level is classified in the easy level and vice-versa. This preliminary study should be examined in greater detail.

PA	Visor.	Easy	Mid	Diff	Very diff	Extrem. diff
Very easy		2		1		
Easy		4	6			
Mid.		5	15	7		
Diff.		1	7	9	2	
Very diff.						

Table 4. Confusion matrix for difficulty level of the route between Visorando and PA routes

For landmarks, for each data matching a confusion matrix is computed. For C2C-BDTopo data matching, it can be noticed that summits in C2C are often matched with summits (547 features), pics (259), ridges (29 features), rocks (29 features). Another relevant rule concern the landmarks having the type equal to access. This landmarks are matched to parking (40), passes (33). It confirms the definition of access type in C2C (i.e. represents real world objects linked to the access to a given position or as started point from a position A to a position B). Second, for Refuges.info, it can be noticed that passages are always matched with passes. Another interesting matching such as summit/pic (16 landmarks) and water-points/waterfalls (5) are found. A water-point in Refuge.info represents a water source but also objects such as bridges. These landmarks are rightly not-matched. Finally, the analysis of PA-BDTopo confusion matrix gives very surprising results. For example, landmarks having natures such as geology, history are

matched with canyons, passes, caves, and shelters. This shows the complementarity of PA data with respect with the other data sources. Moreover, point of view landmarks are often matched with summits and passes.



Figure 8. Links between routes and landmarks.

With respect with using a combination of landmarks and routes, Figure 8 illustrated how landmarks can be linked to routes.

Figure 9 represents a real object represented in all data sources (“Lake de la Muzelle”), linked with two homologous routes.



Figure 9. Landmarks representing a lake linked with two homologous routes.

6. Conclusion and perspectives

This paper studies the potential of crowdsourced data for integrating heterogeneous spatial information for mountain rescue tasks. In this paper, we focus on routes and landmarks. To achieve the goal, we define a methodology to compare and evaluate different data sources based on matching techniques.

The approach is tested on three and four data sources for routes and landmarks respectively. Our results highlight the richness and the complementarity of different data sources and a good coverage. Following this analysis, new knowledge regarding the semantic of data (classification of landmarks), geometry and level of routes is generated. Integrating data through identified data sources is not an easy task. The next step is to formalize this new knowledge of different sources via metadata.

Concerning the relation between landmarks and routes, as already exist in some sources, we studied the feasibility to define the relation for all sources by using matching links. The first results, as shown in Figure 8 are promising. Thus, thanks to the matching links, landmarks from a route (purple in figure) can be transferred to another route (orange in the figure) and vice versa. Further work consists to design a new global schema in order to define the relations between landmarks and routes in different sources.

From a technical point of view, this work contributed to a QGIS PlugIn for downloading all data provided by API and a new algorithm to construct the routes⁴.

In addition to specify reference spatial relation between landmarks and routes for optimizing the query, cases when the same object from the real world is represented in many sources occur. In this case, the question is which feature should return the web service? A multiple representation or a fuzzy representation for matched features seems to be an interesting solution. More generally, how the spatial context can be taken into account by the web service in order to return the information as complete as possible?

In this paper, for identifying landmarks in mountain area, many types of landmarks are not represented in authoritative gazetteer. Thus, a further research is to identify all types of landmarks not presented in authoritative data and to expand the ontology of authoritative data (Mustière et al., 2011).

Routes matching results indicate some weaknesses in our definition. For example, two routes very close can't go through by any same waypoint or can have a major different vertical drop. By improving some criteria of data matching, such as by weighting with terrain condition (difference in elevation) in the spatial criteria or by weighting place-names identified in the title, routes matched definition can be extended to : two routes are equivalent when there have some same waypoints and terrain features are similar between two waypoints.

In this work, we have focused on a test area in the French Alps. We believe that our methodology can be apply in other mountain countries. It depends, of course of the availability of data. Concerning the crowdsourced data, some of the used data such as C2C has features outside France (e.g. Italian and Swiss Alps, Spain Pyrenees, Germany, Nepal). Other existing crowdsourced initiatives having world spatial extent (Wikiloc) or global gazetteers (GeoNames, DBpedia) also exist. For authoritative data national gazetteers exist in most of developed countries. To extend the method to others mountains countries, the matching methods might be adapted to the characteristics of data typically in the choice of criteria and types of distances.

Finally, in future research, we plan to use a multi-source matching algorithm. Indeed, the multi-criteria algorithm compares a given dataset to a reference or to the biggest dataset (Visorando for routes). The matched links cardinalities corresponds to 1:1 or 1:n (by post-processing).

For n data sources and with a reference dataset, we need to run the algorithm $n*(n-1)$ times, which is not an optimum solution.

