The ASF's Web-based National Karst Index Database

Implementation & Implications

Michael Lake

ASF Conference, Bathurst
January 2000

Abstract

The ASF's Web-based National Karst Index Database
Mike Lake and Jill Rowling

In 1985 the ASF produced a landmark document; the *Australian Karst Index 1985* [1]. The 480 pages and coverage of around 6,500 caves in all states of Australia ranks this as an important reference work and an exemplary piece of work by the members of the ASF.

Since then, the Karst Index Database (hereafter referred to as KID) has been developed using the DOS-based Paradox database. Although it was basically functional, time has overtaken it in that few people are prepared to use the DOS-based system. Additionally setup can be tricky for users with newer MS Windows systems, the database can only be run by users with systems that support or emulate MS-DOS and users must obtain and install the software to run it.

Towards the end of 1999 Mike Lake and Jill Rowling put forward a proposal to the ASF to put the Karst Index Database on the Web, based on Open Source software. The general concept was accepted at the January ASF council meeting, and specifications refined over the year. In Spring 2000, tenders were drafted and circulated amongst the ASF Executive, the final request for tender was advertised nationally and a programmer was selected from many tenderers. A working system was developed within 2 weeks. The Web-based Karst Index Database was running on the ASF's Web site by December and ready for ASF's January Conference.

You can find the KID at ASF's Web site `http://www.caves.org.au`.

This paper covers:

- A brief history of the web-based KID and it's development
- Advantages of a Web-based KID
- Privacy and Security aspects
- Implementation and use of open source software
- Implications of open sourcing the KID software
- Where do we go from here?

A separate workshop was held at the conference to demonstrate the Web-based KID, how it can be used and to gauge feedback from cavers and speleologists.

1 History & Project Planning

A proposal to the ASF Executive in October 1999, by Michael Lake and Jill Rowling, ``Suggestions and Comments on the Karst Index Database" [5] suggested:

``...the Karst Index Database should be implemented as a Web-based open-source application. The use of open-source software would mean that the Karst Index Database is firmly based upon publically available, open standards thus ensuring future compatibility, lowest cost-of-ownership for the ASF and end users and the highest level of support and review. A Web-based approach to the user interface would allow controlled access to the Karst Database Information from all modern operating system platforms via a Web browser for ease of use, yet still allow sophisticated queries."

and...

``Both the program and the data should be freely available - the only charge being for the media as the 1985 Karst Index Book is today. The ASF should capitalise on the Karst Index Database through providing expertise in data mining, integration and interpretation services."

The ASF was supportive of this proposal and it was moved and carried at the January 2000 ASF Conference in Canberra. The ASF obtained its own domain, `caves.org.au`, in February 2000. This is in keeping with the domain names chosen by the National Speleological Society of USA (NSS) at `http://www.caves.org` and the British Cave Research Association (BCRA) at `http://www.caves.org.uk` There is no cost in obtaining or maintaining a `.org` as the management of these domains is done voluntarily.

A summary of the schedule indicates the speed at which development subsequently occurred:

- 29 January 2000, proposal accepted at Canberra ASF meeting and a budget of $5000 allowed.
- February, ASF applies for and granted the domain `caves.org.au`
- May-Sep, specifications developed and circulated
- September, hosting provider chosen from among several choices
- October, tenders called (closing date 3rd November 2000)
- November, programmer selected, contract signed
- 10th November, work begins
- 20th November, first working version
- December 2000 to February 2001, testing and warranty period

The programmer finally chosen was Rick Welykochy of Praxis Services Pty Ltd in Sydney [9]. Rick Welykochy actually produced a working prototype based on the tender specification, filled the data fields with dummy cave data, and supplied us with a username and password to try it out before the closing date of the tenders!

There were several main reasons why the ASF believed we would be able to accomplish this in the timeframe and why the database was finished on-budget and on-time.

- Specifications for the database table structures and data fields had already been defined by the work of Peter Matthews and the Informatics Commission of the Union International Speleologie [4]. In fact the database tables were created by the programmer writing a simple program which read in Peter Matthew's HTML documentation of the tables and fields and automatically created the database structures.

