Observation feedback archiving in MARS

Piotr Kuchta

Data and Services Section, ECMWF

P.Kuchta@ecmwf.int

Acknowledgement:
Manuel Fuentes, Baudouin Rault, Erik Andersson, Anne Fouilloux, Lars Isaksen, Drasko Vasiljevic

ECMWF

Outline

- **Introduction**
  - **Observation Feedback – definition, usage**
  - **Requirements for improved handling of Observation Feedback**
  - **ODB**
  - **MARS**
  - **Handling of observations at ECMWF**
- **Proposed handling of Observation Feedback at ECMWF**
- **Work in progress**
  - **Governance of observational codes**
  - **ODB archive file format, library & tools**
  - **MARS extensions**
- **Future plans**

Observation feedback

- **Observation feedback:**
  - information generated by monitoring and assimilation of atmosphere, ocean and land-surface observation data
- **At ECMWF generated by**
  - Atmospheric 4DVar system (Operations)
  - RD experiments
  - ERA climate reanalysis, GEMS, ocean and land-surface analysis (planned)
- **Used for monitoring of performance and quality of main components of the assimilation system:**
  - observations
  - forecasts
- **Examples of feedback data:**
  - Blacklists, quality checks, other flags (e.g. duplicate report)
  - First guess departure
  - Analysis departure

Processes producing feedback information

Data extraction

↓

- Check out duplicate reports
- Ship tracks check
- Hydrostatic check

↓

Thinning

- Some data is not used to avoid over-sampling and correlated errors
- Departures and flags are still calculated for further assessment

Blacklisting

- Data skipped due to systematic bad performance or due to different considerations (e.g. data being assessed in passive mode)
- Departures and flags available for further assessment

↓

Model/4D-Var dependent QC

- First guess based rejections
- VarQC rejections

Used data → **Increments**

↓

Analysis

Observation data count for one 12h 4D-Var cycle 0900-2100UTC 3 March 2008

Screened			Assimilated		
Synop:	450,000	0.3%	Synop:	64,000	0.7%
Aircraft:	434,000	0.3%	Aircraft:	215,000	2.4%
Dribu:	24,000	0.02%	Dribu:	7,000	0.1%
Temp:	153,000	0.1%	Temp:	76,000	0.8%
Pilot:	86,000	0.1%	Pilot:	39,000	0.4%
AMV's:	2,535,000	1.6%	AMV's:	125,000	1.4%
Radiance data:	150,663,000	96.9%	Radiance data:	8,207,000	91.0%
Scat:	835,000	0.5%	Scat:	149,000	1.7%
GPS radio occult.	271,000	0.2%	GPS radio occult.	137,000	1.5%
TOTAL:	155,448,000	100.00%	TOTAL:	9,018,000	100.00%

We need to archive everything!

Only 6% of screened data is assimilated.

99% of screened data is from satellites

96% of assimilated data is from satellites

ECMWF

Requirements for improved handling of Observation Feedback at ECMWF (1)

- **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**
  - **MARS will stay the primary archiving facility**
  - **Minimize efforts in connection with frequently occurring upgrade actions**
- **Make archive of obs. 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 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**

Requirements for improved handling of Observation Feedback at ECMWF (2)

- **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**
- **The new feedback data archive (*binary compressed feedback*) will sit alongside the existing archive in MARS**

Introduction to ODB

- **ODB: Observational DataBase**
  - **In-memory, parallel source of input (observations) and output (feedback) for IFS (Integrated Forecast System)**
  - **Data queries in SQL (Structured Query Language)**
- **See presentation of Anne Fouilloux**

Introduction to MARS

- **MARS: Meteorological Archival and Retrieval System**
- **System in use at ECMWF for more than 20 years**
- **Managed Archive**
  - **Data Collocation**
  - **Scalability**
- **Post-processing facilities**
  - **Field Interpolation**
  - **Sub-area extractions**
  - **Data filtering**
- **Data formats supported are self-described: GRIB, BUFR**

Introduction to MARS: MARS language

- **Data described in meteorological terms (date, latitude/longitude, parameter, pressure level, etc)**
- **User does not need to know where the data physically resides**
- **Simple syntax**
- **Example:**

```
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
```

Introdu

- A Web
  - Cat
  - Dat
  - Par
- After fi
  - Vie
  - Esti
  - Ret
  - Plo

Observation F

MARS, the ECMWF archive - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ecmwf.int/services/archive/

Google

Pressure levels - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.ecmwf.int/services/archive/d/catalog/class=od/stre

Google

ECMWF

Home Your Room Login Contact Feedback Site Map Search:

About Us
Overview
Getting here
Committees

Products
Forecasts
Order Data
Order Software

Services
Computing
Archive
PrepIFS

Research
Modelling
Reanalysis
Seasonal

Publications
Newsletters
Manuals
Library

News&Events
Calendar
Employment
Open Tenders

Home > Services > Archive > Catalogue > Operational archive > Atmospheric model > 1 > Analysis > 2009 > October > Pressure levels >