7. References

- Acheson, E., De Sabbata, S. and S Purves, R. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers Environment and Urban Systems* 64, pp. 309–320.
- Bégin, D., Devillers, R.; Roche, S (2013). Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours. In *Proceedings of the 8th International Symposium on Spatial Data Quality*, Kowloon, Hong Kong, China; pp. 149–154.
- Brosset, D., Claramunt, C. and Saux, E. (2008). Wayfinding in natural and urban environments: A comparative study. *Cartographica* 43, pp. 21–30.
- Denis, M. (1997). The description of routes: A cognitive approach to the production of spatial discourse. *Cahiers de Psychologie Cognitive* 16, pp. 409–458.
- Derungs, C. and S. Purves, R. (2013). From text to landscape: Locating, identifying and mapping the use of landscape features in a swiss alpine corpus. *International Journal of Geographical Information Science* 28, pp. 1272–1293.
- Duckham, M., Lingham, J., Mason, K. and Worbosy, M. (2006). *Qualitative Reasoning About Consistency in Geographic Information, Information Sciences—Informatics and Computer Science, Intelligent Systems, Applications: An International Journal* vol. 176 (6), pp. 601-627.
- Egenhofer, M. J. and Mark, D. M. (1995). Naive geography. In: A. U. Frank and W. Kuhn (eds), *Spatial Information Theory A Theoretical Basis for GIS*, Springer Berlin Heidelberg, Heidelberg, pp. 1–15.
- Ferrara, A., Nikolov, A. and Scharffe, F. (2013). Data linking for the semantic web. *Semantic Web: Ontology and Knowledge Base Enabled Tools, Services, and Applications*, pp. 326/
- Girres J.-F. and Touya G. (2010). Quality Assessment of the French OpenStreetMap Dataset, *Transactions in GIS* 14(4), pp. 435–459.
- Goodchild, M. and Li, L. (2011). Formalizing space and place. In: *CIST2011 - Fonder les sciences du territoire, Fonder les sciences du territoire, Collège international des sciences du territoire*, Paris, France, pp. 177–183.
- Ivanovic, S, Olteanu-Raimond, A-M, Mustière S. and Devogele, T. (2016). Detection of outliers in crowdsourced GPS traces, *Spatial Accuracy 2016 Symposium*, July 2016, Montpellier, France.
- Halilali, M., Gouardères, E. Devin, F. and Gaio, M. (2018) Plateforme logicielle pour l'intégration et la composition de services géospatiaux. In *Proceeding of SAGEO'2018 (Spatial Analysis and GEOmatics)*, November 6-9, 2018, Montpellier, France.
- Hill, L. and Zheng, Q. (1999). Indirect Geospatial Referencing through Place Names in the Digital Library: Alexandria Digital Library Experience with Developing and Implementing Gazetteers, in *Proceedings of the 62nd Annual Meeting of the*

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

- American Society for Information Science, Washington, DC, USA, pp. 57–69.
- Kasemsuppakorn, P. and A. Karimi, H. (2013). A pedestrian network construction algorithm based on multiple gps traces. *Transportation Research Part C: Emerging Technologies* 26, pp. 285–300.
- Kerouanton, C., Duparc, A., Jolivet, L., Perrin-Malterre, C. and Loison, A. (2017). Outdoor recreationists, how to find their pauses by using ST-DBSCAN clustering algorithm. In *treizièmes Rencontres de Théo Quant*, Besançon, France.
- Lynch, K. (1960). *The Image of the City*. MIT Press.
- Moncla, L., Renteria-Agualimpia, W., Nogueras-Iso, J. and Gaio, M. (2014). Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. In: ACM(ed.), *ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2014)*, Dallas, Texas, United States.
- Mustière, S., Abadie, N., Aussenac-Gilles, N., Bessagnet, M.-N., Kamel, M., Kergosien, E., Reynaud, C., Safar B. and Sallabery, C. (2011). Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique, *Revue Internationale de Géomatique*, 21(2), pp. 155–179.
- Neis, P. and Zielstra, D. (2014). Recent developments and future trends in volunteered geographic information research: The case of Openstreetmap. *Future Internet* 6(1), pp. 76–106.
- Newson, P. and Krumm, J. (2009). Hidden markov map matching through noise and sparseness. In: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*, ACM, New York, NY, USA, pp. 336–343.
- Olteanu-Raimond, A.-M., Mustière, S. and Ruas, A., (2015). Knowledge formalization for vector data matching using belief theory. *Journal of Spatial Information Science*, No 10.
- Pitman, A., Zanker, M., Gamper, J. and Andritsos, P. (2012). Individualized hiking time estimation. pp.101–105.
- Samal, A., Seth, S. and Cueto, K. (2004). A feature-based approach to conflation of geospatial sources. *IJGIS*, vol. 18, n°5, p. 459-489
- Schockaert, S., Cock, M. D. and Kerre, E. E. (2008). Location approximation for local search services using natural language hints. *International Journal of Geographical Information Science* 22(3), pp. 315–336.
- Timpf, S. (2005). *Cognitive Wayfinding Agents in Public Transportation Networks*, Geographic Information Science Center, University of Zurich, Switzerland.
- Teulade-Denantes, J., Maudet, A. and Duchêne, C. (2015) Routes visualization: Automated placement of multiple route symbols along a physical network infrastructure. *Journal of Spatial Information Science*.
- Vias, J. and Ocana Ocana, C. (2014). Multicriteria evaluation by GIS to determine trail hiking suitability in a natural park. *Boletín de la Asociación de Geógrafos Españoles* pp. 323–339.
- Wu, Z., and Palmer, M. *Verb Semantics and Lexical Selection* (1994), In: *Proceedings of the 32nd Annual Meetings of the Associations for Computational Linguistics*, pp. 133-138.
- Zhou, S., Winter, S., Vasardani, M. and Zhou, S. (2017). Place descriptions by landmarks. *Journal of Spatial Science* 62(1), pp. 47–67.