DATA MANAGEMENT AND ARCHIVING

J.K. Gibson

European Centre for Medium-Range Weather Forecasts
Reading, U.K.

1. INTRODUCTION

Figure 1.1 illustrates the component sub-systems and data concentrations within ECMWF's Meteorological and Operational Systems (EMOS).

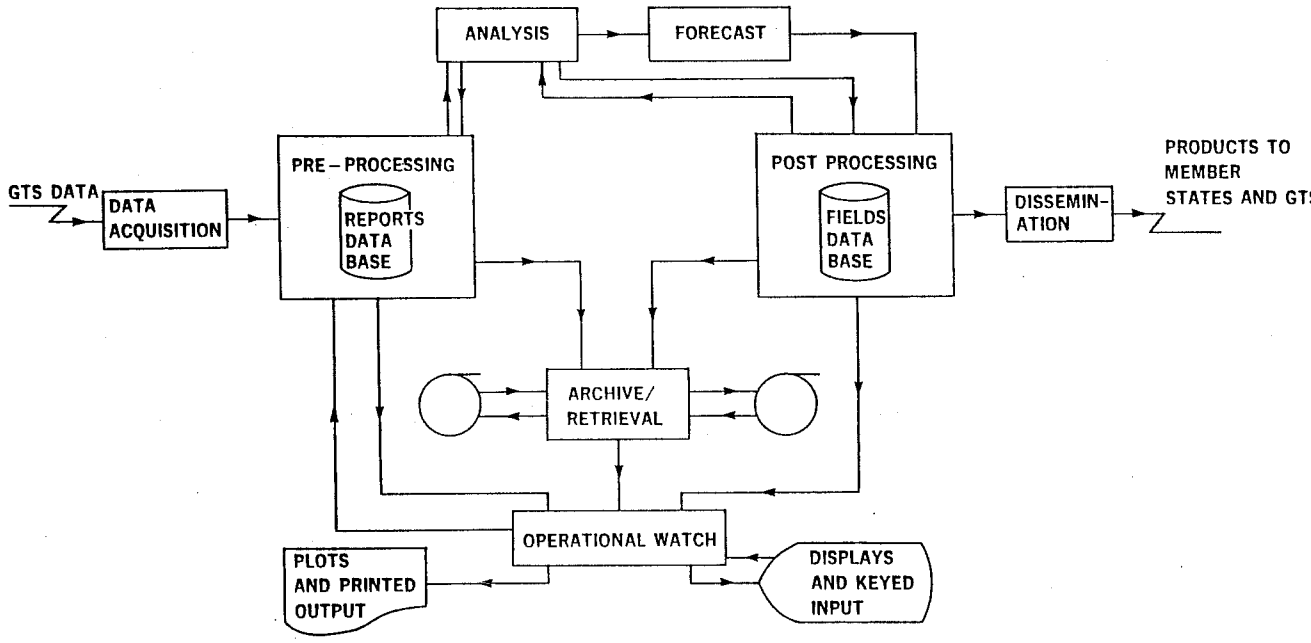

Fig. 1.1: EMOS-ECMWF's Meteorological Operational System

In the first generation system, many forms for the representation of data were developed in an ad-hoc fashion. One characteristic, common to several types of processed data, is the packing or compaction technique where data are represented with reference to a minimum value such that

$$V = R + SI$$

where V = original value

R = minimum of all values to be represented

S = a scale factor

I = a packed value in integer form

Additionally, a rigid operational archive was created which took no account of the requirements of research, or of the demands of external access to ECMWF data in the form of data services.

2. THE MARS (METEOROLOGICAL ARCHIVE AND RETRIEVAL SYSTEM) CONCEPT

2.1 Overview

To improve data handling in general, and to address the particular problems outlined above, a special project was set up. As part of this project, it was decided to

- move data handling and storage to a dedicated mainframe
- introduce a hierarchy of storage media - initially disks, MSS, and magnetic tape
- aim to make use of new storage devices and media as new technology became available
- manage the file positioning by means of the Los Alamos Common File System (CFS)
- manage the meteorological archives by means of a higher level software package (MARS) interfacing to CFS
- enable the management of all meteorological data within MARS - operational and research.

2.2 MARS Design

MARS was designed to

- rationalise data formats
- simplify data retention
- optimise hardware support, and
- generalise the user interface

Thus, instead of a data format designed for each application, it is intended to assist the development of internationally agreed forms for binary representation, and to use these in all applications. To this end ECMWF have taken part in the development of the two WMO binary forms, FM 92 GRIB, and FM 94 BUFR. Whereas traditionally compressed forms were used for long-term retention, and the fully expanded data were retained medium-term for experiments, the MARS design envisages compressed, standard forms for both medium and long-term retention. To optimise hardware support, the standard compressed forms only will be transferred between computers; expansion will be performed on the applications computer. The data handling machine will be allocated the rôle of performing area extraction, horizontal interpolation, etc. The MARS user will not need to know details of file names, storage media, and the like. A simple command interface will allow the user to define in meteorological terms the required data - MARS will map such requests onto the physical storage media. Figure 2.2.1 contains an example of a simple retrieve request.

```
RETRIEVE,  CLASS=OPERATIONS,
           TYPE=FORECAST,
           STREAM=DAILY ARCHIVE,
           LEVTYPE=PRESSURE LEVELS,
           LEVELIST=850/700/500,
           REPRES=SPHERICAL HARMONICS,
           PARAM=GEOPOTENTIAL,
           DATE=850930,
           TIME=12,
           STEP=24/TO/240/BY/12,
           TARGET="MYFILE",
           FORMAT=UNPACKED,
           RESOLUTION=106,
           AREA=GLOBAL.
```

