Report on the COPE technical meeting
held at ECMWF, Reading 9-12, June 2014

Alena Trojáková

1 Background

The ECMWF initiated the Continuous Observation Processing Environment (COPE) project to support the more timely delivery and better monitoring and control of meteorological observations for ECMWF's operational assimilation system.

The main goals are:

- to perform a substantial part of the observation processing earlier, so that it takes place before the cut-off time instead of after;
- to perform most of observation processing tasks using the ODB rather than BUFR format;
- and to enhance early detection and handling of observation anomalies that could cause failures in the operational suite.

Status of COPE:

- project is ongoing, but the full design is not yet set, see Fig 1, mostly infrastructure is being build. By the end of 2014 the technical infrastructure and tools to support the OBS2ODB conversion and simple filters should be available. The ECMWF will probably implement some components in operations with the next IFS cycle, but that this would not impact Météo-France because they do not use this part of the IFS (the makecma replacement code);
- project will require a long term is development. More advanced components that rely on background information would only be tackled in the coming year(s). In this time frame it is expected that COPE will become part of ECMWFs scalability programme;
- to explore more options of software development the COPE is planned as "a community project", using Apache-2 as software license; The ECMWF is ready to facilitate the development, but the project will not necessarily be mainly driven by ECMWF staff. Contribution by other partners is more than welcome.

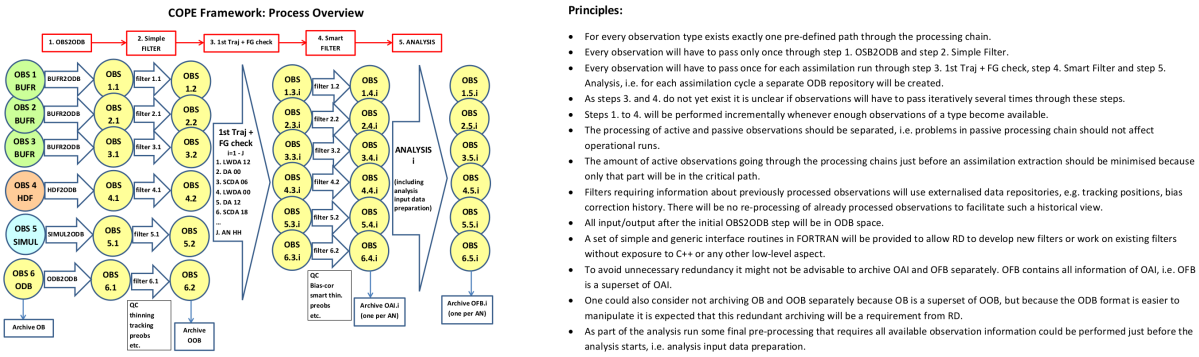


Fig. 1: COPE framework

2 Summary of the technical meeting

To intensify COPE collaboration a kick-off meeting about practical issues related to COPE cooperation and coordination was proposed for 2-3 June in Toulouse. Unfortunately, HIRLAM people could not attend it. Therefore, one week later the second meeting was organized at ECMWF. Finally I attended the COPE meeting at ECMWF together with HIRLAM colleagues Eoin Whelan (Met Ireland), Mats Dalhborn, Bjarne Amstrup (DMI). Drasko Vasiljevic, Enrico Fucile, Tomas Kral, Alfred Hofstadler, Peter Kuchta, Peter Lean (ECMWF). Both meetings have similar agenda, the second one included more time for practical testing of available tools.

The first part of the COPE technical meeting was devoted to general presentations.
E. Fucile: Presentation of the new data acquisition environment at ECMWF - SAPP
Status:

- Scalable Acquisition and Pre-Processing (**SAPP**) system for observations;
- internal ECMWF development to build a frame-work for observation handling;
- designed to support COPE and to provide full real-time monitoring of observation data flow;
- operational since mid June 2014, but with "old" decoders (based on BUFRDC) and some simple QC and conversions;
- scalability improved performance, e.g. 1 year of obs pre-processing with the old system took about24hours, while now it takes about 1hour (on 10 parallel processes).

Plan:

- "old" decoders to be replaced by **ecCodes**;
- simple QC and conversions will be externalized and to be applied via **filters**.

Relevance for ALADIN/LACE:

Data acquisition (including decoding and monitoring) is usually done by observation department of NMS, not by NWP group. NWP group (or OPLACE) takes care of format conversions to the suitable format for BATOR program. So there is no direct relevance for OPLACE.

The SAPP system might be of interest for ALADIN/LACE countries, but at the moment is not ready for remote implementation, although it should not be too complicated (SAPP is based on Linux server, python, mysql, push service/tool, Django). We should keep in mind the new structure - separate pipe-line based processing of each observation type, and consider eventual preparation/adaptation of local observation handling for this approach.

