PROJECT FINAL REPORT

OpenDataSyris
Open Data pour Syris: Système de Recommandation pour Itinéraires pédestres de santé

Appel à projets Open Data HES-SO

Alexandre Cotting, Jean-Paul Calbimonte
Institut Informatique de Gestion

HES-SO Valais-Wallis

Sierre
November 2, 2021

HES-SO Domaine Economie & Services

Abstract

The popularity of hiking has steadily increased in the latest years, across different segments of the population. Although there is considerable evidence of the benefits for hikers regarding physical and mental health, the inherent risks of these outdoor activities cannot be underestimated. Accident prevention and an increase of awareness about possible risks are necessary to minimize hiking and pedestrian tourism's negative consequences. In most hiking information maps and interactive applications, there is usually not enough information about difficulty points or the granularity level required to provide tailored recommendations to hikers with physical or psychological limitations. This report documents the process of preparing and opening the data of the Syris project, a geo-information system for hiking itineraries that incorporates Points-Of-Difficulty to assess the level of effort, technique, and risk of hiking trails. This dataset published in Zenodo[1], is a pioneering effort to offer comprehensive information about detailed difficulty and effort levels in hiking paths at Val d'Anniviers in Switzerland, with the long term goal of expanding the database to the rest of the country.

1 Introduction

In the latest decade, pedestrian tourism has experienced increased interest, motivating the emergence of online services and products for hiking activities [10, 3]. Nowadays, the availability of mountain paths and hiking trails reaches a wide range of population segments, from novice to expert users. The heterogeneity of the potential hiker profiles leads to several challenges, especially regarding safety and risk concerns, which have not been fully addressed by existing information platforms and applications. Indeed, in most mountain tourist destinations, there is only superficial information about trail difficulty and risks. Moreover, it is often not standardized nor reflecting specific elements that can drastically affect the overall hiking experience.

This report describes the process of opening the data collected for Syris, a comprehensive platform for management and recommendation of pedestrian tourism trails, including the assessment of difficulties and risks. Syris, developed by the eHealth unit at the Institute Of Information Systems HES-SO Valais-Wallis, follows a methodology and data model for describing and representing points of difficulty, according to well-defined criteria [5]. Moreover, it includes visualization and filtering functionalities, aiming at providing tailored recommendations for users according to their physical skills and preferences. Furthermore, the platform includes a mechanism for data acquisition of points-of-difficulty through a

[1] `https://zenodo.org/record/5527030`

mobile application, leveraging knowledge, and experience from expert hikers on-the-ground. Syris has been deployed[2] and tested in the mountain region of Val d'Anniviers in Switzerland, including the acquisition of real data on more than 70 pedestrian tourism trails of different characteristics, in collaboration with local tourism offices. This data was acquired by volunteers following a bottom-up approach, and is now made available to the scientific community and to the public in general.

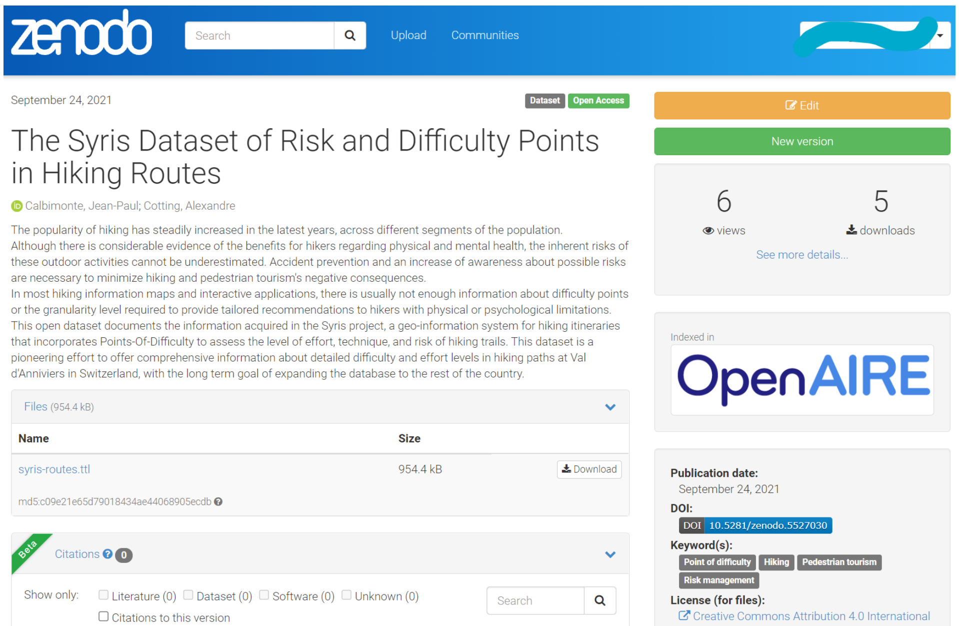

Figure 1: Zenodo.

2 Data description

Before introducing the details about the data we provide a description and context of the system that is used to manage, collect, and exploit them. Then, we proceed to describe the different models, data formats, volume and other relevant information.

2.1 Syris: system overview

The data described in this report, and made available as Open Data, is managed and exploited by the Syris platform [4]. Syris has been conceived as a prototype

[2] Available at: http://syris.iigweb.hevs.ch/.

that integrates both geographical aspects of hiking trails, as well as detailed risk and difficulty information. To do so, Syris relies on a well-defined model for both hiking trails and their difficulties, and user profile models, as described in [5]. This model considers the complexities due to the subjectivity of difficulty information, related to the perception of risk from heterogeneous users. As it has been evidenced in previous initiatives, through questionnaires and workshops with key actors in the Swiss tourism sector [6], difficulty and risk details are continuously demanded by end-users and tourism providers.