Pressure levels

Archive

Overview
Catalogue
Data Finder
Last Update
Server Activity

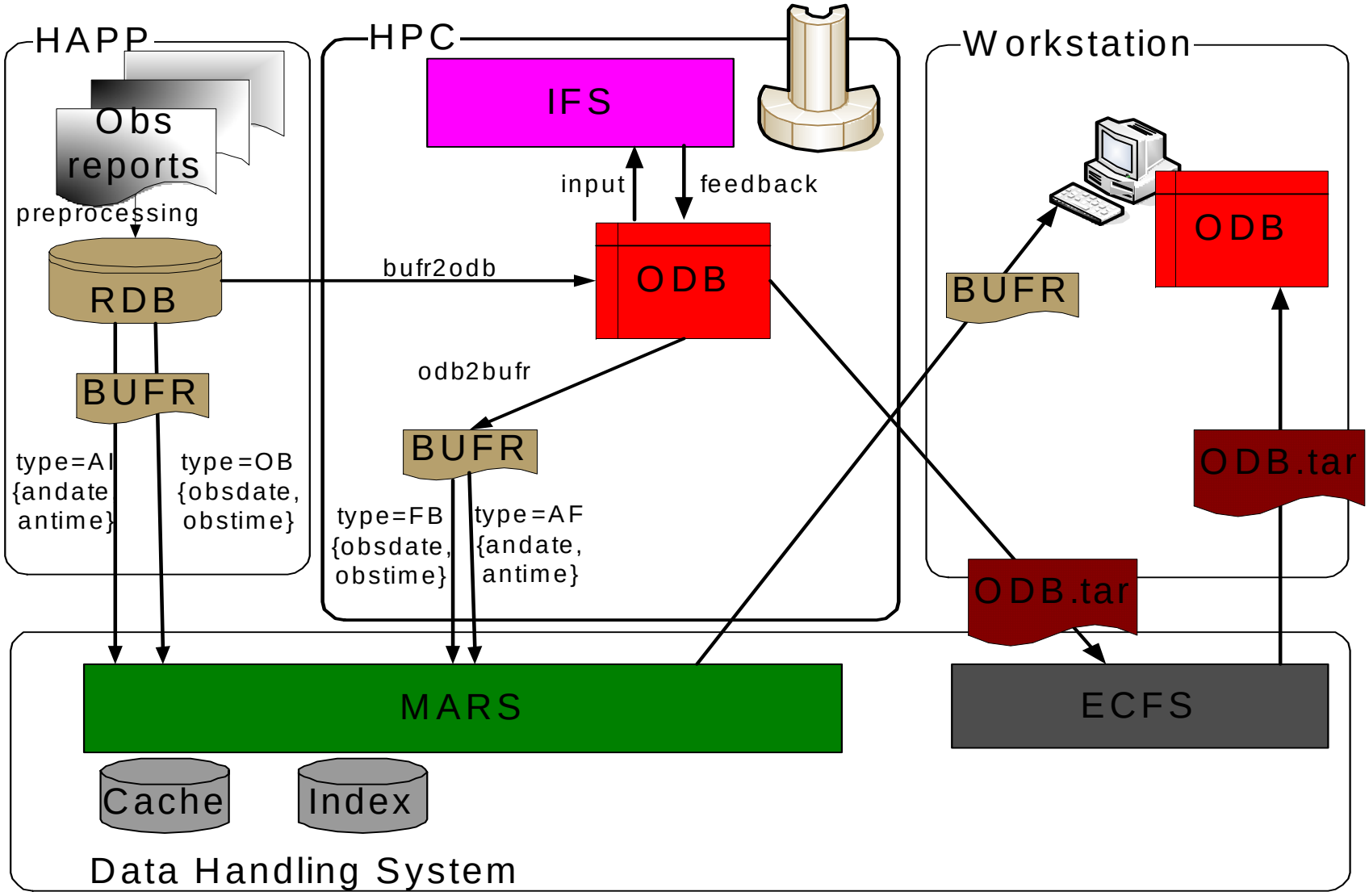
Personal

Your Account
Your Results

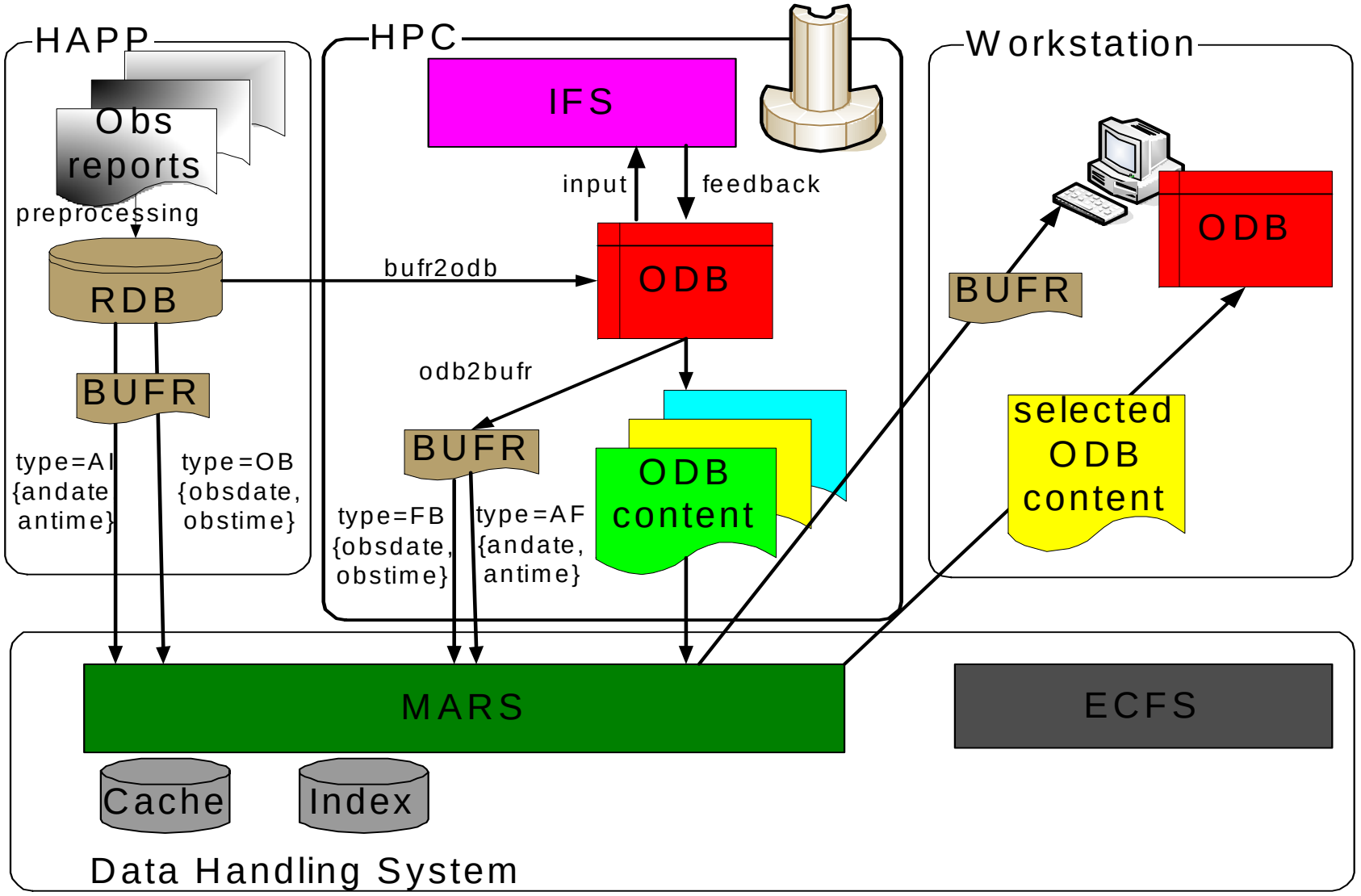
Date (31 values)	Time (4 values)	Level (25 values)	Parameter (11 values)
2009-10-01	00:00:00	1	Divergence
2009-10-02	06:00:00	2	Geopotential
2009-10-03	12:00:00	3	Ozone mass mixing ratio
2009-10-04	18:00:00	5	Potential vorticity
2009-10-05		7	Relative humidity
2009-10-06		10	Specific humidity
2009-10-07		20	Temperature
2009-10-08		30	U velocity
2009-10-09		50	V velocity
2009-10-10		70	Vertical velocity

- Check for availability
- View the batch request
- Estimate request cost
- Retrieve the selection in GRIB or NetCDF (experimental)
- Plot the selection

