Setting standards for global exchange of real-time data

Simon Flower,
British Geological Survey, West Mains Road, Edinburgh EH9 3LA

Abstract

The British Geological Survey is leading work with both the World Data Centres for geomagnetism and INTERMAGNET in decreasing the time taken between data being recorded and delivered to customers. This poster will show the techniques that have been used to securely transport data around the world. It will also describe the real-time performance of a prototype system.

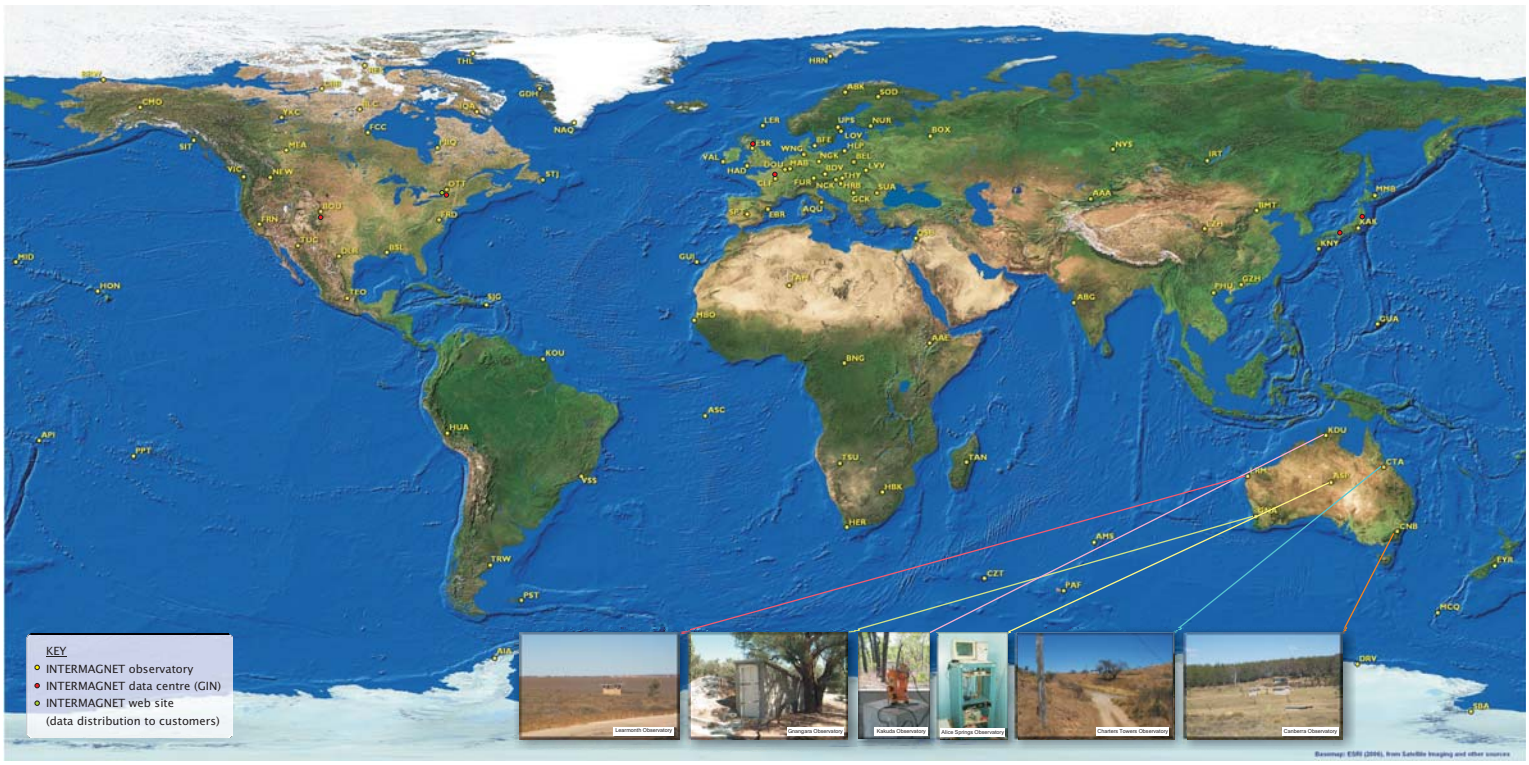
INTERMAGNET organisation

INTERMAGNET is a global network of about one hundred observatories, monitoring the Earth's magnetic field, constituted from teams in existing institutes whose primary task is one of geomagnetic measurement. The infrastructure at each observatory varies from state of the art to very basic, for example some observatories do not have INTERNET connections. Six regional data centres (called Geomagnetic Information Nodes or GINs) accept data from observatories and forward it to the INTERMAGNET web-site, from where the scientific community can access it.

