KarstLink : organise the sharing of karst data

Eric MADELAINE (1), Peter MATTHEWS (2) & Frédéric URIEN (3)

- (1) Comité Départemental de Spéléologie des Alpes Maritimes, France (eric.madelaine@inria.fr)
- (2) President of the UIS Informatics Commission (http://uisic.uis-speleo.org/)
- (3) President of the Wikicaves association (president@wikicaves.org)

Abstract

You've probably wondered how to find cave data? How to relate the caves to the documents which mention them? How to create links between the scientific observations, the measurements carried out and the caves in which these observations and measurements were carried out?

This is what the KarstLink project plan to offer you. The organization set up by the UIS has made it possible to mobilize skills in many countries and to advance this question, which was raised at the UIS Congress in Sydney in 2017. Using the tools of the semantic web, an ontology was developed. This common vocabulary makes it possible to share data without changing the structure of existing databases, thanks to tools that have been set up by various partner structures. The 2021 congress is an opportunity to check the progress of this project, to present a prototype for the interconnection of several databases, and to present the links with other subjects that mobilize people interested in the underground environment. We will also indicate the steps we want to take together in the forthcoming years.

Résumé

KarstLink : organiser le partage des données liées au karst. Vous vous êtes sans doute demandé comment trouver des données sur les cavités ? Comment relier les cavités aux documents qui les mentionnent ? Comment créer des liens entre les observations scientifiques, les mesures effectuées et les cavités dans lesquelles ces observations et mesures ont été réalisées ?

C'est ce que vous propose le projet KarstLink. L'organisation mise en place par l'UIS a permis de mobiliser des compétences dans de nombreux pays et de faire avancer cette problématique qui avait été évoquée au congrès UIS de Sydney en 2017. En utilisant les outils du web sémantique une ontologie a été élaborée. Ce vocabulaire commun rend possible le partage de données sans changer la structure des bases de données existantes, grâce à des outils qui ont été mis en place par différentes structures partenaires. Le congrès 2021 est l'occasion de faire le point sur l'avancement de ce projet, de présenter un prototype d'interconnexion de plusieurs bases de données, et de présenter les liens avec les autres sujets qui mobilisent les personnes intéressées par le milieu souterrain. Nous indiquerons également les étapes que nous voulons franchir ensemble dans les années à venir.

1. Motivation

Data sharing is a problem that has interested scientists for many years. Researchers working on questions related to karst are naturally concerned by this subject which is taking an increasingly central role in their daily work.

In the field of caving databases, the tools we have today are much dispersed; there are many tools built using different technologies, covering various territories, and even in the same territory different sources of information. We are talking about cavity databases here, but also documentary databases, or bio-speleological databases, for example. Often, too, database developers find easier to incorporate all types of information into the same database, eventually duplicating data through different tools. This increases individual tool development efforts,

2. Sydney: already 4 years ago

Jason Boczar and George Veni (Boczar and Veni 2017) took part in the Sydney congress to present the Karst Information Portal. They presented the path that should be but also the risk of incoherencies, and difficulties to maintain evolution of data through several tools. Solutions based on modular databases and linked data would be much better.

Beyond developing databases, their access by users is done through very diverse methods and formats, both for feeding data in the databases and for searching existing data through several different databases.

We think that users (both cavers and researchers) would be very interested to have compatible interfaces with several databases, and to send complex requests to many different databases. It is therefore essential to provide tools that provide a link between the producers of information and those who wish to use it.

followed to use the tools of the Semantic Web and to organize the sharing of karst-related data on a very large scale.

In France, the FFS databases working group started debating about WC3 data exchange technologies more or less at the same time. Several other experiments appeared in publications (see the bibliography section below).

Discussions on the subject became more intense in 2019. After obtaining the support of the European Federation of Speleology and the French Federation of Speleology which participated in the promotion of KarstLink, the operations were launched on November 23, 2019 during a meeting of the FFS scientific commission in Courthézon.

A Sub-Commission was set up in January 2020 under the aegis of the UIS Informatics Commission both to support

3. The semantic web

The semantic web or web of data is a set of tools standardized by the World Wide Web Consortium (W3C) that allows data available on heterogeneous resources to be linked, without having to modify the technical infrastructure that makes this data available.

The foundation stone for organizing this information sharing is what is called an ontology. This is like a machinereadable dictionary defining the meaning of the terms of interest and also any relations to other terms. In order for the exchange to be possible, each of these elements is associated with a keyword, a definition that allows humans to fully understand the role that each of these keywords plays. Often ontologies are multilingual, allowing uniform access to data in different languages.

The second stone is a server that provides access to the ontology, either as a web page for us to read, or as a machine-readable file for servers that want to use the ontology directly.

The next step is to set up servers that are able to collect and deliver data that has been made available by reference to the KarstLink ontology. Everyone can naturally implement and provide a server, and a web portal, but it is essential that there is at least one.

Finally, the structures which have data to share must format them so that the exposed files are "translated" according to the rules set by KarstLink.

