A PROPOSAL FOR THE REGISTRY

ABSTRACT

This is a proposal for the National Cave, Karst and Mines Registry. The aims and design are covered. The initial aim is to produce a list of all the sites of interest. The idea is that the list will act as an index to the other sources of information such as the Local Registries or other speleological databases. Giving each site a public unique identifier and agreeing to use that identifier in each database will achieve this. Your comments and criticisms are invited.

INTRODUCTION

To create the finest possible National Cave, Karst and Mines Registry (The Registry) I need your views and comments. To this end I'm borrowing an idea from the early development of the Internet and issuing this paper as a Request For Comment or RFC for short. Those comments will be used to produce future more refined and detailed RFC's, which will effectively become a specification of The Registry. I hope this approach will encourage a wide discussion and involvement so that the best system can be developed.

This paper outlines my personal view of how The Registry should work. They have been arrived at after some consultation and considerable thought on my part drawing on my experience designing business systems and databases. Although I'm normally considering invoices or stock records, the same principles apply. This is a system. It is more than deciding what information to store. Thought must also be given to how information flows around the system, how it is collected, added, updated or retrieved, who is responsible for each task, security, etc. Now is the time for wider consultation. Some of the ideas may be controversial, but that is the recipe for a stimulating discussion …

This paper can be found at http://www.cave-registry.org.uk. The website also has details of how to join The Registry mailing list so that you can take part in the debate. I would also be glad to receive your comments by email at coordinator@cave-registry.org.uk or by post at The Cave Registry Co-ordinator, 3 Starrs Close, Axbridge, Somerset, BS26 2BZ.

THE AIMS

The aims that I have set for The Registry are as follows.

1. To promote and support the systematic collection, recording and publication of information about the subterranean world.

2. To maintain a comprehensive list of all sites within the British Isles of interest to cave and mine explorers.

3. To act as an index or link into the other sources of information, regional or otherwise.

4. To record other data about sites as deemed useful and practical.

5. To make The Registry's information available as widely as possible with due regard to the intellectual property rights of the information providers and possible abuses of the information by commercial organisations or otherwise.

6. To set and adopt standards in conjunction with other similar organisations.

7. To recognise that a single centralised system is impracticable and that a distributed system of collection and recording is essential.

Initially I want to concentrate on the second aim, which is to produce a list of all the sites of interest. This implies I want to hold little more data than is necessary to identify a particular site. I want to leave the forth aim, which is to expand the amount of information held on each site, for the future. The reason for this is to keep it simple; to learn to walk before we run. Although having said that, 'just' producing a simple list will require all the major administrative and technical problems to be solved.

The third aim is achieved by giving each site a single unique identifier and agreeing to use that identifier in each database. The unique name can be used to relate the information in the different databases together. The National Registry is ideally placed to assign the unique identifier. This is the National Registries contribution to the community. This is what sets the National Registry apart from all others registries.

When contemplating a National Registry it would be tempting to design a monolithic centralised database that tries to be all things to all men. Aim seven is there to avoid this trap. The National Registry will be designed to be at the heart of a network databases with most of the data held in distributed local databases. This approach has several advantages. The work is shared between many rather than a few people. New local databases can be incorporated into the network with a minimum amount of effort. A network is more robust and safer than a

centrally held database. Local Registries retain and therefore control the use of their more sensitive data. Local databases are free to use their own proprietary software and design. Advances in technology such as the Internet make this approach more practical than in the past.

THE DESIGN OF THE REGISTRY

The initial aim is to produce a list of all the sites of interest. A seemingly simple objective, but to achieve this all the major administrative and technical problems will have to be solved.

The idea is that the list will act as an index to the other sources of information such as the Local Registries or other speleological databases.

This implies I want to hold little more information than is necessary to find and identify a particular site. My initial list of fields would be:

- Unique Identifier, Name, Location, Short Description, Length, Depth, and Altitude.

Other fields will be needed to manage the data. For example which registry submitted the record and is therefore permitted to amend it. There are also issues such as a cave having more than one name to be accounted for. But fundamentally I want to stick to the above list. Any extra fields will have to justify their existence. A precise specification of what information is to be recorded is a big topic in its own right that I will leave for another paper.

The information for the list is to be gathered from the various Local Registries from around the country into a central registry. Each site that the Local Registry submits will create an entry in a database. A fundamental design philosophy is that the Local Registries are in control of their submitted data. They can add, update or delete it as they wish. To do this each entry will need a unique identifier so that the correct entry can be found and altered. It is called the Local IDentifier or LID for short. The LID is assigned and is under the control of the Local Registry, with the golden rule that once assigned to a site it is never re-assigned.

It is quite possible that different registries will submit information about the same cave, which gives rise to a problem. Our list may have several entries for the same cave with similar but not identical information. One could attempt to solve this by amalgamating the information into one entry, but which registry would control the entry? Or one could insist that registries do not geographically overlap. Many countries do it this way but it is restrictive and doesn't recognise the situation that has developed in Britain. The solution is to have a second, nationally unique, Speleo-IDentifier or SID for short. A site can only have one SID but may have several LID's. The same golden rule applies to SID's as LID's; once assigned they can never be re-assigned. Reliably assigning LID's to SID's will require expertise and local knowledge. Each geographical area will need to have a single regional coordinator for this task.