Fig. 2.2.1: A MARS RETRIEVE request

2.3 MARS Development

An IBM 4341 was chosen to explore the feasibility of the over-all concept. The CFS software was adapted and installed, and the MARS software written. As a temporary measure, all CFS access was performed from computers remote from the IBM. Since an address space interface to CFS on the IBM was to be developed, MARS was designed to use this feature when available. Thus for temporary initial development, a MARS kernel on the IBM was implemented to map the meteorological data to the physical data, but the actual data retrieval and subsequent processing would be performed on the applications computer requesting the data.

By July 1985 sufficient progress had been achieved to enable the full operational archiving of EMCWF's forecast and analysis products. A primitive retrieval service was quickly established, and tested by in-house users before being offered to ECMWF's Member States users. Work continues on the provision of data processing options, the management of additional data, making use of enhanced system features, and increasing the services offered to users.

On the MARS software front, spectral to grid, and grid to grid interpolation, and the extraction of sub-areas are now fully supported. It is intended to add parameter derivation (eg u- and v- winds from vorticity and divergence), a FORTRAN callable interface, and access facilities from the on-line Fields Data Base.

Data enhancements include the addition of back data, which is currently in progress, and the provision of a special archive for TOGA (Tropical Ocean Global Atmosphere). Future developments will address the addition of monthly means, a time series archive, and archive facilities for observational data.

Recent systems enhancements include the replacement of the IBM 4341 by an IBM 3090 150-E with a vector processor. A SUPERLINK connection between the IBM and the CRAY X-MP, and an address space IBM MAR-CFS interface are planned to further improve the service.

Users so far have been provided with a MARS manual which is designed as a reference manual. As such, it does not make the introduction of new users to MARS an easy task. To counteract this, it is intended to introduce a purpose written MARS Introductory Manual to complement the existing manual.

3. DATA IN SUPPORT OF INTERNATIONAL RESEARCH

ECMWF has agreed to become a TOGA level IIIA Atmospheric Data Centre. To support the International Community in this respect an ECMWF/WRCP level III Atmospheric Data Set is being constructed. This will comprise

- a basic data set, for general users
- an advanced operational data set
- a supplementary fields data set

The latter two data sets will contain data at operational forecast model resolution, of interest to users with considerable computational facilities and modelling expertise.

A unique feature of the data services offered in support of these data will be the tailoring of the data to the user's specific needs. Thus

- area extraction
- spectral truncation, and
- horizontal interpolation

will be offered.

Figures 3.1, 3.2, and 3.3 illustrate the contents of this archive. Currently, data from July 1985 to date are stored. The back-archive to 1 January 1985 will be added in due course. It is planned that the data will be accumulated at least until the end of the TOGA period (December 1994).

The basic data set

- 00 UTC + 12 UTC
- Uninitialized analysis
- 2.5° x 2.5° global
- 14 pressure levels
- z, u, v, w, T, RH
- P_s, T_s, P_s, P_{msl}, u_{10}, v_{10}, T_2, Td_2, Z_0, L/S mask

Fig. 3.1: ECMWF/WRCP Level III Archive - Basic Data set

Advanced operational data set

- 00, 06, 12, 18 UTC
- Uninitialised analysis
- T106/N80 global*
- 14 pressure levels
- z, u, v, w, T, RH, P_s, T_s, P_{msl}, u_{10}, v_{10}, T_2, Td_{msl}, snow, cloud
- Soil/deep soil parameters
  Z_0, L/S mask, roughness, albedo,
  climate deep soil parameters

* Subject to change if model resolution changes

Fig. 3.2: ECMWF/WRCP Level III Archive - Advanced Operational Data Set

Supplementary fields data set

- 00 UTC, 12 UTC
- First-guess
- Gaussian N80, global*
- Surface fluxes - sensible heat, latent heat
  Net radiation (shortwave and longwave surface and top)
  Surface wind stress

* Subject to change if model resolution changes

Fig. 3.3: ECMWF/WRCP Level III Archive - Supplementary Fields Data Set

4. DATA MANAGEMENT - SOME CONCLUDING REMARKS

The development and application of international standard binary forms is seen as the key to simplifying data processing - BUFR for observed data, and GRIB for products. Further standardisation of expanded forms will enable applications to be written in a transportable manner. Thus software exchange will become a possibility whereas different forms of data interface precluded such exchange in the past.

Above all, there are vast volumes of meteorological data which exist, but which are so expensive to transform that they cannot be used. By archiving ECMWF's data in an agreed standard form it is believed that the usefulness of the resulting archive will be enhanced, and its potential maximised.