The Syris platform incorporates difficulty and risk information in a Web interface that includes search and visualization capabilities (Figure 2). This interface also provides basic recommendations to users, according to their physical condition and perception of risk.

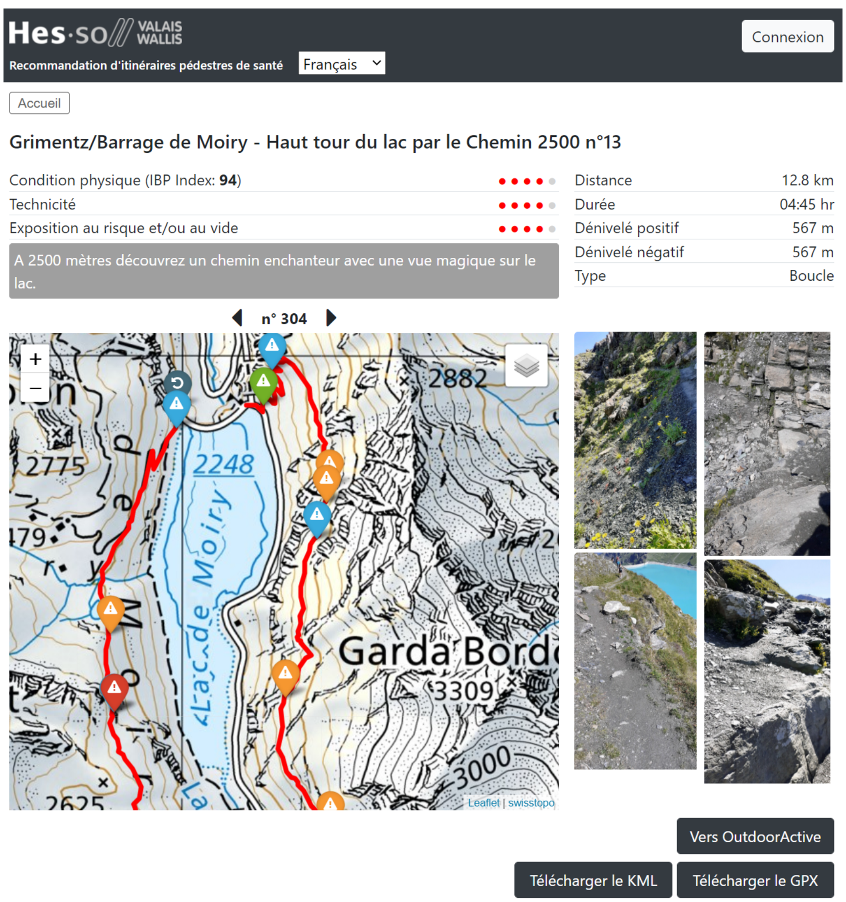

Figure 2: The Syris Web interface: difficulty and risk information integrated within the geographical and attributive data of the available hiking trails.

The main functionalities of the Syris Web application can be summarized as follows:

- *Route geographic visualization:* The system displays the available hiking trails on a Web map, including features as: zoom + drag, different map overlays (OpenStreetMap, OpenTopoMap, and SwissTopo), and route clustering according to proximity.

- *Route filtering:* Routes can be filtered according to different criteria, e.g., effort, technique, and risk. The difficulty according to the three main dimensions is differentiated with colored icons. Further information about the routes are available by clicking on these icons. A dynamic list of routes is presented along with the map.

- *Health profile:* Registered users can save their profiles, including their physical condition (skills), technique, and risk perception. These parameters are obtained by filling a version of the Baecke questionnaire [1].

- *Management of points of difficulty:* Specific difficulty points are highlighted in the map visualization. The itinerary's start and endpoints are also highlighted, as well as additional information about the difficulty and other route attributes. Routes and points of difficulty can be exported as GPX and KML. An administrator can validate or control the quality of any point of difficulty.

The Syris data is stored in a geospatial PostgreSQL database with PostGIS extensions. Both the database and Web servers have been deployed as docker instances. The points of difficulty are fed dynamically from an external Data Acquisition mobile application (see Section 2.3), and effort scores of the itineraries are taken from the IBP index[3] API.

2.2 Data Model

The Syris data model (see Figure 3) is based on our previous research on semantic representation of hiking trails [5]. This data model considers the assessment of difficulties as a combination of three different dimensions, as developed by the classification methodology of the French Hiking Federation[4]:

- Effort: this dimension refers to the physical energy required to complete a given itinerary. The calculation of this factor takes into account different parameters, such as the total distance, slopes, altitude, descent, and slope changes. More specifically, Syris uses the IBP index, a numerical scale for representing the human effort in hiking and biking tracks.

[3] `https://www.ibpindex.com`

[4] Fédération Française de la Randonnée Pédestre (FFR) `https://www.ffrandonnee.fr`

- Technique: this dimension refers to the ability and motricity skills required to overcome obstacles present in the track. Examples of technique-related difficulties include rocks that require specific skills to be overtaken, e.g., raising the leg to the knee or hip level, or even using the hands to overcome the obstacle. Given the lack of automatic calculations of an index for this dimension, experts usually perform the assessment on the itinerary.
- Risk: this dimension refers to psychological difficulties related to situations to which the hiker is exposed and could lead to potential accidents or dangerous conditions. As with the technique, Syris requires the intervention of experts in the field to assess the risk level. This score is typically associated with the risk of falling or losing control while climbing.