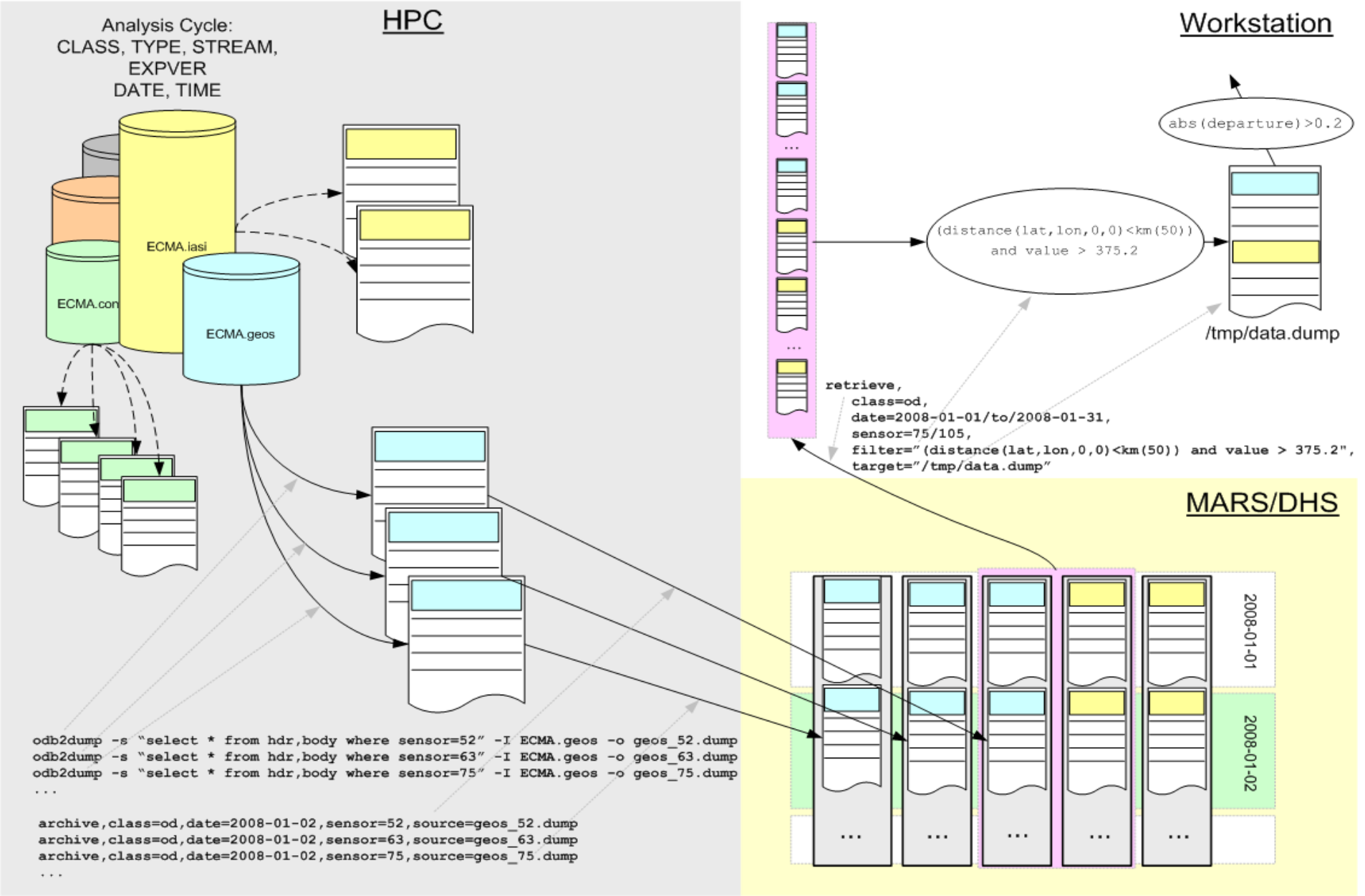
Current observation data flow at ECMWF



Proposed observation data flow at ECMWF



Proposed archiving of ODB content in MARS



Work in progress: Governance of observation codes

- **There are no guidelines on how to add new data types to ODB**
  - **New table ?**
  - **New column ?**
  - **What code to use for a satellite, sensor, etc… ?**
- **Review ODB contents**
  - **Re-organisation of ODB tables & ODB column names**
  - **Use WMO codes**
  - **Define infrastructure for maintenance: databases, web applications and procedures**
- **Definitions must not change over time, once defined**

WMO codes

- **Use standard codes from WMO tables:**
  - **Common code table C-5: Satellite identifier**
  - **Common code table C-8: Satellite instrument**
  - **Common code table C-2: RadioSonde/Sounding system**
  - **Satellite sensor indicator**
  - **Satellite Derived Wind Computation Method**
  - **BUFR types and subtypes, etc…**

Work in progress: Report Type

- **ReportType: Unique mapping of a series of attributes of observation data**
  - **Group, Sensor, Platform, Bufr Type, Subtype, Obstype, Codetype**
- **Report types are *grouped* by type of data**
  - ***Group* seems to define naturally the organisation of the archive**
    - **Conventional**
    - **Satob**
    - **Hirs**
    - **AMSUA**
    - **AMSUB**
    - **GEOS**
    - **…**

