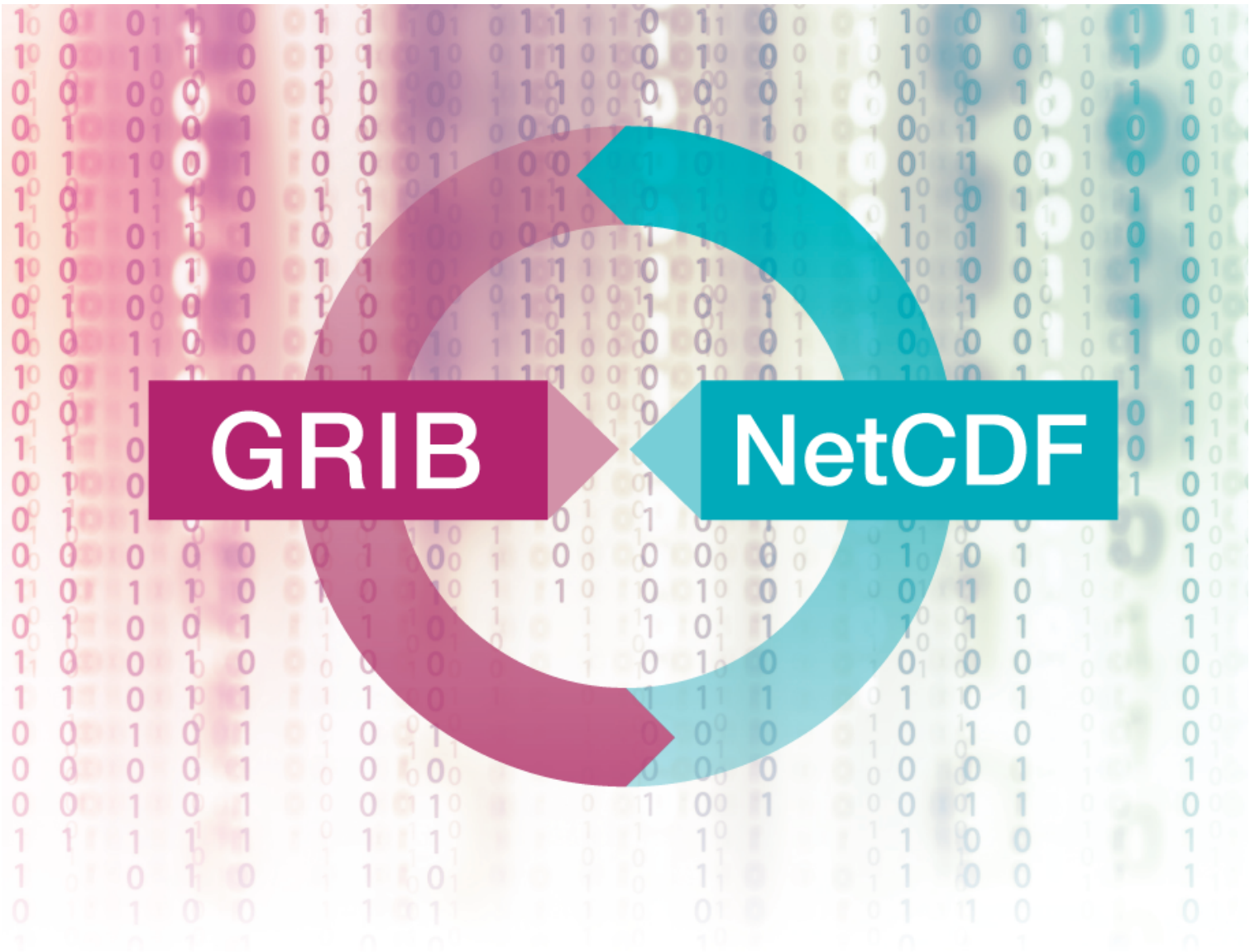

Closing the GRIB/NetCDF Gap Workshop Proceedings

24–25 September 2014

M. Fuentes, K. Marsh

ECMWF

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

1. Introduction

This document presents a summary of the "Closing the GRIB-NetCDF Gap Workshop" which was held at ECMWF in September 2014. Please note that the views expressed herein are those of the attendees themselves and may not reflect those of any organisation to which they may belong.