It is important to consider that a single hiking trail may be divided into different sections, with possibly very different difficulty characteristics. A pathway may start with a very easy sector in terms of risk and effort, while afterwards, there might be points of difficulty that require advanced skills to be crossed. In Syris, all major points of difficulty of a trail are reported and displayed. However, the highest-ranking points are taken into account to have an overall characterization of the trail.

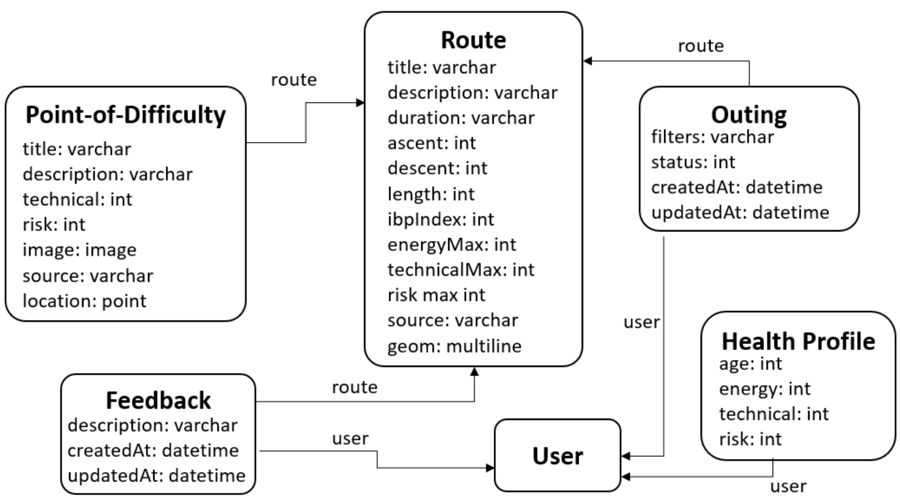

Figure 3: Syris Data model.

The data model of Syris, as depicted in Figure 3, includes the following main elements:

- Routes. Attributive data comprises the route title, short description, long description, duration, height difference, length, type (loop, round trip, traverse). It also includes route metadata regarding the risks, as described previously, i.e., IBP Index, energy, technical, risk. The route is also associated with a geometrical object, represented as a multi-line.

- Points of difficulty: includes attributes including title and description, as well as the risk and technical evaluation scores. It may also include an image of the point of difficulty, and the geometrical object, typically represented as a point in space.
- User information: which includes not only the user’s basic information, but also the health profile (i.e., according to the Baecke questionnaire), and the risk/techniques scores. Any filtering and saved search are also stored, along with feedback provided about specific hiking routes.

For the open data initiative we focused only on the routes and points of difficulty, and we excluded all user-related data, due to the sensitive nature of this type of information.

Data from Syris can be exported in GeoJSON format. This format is syntactically a JSON file in which all geo-referenced objects are represented as *Feature* objects. GeoJSON allows any number of properties in a given feature (e.g., title, description, duration, etc.), while storing the geographical location in a geometry object. The example in Listing 1 represents a hiking trail, including its multi-line geo-coordinates, and all its associated attributes.

```
{
  "type": "Feature",
  "properties": {
    "title": "Vissoie - Hotel Weisshorn par Gillou",
    "subtitle": "https://www.outdooractive.com/fr/route/randonnee/sierre/vissoie-hotel-weisshorn-par-gillou-n-9/38445310/",
    "description": "Cette randonnÃ©e, avec une denivellation considÃ©rable, permet de relier Vissoie Ã  l’Hotel Weisshorn, considere encore aujourd’hui comme l’une des perles du Val d’Anniviers.",
    "duration": "03:00",
    "ascent": 1133,
    "descent": 0,
    "tour_length": "5.7",
    "typeof": "Traversee",
    "ibp_index": 85,
    "energy_max": 4,
    "technical_max": 3,
    "risk_max": 3,
    "pk": "2919"
  },
  "geometry": {
    "type": "MultiLineString",
    "coordinates": [[[7.585609, 46.215201], [7.585611, 46.21518], [7.585626, 46.215141], [7.585651, 46.215109], [7.585679, 46.215083], [7.585712, 46.215065], [7.585756, 46.21504], [7.585816, 46.215011], [7.58587, 46.214982], ... ]]
  }
}
```

Listing 1: Data of a hiking route from Syris in GeoJSON format.

In the same way, all points of difficulty are also represented as *Feature* objects. Each point has simple coordinates, as well as a description and characterization of difficulties. Each point of difficulty may have an associated picture, as in the example in Listing 2.

```
{
   "type": "Feature",
   "properties": {
     "title": "Racine",
     "description": "Racine contournable",
     "technical": 3,
     "risk": 2,
     "img": "pods_pics/ec8c1d91.jpg",
     "pk": "289"
   },
   "geometry": {
     "type": "Point",
     "coordinates": [7.607387, 46.210189]
   }
}
```

Listing 2: Data of a hiking route from Syris in GeoJSON format.

2.3 Data Acquisition

Difficulty points are acquired in the field via a mobile application (Web or native) that allows the user to take a picture of the point of difficulty, obtain its geographic coordinates, and add a few attributes characterizing the type of difficulty. For the mobile application, we opted for using an existing mobile data acquisition platform: Appsheet[5], which allows us to easily create a small mobile Web application with a form including the possibility to take pictures or to geo-locate the difficulty (Figure 6). The data is saved and then imported into the Syris database through a Python script. Each time a point of difficulty image is imported into the Syris, it will be compressed, resized, while the latitude and longitude will be extracted, providing the location. Once the import is completed, recalculating the technical and psychological difficulties of the routes according to the newly imported point of difficulty is imperative. We have chosen to exclude points beyond a threshold distance from the route in the calculation of the difficulties score.