Work in progress: Report Type table

group		reportype		Local sensor		WMO sensor		platform		Comp. Method		bufrtype		subtype		obstype		codetype	
No	Name	No	Name	No	Name	No	Name	No	Name	No	Name	No	Name	No	Name	No	Name	No	Name
1	HIRS	1	TIROS-N HIRS Radiances	0	HIRS	605	HIRS; High Resolution Infra R(red) Sounder/2 (3=606; 4=607)	708	TIROS-N	N/A	N/A	3	Satellite Sounding	55	ATOVS	7	SATEM	210	Radiances
1	HIRS	2	NOAA 6 HIRS Radiances	0	HIRS	605	HIRS; High Resolution Infra R(red) Sounder/2 (3=606; 4=607)	706	NOAA 6	N/A	N/A	3	Satellite Sounding	55	ATOVS	7	SATEM	210	Radiances
1	HIRS	3	NOAA 7 HIRS Radiances	0	HIRS	605	HIRS; High Resolution Infra R(red) Sounder/2 (3=606; 4=607)	707	NOAA 7	N/A	N/A	3	Satellite Sounding	55	ATOVS	7	SATEM	210	Radiances
1	HIRS	4	NOAA 8 HIRS Radiances	0	HIRS	605	HIRS; High Resolution Infra R(red) Sounder/2 (3=606; 4=607)	200	NOAA 8	N/A	N/A	3	Satellite Sounding	55	ATOVS	7	SATEM	210	Radiances
1	HIRS	5	NOAA 9 HIRS Radiances	0	HIRS	605	HIRS; High Resolution Infra R(red) Sounder/2 (3=606; 4=607)	201	NOAA 9	N/A	N/A	3	Satellite Sounding	55	ATOVS	7	SATEM	210	Radiances
1	HIRS	6	NOAA 10 HIRS Radiances	0	HIRS	605	HIRS; High Resolution Infra R(red) Sounder/2 (3=606; 4=607)	202	NOAA 10	N/A	N/A	3	Satellite Sounding	55	ATOVS	7	SATEM	210	Radiances

Work in progress: ODB Archive format

- **New file format: ODA (ODB Archive format)**
  - -Simple, well defined, machine independent
  - -Self-described (meta data in header)
  - -Extendable: new codecs (compression alg.) can be added
- **An ODB archive file represents a flat table**
  - -Every column has: name, type (REAL, INTEGER, STRING, …), value of missing value (NULL), codec (invisible to the user)
  - -Data layout: report oriented (all data belonging to one report stored on 1 line)

reptype	class	andate	antime	lat	lon	statid	obsvalue	blacklist	status
14	‘od’	20090730	0	0.364777	0.986071	223	235.009995	0	01100
14	‘od’	20090730	0	0.367758	0.972175	223	235.949997	0	01100
14	‘od’	20090730	0	0.370347	0.959772	223	236.880005	0	01100

ODB Archive format features (1)

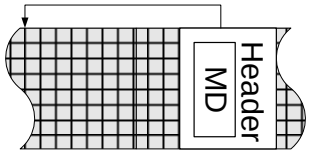
- **Handles data types and metadata of ODB**
  - **Changes to ODB like addition of new columns propagated automatically to archives**
- **Open for future extensions**
  - **New codecs can be easily added without breaking compatibility**
- **Small memory footprint needed to process large datasets**
- **Simple yet efficient and fast encoding/decoding**

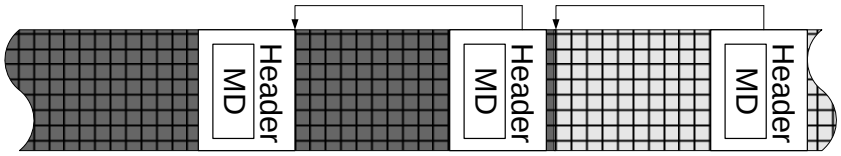
ODB Archive format features (2)

- **New data can be appended to existing files**
  - -Efficient usage of storage medium (tapes)
  - -Can retrieve long time series in one request

reptype	class	andate	antime	lat	lon	statid	obsvalue
14	‘od’	20090730	0	0.364777	0.986071	223	235.009995
14	‘od’	20090730	0	0.367758	0.972175	223	235.949997
14	‘od’	20090730	0	0.370347	0.959772	223	236.880005

reptype	class	andate	antime	lat	lon	satid	sensor	obsvalue
14	‘od’	20090731	0	0.364777	0.986071	222	3	235.009995
14	‘od’	20090731	0	0.367758	0.972175	222	3	235.949997
14	‘od’	20090731	0	0.370347	0.959772	222	3	236.880005





ECMWF

Work in progress: ODB Archive API

- **Library for writing, reading and filtering ODB archives**
  - Designed to work well with MARS, but does not depend on it
  - SQL filtering engine supporting ODB/SQL syntax included
  - API for C, Fortran, C++
  - Examples of users: oda2odb, obstat, Metview, Magics++

```
string sql = "select lat, lon, obsvalue from \"in.oda\""
             " where obsvalue > 0.5 ";
ODA oda("");
for (ODA::select_iterator it = oda.select(sql);
     it != oda.selectEnd();
     ++it)
     cout << it->data(0) << " " << it->data(1) << " “ << it->data(2) << endl;
```

