British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

The British Geological Survey's new geomagnetic data web-service

Ewan Dawson[1] (ewan@bgs.ac.uk), Jonathan Lowndes[1], Pragna Reddy[2] and Ciarán Beggan[1]
[1]British Geological Survey, West Mains Road, Edinburgh EH9 3LA, United Kingdom; [2]Robert Gordon University, Aberdeen, UK

Introduction

The British Geological Survey (BGS) operates eight geomagnetic observatories around the world. The data from these observatories are transmitted to Edinburgh, where they are processed and subjected to rigorous quality control procedures. The data are then disseminated to the community over the internet via a number of channels, including the Edinburgh INTERMAGNET Geomagnetic Information Node (GIN), and BGS's own public website.

Increasing demand for real-time or near-real-time observatory data means there is a requirement for institutes to have a robust and scalable data-processing architecture capable of delivering geomagnetic data products over the internet in a variety of commonly-used formats. As a consequence, BGS has spent the past year developing a new web-service system for the processing and distribution of our geomagnetic observatory data.

Why develop web-services?

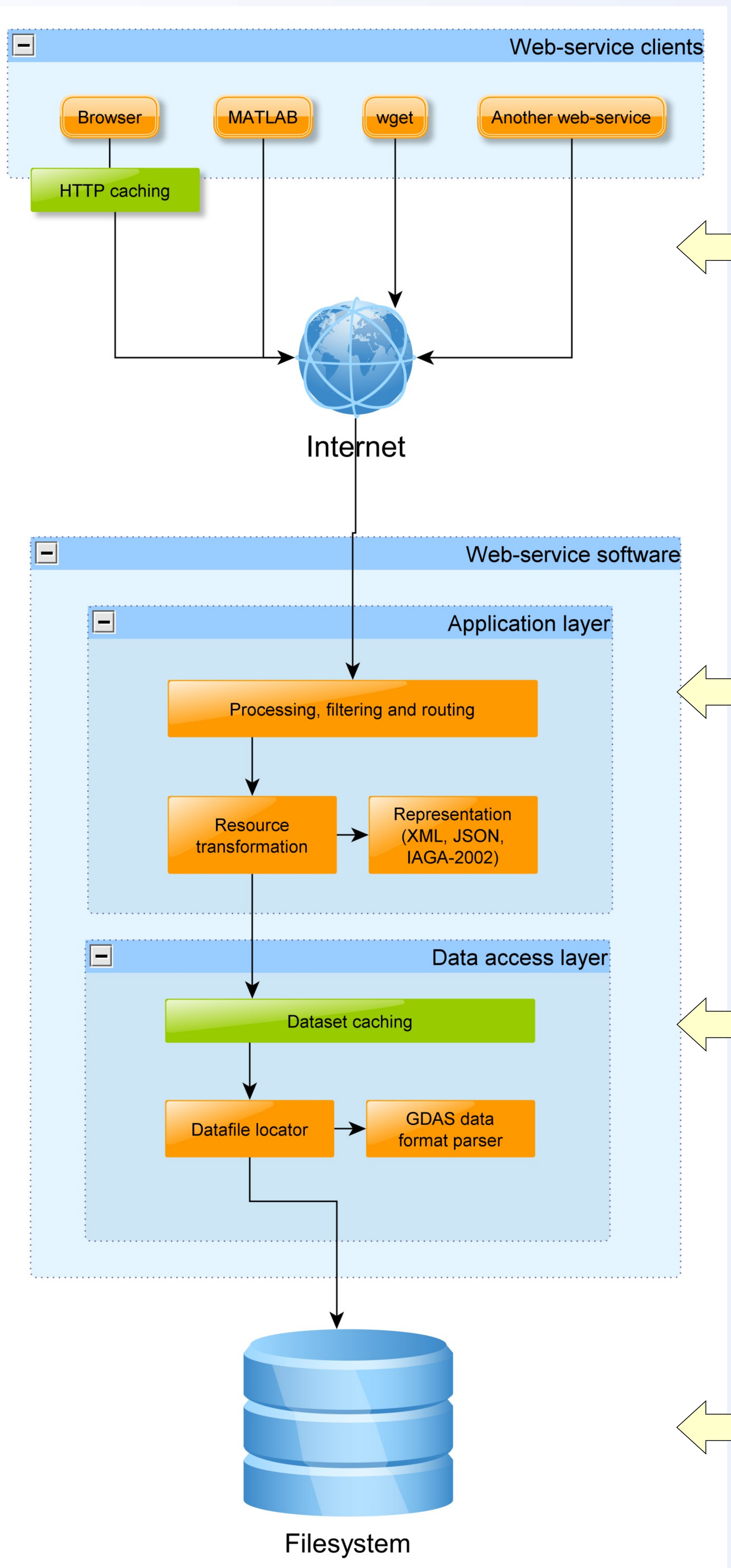
A web-service is simply a way for computers to communicate information to each other over the internet. The communication is usually done using the HTTP protocol (as used by web-browsers), and the information transmitted is usually in some well-specified format, typically XML.