Enrico Fucile: Presentation of the new data decoding environment at ECMWF : ecCodes

- decoding/encoding software for binary and text codes;
- based on the key/values concept as grib_api;
- with same functions and tools, python, C and Fortran interface as grib_api.

Status:

- a prototype currently supports METAR and ECMWF BUFR decoding;
- new tools available:

  bufr_new_from_file (C, Fortran, python)

  bufr_dump (new option j for json dump),
  bufr_filter, bufr_copy, bufr_ls, bufr_get
- No documentation available. Please use grib_api documentation.

Plan:

- Import the code tables in the database and produce the def files;
- Implement decoding for multi-subset uncompressed messages (already implemented for compressed);
- Implement encoding:
- Implement few missing operators;
- Implement a web interface for the database of elements, descriptors and code tables;
- Finalize the abbreviations for elements and code tables;
- Integrate grib_api into ecCodes.

Relevance for ALADIN/LACE:

Current data processing is rather complicated and a common tool to decode "any format" can be very helpful for obs teams or small NMSs, where everybody does a bit of everything. The ecCodes will be most probably part of "a new common conversion to ODB" and following areas of possible collaboration were identified:

- unique definitions of key names for BUFR;
- an "engine" exists, but local BUFR decoding requires work on configuration files (ASCII);
- decoding rules for the conventional data type (e.g. SYNOP) in ecCodes;
- testing and feedback on use of the system and documentation;
- extension/development of other formats (GRIB need checking, local BUFR decoding, TAC, HDF5).

Tomas Kral: Development of observations filters in COPE

- conceptually, observation processing can be seen as a sequential application of various transformations on each report in the observation database;
- the idea is to break the complex processing task into smaller, manageable steps that can be chained one after another, e.g.

- different quality checks, unit conversions, computation of derived variables, observation error assignment, bias correction
- thinning, blacklisting, first guess check
- redundancy checks (too difficult to externalized)

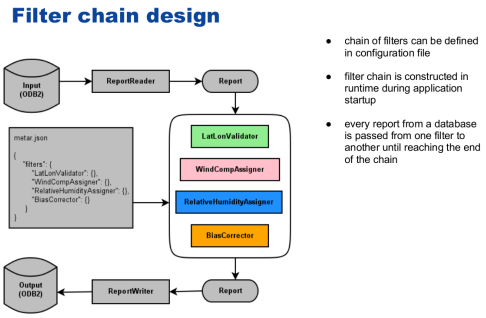


Fig. 2: COPE filters

Benefits

- Better code reuse. Filters implement only a single functionality and can be easily used to construct different solutions.
- Reduced coupling. Filters do not share state and are unaware of other filters.
- Improved flexibility. Filter chains can be constructed dynamically at runtime

Disadvantages

- Identifying reusable filters in a very complex system can be challenging task
- Implicit order dependencies can play important role in choosing the proper sequence in which the filters have to be applied (e.g. observation errors are depending on the vertical coordinate)

Status:

- presently filters can only handle ODB-2 input and have to be written in C++.

Plans:

- to develop Fortran interface;
- assess whether PREOB jobs and thinning tasks can be implemented in terms of filters;
- externalize first guess checks.

Relevance for ALADIN/LACE:

The filters are currently embedded in various observation pre-processing components. Collaboration is possible on externalization/identification of "filters" included in the ALADIN/LACE pre-processing (OULAN, BATOR, screening) chain, e.g. geographical selection, specific treatment for new observations.

Alena Trojakova: Observation processing in ALADIN/LACE

- ALADIN/LACE observation processing is closely linked with Météo-France, see Fig 3.