The form in which data are made available to different user communities can sometimes be a barrier to maximising data use and reuse. Data presented in GRIB (GRIdded Binary) may not be used by groups who work primarily with NetCDF (Network Common Data Form), and vice versa. The quantity and quality of the metadata provided can also be variable, and restrict interoperability.

The Meteorological Archival and Retrieval System (MARS) archive at ECMWF has around 50 PB of data stored, much of it in GRIB format. There is a growing need to make more of these data available in NetCDF format to encourage their usage within new user communities. Also, while GRIB will remain the main data format in MARS, new datasets are being generated at ECMWF which are in NetCDF (e.g. ocean model outputs). These will also require archival in MARS, and a clear "roadmap" for the way forward needs to be developed to enable this to happen.

The "Closing the GRIB-NetCDF Gap" Workshop at ECMWF was an ambitious first step in formalising the discussions around this important issue and formulating that "roadmap". An initial outline of a possible "roadmap" based on the plenary discussions of the findings of the working groups is also presented within this document.

2. Workshop aims

The development of this "roadmap" was one of the main aims of the workshop. It was decided that this could best be facilitated by bringing experts from the various domains together to share knowledge and collaborate on finding solutions. These included specialists in the Climate and Forecasts (CF) Metadata conventions, WMO experts in Table Driven Code Forms (GRIB) and NetCDF. The issues to be discussed included many issues such as how mappings between GRIB and the CF Metadata convention can be established and maintained, and the governance models of the two communities.

Another important outcome from the workshop is to establish how these governance models can work together. This will also allow us to ensure that CF-NetCDF data are fully supported within MARS at ECMWF, and that the data produced are used to their maximum potential.

3. Workshop structure

Over 30 experts were invited to the two-day workshop, and invitations were also sent to members of ECMWF's Technical Advisory Committee (TAC) so that they could nominate delegates to attend the meeting. In most cases where an invited expert from a given institute was unable to attend, they were able to suggest a colleague with a similar level of expertise to represent them at the workshop. In addition, numerous ECMWF staff attended the presentations at various times and took part in the Breakout Group discussions. The full list of attendees is given in Appendix 1 of this document and the programme is in Appendix 2. Please note again that the views expressed by the participants contained within this document are their own personal opinions.

The presentations were streamed live via WebEx on both days to expert colleagues at UNIDATA in the United States. All of the presentations in PDF form along with additional workshop information are available at: *http://www.ecmwf.int/en/workshop-closing-grib/netcdf-gap*.

4. Workshop summary

At the start of Day 1, a welcoming address was given by Florence Rabier, Director of Forecasts at ECMWF. This outlined the goals of the workshop, and was followed by a presentation from Baudouin Raoult of ECMWF which 'set the scene' for the meeting. Over the next few hours, there were 13 invited presentations from various domain experts which mainly dealt with aspects of governance and data standards, and which promoted lively discussions.

The meeting was then split into two working groups to begin discussions on the chosen themes for the Workshop – "Mappings" and "Governance", which were chaired by John Caron (UNIDATA) and Simon Elliott (EUMETSAT). These broad themes were selected as it was felt that they reflected the "Present" and "Future" of GRIB-NetCDF interoperability.

Before the workshop, an on-line questionnaire had been sent to the external delegates asking them several questions, including which issues they felt were the most important to address at the meeting. It was these responses which formed the basis of the discussions under the guidance of the chairs of the working groups. These questions are listed in Appendix 3.

A plenary session was held at the end of Day 1 to present the initial findings from the working groups and discuss the issues raised so far.

Day 2 followed a similar structure, with eight presentations describing issues such as "real-world" implementations in archives, and software tools which are useful when converting data between GRIB and NetCDF. These presentations also stimulated further discussions between the attendees.

The remainder of Day 2 was taken up by further working group discussions. The "Mappings" group decided that there were several important areas of interest, and so subdivided themselves into groups concerned with the issues of "Parameters/Standard_Names/Units and Statistical Processes/Cell Methods", "Tools", and "Mappings".