Providing a web-service as the primary means of accessing observatory data, both publicly and internally, brings a number of benefits:

- **Ease of access**: clients need only know how to access a URL and parse the data received; they need know nothing of how the data is formatted and stored in the repository.
- **Reduction of code duplication**: low-level data access code is isolated in the web-service software – client software need not duplicate this code. This leads to faster and more reliable client development.
- **Increased resilience to change**: since the low-level data access details are abstracted away by the web-service, changes to the data repository (location, storage format, structure) need only be reflected in a single place – the web-service – while clients using the data are unaffected.
- **Increased security**: client authentication and authorisation can be managed at a single point, simplifying the protection of commercially sensitive data.
- **Interoperability**: the web-service can provide data in a number of standard formats (e.g. XML, JSON), so it becomes much easier to integrate BGS geomagnetism data into existing software and systems.

System architecture

The web-services software was developed as a Java web-application, using the Restlet framework. The Restlet framework provides an abstraction on top of the standard J2EE Servlet 2.5 specification, and enables the rapid development of RESTful web-services applications. The entire web-service runs in a standard Tomcat web-application container on a dedicated Linux machine.



The HTTP headers generated by the web-service contain all the information required for clients to safely cache the response data. Cache-capable clients (such as web-browsers) can issue a "conditional GET" request to check if their cached data are still fresh. The server will perform a check, and if the data have not changed since the client's last request, no new data will be sent. This feature can significantly speed up data access in certain scenarios.

Not only can clients cache responses from the web-service, but any proxy server *en-route* may also safely cache the data, without risk of sending stale data to the client. This includes any reverse-proxy/web-accelerator type caches which may be deployed on the network hosting the web-service.

The system is divided into two loosely coupled layers:

- **Application layer** - consists of the following components:
  - **Processing, filtering and routing** - includes a number of stages of request handling, such as routing (based on request URL), extraction of query parameters, and request authentication and authorisation.
  - **Resource transformation** - interfaces with the data access layer to retrieve the data, before performing any necessary transformations, such as applying a baseline or transforming the raw data into one of the derived quality control data products (e.g. comparison of co-located GDAS systems).
  - **Representation** - serializes the data resource into a byte stream of the requested data format, ready to be sent to the client in the response.
- **Data access layer** - handles the low level interface with the data repository (in this case a filesystem on the BGS Storage Area Network).
  - **Dataset caching** - finding, loading and parsing the data files required to respond to a particular request is relatively time-consuming, so the data access layer makes use of an in-memory cache, which stores frequently accessed data to enable faster response times. The cache first checks that none of the relevant data files have been modified since the data was cached; this ensures that stale data are never sent to the client.
  - **Datafile locator** - a single request may require data from a number of files (GDAS data is stored in one file per-day-per-instrument), so the data file locator assembles the requested data from the filesystem, using the GDAS data format parser to read the files.

Because only the data access layer knows anything about the nature of the data repository, if any changes are made to the network location or layout of the filesystem, or even if the filesystem repository is replaced by an relational database or FTP server, then only code in the data access layer of the web-service need be changed.

Users of the data, and in fact the application layer of the web-service itself, are insulated completely from the low-level details of the data repository.