Local Registries will be able to add records to the National Registry, once added they would be able to update or delete them as they wish. For this to work effectively the process needs to be automated. I envisage that a standard format text file will be submitted, which will then be processed by an import routine.

This format will need to be defined precisely. Initially I favour a form of comma separated variable (CSV) format. CSV is widely supported by many databases and spreadsheets. A version in XML format could be developed for the future.

Some consideration needs to be given to security so that only authorised people can submit text files.

The National Registry will be written using Microsoft's Access Database. The software is powerful enough for the task, it is not expensive and it is widely understood so finding the next co-ordinator with the appropriate skills should not be impossible.

LIDs AND SIDs

For my proposed design two kinds of unique identifiers are required. A locally assigned Local IDentifier (LID) and a nationally assigned Speleo-Identifier (SID). The identifiers must meet two requirements:

1. They must be globally unique. I.e. no other site in the world will have the same identifier.

2. They must be permanent. I.e. once assigned to a site it is never re-assigned.

It is also a good idea to assign them in a systematic way and to make them mnemonic, i.e. have some meaning to a human.

The format I propose for the LID is of the form '*GBOOO99999*', for example GBMEN00001.

Where:

GB is the 2 letter ISO country code. This would be GB = Great Britain.

OOO is the 3 letter organisation coded indicating the organisation which originally created the LID. For example MEN = Mendip Cave Registry. The organisation codes will need to be nationally agreed.

99999 is a serial number left padded with zeros and is unique within *GB* and *OOO*.

This format meets The International Union of Speleology (UIS) standard for creating record identifiers in their 'Proposal for Cave and Karst Data Exchange Standards'. Which can be viewed at http://rubens.unimelb.edu.au/~pgm/uisic/exchange/exchprop.html.

The format I propose for the SID is of the form '*GBCCAAA9999*' for example GBSOEME0001.

GB is the 2 letter ISO country code. This would be GB = Great Britain.

CC is the 2 letter county code indicating the county the site is in. For example SO = Somerset.

AAA is the 3 letter area code indicating the area the site is in. For example EME = East Mendip. The area codes will need to be nationally agreed.

9999 is a serial number left padded with zeros and is unique within *GB, CC* and *AAA*.

It is worth comparing this format with the similar Australian Speleological Federation standard for Numbering Codes. It can be viewed at http://www.caves.org.au/standards/set_standards.htm

The LID is designed for the internal use of a registry. It is less mnemonic but has the advantage it can be quickly assigned with no reference to the National Registry. A Local Registry simple numbers its first site 1, the second site 2 on so on.

The SID is designed for external or public use and carries more meaning, chiefly an indication of the area that it is in. The National Registry will assign SID's, although this might take some time, as checks must be made that a proposed new site is genuinely new and has not been assigned a SID already. Each geographical area will need to have a single regional coordinator responsible for doing this task.

If all the various speleological databases agree to use SIDs to identify their entries it becomes a simple task to search all those databases for information about a particular site, even if those databases are of different types and design. This simple idea is very powerful as it potentially allows a multitude of databases to be linked together.

PUBLICATION

Having described how information is to be put into a National Registry, some thought needs to be given to publishing it. Which means we must also consider copyright. The Registry will obtain permission from the copyright holder before any kind of publication. I'm using publication in the widest sense to mean whatever method, be it by paper, book, website, or even giving details out in a phone call.

With luck data will be submitted to the National Registry with copyright permission. I have some hope that this will be the case since so few fields are being recorded, most of which are in the public domain anyway. The only field that I think might cause a problem is the location field. A way around this would be to only allow publication of the location field to, say, the nearest kilometre.

At the moment I don't have a good feel for Local Registries opinion on copyright so I'll produce a paper on methods of publication after further consultation.

It will be a condition of the National Registry that you have the permission of the copyright holder for any data submitted to the National Registry.

CONCLUSION

I hope there will be broad agreement that the aims and design are good and that this proposal is the way forward even if some of the detail needs amending. I look forward to your comments and criticism.

David G Cooke
The Cave Registry Coordinator
coordinator@cave-registry.org.uk

GLOSSARY

Field	A field stores a single item of data, for example a sites name.
Record	A record is made up of related fields, for example a record could contain all the information about a particular site.
Table	A table is made up of records. Tables logically divide the information, for example there might be a table of sites and another table of information about authorised users.
Database	A database is made up of tables and contains all the information stored.
RFC	A publicly published paper about the system that invites Requests For Comment. The RFCs effectively document the system in a way that is open to all
LID	A Local IDentifier uniquely identifies a site. The LID is designed for the internal use of a Local Registry
SID	A Speleological IDentifier uniquely identifies a site. The SID is designed for external or public use. A site can only have one SID but may have several LIDs if information about it is kept on several Local Registries.
CSV, XML	Standard text file formats. Comma Separated Variable and Extensible Mark-up Language respectively.
National Registry	The British Caving Association's National Cave, Karst and Mines Registry
Local Registry	The Regional Registries or other speleological database that provide data for the National Registry.