Work in progress: ODB archive tools

- **Command line tools**
  - tool for exporting ODB contents to ODB archive files
  - tool for recreating ODB from ODB archive files
  - various tools for comparing, viewing, filtering & transforming (using SQL) ODB archive files

```
$ oda sql select lat, lon, obsvalue from \"in.oda\" where obsvalue \> 0.5

$ oda sql select lat, lon, obsvalue \
        into \“out.oda\"  \
        from \"in.oda\" \
        where obsvalue \> 0.5

$ oda sql query.sql
```

Work in progress: MARS extensions

- **Support for ODB contents**
  - **Extension to MARS language (MARS server and client)**
  - **Configurable mapping between ODB columns and MARS language keywords (MARS server and client)**

```
TIME: antime@hdr
DATE: andate@hdr
…
```

  - **Flexible indexing (MARS server)**
  - **`oda2request` tool (builds MARS ARCHIVE request from ODB archive files)**
  - **Validating requests against data when archiving (both server and client)**
  - **Filtering (SQL) on the MARS client**

MARS language extensions

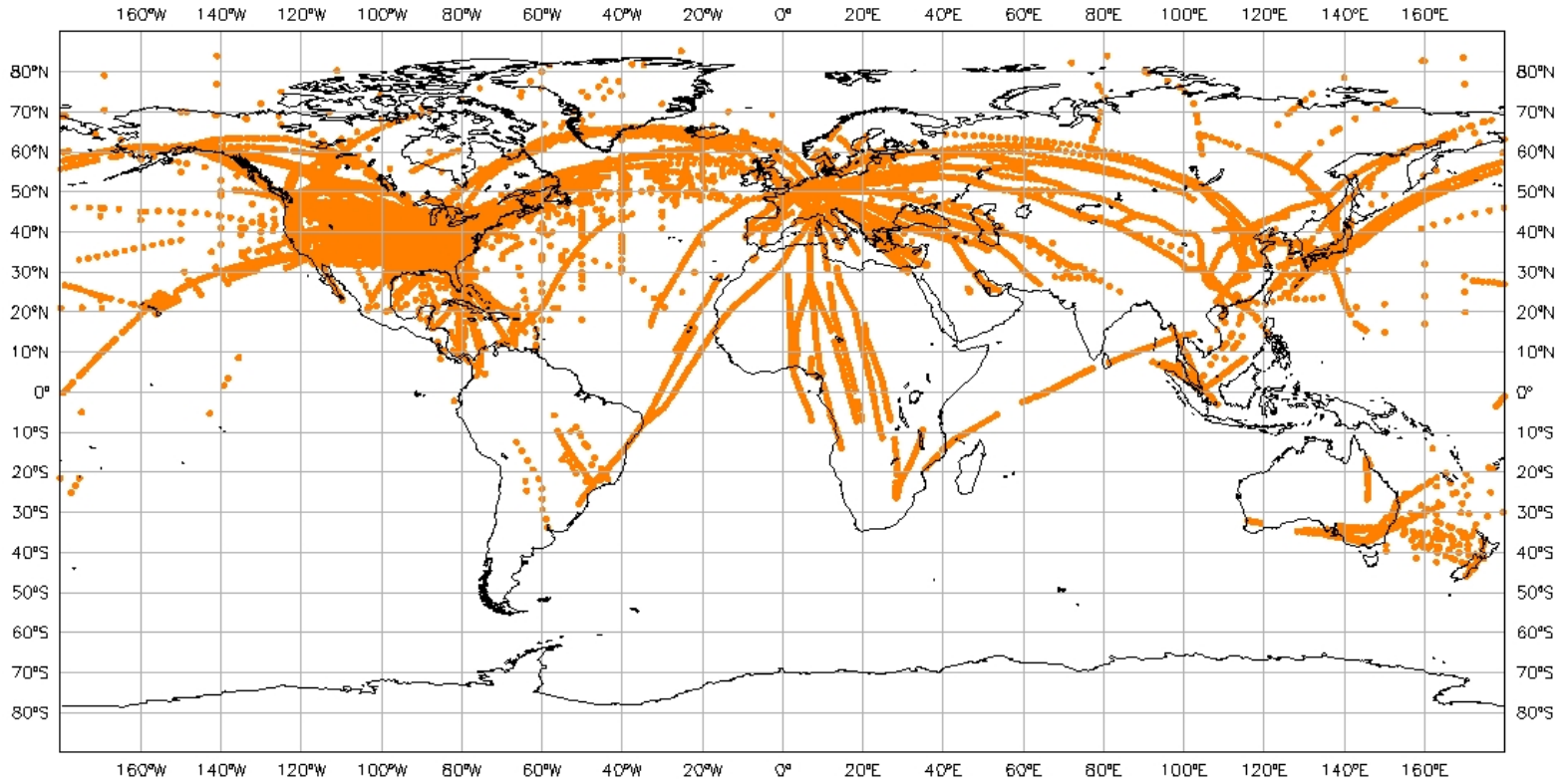
- **SQL capabilities incorporated into MARS syntax**
- **Example:**

```
RETRIEVE,
OBSGROUP    = hirs,
DATE        = 20090728,
TIME        = 0,
TYPE        = FB2,
REPORTTYPE  = 13,
SENSOR      = 0,
FILTER      = "distance(lat, lon, 0, 0)<km(50) and value>375.2",
TARGET      = "ECMA.hirs.13.oda.from_mars"
```