The data acquisition process has been validated by a set of volunteer beta-testers who have hiked through the selected 79 routes of the region of Anniviers to collect difficulty points and populate our database. The points of difficulty collected have been validated by terrain experts, although it has been acknowledged that there is a need for a standardized procedure for such type of validation, potentially including the Swiss Council for Accident Prevention [6]. Although the acquisition App is limited to the collection of points of difficulty, it has been evidenced that its usage could be extended in order to provide data quality and validation features. Following a citizen-science approach, a collaborative methodology can be put in place in order to allow users to contribute difficulty and risk data according to an established protocol.

[5] https://www.appsheet.com/
[6] https://www.bfu.ch/

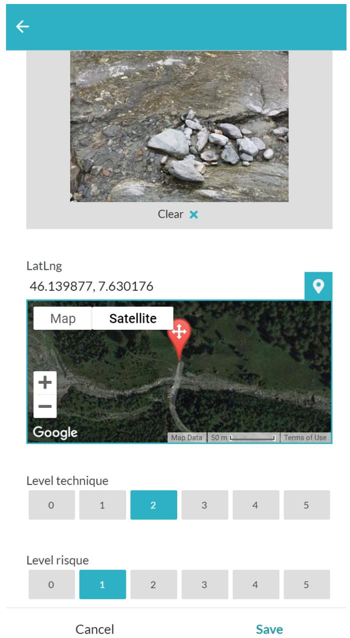

Figure 4: Syris Acquisition App based on Appsheet.

3 State of the Art

Given the rising popularity of outdoor activities, there are several commercial solutions that offer hosting and managing hiking trail information, often offered to local tourism offices. For example in Austria, the local tourism office of Vorarlberg[7] maintains an online database of hiking trails. The Swiss canton of Valais delegated the management of official hiking trails to Valrando[8], while in local municipalities other commercial services are used, such as Outdooractive[9].

Although a difficulty level is shown in these platforms, it follows a generic classification which is not always easy to interpret. If the difficulty is related to physical effort (e.g. due to the total distance), a hiker may decide to just take a portion of the trail. However, if the difficulty is related to the technique, the user has no way to know where this difficulty is located, or if there are other factors (e.g., risk & safety) which could play a role in deciding what actions to take. Although we have taken Outdooractive as an example, other platforms have similar limitations regarding the completeness of safety information, as they rely on the existing sources, which do not go deeper into modelling this kind of data.

In Switzerland, there are different efforts to characterize mountain trails, with different levels of adoption and standardization. The classification provided by

[7] `https://www.vorarlberg.travel`
[8] `https://www.valrando.ch/`
[9] `https://www.outdooractive.com`

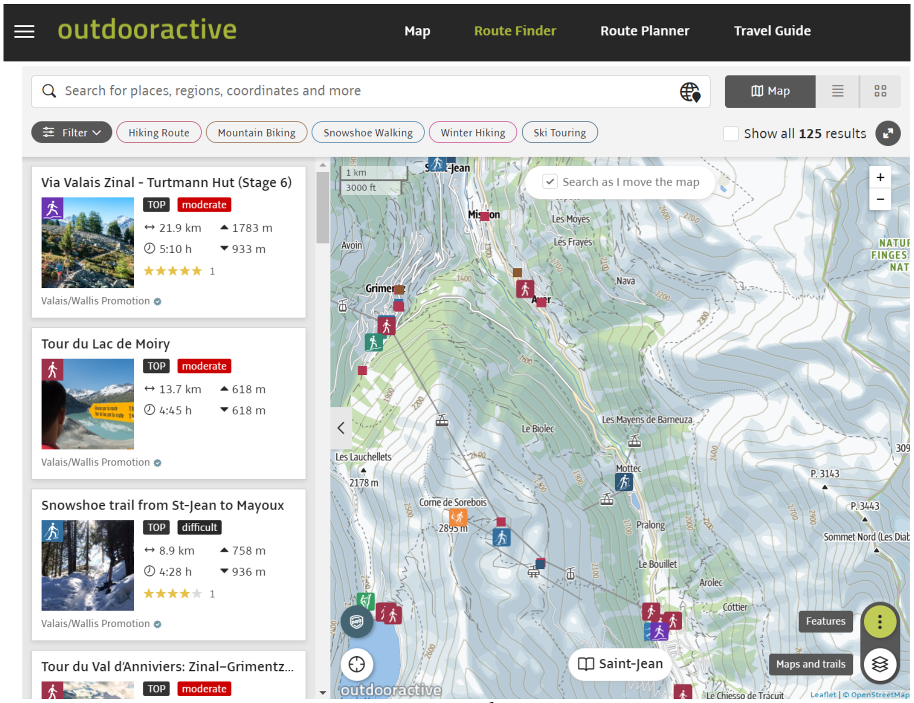

Figure 5: Outdooractive.