  We had to work out what queries we wanted to make, have some idea of what our user interface should look like, and how it should behave.

- The ASF had chosen to pay a programmer rather than relying on the voluntary efforts of ASF members.

- Today there is a considerable range of software tools that enable programmers to develop databases to cater for specific needs (such as a cave database) rather than using off-the-shelf products.
- Using open source software [7] the programmer could build our database using freely available software components - our programmer did not have to ``re-invent the wheel''. This results in faster development speed and lower production costs.

2 PC-based versus Web-based Database

2.1 PC-based Database

A PC-based KID system is one where the database and program resides on a personal computer. Everybody who wants to use the database installs a copy of it on their own PC. The installation is done by downloading the database from either a web site or by installing from some physical media such as floppy disk or CD-ROM. The data for the KID fits on one 1.44 MB floppy disk.

Advantages

- Standalone running of the database is possible. It can be run on a laptop in the field or on a cheap PC in a cavers hut.
- There is a perception of private ownership of the database. Some users prefer this.

Disadvantages

- There are security concerns with unauthorised copying of the database as the copy protection available to the ASF can be easily circumvented even by inexperienced persons. It just takes a little web searching. The unencrypted data and cracked program could then be redistributed to others.
- There is no audit trail of a copied floppy disk or CD-ROM.
- Duplication and distribution costs have to be borne by the users if physical media is used. Although a web site could used to distribute the data and program.
- Updating issues - who do we send updates to and how often?
- There is great difficulty in supporting several different platforms; we should really support at least the widely used Windows 98, Windows NT, Windows 2000, Linux and Macintosh. The cost in keeping up to date with compilers to produce executables for all these platforms is considerable. There are other, lesser used platforms, that we would not be able to support.

2.2 Web-based Database

A Web-based is one in which the program and data reside on a web server and the user accesses the information through a web browser.

Advantages

- There is only one database to update and maintain.
- Users don't have to install any software.
- A Web-based system can provide immediate access to more users than any other method.
- Overseas users can see the achievement of the ASF with our National KID.
- The ASF can remove access to any user at any time.
- There is an audit trail of access. This is the log file of the web server.
- It is a Federal offence to break into a computer system. This in itself discourages those persons that might know enough to be able to crack into the system. The reward versus the risk is simply too high.

Disadvantages

- Some people have security concerns about crackers breaking into the web server and gaining access to the data.
- The WWW is not accessible in some areas or to some persons so it discriminates against those that do not have web access at their home or work (note: all public libraries in Australia have web access).

3 Privacy and Security of a Web-based System

All web sites that store usernames, passwords and other information about users should clearly display a privacy statement that tells the user what information is kept about them on the server. The ASF's KID has such a privacy statement.

The Web-based KID;

- does not use cookies
- does not use Java, Javascript, Visual Basic Script or any other client-side scripting (ie. scripts that run on your own machine)
- only stores a users username and password
- only stores with your username search-related data such as your individual UIS allow fields list and your saved searches
- no personal data is stored.

Security of the karst index data had to be carefully considered as there is sensitive information stored in the database. Information such as the existence of aboriginal carvings and paintings, stone flakes, or human remains in particular caves is listed. Even species lists collected from caves may be considered to be sensitive data if that species is rare. For this reason several fields are in the database are not available to the general ASF membership or the public. ASF members, cave managers and speleological researchers with legitimate reasons can apply for access to this data.

The hosting provider's expertise, the operating system and web server used by the host and the way that access for the KID users and administrators is implemented all impinge on the security. Security considerations commenced with our choice of hosting provider and developer and, during development of the database, involved discussions with the programmer in the choice of scripting language, user access and authentication, encryption and many other aspects of the database and associated software. The hosting provider was consulted on a number of these issues.

In summary some of the salient points with regards to the security are:

- Security depends firstly on the system administrator and secondly on the server software
- There is no cave location information in the Web-based KI data
- The KI data is not in the ASF accessible directory on the web server
- Access is logged and traceable back to an ISP and in many cases back to an individual computer
- Breaking into a commercial web server is a federal offence
- Access down to individual fields of the database can be specified on a per user and a per State basis.