Figure 1 shows a simplified view of the KarstLink principles using semantic web concepts: at the toplevel, a database or a human, through a portal, will send requests to one or several KarstLink-enabled databases. These can be cave databases (in grey here), or document DBs (orange), biological DBs (green), etc.

3. Project organisation

There are three guidelines that we have set for building the ontology; they have served as a common thread since the launch of the project:

- All organizations wishing to provide data using KarstLink ontology must be able to find the concepts meeting their needs. It is for this reason that we have chosen to use entities that correspond to the most general concepts the discussion and to provide the technical infrastructure necessary to carry the project. It is managed by the authors of this article and thanks to Peter Matthews it has an evocative and sympathetic name: KarstLink.

We have decided to set ourselves the goal of offering, at the UIS 2021 congress, 4 years after that of Sydney, an infrastructure, tools and functional examples. This prototype should serve as a demonstration to attract other players, and a solid base to enrich the project.



Figure 1 : KarstLink: Our dream

Both the requests, and the databases interfaces, will make use of Entities defined by ontologies, either ontologies from the KarstLink project (first line), or already existing ontologies (second line).

possible. Thus we propose an Under-groundCavity object which must be able to describe a cavity whatever the definition that this cavity has for the group that wishes to share its work. It can be a rock shelter, a spring, a mega sinkhole or any kind of natural or artificial cave.

- It was also necessary to take into account the organization of the data which is not always identical in all

the databases. So often the data corresponding to the cavities is stored in a dedicated table, but there are also web services that rely on a data organization where the cavity and the entrance to the cavity are 2 different entities. The model that was built makes it possible to have these 2 types of data organization.

- It was also important to find the broadest possible consensus, for this we organized a major information campaign. Thanks to the network of contacts of the UIS, the FSE and the FFS, we were able to reach many of those who might be interested in KarstLink. A large number of other federations and structures then promoted KarstLink and we are very grateful for this very warm movement. A Wiki has made it possible to collectively build the ontology, to share resources. A mailing list has been set up by the UIS, it brings together more than fifty people who were able to participate in a vote organized from July 19 to 26, 2020 and which fixed the first elements that appear in the ontology. The choice was also made to rely whenever possible on already existing ontologies very widely used by other projects. For example DCMI (Dublin Core) was chosen for documents, Darwin Core for biospeleological observations, Geo for points and geographic coordinates, Geonames for geographic identities and Schema for postal addresses. The number of Classes and relations created specifically within the framework of KarstLink is therefore as limited as possible and corresponds, for the most part, to only the karst-related parts of our project.

Integrating into KarstLink all the elements allowing us to describe a cave, an underground river, a Paleolithic deposit, a colony of bats or a speleothem is a very long process. We have chosen to initially work only on the basic entities and to select only the essential elements for each of these entities. So we think we can meet the needs expressed by the greatest number of databases creators and the fact of working on a limited number of elements has given us the time to work in depth on the best solutions for this kernel. An ontology is a living object and KarstLink will necessarily expand later to allow new possibilities. This will be done in view of the needs or difficulties that will arise after a few months of actual use.

4. KarstLink at your service

Gouffre Jean Bernard		
name		Gouffre Jean Bernard
alternate name		Jean Bernard (Réseau)
contained in place		
latitude		46.1030
longitude		6.78
altitude		1600
coordonates precision		-1
length		25819
extent above entrance		
extent below entrance		
vertical extent		1625
country code		FR
discovered by		Groupe Spéléologique des Vulcains
Josiane Lips		
Para and a second second		
hrst name	Josiane	
first name last name	Josiane	
first name last name nickname	Josiane Lips Jo	
hrst name last name nickname countryCode	Josiane Lips Jo FR	
first name last name nickname countryCode visited	Josiane Lips Jo FR Gouffre Jean Bernard V4	
hrst name last name nickname countryCode visited member	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vulca	ins Fédération Française de Spéléologie
hrst name last name nickname countryCode visited member rights	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vulca attribution name	ins Fédération Française de Spéédologie Frédéric Unien
Inst name last name nickname countyCode visited member rights	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vulca attribution name Elicence type	Ins Fédération Française de Spélédologie Frédéric Urien ODC-By
Inst name last name nickname countryCode visited member rights	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vulca attribution name licence type created	Ins Fédération Française de Spéléologie Frédéric Unien ODC By 10/23/2020
Inst name last name nickname countyCode visited member rights	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vuics attribution name licence type created modified	Int Fédération Française de Spéédologie Frédéric Unien ODCey 10/23/2020 10/30/2020
Inst name Iast name idxname countryCode visited member rights	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vulca attribution name licence type created modified attribution URL	Int Fédération Française de Spéléologie Frédéric Urien ODCey 10/23/2020 10/30/2020 https://ortology.uis-spéleo.org/example/
Inst name last name nickname countryCode visited member rights	Josiane Lips Jo FR Gouffre Jean Bernard V4 Groupe Spéléologique des Vuico attribution name licence type created modified attribution URL	ins Fédération Française de Spéléologie Frédéric Urien ODC&y 10/23/2020 10/30/2020 https://ontology.uis-spéleo.org/ecample/

Figure 2 : Exemple using the core KarstLink Ontology

The ontology directory at *https://ontology.uis-speleo.org* gives access to the ontology with an xhtml file for human beings, the server having been configured to refer to the ontology which is described in a ttl (Turtle format) file

when the request comes from a machine which expects this type of response.