Suisse Rando[10] is widely used in a large set of Swiss hiking trails, at the physical yellow signposts and painted over rocks along the paths. There are mainly three categories, which are mainly used to describe the type of trail. Although this classification is targeted more towards characterizing the type of trail, it inherently integrates certain notions of difficulty. In a very coarse-grained manner, this classification corresponds to an easy-moderate-and-hard difficulty levels. At the same time, the Club Alpin Suisse (CAS) has established a 6 level scale as follows:

- T1 Randonnée
- T2 Randonnée en montagne
- T3 Randonnée en montagne exigeante
- T4 Randonnée alpine
- T5 Randonnée alpine exigeante
- T6 Randonnée alpine difficile

There exists an equivalence table for the CAS and Suisse Rando classification:

- Suisse Rando : jaune ; CAS : T1

[10] https://www.randonner.ch

- Suisse Rando : blanc-rouge-blanc ; CAS : T2 and T3
- Suisse Rando : blanc-bleu-blanc ; CAS : T4 and T5

There is also guidance criteria to link difficulty levels to route distance, duration and altitude differences:

- Facile (long ¡= 15 km ; ¡= 300 m montée, max 4h)
- Moyen (max 20km ; max 800m montée, max 6h)
- Difficile (more than 20 km, more than 800 m montée ; more than 6h)

Nevertheless, it is often hard for hikers to position themselves with respect to these parameters. Furthermore, they only provide a high-level classification disregarding punctual risks and difficulties that may entirely change the perception of the hiking trail. In a wider scope, SuisseMobile[11] provides an online application for displaying and creating pedestrian trail maps. Collecting and maintaining up to date information about hiking trails is expensive and a human-intense activity [14, 7]. Nevertheless, this detailed information is crucial to provides hikers with crucial indications about the trails they are willing to undertake.

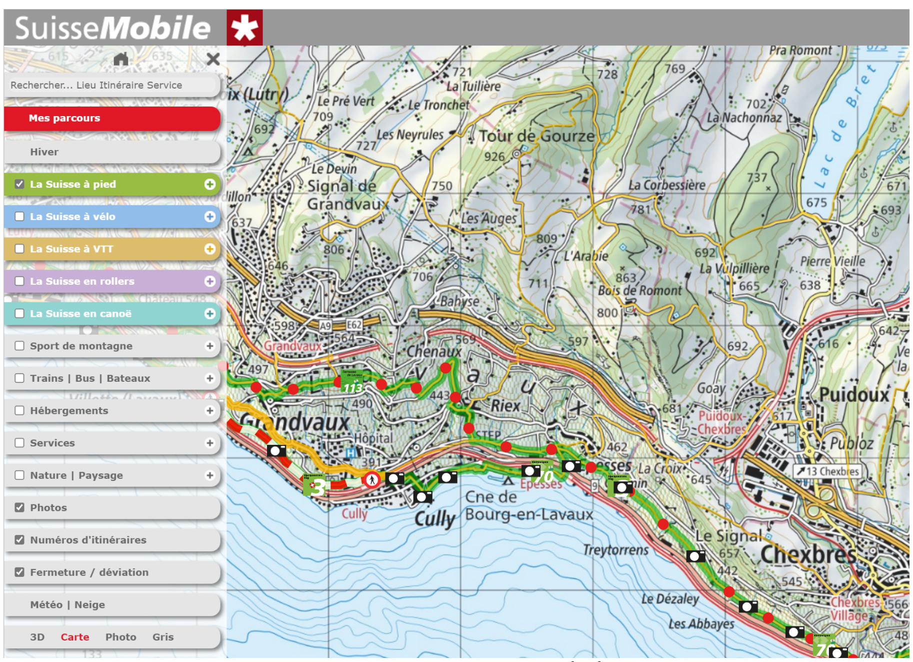

Figure 6: Suisse mobile.

Although these public and commercial data repositories and systems provide a large amount of data regarding hiking paths, there is still large room for improvement in terms of risk & difficulty information. From an economical point of

[11] https://www.schweizmobil.ch

view, the large number of mountain hikers encouraged more investments in maintaining and promoting hiking sites, which in turn produced a remarkable return of interest [21, 20]. From the prevention and healthcare perspective, the need for improving existing guidelines and their accessibility has attracted the attention of health institutions and associations [9, 12, 15], who promoted initiatives to raise awareness about hiking trail difficulties [13]. Studies assessing difficulty in terms of energy expenditure [11], hikers' preparedness [16], and collaborative annotation of tourism objects[12] can be considered pioneering initiatives. More recent contributions span from preference-based recommendations for groups of users [8] to more technological ones employing recommender chatbots [17].

However, providing suggestions for hiking trails has only been addressed partially. For example, some recommendation systems rely on hiking time estimation [2, 19] or on questionnaires to estimate the users' physical condition [6]. On the one hand, although some studies introduced elements relevant for evaluating track difficulties, all of them overlooked several factors included in Syris. On the other hand, several standards for representation and modeling of tourism information have gained adoption. For example, Open Travel Schema developed by the Open Travel Alliance (OTA)[13] is a standard model for travel objects. However, it mostly focuses on booking information and neglects to current ontology modeling standards. The WTO (World Tourism Organization) [18] developed the Thesaurus on Tourism and Leisure Activities, used mainly for indexing concepts in the tourism realm. Similarly, a set of standards, including different messaging and tourism agent specifications, has been provided by the Travel Technology Initiative[14]. Solely focusing on booking operations and availability, and less on outdoor activities, such standards can guide the integration of tourism information. Nevertheless, the more significant limitation is the lack of data description models for pedestrian roads and the different aspects tackled in Syris (e.g., description and specification of difficulty types). As we will see in the next section, the data made available through Syris considers metadata as a key point in order to maximize the understandability and re-usability of the hiking trails difficulty data.

4 Implementation of the FAIR principles

The process of making the Syris data available as an open dataset adopts the FAIR principles: Findable, Accessible, Interoperable and Reusable. We have made data available through the well known open data repository Zenodo[15], and indexed in

[12] https://www.apidae-tourisme.com/
[13] https://opentravel.org
[14] https://www.tti.org/
[15] https://zenodo.org

the OpenAIRE[16] platform. In the following, we explain how each of these principles is addressed.