SQL engine used for filtering (WHERE clause of SELECT statement)

SQL filtering, example 1

select lat, lon, obsvalue from hdr,body where obstype = $airep and varno = $t and status.active@body = 1



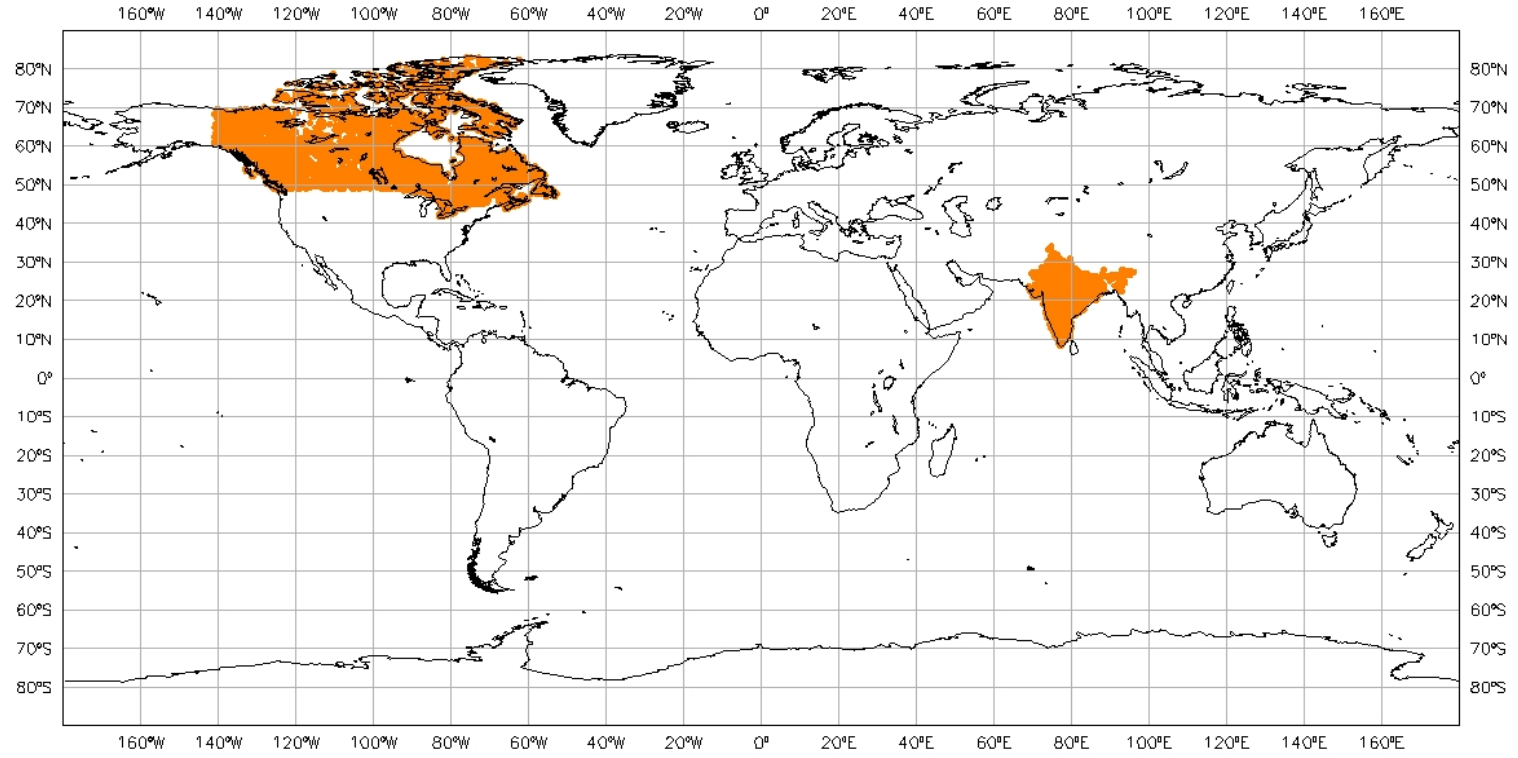
ECMWF

SQL filtering, example 2

select degrees(lat),degrees(lon),obsvalue from hdr,body

WHERE inside("canada")

OR inside("india")



