

Baudouin Raoult*, Manuel Fuentes, Matthew Manoussakis

Operations Department, ECMWF

1. INTRODUCTION

Over the years, ECMWF has built one of the largest archives of numerical weather prediction model outputs. The archive holds today more than 7 billion meteorological fields from ECMWF operational forecasts, various scientific projects, research experiments and two reanalyses, one of which (ERA-40) covers a period of 45 years, starting late 1957.

In the last couple of years, ECMWF has developed two online services to provide access to these data.

2. PUBLIC DATA SERVER

This data server is principally aimed at research users and is for non-commercial use.

2.1. Datasets

2.1.1. DEMETER

The first dataset that was made available was the results of the DEMETER project, a multi-model seasonal prediction experiment. DEMETER is an EU funded cooperation between ECMWF and other research institutes.

The various partners had a need to exchange and compare their results. The data server has proved to be a success in achieving that goal.

The DEMETER project has currently 22 million fields on the server, totalling about 430 Gbytes of data. More will be added as various results continue to be produced.

2.1.2. ERA15

At the end of the ECMWF 15 years reanalysis project, a CD-ROM was created that contained a selection of fields on a 2.5° grid. This dataset was loaded onto the data server. It has only 82,000 fields for a total of 1Gbyte.

2.1.3. ERA40

The previous dataset is now superseded by ERA40, ECMWF's 45 year reanalysis (from September 1957 to August 2002). A substantial selection of the reanalysis results were selected and interpolated to a 2.5° grid.

The dataset contains 54 surface parameters and 11 upper-air parameters on 23 standard pressure levels, 4 times a day for 45 years. This is complemented by the corresponding monthly products.

This dataset contains about 21 million fields, for a volume of about 413 Gbytes.

2.2. User interface

The use of this service had to be as simple as possible for the end users. There is a single web page for each dataset from which the user can select any combination of fields. Once fields are chosen, the user is presented with post-processing options such as resolution changes and sub-area extraction. A third web page presents the resulting data for downloading.

2.3. Data formats

2.3.1. GRIB

All fields are stored on the server in GRIB form, which is the native format of all the fields produced at the Centre. GRIB is a WMO standard and decoders are now freely available from the Centre's web site.

Although all fields are global, the user can ask for specific areas and resolutions. The system will interpolate the data on the fly, and return the result.

2.3.2. NetCDF

Researchers from institutes other than National Meteorological Services, such as universities, are usually not familiar with GRIB formats. It was acknowledged that NetCDF was a popular data format in these communities, and that the data server should provide data in this format.

A GRIB to NetCDF converting tool was developed, trying to follow established conventions

* Corresponding author's address: European Centre for Medium-Range Weather Forecasts (ECMWF), Shinfield Park, Reading, Berkshire RG2 9AX, U.K.;

as much as possible. This tool is still at an early stage and will be developed further.

2.3.3. *Plots*

Using Metview, the data server is able to plot the requested fields on a selection of projections and geographical areas. Plots can be viewed and animated on the web or downloaded in Postscript format. This facility is useful for a user with small Internet bandwidth, as plots are generally smaller in volume than the raw data.

This service provides very little control to the user regarding the choice of plotting attributes, such as contouring, shading, colour selection and legend creation. Work is currently being done to overcome these limitations.

3. ONLINE ORDERING SYSTEM

ECMWF runs a Data Services group that deals with data orders from researchers and commercial institutions.

Until recently, in order to get data from the Centre, clients had to make their selection from a fixed catalogue, download a PDF form, fill it and fax it to the Data Services group that would deal with the order in a semi-automatic manner. The catalogue had to be manually edited from time to time.

To simplify the management of data orders and to cater with the ever increasing diversity of data available from the Centre, a web based application was developed that allows the full automation of data order processing, from catalogue selection to data delivery.

3.1. *Automatic catalogue*

The first requirement was to establish a catalogue of the available data that would be always up-to-date and reflect the actual content of the archive. This information is collected daily and stored in a database that provides the basis for the generation of dynamic web pages from which the user will select data.

3.2. *Data selection*

Because of the size and complexity of the Centre's archive, the challenge to solve was to build a simple and intuitive user interface.

The data selection is made by following a series of steps comprising the selection of the dataset, then the desired period followed by the choice of meteorological parameters, and

then by the specification of any post-processing such as change of resolution or sub-area extraction.

The volume and cost of the selected data is presented to the user. He or she can then save a data request for the selection in a "shopping cart" and iterate the process to add more data requests from other sources. Once all data as been selected, the user can proceed to the "checkout" where his or her details are provided and the request is submitted.

Because external users are not familiar with the Centre's data, a comprehensive context sensitive help can be accessed at any time during all steps of the selection. All pages contain links to relevant documentations. A search facility helps the user to quickly find in which dataset specific meteorological parameters can be found, as well as the extent of the datasets.

One particular difficulty was to cope with changes and defects in the archives. A database was set up containing these events that will be presented to the user throughout the selection process.

3.3. *Users and orders database*

A database was designed to manage users and orders, as well as various information such as media characteristics. As the processing of an order progresses, the database is updated so that the user can be informed.

3.4. *Automatic order processing*

Automatically managing a user's request involves the generation a suite of tasks that retrieve data from the archive in an optimum fashion, taking into account the available resources and the characteristics of the media on which the data will be delivered.

To achieve this, a complex algorithm was developed that analyses the user's data requests and decides on the most efficient strategy. The output of this algorithm is the definition of a series of tasks to perform that is fed into SMS, the ECMWF scheduler. The tasks are then run under the supervision of this scheduler. The data is retrieved from the archive; some processing may be performed on the data. Resulting files are written to the desired media or made available for Internet download and a PDF report is generated and sent to the user, describing the content of each media and of each file.

3.5. Payment

This is the only stumbling block to reach a fully automatic ordering system. The Centre is not currently ready to implement and support online payment. Solutions such as prepayment or subcontracting to online payment companies are being investigated.

4. CONCLUSION

The public data server is a very successful service. Several thousands of users from all around the world have downloaded data at a rate of 1 Terabyte a month. As the Centre will take part in future research projects, more datasets will be made available on the server.

The online ordering system is operational but still needs some further development. User feedback has been very positive, and the experience gain in developing this application, especially in finding ways to represent large and complex archives on the web, will be valuable for future projects.

The ECMWF public data server can be found at:

<http://data.ecmwf.int/data>

and the online ordering system at:

<http://www.ecmwf.int/products/data>.