4.1 Findable

As explained in the FAIR principles: *The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.* In order to comply with this principle we have modeled the data and used well structured metadata, making it easier to describe, index and search the data. More precisely, the Findable principle is divided in the following four aspects for which we provide further details.

F1. (Meta)data are assigned a globally unique and persistent identifier In order to comply to this principle the whole dataset has a public unique identifier, i.e. a DOI provided by Zenodo:

DOI: 10.5281/zenodo.5527030

The DOI can be de-referenced through its URL: `https://doi.org/10.5281/zenodo.5527030`.

Moreover, each element in the dataset is provided with its own unique identifier. For instance a route with the ID 2886 has the following unique IRI (Internationalized Resource Identifier):

```
<https://hevs.ch/santour/anniviers/route_2886>
```

Listing 3: IRI.

Similarly, each point of difficulty has its own IRI:

```
<https://hevs.ch/santour/anniviers/pod_189>
```

Listing 4: IRI.

F2. Data are described with rich metadata Rich and semantic metadata is used to describe the Syris dataset. This metadata is based on the very well known and widely used schema.org vocabulary[17]. This vocabulary is used in major websites and is used by Web crawlers for optimizing data search. For instance, each route and each point of difficulty are also instances of a schema:Place according to schema.org. The schema indicates that a Place can have a geographic location, specified as a schema:GeoShape. It also allows to establish containment relationships (e.g. between a point of difficulty and a route), as well as general parameters such as description, title, etc.

[16] https://www.openaire.eu/

[17] `https://schema.org`

Regarding the metadata of the dataset itself (e.g. authors, version, release date, institution, etc.), Zenodo is compliant with the DataCite Metadata[18] schema. See Listing 5 for an excerpt of this metadata.

```
{
  "inLanguage": {
    "alternateName": "fra", "@type": "Language", "name": "French"
  },
  "description": "<p>The popularity of hiking has steadily increased in the latest years, across different
        segments of the population. <br>\nAlthough there is ...,"
  "creator": [
    {
      "affiliation": "University of Applied Sciences and Arts Western Switzerland HES-SO",
      "@id": "https://orcid.org/0000-0002-0364-6945",
      "@type": "Person", "name": "Calbimonte, Jean-Paul"
    },
    {
      "affiliation": "University of Applied Sciences and Arts Western Switzerland HES-SO",
      "@type": "Person", "name": "Cotting, Alexandre"
    }
  ],
  "url": "https://zenodo.org/record/5527030",
  "datePublished": "2021-09-24",
  "version": "0.0.1",
  "keywords": [ "Point of difficulty", "Hiking", "Pedestrian tourism", "Risk management" ],
  "@context": "https://schema.org/",
  "distribution": [
    {
      "contentUrl": "https://zenodo.org/api/files/c964b774-2bc7-4941-9000-1dd6ce7122ba/syris-routes.ttl",
      "encodingFormat": "ttl", "@type": "DataDownload"
    }
  ],
  "identifier": "https://doi.org/10.5281/zenodo.5527030",
  "@id": "https://doi.org/10.5281/zenodo.5527030",
  "@type": "Dataset",
  "name": "The Syris Dataset of Risk and Difficulty Points in Hiking Routes"
}
```

Listing 5: IRI.

F3. Metadata clearly and explicitly include the identifier of the data they describe As seen in the metadata in Listing 5, the DOI identifier of the dataset is explicitly included.

F4. (Meta)data are registered or indexed in a searchable resource The Syris metadata is indexed in OpenAIRE, as seen in Figure 7, which makes this data available for major research database crawlers and engines worldwide.

4.2 Accessible

As stated in the FAIR principles, accessible data refers to the following: *Once the user finds the required data, she/he/they need to know how can they be accessed,*

[18] https://schema.datacite.org/

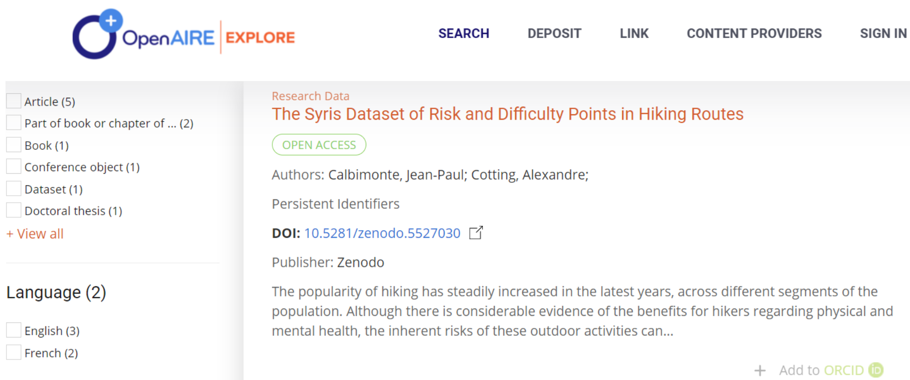

Figure 7: OpenAIRE.

possibly including authentication and authorisation. In this regard, thanks to the indexing of the dataset in Zenodo, we guarantee that data access protocols and standardized access is provided.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol The dataset records are accessible using the OAI-PMH protocol, provided by Zenodo. The metadata is also available through a REST API. In particular:

A1.1 *The protocol is open, free, and universally implementable.* Both OAI-PMH and REST are open, free and universal protocols.

A1.2 *The protocol allows for an authentication and authorisation procedure, where necessary.* Metadata are available in the public domain through Zenodo, and authorization is not necessary.