4 What did the ASF Get?

A Karst Index Database that puts Australia in the forefront of using modern information technology to assist ASF cavers, cave managers and speleological researchers.

4.1 Features of the Web-based KID

- ASF data compatible and UIS field compatible
- user administration and access control system
- security features appropriate for the data

- open source based
- well documented
- works with text based browsers - important for visually imparied users
- data import and export capability
- scalable to cope with many more caves than we will ever have

Ease of Use

The database is relatively easy to use and there is web based help available. Different types of users have been addressed by providing a range of pre-configured searches. These range from simple searches where one only needs to enter a cave area and a cave number or the ``top 10" search where the top ten longest and deepest caves are returned to the advanced search. The advanced search allows each field to be queried but requires some understanding of how the search queries the database for best results.

Web-based User Administration

To manage user access the KID has an excellent web-based user administration facility. This allows the KID administrators to create new users or delete users. They can modify a users access such as restricting information to particular States or to particular UIS fields.

This administration is done entirely though a web based form interface that is very easy to use. Administrators must log-on to the ASF site and after the password is authenticated they are presented with an Administrators page which is a simple table listing the users name and access information. Editing a user is done by clicking on the users name; a form is presented to the administrator with drop down lists to select some access options and multiple radio buttons to select UIS fields to allow or deny. Once the administrator makes any changes clicking the submit button will enact those changes.

Usability by Vision Imparied Users

One of the requirements for the developer was useability of the database from text based web browsers such as Lynx and w3m. Text based browsers are used by vision impaired users for web browsing and thinking about this from the start means that it is less likely that features are incorporated that would prevent such users from easily accessing the information in the KID.

5 What Runs It?

The software technology behind the KID comprises;

- A server running the GNU/Linux operating system (RedHat distribution)
- The Apache web server with mod_perl and blowfish encryption
- The Perl[1] programming language [8] and several CPAN[2] [2] modules
- Open source relational database interface modules developed by Praxis for rapid application development (see the advantages of open source!)
- A MySQL relational database [6]

6 Open Source?

The basic idea behind open source is simple. When programmers on the internet can read, redistribute, and modify the source for a piece of software, it evolves. People improve it, people adapt it, people fix bugs. The open-source community have learned that this rapid evolutionary process can produce better software than the traditional closed model that produces proprietary software. This process of peer review of open source software helps in finding any bugs, or as open source programmers like to say; *many eyes make all bugs shallow*.

Everybody who sends email or uses the Web is using open-source software. The running gears of the Internet (its mail transport, web and FTP servers, even the domain name system itself) are almost all open source. Scripting languages such as Perl, used by this database, are behind most `live' content on the Web. CGI scripts are an example; most are written in Perl.

Open-source:

- provides faster development speed
- provides lower software production costs
- open source authors frequently find themselves receiving, for free, improvements, bug fixes and additional functionality
- does not rely on ``Security through obscurity''
- open source promotes software reliability and quality by supporting *independent peer review* and rapid evolution of source code
- to be open-source certified, the software must be distributed under a license that guarantees the right to read, redistribute, modify, and use the software freely.

Open source code written by programmers is usually released under either the General Public License (GPL) [3], Artistic Licence or other similar license. The source code for the ASF's KID software is available as open source under the GNU General Public License.

6.1 GNU General Public License

This is commonly referred to as the GNU GPL or the GNU Copyleft [3]. A simple quote from the start of the GPL will provide an idea of the philosophy behind it.

``The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software-to make sure the software is free for all its users.''

In the GPL free software refers to freedom, not price. The GPL is a legal document designed to make sure that you have the freedom to distribute copies of free software, that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

6.2 Implications

The use of open source in developing the database and the decision to open source the code places some obligations upon the ASF. We need to provide access to the source code so that the ASF acts in the spirit of the GNU GPL, acknowledging our use of free software, and giving back to the open-source community.

