

# Observation feedback archiving in MARS

Piotr Kuchta, ECMWF

## Abstract

ECMWF's observation monitoring and assimilation activities of atmosphere, ocean and land-surface data generate a large amount of valuable information. This information is collectively known as observation feedback data. It is extensively used for performance monitoring of several main components of the forecasting system: the observations, the assimilation systems and the forecast models.

We describe here a work plan and its progress for the implementation of an improved archiving facility for observation feedback data.

## Introduction

At ECMWF observation feedback data is generated by:

- Atmospheric 4DVar system (Operations),
- RD experiments,
- ERA climate reanalysis, GEMS, ocean and land-surface analysis (planned).

Observation feedback is used for monitoring of performance and quality of main components of the assimilation system: observations and forecasts. Examples of feedback data:

- Blacklists, quality checks, other flags (e.g. duplicate report),
- First guess departure,
- Analysis departure.

Currently, the observation feedback data is stored at ECMWF in two forms, either as ODB tar-files in ECMWF's file storage (ECFS) or as BUFR-feedback files archived in MARS.

## Requirements for improved handling of Observation Feedback at ECMWF

In response to strong user requirements there was a necessity to improve ECMWF's archiving of observational feedback data. These include:

- Systematic, long-term (decades), secure archiving
- Keep long-term maintenance effort at a minimum
  - Archive should rely on IFS/ODB for its definitions of data
  - ECMWF's MARS system will stay the primary archiving facility
  - Minimize efforts in connection with frequently occurring upgrade actions
- Make archive of observational data more 'aware' of its content
  - Enable data discovery by browsing the web-based catalogue
  - Catalogue with more entries than in the current BUFR-feedback archive in MARS, e.g. make satellite instrument (or sensor) a high level categorization of archived observation data
  - Support refined data queries to allow fairly fine-grained extraction of subsets of data
- Performance improvements
  - Improve data-latency of typical users' data retrieval requests for specific sub-sets of feedback data
  - Flexible and fast retrieval, e.g. make retrievals of several months of data a realistic prospect
- Practical considerations
  - Data format of the feedback archive to be self-described
  - Data stored in large files spanning several days to a month, depending on observation type (better collocation)
  - Archive volume is not the primary design driver, but performance, resilience, flexibility and ease of use are
  - Provide built-in ability for future extensions (new data types and compression algorithms, new data items)

- Input observation data will continue to be archived in BUFR format
- The new feedback data archive (binary compressed feedback) will sit alongside the existing archive in MARS

## Introduction to ODB

ODB (Observational DataBase) is an in-memory, parallel source of input (observations) and output (feedback) for IFS (Integrated Forecast System). Data in ODB can be accessed using SQL (Structured Query Language) queries.

For more information on ODB see presentation by Anne Fouilloux ODB (Observational DataBase) and its usage at ECMWF.

## Introduction to MARS

MARS (Meteorological Archival and Retrieval System) is a system in use at ECMWF for more than 20 years. It is a scalable, managed archive facilitating highly configurable data collocation, data post-processing, field interpolation, sub-area extractions and data filtering. MARS currently supports GRIB and BUFR data formats.

Users describe data to be retrieved from MARS in meteorological terms (date, latitude/longitude, parameter, pressure level, etc.) and do not need to know where the data physically resides. Example query in MARS language syntax:

```

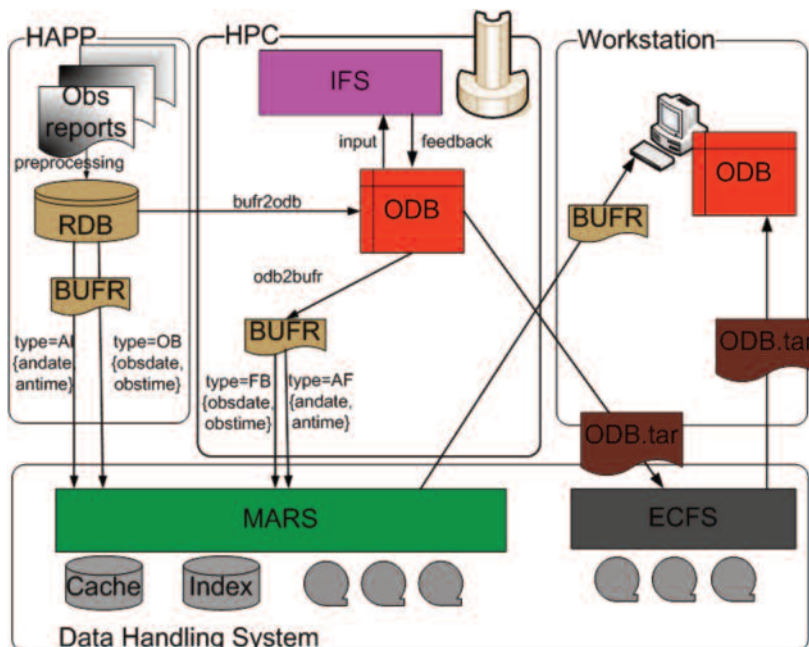
retrieve,
date      = 20010101 /to/ 20010131,
parameter = temperature /geopotential,
type      = forecast,
step      = 12 /to/ 240 /by/ 12,
levels    = 1000 /850 /500 /200,
grid      = 2 /2,
area      = -10 /20 /10 /0