A2. Metadata are accessible, even when the data are no longer available Zenodo is hosted by CERN, and it guarantees the lifetime of the repository for the lifetime of the institution. It is also harvested in OpenAIRE, which holds the metadata in its Knowledge Graph, and will continue in the long term.

4.3 Interoperable

Regarding interoperability, Zenodo offers several key features that the Syris dataset profits from. The FAIR principles refer to this point as follows: *The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.* More precisely we address this point as follows:

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. Zenodo internally uses JSON schema, and allows exporting the dataset metadata to several other standards, such as Dublin Core, MARC XML, or JSON-LD, as shown previously in Listing 5.

Regarding the data itself, it uses schema.org as its main data model, with a semantically rich and machine interpretable data representation. As an example, in Listing 6 we present an extract of this model.

```
@prefix anniviers: <https://hevs.ch/santour/anniviers/> .
@prefix qudt: <http://qudt.org/vocab/unit/> .
@prefix santour: <https://hevs.ch/ontology/santour#> .
@prefix schema: <https://schema.org/> .

anniviers:route_2886 a santour:Route,
        schema:Place ;
    schema:identifier "2886" ;
    schema:name "Fang - Vissoie par le chemin historique n5" ;
    schema:description "Cette randonnÃ©e est l'occasion de dÃ©couvrir le petit village de Fang, vÃ©ritable joyau du Val d'Anniviers." ;
    schema:url <https://www.outdooractive.com/fr/route/randonnee/sierre/fang-vissoie-par-le-chemin-historique-n-5/38445048/> ;
    santour:hasPointOfDifficulty anniviers:pod_189, anniviers:pod_193,
        anniviers:pod_194, anniviers:pod_195 ;
    schema:amenityFeature [ schema:propertyID santour:effort ; schema:value 2 ],
        [ schema:propertyID santour:risk ; schema:value 2 ],
        [ schema:propertyID santour:ibpIndex ; schema:value 35 ],
        [ schema:propertyID santour:technique ; schema:value 1 ],
        [ schema:propertyID santour:ascent ; schema:unitCode qudt:M ; schema:value 271 ],
        [ schema:propertyID santour:descent ; schema:unitCode qudt:M ; schema:value 122 ];
    schema:geo [ a schema:GeoShape ;
            schema:line "7.597375 46.220503, 7.597519 46.220537, 7.597507 46.220571, 7.597457 46.220631, 7.597315 46.220718, 7.597304 46.220765, 7.597403 46.220899, 7.597502 46.221112, 7.597548 ... " ] ;
    schema:potentialAction [ a schema:ExerciseAction ;
            schema:distance [ schema:propertyID schema:Distance ;
                    schema:unitCode qudt:KiloM ; schema:value 3.5 ] ;
            schema:exercisePlan [ schema:activityDuration "P0DT1H30M0S" ] ] .
```

Listing 6: Data of a hiking route from Syris in GeoJSON format.

I2. (Meta)data use vocabularies that follow FAIR principles The metadata managed by Zenodo use FAIR-compliant vocabularies, as mentioned before. Regarding the vocabularies used for the data, schema.org is used, which is also findable, accessible, interoperable and reusable.

I3. (Meta)data include qualified references to other (meta)data Given that the data is described in RDF, it natively supports referencing other data through universal identifiers (in this case, IRIs). As we showed in previous examples, it is possible to cross-reference points of difficulty and routes through their IRIs.

4.4 Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

- *R1.1. (Meta)data are released with a clear and accessible data usage license*. The data has been released under a Creative Commons Attribution 4.0 license [19]. This allows reuse of the data by third parties, with the sole condition of a proper citation of the dataset and its authors.
- *R1.2. (Meta)data are associated with detailed provenance*. The semantic metadata embedded within the dataset provides details about its origin, where it was produced and when it was added.
- *R1.3. (Meta)data meet domain-relevant community standards*. We have followed the widely known schema.org standard that is used for major touristic information websites throughout the world.

5 Ethical and legal aspects

Regarding the ethical and legal aspects for making this dataset openly available, we have taken the following considerations. First, we analyzed the aspects related to confidentiality, privacy and sensitiveness of the data. In principle, data about the itineraries and their difficulties do not include sensitive or private data, so at this level there were no barriers or need to anonymize the information. However, all data related to the participants who contributed the data was completely excluded from the Open Data process. It is not relevant nor appropriate to include detailed information of who assessed a point of difficulty or a specific trail. However, in the future we may need to include different levels of assessment expertise, even if we completely hide any identity-related data of the people who perform the evaluation. Second, regarding the acquisition of the data itself, each participant has agreed to the publication of the data. In the future, a more elaborated process of data acquisition consent may be required, if this process is expanded towards a massive citizen science approach. Third, the pictures may contain metadata that might have contain personal information to a certain degree (e.g. camera specifications, etc.). If pictures are uploaded into Zenodo they should be stripped of any

[19] https://creativecommons.org/licenses/by/4.0/legalcode

metadata that may include personal or identifying information. Fourth, regarding the license, as stated in the previous section, we opted for a Creative Commons Attribution license. This opens the way for reuse in a very large scope, only requiring proper citation of authorship.