Association: IAGA
Symposium: A131
Poster: 5706

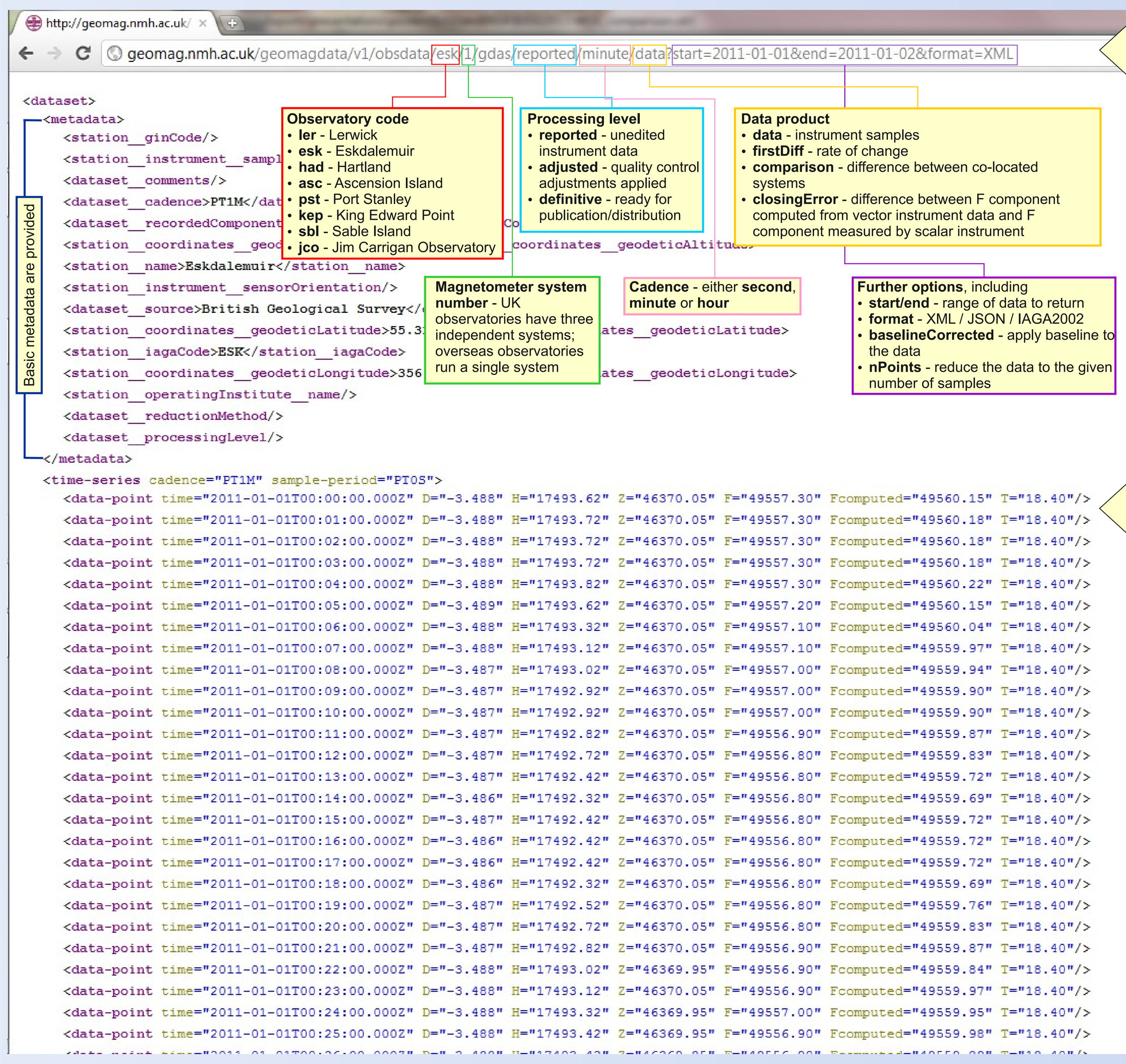
IUGG Melbourne, Australia
28th June - 7th July 2011

Using the web service

Our geomagnetic data web service exposes a number of data products:

- Second-, minute- and hour-cadence data from the observatory vector and scalar magnetometers.
- Derived quality control products, for example the comparison of two co-located systems, the rate of change in each component, etc.
- For each data product, we can access the raw variation data or the absolute data (variation data with a baseline correction applied).

Each data resource is identified by and retrieved via a URL. The data may be retrieved by any software capable of making an HTTP request. This provides a very simple interface to the data, accessible by a wide range of systems. The data can easily be made accessible to both observatory operations staff and public users wherever there is access to the internet.



- This screenshot shows an example of accessing the web-service using a browser. The data resource is completely specified by the URL.
- In fact, any software capable of making an HTTP request can be used with the web-service, e.g. the command-line tools wget and cURL, MATLAB, R, Excel, etc.
- You can write your own software to access the data in real-time, using any programming language which has an HTTP client available - Java, JavaScript, Python, C...

Client software may request the data in any of a number of formats, using the *format* URL query parameter. Currently supported formats are:

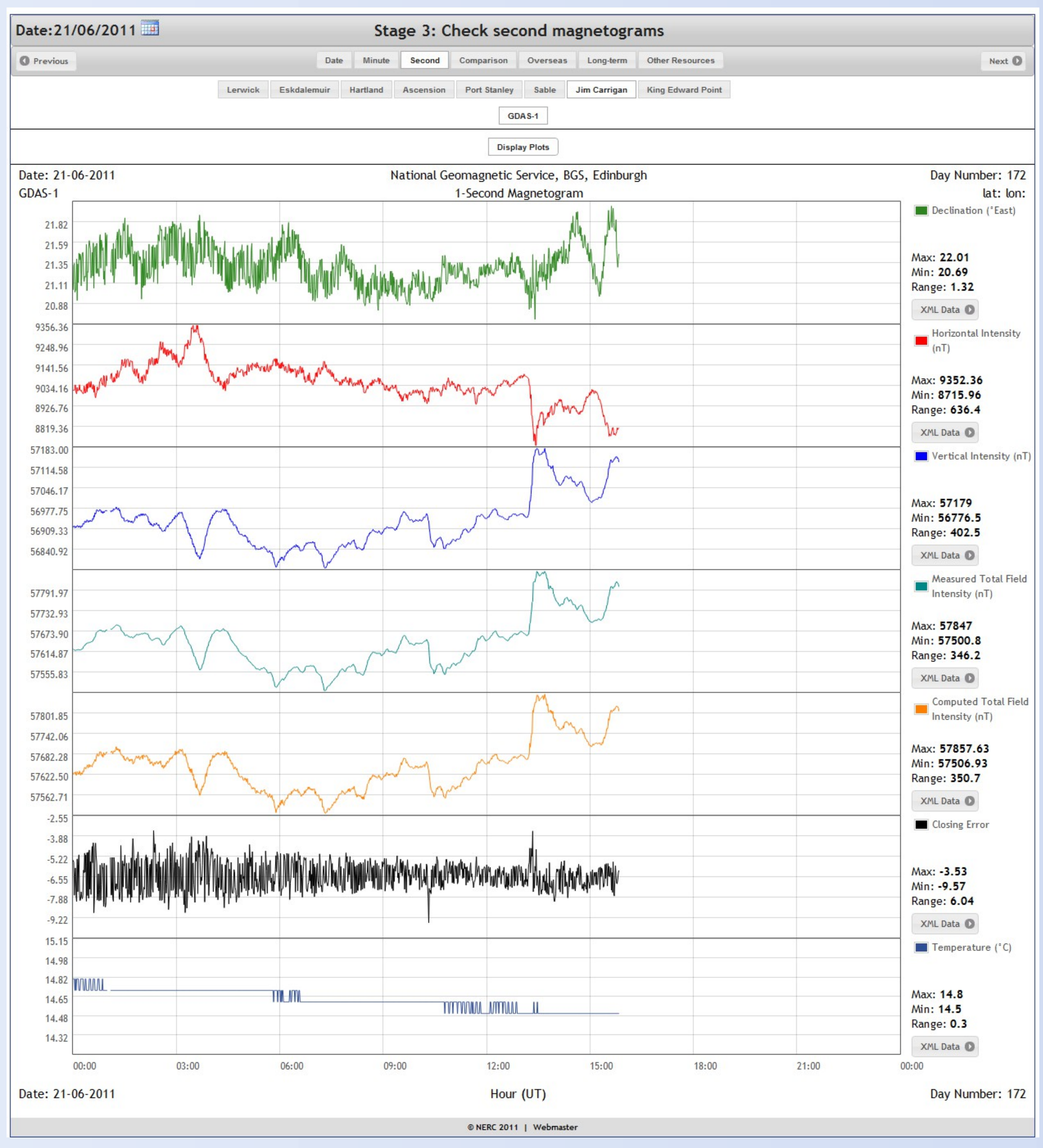
- **XML** - (pictured) as XML is a standard data exchange format, this allows interoperability with a wide range of systems.
- **JSON** - used by browser-based plotting tools such as *flot* and Google Charts.
- **IAGA-2002** - common data format in the geomagnetism community. Also easily parsed by MATLAB and R.

We plan to extend the web service to deliver data in other commonly-used data formats, such as WDC, INTERMAGNET binary, and NetCDF.

Data analysis and QC software

We have already taken advantage of the ability to rapidly build and deploy applications utilizing web-services by developing software to aid observatory operations staff in data analysis and quality control.

We developed this browser-based application using the JavaScript libraries *jQuery* (for the user interface) and *flot* (for plotting).



When the user navigates between the different data products using the buttons at the top of the screen, the application makes the appropriate requests to the web-service and plots the responses.

Because the web-service generates the data dynamically in response to each request, the latest data are always available.

Thanks to the various performance optimisations (see "System architecture"), the time taken from mouse-click to the data being plotted on-screen is typically less than 1 second, even when displaying a day of second-cadence data with numerous transformations applied.

Next steps

The web-service is currently not available publicly. We plan to make a subset of the data products described above available to the geomagnetism community in the near future via the BGS website. In addition, we will expand the range of formats in which the data may be accessed, to include WDC, INTERMAGNET CD-ROM binary (IAF), and NetCDF.

Reference

Reddy, Pragna (2011). Geomagnetism Web Services. Internal Report for British Geological Survey.