```

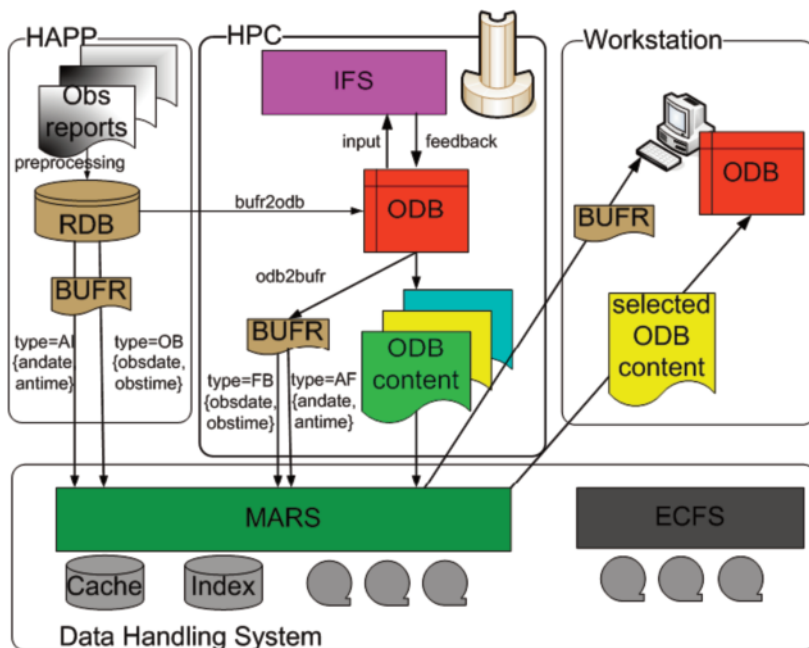
A web interface allows users to browse a catalogue of all MARS archive content. After finding required data user can:

- view the batch request (to be used with the command line client),
- estimate request cost (number of fields, tapes, data volume),
- retrieve or plot selection.

## Current observation data flow at ECMWF



## Proposed observation data flow at ECMWF



## Governance of observation codes

As the archive is expected to be operating for several decades, it is essential that the data contents are self-described, accurate, permanent and tightly governed. Feedback information is specific to the applications generating the data. Therefore, the governance of the archive descriptors will become an essential element for creating new releases of the forecast system.

The governance will include guidelines on how to add new types of data, tables and columns onto ODB, what codes should be used for a satellite, a sensor, etc. Where possible these should rely on internationally recognized definitions as drawn up and agreed by the WMO. Infrastructure for maintenance of archive descriptors and codes is being developed, this including databases, web applications and procedures.

It was decided that in order to resolve any possible inconsistencies, a review and re-organization of current data definitions in operational ODB is needed. This should precede the archival of operational ODB that uses the new MARS facilities.

## ODB Archive format

A new file format used internally by MARS for storing observation feedback was developed. ODA (ODB Archive) is a simple file format, well defined, machine independent, self-described (meta data in header), and extendable with new codecs (compression algorithms).

An ODB archive file represents a flat table. Every column is described by name, type (REAL, INTEGER, STRING, etc.), value of missing value (NULL) and codec (invisible to the user). It is designed to handle data types and meta-data of ODB.

The ODA format has several characteristics imposed by requirements for data archived in MARS, including:

- simple, yet efficient and fast encoding/decoding,
- small memory footprint needed to process large datasets,
- new data can be appended to existing files,
- efficient usage of storage medium (tapes).

A software library for writing, reading and filtering ODB archive files was developed. It was designed to work well with MARS, but does not depend on it and can be used in other applications. It has a SQL filtering engine supporting ODB/SQL syntax. The library was written in C++ but has also an API for C and Fortran. Examples of applications using the ODA library at ECMWF are oda2odb, obstat, Metview and Magics++.

Various UNIX command line tools were developed for exporting ODB contents to ODB archive files, recreating ODB from ODB archive files, as well as for comparing, viewing, filtering & transforming (using SQL) ODB archives.

## Work in progress: MARS extensions

In order to support archiving of ODB contents, several new MARS features were developed:

- extensions to MARS language, including:
  - configurable mapping between ODB columns and MARS language keywords, and
  - SQL filtering on MARS client facilities (new keyword FILTER).
- flexible, configuration file driven indexing (MARS server); any ODB column can be used as an index, for example: Observation / Analysis date / time, Report type, Platform (satellite), Sensor (instrument), Computation method, Obstype, Codetype, Varno, Channel, Version of data (reprocessed data, ...), Source of data.

Each type of observational data can be indexed by different attributes (ODB columns), examples: for SATOB: large number of report types, so index by computation method, for IASI: very large volume of data, so index by channel.

A tool for building MARS ARCHIVE request from ODB archive files was developed. Before the file is actually archived the request is validated against data, first by the MARS client, then MARS server.

Changes to ODB, like addition of new columns are propagated automatically to archives.

Example of filtering that uses the new keyword filter and SQL on the MARS client:

```
retrieve,  
obsgrout  =  hirs,  
date      =  20090728,  
time      =  0,  
type      =  FB2,  
reportype =  13,  
sensor    =  0,  
filter    =  "distance(lat, lon, 0, 0)<km(50) and value>375.2",  
target    =  "ECMA.hirs.13.oda.from_mars"
```

## Current status

The following has been achieved so far:

- definition of ODB archive format,
- working version of library for handling ODB archives,
- tools for archiving ODB content and reading it back to ODB,
- integration with Magics++, MetView and Obstat,
- MARS server support for ODB content (more testing needed),
- basic support for ODB content in MARS client.

Remaining work includes:

- finalization of data organization in MARS,
- completion of ODB re-organization, making use of WMO codes,
- back-archiving of observation feedback, currently archived in BUFR,
- testing and tuning of filtering on MARS client,
- easy monitoring, alarm system (obstat).