6 Discussion

This report presents a novel open dataset that provides information for assessing points of difficulty in hiking trails in the region of Anniviers, considering different dimensions related to effort, technique and risk. The implementation of the data model materialized in this dataset opens the door for providing recommendations based on the matching of the health profile of a user with the characteristics of available routes. The Syris prototype and this dataset include several important steps in the data life-cycle of pedestrian and mountain routes, from data modelling, acquisition, storage, exploitation, and validation. In this regard, it has the potential for empowering different types of users, according to their needs. For example, regional tourist offices and local development authorities can use it as a basis for tailored recommendations for tourists, reaching segments of the population such as older adults and people with disabilities. Health professionals can also use the Syris dataset for helping patients to choose a trail according to physical activities that are appropriate to their profile. For mountain guides, it can help as a tool for assessing and validating difficulty points, and choosing the best itinerary for a group of customers. Moreover, the Syris data can be used as an instrument of prevention, with the indirect target of decreasing the number of accidents thanks to a better understanding of risks and difficulties.

Although the Syris dataset was conceived fundamentally for the assessment of difficulties in pedestrian routes, it is easy to see that the concept can be expanded towards a Citizen Science approach for acquiring, managing and validating outdoor activity data. This concept may extend also to other sports such as mountain biking or even skiing, although with additional constraints. The data acquisition approach, which is currently implemented simply for difficulty points, can also be expanded to gather other types of relevant information, e.g., pollution indicators, weather-dependent risks, or difficulties according to age. In the future we plan to extend this dataset beyond the itineraries of Val d'Anniviers, in order to have a larger and more heterogeneous set of scenarios and point of difficulty.

References

[1] J. A. Baecke, J. Burema, and J. E. Frijters. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *The*

American journal of clinical nutrition, 36(5):936–942, 1982.

[2] M. Boerger. Hiking suggestions and planner, Apr. 1 2014. US Patent 8,688,374.

[3] F. Boller, M. Hunziker, M. Conedera, H. Elsasser, and P. Krebs. Fascinating remoteness: The dilemma of hiking tourism development in peripheral mountain areas. *Mountain Research and Development*, 30(4):320–332, 2010.

[4] J.-P. Calbimonte, S. Martin, D. Calvaresi, and A. Cotting. A platform for difficulty assessment and recommendation of hiking trails. In W. Wörndl, C. Koo, and J. L. Stienmetz, editors, *Information and Communication Technologies in Tourism 2021*, pages 109–122, Cham, 2021. Springer International Publishing.

[5] J.-P. Calbimonte, S. Martin, D. Calvaresi, N. Zappelaz, and A. Cotting. Semantic data models for hiking trail difficulty assessment. In *Information and Communication Technologies in Tourism 2020*, pages 295–306. Springer, 2020.

[6] J.-P. Calbimonte, N. Zappellaz, E. Hébert, M. Simon, N. Délétroz, R. Hilfiker, and A. Cotting. Santour: Towards personalized recommendation of hiking trails to health profiles. In *International Conference on Web Engineering*, pages 238–250. Springer, 2018.

[7] L. Carvalhinho, P. Rosa, and F. Gomes. Hiking trails evaluation in the natural park of serras de aire e candeeiros, portugal. *European Journal of Tourism, Hospitality and Recreation*, 6(2):139–156, 2015.

[8] A. Delic and J. Neidhardt. A comprehensive approach to group recommendations in the travel and tourism domain. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*, pages 11–16. ACM, 2017.

[9] W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, and A. Bauman. Physical activity and public health: updated recommendation for adults from the american college of sports medicine and the american heart association. *Circulation*, 116(9):1081, 2007.

[10] T. A. Heberlein, P. Fredman, and T. Vuorio. Current tourism patterns in the swedish mountain region. *Mountain Research and Development*, 22(2):142–150, 2002.

[11] M. L. Hugo. Energy equivalent as a measure of the difficulty rating of hiking trails. *Tourism Geographies*, 1(3):358–373, 1999.

[12] J. Jackson and P. Murphy. Tourism destinations as clusters: Analytical experiences from the new world. *Tourism and hospitality research*, 4(1):36–52, 2002.

[13] K. V. Kortenkamp, C. F. Moore, D. P. Sheridan, and E. S. Ahrens. No hiking beyond this point! hiking risk prevention recommendations in peer-reviewed literature. *Journal of outdoor recreation and tourism*, 20:67–76, 2017.

[14] S. Y. Lee, C. Du, Z. Chen, H. Wu, K. Guan, Y. Liu, Y. Cui, W. Li, Q. Fan, and W. Liao. Assessing safety and suitability of old trails for hiking using ground and drone surveys. *ISPRS International Journal of Geo-Information*, 9(4):221, 2020.

[15] D. Lloyd-Jones, R. Adams, T. Brown, et al. Health benefits of hiking. *Circulation*, 121:e1–e170, 2010.

[16] R. C. Mason, S. Suner, and K. A. Williams. An analysis of hiker preparedness: a survey of hiker habits in new hampshire. *Wilderness & environmental medicine*, 24(3):221–227, 2013.

[17] T. N. Nguyen and F. Ricci. A chat-based group recommender system for tourism. In *Information and Communication Technologies in Tourism 2017*, pages 17–30. Springer, 2017.

[18] W. T. Organization. *Thesaurus on Tourism and Leisure Activities*. UNWTO, 2001.

[19] A. Pitman, M. Zanker, J. Gamper, and P. Andritsos. Individualized hiking time estimation. In *2012 23rd International Workshop on Database and Expert Systems Applications*, pages 101–105. IEEE, 2012.

[20] R. Scarpa, M. Thiene, and T. Tempesta. Latent class count models of total visitation demand: days out hiking in the eastern alps. *Environmental and Resource Economics*, 38(4):447–460, 2007.

[21] M. Thiene and R. Scarpa. Hiking in the alps: exploring substitution patterns of hiking destinations. *Tourism economics*, 14(2):263–282, 2008.