Standardised data representation - power of reproducible work-flow

Marijana Crepulja
marijana.crepulja@ecmwf.int

Forecast Department – Observation team
Introduction

- Hight level – ECMWF Workflow
- Earth Observations
- Observation data governance
- Conclusion
ECMWF Workflow - from observation to weather forecast

- Acquisition
- Pre-processing
- IFS Analysis and Forecast
- Product Generation & Dissemination

Member States And Customers

BUFR
ODB
Fields
MARS Archive
Earth observations - key to understanding weather, climate and hydrology

- Observations are key to our understanding weather, climate and hydrology.
- Combining the models with the observations we can identify
  - uncertainty in the observations
  - uncertainty in what we predict
  - discover systematic errors (bias)
- Improve instruments and refine the models.
- Adding new and better-quality observations successively enhance our forecast.
Earth observations pre-processing

The **Binary Universal Form for the Representation** of meteorological data is a **binary** data format maintained by the **World Meteorological Organization** (WMO).
BUFR - WMO binary code

- **Binary Universal Form for Representation** of meteorological data governed by WMO.
- Continuous bit stream made of sequence of octets.
- Used to encode in situ and satellite observations.
- Self-descriptive code and machine independent.
- Compression available for improved transmission speed.
- Table driven data format. Authoritative definitions with encoding information unit of measure and precision, derived from ‘scale’, ‘reference value’ and ‘data width (bits)’.
- A new version of the tables which are part of the manual is released externally twice a year.
Collaboration and Data governance

Internal and external stakeholders

- ECMWF works closely with space agencies and other data providers with respect to observations data governance to meet ECMWF and Member and Co-operating States requirements.
- In particular, concerning the content and format of new observations.

Inter-Programme Expert Team on Codes Maintenance

- New definitions of data representation
- Regulations on data exchange
- New editions
- Validations of proposed data models
ECMWF Data governance

- Support current international standards and develop those which will be used in the future.
- Dedicated space on Confluence for collaboration with Research Department and external stakeholders (data producers and data users).
Observations - standardised representation

• Exchanging data into dedicated develop data model to enhance seam-less use of data in NWP operation, verification and archiving. Efficient use/re-use data in pre-operational new model cycle, reanalysis.

• One software application can handle data from various data providers, satellite/instruments represented in standardised format.

• Originating centres providing Radio occultation observations
  - EUMETSAT for GRAS data
  - DMI for GRAS data
  - UCAR for COSMIC and KOMPSAT-5 data
  - GFZ for TerraSAR-X and TANDEM-X data
  - CMA for FY-3C/D data
  - ISRO for Megha-Tropiques data

• Conventional data SYNOP

Data volume increase - challenge

Data acquisition → Data pre-processing

- 10 x more Observations per day

Data Assimilation → Forecast run

- 2000x more data per model time step

Product Generation → Dissemination

- 30 x more data disseminated per day
- 100x more data archive per day

BUFR ODB GRIB

10 x more Observations per day

2000x more data per model time step

100x more data archive per day
Conclusion

• Standardised observations to ensuring reproducibility workflow.
• Exchanging and processing standardised data
  - Better quality, seamless introduction in operation
  - Use/re-use data in pre-operational new model cycle, reanalysis
• Facing high volume diverse conventional and space-based observations in future
  - Develop different software application for handling data from different providers in different data formats (time consuming, readiness of data archiving, re-using data)?
  - Is it better to invest in data standardisation?
THANK YOU