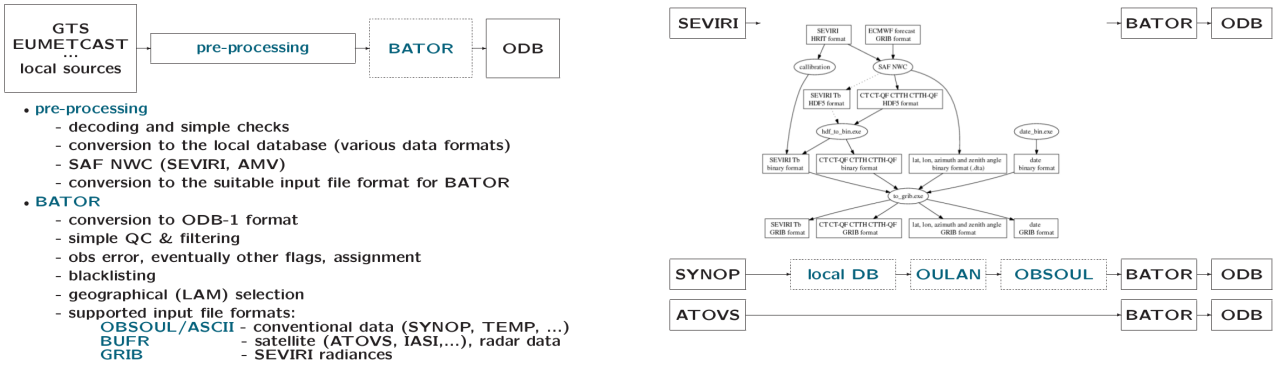


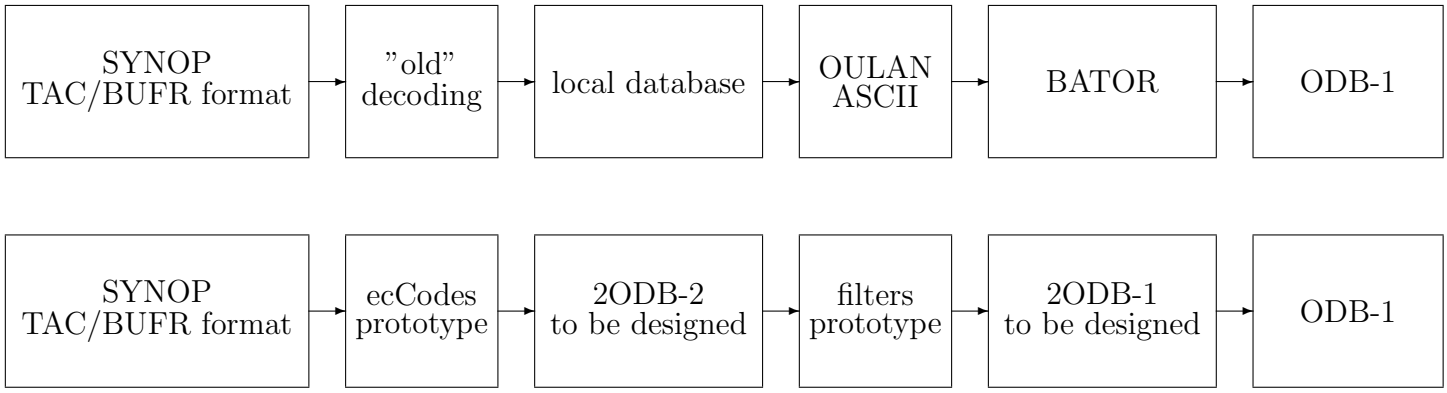
Fig. 3: ALADIN/LACE observation handling

- Observation Pre-processing system for LACE (**OPLACE**)
  - common system for LACE Members (AT,CR,CZ,HU,SI,SK,RO) operated at Hungarian Meteorological Service since 2009
  - aimed to provide real-time obs for NWP and verification purposes and to share efforts on installation and maintenance
  - observations provided in hourly (+/- 30 min) time-slots with a quasi-continuous update, in appropriate formats for BATOR

The rest of the meeting was dedicated to informal discussions, installation and testing of available tools: ecCodes, cope filters, Odb-1.0.0-Source and odb_api.

There are still open questions regarding ODB (dependency of IFS cycles, format). Currently there are two ODB formats. The ODB-1 is column based database (directory with many files), used in IFS because of efficiency in memory and paralelization. ODB-2 is row based (single file), used for archiving because is efficient for disc access. Conversion from ODB-1 to ODB-2 (so-called transposition) is rather easy, while ODB-2 to ODB-1 (inversion) is more difficult. There is ongoing work on ODB libraries harmonization and an interface between ODB-1/2 formats will be developed, which will be hopefully transparent for users.

There was a proposal to test the design by building two examples of SYNOP and satellite obs within the new framework, e.g.



3 Summary

In 2014 the COPE project will focus on building infrastructure components, and create a proof-of-concept prototype for two observation types (one conventional and one satellite data type) from the incoming telecom message up to the use in the assimilation and archiving. It will not produce an operational system, nor will it renew all the existing decoders, filters and ODB loading software. The main infrastructure components are:

- ecCodes: supporting GRIB, BUFR, and try to include ODB,
- Harmonized ODB libraries and interfaces, ideally following ecCodes concepts,
- Simplified filter interfaces, and
- MARS interfaces.

With these components the processing chains for the two chosen data types will be built. These contain:

- Decoders (telecom message to BUFR),
- obs2ODB (bufr2ODB), and
- Filters, facilitating pre-obs and quality control.

Météo France showed interest in collaborating with two persons. These persons would be working under the ECMWF work package managers (Enrico and Ioannis). They would be mainly working on the following items:

- Unique name definitions for BUFR (and ODB),
- Decoding rules for the conventional test data type (e.g. SYNOP) in ecCodes,
- bufr2ODB for the conventional test data type in FORTRAN + work on factorisation to develop abstract framework, and
- Filters: analyse pre-obs and QC in existing systems (BATOR, ECMWF decoders, pre-obs, IFS) and try to harmonise between partners to define common filter functionality.