We provide an example currently containing cave, organisation and person entities in various formats (html, csv, ttl, rdf, n3. It will be extended with all other entities approved by Karstling, e.g., points, areas, documents or biologicalobservations.

In Figure 2, we show the example in a human readable form. It shows an excerpt of a cave entity (The Gouffre Jean-Bernard in the Alps) with some of its related data. In red or blue boxes you can access related entities, e.g. here one of the Persons involved in the exploration, and the organization they belong to. All entities may include a "rights" object specifying the relevant license.

In order to collect the data and make it available to the community, a server has been set up by the Wikicaves association which publishes the site grottocenter.org. It is based on the Semantic-Forms software, developed by Jean Marc Vanel who was involved in the project and brought all his skills in the field of semantic web. This server will be searchable from the grottocenter.org site but it can also be searchable from any site using the APIs that we make available.

The precise list of available entities is evolving fast in this starting phase of the project, please refer to the KarstLink wiki if you are interested.

To date the APIs of grottocenter.org have been modified in order to respect the JSON-LD formalism which is one of the formats of the semantic web.

5. Conclusion

In less than two years, thanks to the involvement of UIS, FSE, FFS, and many other structures, KarstLink was able to achieve its goal of being present at the 2021 UIS congress.

A large community showed interest and followed the progress of the work but the real involvement in the project was not as great as we had hoped. The difficulty that we had not measured by embarking on the adventure is that the concepts associated with the semantic web are far from being mastered by a very large majority of speleologists, or even database developers, even if they are present in a very large number of sites and more particularly those of the large institutions offering scientific resources. Clearly we will need to implement one of our original goals which was to produce documentation, tools, on-line tutorials and a go-to enquiry/advice service to assist people to take advantage of our new facilities.

Beyond this understanding of general concepts, it will be necessary to form a larger group of computer scientists, of people with the technical skills to assemble new bricks and bring KarstLink to life, which has already made a very good start. The chosen organization and the tools put in place, in particular with freely accessible code on the Github platform, make it possible to ensure this development. Another aspect important to allowing full scientific use of shared data is that of data quality and reliability. Receivers of data need to know its provenance. We aim to follow the guidance of the <u>World Data</u> System of the International Science Council for this aspect of trusted data.

Beyond the UIS 2021 congress, we will have to unite new players, interface other databases with the project, but also extend our ontology and our capabilities to other related fields. Some already have existing databases, some are accessible with semantic web formats (sensors and measurements, meteorology, geology, etc.). Others are yet to be invented or specialized, but are of great interest to some of us, for example in the field of topographic databases, or specific data sensors and observations (temperature, pressure, flow rates, gas, etc.).

For most people KarstLink will become a practical and easy-to-use tool for providing or researching data, and that is the key.

Acknowledgments

We warmly thank the UIS for its support, both in terms of technical infrastructure, and for providing us with worldwide visibility and connections. We also thank all contributors and participants to the KarstLink mailist and wiki.

References

Jason Boczar, George Veni, Proceedings of the 17th International Congress of Speleology p. 389-391 Linked Data, the Semantic Web, and the Karst Information Portal, 2017 https://uis-speleo.org/wp-content/uploads/2020/08/Proceedings-17th-ICS-Vol2-Ed2-screen.pdf

Fagnoni E., 2018. Linked data for libraries. Proceedings of the 3rd Int. Symposium of Speleology, Varenna (Como, Italy), April, 29-30th, 2017. Memorie dell'Istituto Italiano di Speleologia, s. 2, 33: 115-117.

Ferrari G. W., 2018. Cave-O-Zines: a prototype virtual library of Italian cave journals. Proceedings of the 3rd Int. Symposium of Speleology, Varenna (Como, Italy), April, 29-30th, 2017. Memorie dell'Istituto Italiano di Speleologia, s. 2, 33: 111-114. <u>https://www.researchgate.net/publication/329684644_CAVE-O-ZINES_A_Prototype_Virtual Library of Italian Cave Journals</u>

D. Wood et al, Linked Data – Structured data on the Web, 2013 - <u>https://www.manning.com/books/linked-data</u> (French): Online course on the semantic web: <u>https://www.fun-mooc.fr/courses/course-v1:inria+41002+self-paced/about</u>

KarstLink project page: <u>http://uisic.uis-speleo.org/exchange/KarstLink/index-en.html</u> Version française: <u>http://uisic.uis-speleo.org/exchange/KarstLink/index-fr.html</u> Wiki home page : <u>http://uisic.uis-speleo.org/wiki/KarstLink/index.php</u> KarstLink ontology: <u>http://ontology.uis-speleo.org</u>