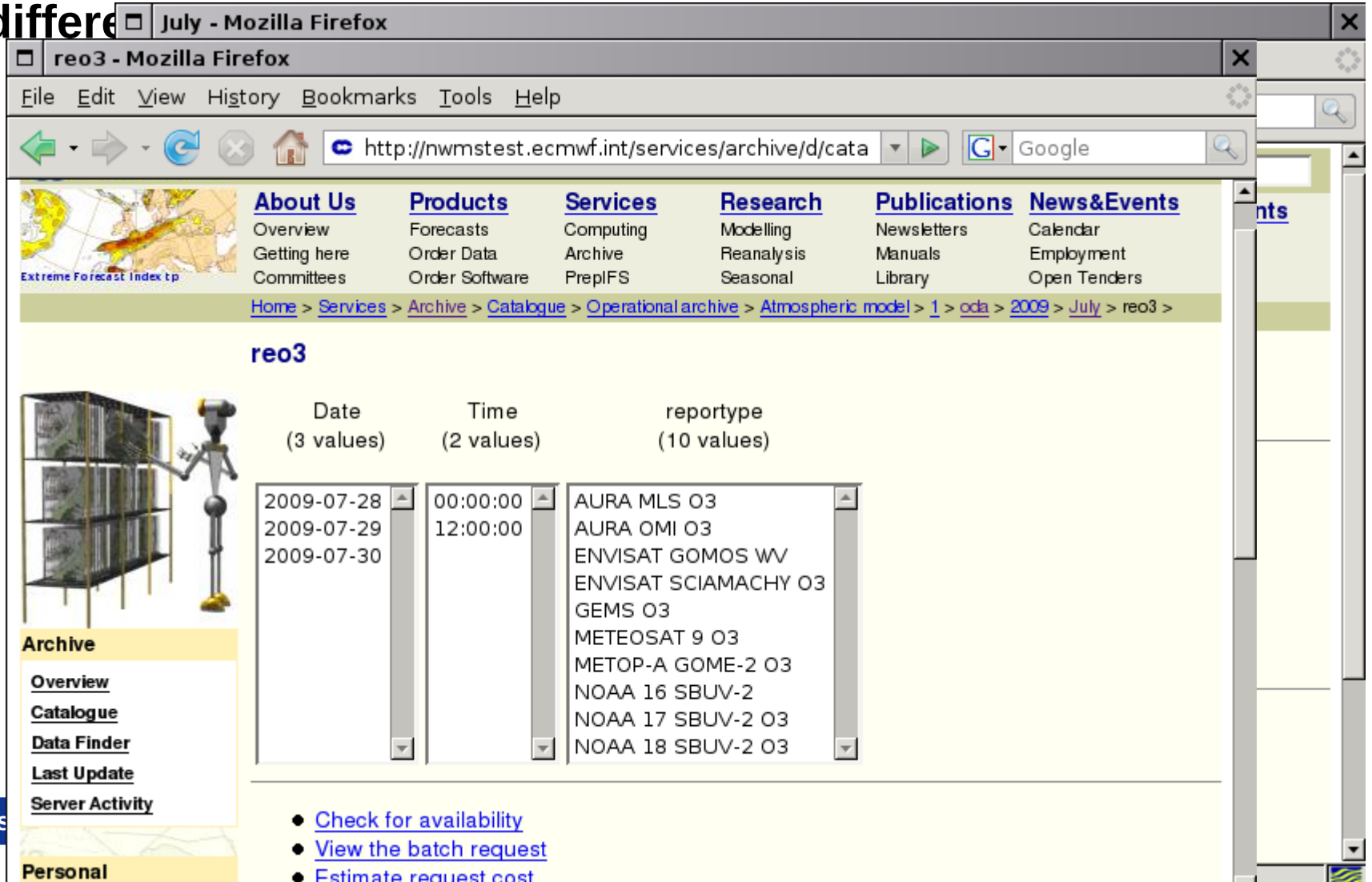
ECMWF

MARS extensions: indexing of ODB in MARS (1)

- **Configuration file driven**
- **Any ODB column can be used as an index:**
  - **Observation/Analysis date/time,**
  - **Report type**
  - **Platform (satellite)**
  - **Sensor (instrument)**
  - **Computation method**
  - **Obstype, Codetype, Varno, Channel**
  - **Version of data (reprocessed data, …)**
  - **Source of data**

MARS extensions: indexing of ODB in MARS (2)

- **Each type of observational data can be indexed by differe**



Current status

- **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**
  - **Obstat**
- **MARS server support for ODB content (more testing needed)**
- **Basic support for ODB content in MARS client**

Future plans

- **Finalize data organization in MARS**
- **Finalize ODB re-organization, make use of WMO codes**
- **Back-archiving of observation feedback currently archived in BUFR**
- **Filtering on MARS client**
- **Easy monitoring, alarm system (obstat)**