HIRLAM and ALADIN/LACE also expressed their interest in collaborating on the same topics. At this time it is not clear which synergies can be found, and what resources can be brought to the main project. In any case these LAM consortia could provide valuable feedback on the ease of use of the COPE framework by extending its use to LAM specific observation sets.

A design specification with an outline time and resource plan will be established during the next two weeks and shared with the partners for information, but also in order to confirm availability of resources. The COPE collaboration is planned as a long-term commitment and therefore certain topics concerning the collaborative software development, software licenses and software distribution will have to be discussed further.

Actions required:

- to investigate feasibility of Apache-2 as software license
- to find manpower/resources for collaboration

4 Addendum

Here follows the summary of ECMWF/Météo France/HIRLAM wrap-up video-conference on 17 June 2014. I didn't attend due to clash with other meeting.

- Météo France commented that there is now a better understanding of COPE and that they will be able to collaborate within the project mainly on the decoders and the filters.
- two persons will work on COPE for the rest of the year: Frank Guillaume 50% and Patrick Moll 25% of his time.
- HIRLAM will also provide the equivalent of half a person for the rest of the year, with Eoin Whelan providing the main resource.
- ECMWF will work over the next two weeks to do some more detailed planning of the work packages and then share these plans with the collaborators to agree on who should work on what. This can either be done via e-mail, or if possible we can organise another video-conf in July. Once we agree on the work we will start some rapid prototyping to try out a few ideas and settle the design.
- On IPR/software licensing Météo France mentioned that there is some reservation on going Open Source with all of the software because some of the developments are seen as strategically important and should not be made available to the general public. Baudouin Raoult explained that Apache-2, which is already used by ECMWF, allows the software to be used in a closed context. This opens up the possibility to use Apache-2 for the main infrastructure components like ecCodes, ODB, but that the collaboration partners will have the option to exclude some of the software from the main distribution, and either share it only between a small number of collaboration partners (similar to the IFS) or keep it completely private within Météo France. HIRLAM is fine with an Apache-2 distribution for all software they contribute to the project.
- Météo France asked if the COPE software would be included in the IFS package, or versioned separately. ECMWF made clear that the software would be developed and versioned separately, but that for each IFS distribution it would be clearly defined which version of the different tools (e.g. ecCodes, ODB) would be required and supported.
- Baudouin Raoult also stressed that Data Governance would be an important aspect for a successful collaboration. Strong governance on the data representation will be required to ensure that the different software developments will work for each of the collaborators. The collaborators will have to agree on the semantics (i.e. what numbers will represent what names). Fredi explained that ECMWF will try to anchor the data representation in the WMO formats (e.g. BUFR) because these formats are the most strongly governed formats. All other representations, like ODB, will be linked back to the BUFR representation. The Data Governance will not come for free and will require some effort on all sides, but ECMWF has some experience and is hopeful that it will work.

Acknowledgment
The author wishes to thank the ECMWF staff for their warm welcome and hospitality. This stay was made possible by the financial support of the LACE (Limited Area model for Central Europe) project.

The list of contacts:

- ECMWF has allocated several persons to work part-time on COPE in 2014
  Alfred Hofstadler, Production Section, project manager
  Drasko Vasiljevic, Production Section, deputy project manager and international coordination
  Enrico Fucile, Production Section, system design, ecCodes
  Cristiano Zanna, Production Section, observation monitoring
  Ioannis Mallas, Production Section, decoders, externalising QC from pre-processing
  Marijana Crepulja, Production Section, decoders
  Tiago Quintino, Development Section, software design, Hub-ODB, ODB, ecCodes
  Shahram Najm, Development Section, ecCodes
  Peter Kuchta, Development Section, ODB
  Peter Lean, Satellite Data Section, ODB
  Tomas Kral, Data Assimilation Section, filters, externalising QC from IFS

- Météo-France:
  Frank Guillaume (frank.guillaume@meteo.fr), ecCodes, filters
  Patrick Moll (patrick.moll@meteo.fr), decoders

- HIRLAM:
  Eoin Whelan, Met Ireland, (eoin.whelan@met.ie), Research and Applications Division, ODB, ecCodes, filters
  Mats Dahlbom, DMI, (mda@dmi.dk), DMI/CMM (Centre for Meteorological Models), ecCodes, filters, decoders
  Bjarne Amstrup, DMI (bja@dmi.dk), DMI/CMM, ecCodes, fiters

- ALADIN/LACE:
  Alena Trojakova, (alena.trojakova@chmi.cz), CHMI/NWP, filters