WEB BASED INTERFACE TO THE ECMWF ARCHIVE

Baudouin Raoult

ECMWF, Shinfield Park, Reading, U.K.

Introduction

Background

The ECMWF Meteorological Archival and Retrieval System (MARS) is a 92 Tbyte archive, containing more than 1.2 billion meteorological fields and 15 years of observations.

We are in the process of starting a project called *Web MARS* that will give access to this large amount of data to all of our Member States users, using Web technologies.

The MARS system is already a server/client application, but what the Web will bring to us is a set of standards and the possibility to serve clients running on different operating systems without having to maintain any specific software.

Metadata

The MARS archive is build on top of an object- oriented database system, used as an index into the actual data archive. This database contains the metadata that describes the physical location of the fields, as well as enough information to be able to address a field in a unique fashion.

This database is always up-to-date and reflects the actual state of the archive. It is permanently online and can be accessed within delays acceptable for interactive use.

The Web MARS system will be based on this database and not the actual data which resides on tape.

Web MARS

The Web MARS system will consist of several components: documentation, help, monitoring, browsing and data retrieval. Some of the components already exist and prototypes have been developed for others.

Documentation

An obvious use of the Web is to provide on-line documentation. Such documentation already exists, based on the training course material and various HTML documents describing the GRIB and BUFR data formats used by MARS to encode fields and observations. This documentation will be extended to include answers to Frequently Asked Questions (FAQ).

Changes to the main archive such as addition of new parameters, extra levels or new data streams will also be documented. This will be done in an automatic fashion, the pages being updated by our operational suites each time there is a significant change.

Finally, various statistics on the usage of the MARS archive will be made available.

Help

Agents

When confronted for the first time by the MARS archive, newcomers often find it difficult to discover what they are looking for. This is because the archive is organised to match the design of our operational suite (analysis, first-guess, forecasts, surface fields, upper-air fields...). To understand the logic of the MARS archive, one has to understand the logic of the assimilation cycles.

To help these users, we will provide series of programs (or "agents") that will answer questions such as *Where can I find cloud fields?* or *How do I retrieve re-analysis data ?*

These agents will scan the metadata database in order to provide the answer.

Assistants

An idea that is very popular among PC users is the concept of trivial programs that assist the user in very common tasks. These programs are often called "wizards". By asking the user very simple questions, they can create complex documents.

In the context of the Web MARS system, such assistants will help the user to build MARS requests based on a set of rules and the content of the MARS metadata database.

Monitoring

The MARS activity is already visible using a web browser. This helps the MARS administrator to monitor and tune the system.

Requests, queues, tape mounts and network transfers can be displayed, as well as various settings such as users' priorities and limits.

We believe that this information will help users develop a good understanding of the MARS system, with its strengths and weaknesses. With this knowledge, the user will be able to retrieve data more efficiently, minimising the number of tape mounts and optimising parallel requests. The outcome will be a better usage of our data handling systems. Allowing internal users to monitor their requests has already shown great improvements in the general throughput of the system.

If we find a secure way to identify a user, we will also provide a way to cancel the user's requests.

Browsing

We would like to offer a vision of more that one billion meteorological fields within the same web page, and with the minimum number of clicks. For that we have built a Java Applet that can display the content of the metadata database.

Users can navigate through the metadata in a hierarchical fashion. They can then select individual fields and view the MARS request to issue to retrieve these fields.

The system will also provides information on sizes and costs in terms of tape mounts and network transfers. This information will be very valuable to estimate the time it will take to carry out the request.

Although no indications are given on the actual data itself (i.e. the kind of weather the selected fields actually represent), we plan to annotate some of the metadata. A user searching for certain meteorological features (cyclones, El Niño...) will be directed to the right data sets.

Data retrieval

Once the user has selected the data in the browser, it seems natural to be able to download it.

Data would be delivered as raw data (GRIB or BUFR) or as a series of plots. The plots would be done using Metview, ECMWF data manipulation and plotting package.

This is the most complex component of Web MARS: it requires computer resources, such as disk space and CPU time.

The HTTP is not designed to handle very long transactions and has very short timeouts. Because it may take hours to actually retrieve and post-process the data, we have to find a way to handle requests asynchronously, make the data available on an FTP server, and notify the user. The user should be able to monitor the progress of the request, and possibly cancel it. The administration of such a system may be complex, especially the disk management on the FTP server.

There are also various security issues to address: we need to identify the user to validate the access to the data and keep track of some accounting information.

Issues

We have noted several issues we must address in order to realise this project:

Compatibility

Although all web browsers should display the same page in the same way, this is not the case. This imposes some limitations upon the design. A lot of testing is required on various platforms and different browsers.

Navigation

Although the various components that will make Web MARS can be described independently, a great part of the design will be their integration into a complete system. The navigation between the components is a key to the success of the project.

Because the Web is still young it is difficult to find anything in the literature to help with the design of a good navigation system.

<u>Security</u>

Most of the services we plan to provide rely on the ability to identify the user. Unfortunately, there does not seem to be a standard and secure system to do so.

<u>Maintenance</u>

When designing this application, we have realised that they were three kinds of page we would have to maintain:

- *Static pages* that are created and edited manually.

- *Automatic pages* that are created by a computed program, in general using a batch method. Once generated, they are copied onto the web server.

- *Dynamic pages* whose content is created on the fly, as the user is loading them.

We will try to minimise the use of static pages, as they are the most difficult to maintain and quickly become out of date.

Conclusion

Apart from some prototypes, the project is really at an early stage. So far we have identified three main issues: the ability to identify the user, the design of a good navigation system and the need to handle very long transactions.

If we succeed in our endeavour, we will give our users the possibility to access one of the largest meteorological archive with great ease.