It also puts obligations on those that use our code. Taking the ASF KID software and distributing it without providing access to source code or informing purchasers of where they can obtain source code is breaking the GPL.

Open sourcing also provides the ASF and it's members with opportunities:

- Other speleological organisations such as the BCRA, NSS, UIS etc. can use the ASF's Web-based KID as a basis for their own KID. This brings kudos to the ASF.
- If other countries wish to use our Web-based KID code the ASF may consider providing hosting for them. This saves them having to setup their own system and maintain it and could at least offset our maintenance costs.
- The ASF can charge money for consultancy services involving KI data mining, interpretation and integration. Others can do this too.

For further information on Open Source visit the Open Source Organisation's Web site [7], or have a look at Eric Raymond's on-line white papers, for example, ``The Cathedral and the Bazaar'' [10].

7 KID Access Policy

This is the access policy decided on at a meeting of the State Documentation Coordinators prior to the ASF meeting of all clubs and societies. This provides balance between reasonable access to information for users and the need to limit information for conservation reasons.

The ASF shall:

1. provide guest access (ie. a published username/password) to all ASF members, speleological researchers and cave management and members of the public. This access will be limited in scope.
2. provide additional Advanced Search functionality for all State Documentation Coordinators.
3. provide additional Advanced Search functionality for genuine speleological researchers and cave managers.
4. provide additional Advanced Search functionality for ASF Members who have a reasonable need for such information.
5. setup a small Web-based KID Access Group (3 persons) to represent user groups and to determine access for users.
6. decide who will have access to the ``User Administration" of the Web-based KID. Those persons will enact the decisions of the Web-based KID Access group.

Carried by the ASF Membership on 2nd January, 2001.

This access policy will provide reasonable access to the Web-based KID to:

- all ASF members as it was they, over many decades, who provided the data
- researchers who provide us with the knowledge which underlies much conservation
- the cave/land managers who trusted us with access.

8 The Next Stage

The current KI data is 15 years old and is in need of updating. The specification for the current KID was for a *read-only* Web-based system and there were a number of reasons for this. Firstly it was a significant step from a DOS based system to Web-based system. Secondly the KID is not a trivial database; although there are not many records in the database (approx. 6,500) the data fields, the database tables and the relationships are quite complex and we believed it was best to `walk' before we `ran'. The brief to the programmer though was that we would wish to provide updating facilities at a later date.

The means by which select users will be able to update the KID are now being considered. These users will be able to update the KID on-line and a mechanism also needs to be provided for those wishing to update information off-line and subsequently transfer that information to the KID. Mechanisms to keep the State Documentation Convener informed of updates and to back out of updates also needs to be addressed.

Bibliography

1 Australian Speleological Federation Inc., *Australian Karst Index 1985*, Edited by Peter G. Matthews ISBN 0 9588857 0 2.

2 CPAN, Comprehensive Perl Archive Network: `http://www.cpan.org`

3 GNU Copyleft license: `http://www.gnu.org/copyleft/gpl.html`

4 International Union of Speleology, Informatics Commission, Specifications for the Field Definitions for the KID. `http://rubens.unimelb.edu.au/~pgm/uisic/exchange/exchprop.html`

5 Lake, M. and Rowling, J., *Suggestions and Comments on the Karst Index Database*, 3rd October 1999., Proposal to the ASF Executive.

6 MySQL database: `http://www.mysql.com`

7 Open Source; Information on what is open source and how it can benefit developers and users can be found at: `http://www.opensource.org/`

8 Perl, Practical Extraction and Reporting Language: `http://www.perl.org`

9 Praxis Services Pty Ltd, Rick Welykochy (Director), email: rick@praxis.com.au

10 Raymond, Eric, *The Cathedral and the Bazaar*, http://www.tuxedo.org/~esr/writings/cathedral.html

Footnotes

... Perl[1]

Perl: Practical Extraction and Reporting Language

... CPAN[2]

CPAN: Comprehensive Perl Archive Network

Karst Index Database discussion Group Arthur Clarke

Joan Crabb at the Bat Party Arthur Clarke