Real-time data exchange - the challenge

INTERMAGNET's original data distribution service was designed in the early 1990s, used e-mail for the transport layer and was based on a daily exchange of data. This service is being upgraded from to a system that can:

- Reliably transport higher volumes of data from around the world in a number of formats to a central repository
- Accept data in as timely a manner as the producer is able to supply it
- Work with the security policies of a number of different institutes
- Give the user access to the system through tools and applications which are available and familiar to all
- Gain international agreement and acceptance

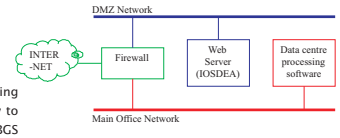


IOSDEA - From observatory to data centre

The INTERMAGNET One Second Data Exchange (IOSDEA) protocol is designed to allow observatories to send their data to data centres. Implementations of the protocol must be reliable, easy for users to use and must conform to the security policies at a number of institutes. IOSDEA uses the HTTP protocol to cache data in a small local store on the data centre's web site. HTTP was chosen as the underlying protocol because it is well understood by security professionals and widely deployed.

The user uses the HTML file upload standard (RFC 1867) to deposit a file in the cache. The protocol is designed to operate interactively as well as a web service (computer to computer), so users can try the system manually before proceeding to automation. BGS has provided examples in both Java and CURL to show how to automate uploading of data . The BGS implementation of the protocol has a number of features designed to protect against malicious use, including:

- HTTP digest authentication (RFC 2617)
- Checks on the type and size of file uploaded
- Daily limits on the amount of data an observatory can deposit



The cache can be searched and data downloaded from the cache using standard HTTP requests. The data centre uses the search facility to download only the most recently received data Once again, the BGS implementation includes examples showing how to automate the download process.

Rsync - From data centre to web site

Data deposited at the data centres is converted into a standard format. The data centres then use Rsync software (as a client) to transfer the data to the INTERMAGNET web site. Rsync is an easy to use, flexible transfer program that is available for a variety of operating systems and allows entire folder structures to be 'mirrored' between two computers across the INTERNET. It provides a very efficient method of exchanging data because it only transmits differences between files, not the entire file. The INTERMAGNET web site operates an Rsync server (the only server needed in the system) on a machine specially configured to handle the security risks associated with Rsync in server mode.

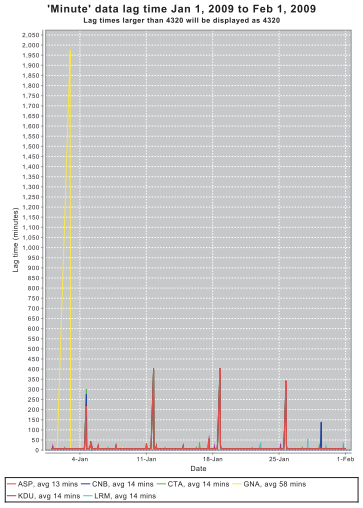
BGS use a database to store INTERMAGNET data. Rsync, which mirrors files, cannot be used directly in this environment. BGS have modified the standard Rsync software to use the database as a data source. A 'virtual file system' has been added to Rsync, so that the database appears to the software as a set of files and folders that do not actually exist in the real file system.

Once data has been transferred to the web site it is immediately made available to users.

Results

BGS deployed the prototype IOSDEA implementation early in 2008. Geoscience Australia, who operate six INTERMAGNET observatories on the Australian mainland, started using the system shortly afterwards. Data from each of the Australian observatories is forwarded to the Geoscience Australia office in Canberra where it is inserted into a database, then forwarded to the BGS IOSDEA in Edinburgh. The rate at which data is forwarded is configurable – currently it transfers new data every three minutes.

As part of its drive for increased real-time performance, INTERMAGNET has been investigating ways of monitoring the time taken for observatories to deliver data. BGS has developed a system to regularly record the 'lag time', that is the difference between current time and the time stamp of the most recently recorded data. The plot opposite shows the lag time (sampled hourly) for the six Australian observatories, as recorded at the Edinburgh data centre, throughout January 2009. The large spike for the GNA observatory probably represents a breakdown in communications between the observatory and the Geoscience Australia office in Canberra.



The BGS data centre is using Rsync to forward data to the INTERMAGNET web site once an hour. This frequency is configurable – in the future it is hoped to increase the update rate between the two sites.

BGS involvement in the project

- Design the protocols, metadata and data formats to support the system
- Create and deploy a prototype implementation of the system at the BGS data centre in Edinburgh
- Make the prototype implementation available to other data centres and help them install and operate the system
- Encourage and help users to start using the new system
- Design and implement tools for testing the performance of the system

References

INTERNET Request for Comments (RFC) 1867 "Form Based File Upload in HTML" www.ietf.org/rfc/rfc1867.txt
INTERNET Request for Comments (RFC) 2617 "HTTP Authentication: Basic and Digest Authentication" www.ietf.org/rfc/rfc2617.txt
Rsync samba.anu.edu.au/rsync/
CURL www.curl.haxx.se
INTERMAGNET web site www.intermagnet.org

Related work

BGS has many years experience in delivering data to the oil industry with very high levels of reliability and very near to real-time. The lessons learned in this commercial environment have been invaluable in the design of a global data delivery network.

BGS is now working with the National Oceanic and Atmospheric Administration (NOAA) in the United States and the University of Kyoto in Japan to update the data exchange and delivery system for the World Data Centres for Geomagnetism.