A final plenary session was held at the end of Day 2 where the initial findings of the four working groups were presented to the meeting by the lead representatives of each group. These are presented below, and it should again be noted that these summaries only reflect the views of those experts who were actually present at the workshop. The workshop closed with Manuel Fuentes of ECMWF giving a summary of the meeting, and thanking all the participants for their valuable contributions.

5. Workshop outcomes – a "roadmap" for a way forward

A 'roadmap' for the possible way forward is outlined below, which arose as a result of the workshop discussions and findings of the working groups.

Set up a dedicated wiki area and mailing list

- *This 'first step' could possibly be done under IS-ENES, and publicised by existing means.*

Select a core set of features initially

- *Such as the master tables of GRIB edition 2 and CF-NetCDF.*

Extract mappings information from existing tools

- *Utilise existing knowledge as another 'first step'.*

Store mappings information in 'METARELATE'

- *Use an existing system to store the information.*

Store additional information on tools and mappings in wiki area

- *So that knowledge can be shared amongst the community.*

Create sample files as examples of 'best practice' and guidelines

- *To encourage a more consistent approach amongst data producers and providers.*

6. Plenary presentations and discussions

The following sections contain the findings of each of the working groups/sub groups as they were presented to the plenary session. These are followed by bullet point summaries of the corresponding plenary discussions; these formed the basis of the proposed "roadmap".

Working Group: Governance

Plenary presentation

Lead/Rapporteur: Simon Elliott (EUMETSAT)

- Best practice recommendations or guidance (not regulations).
- WMO may refer to the best practice.
- Need collaboration tools (wiki, tickets, email).
- Possible stakeholders: WMO ICT-ISS, CF governance panel, WCRP infrastructure panel (WIP), UNIDATA; they would govern the collaborative effort to determine the mapping and the best practices.

To do

- We will start with a bottom up approach to get experience and good understanding of the problem.

Plenary discussion summary

- Engage with the stakeholders to create a group for the governance of mappings. At first, this will be a core set of features, such as the master tables of GRIB Edition 2 and CF-NetCDF.
- Trial 'METARELATE' to store such mappings.
- These on-going discussions should form the basis of "Best Practice" guidelines.
- It is very important to keep sharing relevant information.

Working Group: Mappings (*Subgroup: 'Tools'*)

Plenary presentation

Lead/Rapporteur: Enrico Fucile (ECMWF)

Tools

- grib_to_netcdf (part of grib_api)
  - Mostly ECMWF GRIB
  - Configurable, value packing with offset/
  - Contact: Software.support@ecmwf.int
- NetCDF java library
  - Mostly NCEP GRIB (?)
  - Contact: John Caron
- CDO (climate data operator)
  - CDI
  - Contact: Luis Kornblueh
- Iris
  - Load GRIB/save as NetCDF
  - Contact: Mark Hedley
- CDAT
  - http://www2-pcmdi.llnl.gov/cdat (note cdat is now part of uv-cdat, http://uvcdat.llnl.gov/, Contact: LLNL)
  - BADC
- NCL
  - Contact: Dave Brown at NCAR
  - https://www.ncl.ucar.edu/contributors.shtml

Registries

- http://metarelate.net
- http://codes.wmo.int
- http://reference.metoffice.gov.uk
- http://vocab.nerc.ac.uk

Documents

- http://cfconventions.org
- http://old.ecmwf.int/publications/manuals/grib_api/ (Contains more than just grib_api)
- http://www.wmo.int/pages/prog/www/WMOCodes.html
- http://www.opengeospatial.org/standards/netcdf
- http://cfconventions.org/standard-names.html

To do

- Need to contact the contact listed if a representative was not present at this meeting

Plenary discussion summary

- Existing tools already have some 'knowledge' on mappings. We need to extract some of that knowledge and feed it into 'METARELATE'.
- Luis Kornblueh volunteered to approach IS-ENES to investigate if they would be willing to host a wiki site to support the on-going discussions.
- There is a need to classify the tools which are available, as some of them are able to map both ways between GRIB and NetCDF.
- Mark Hedley volunteered to create an extensive list of tools and to contact the authors.
- Additional tools proposed:
  - Fimex http://fimex.met.no: Contact: Heike Klein.
  - NCL http://ncl.ucar.edu/: Contact: Mary Haley ncl-talk@ucar.edu.

Working Group: Mappings (*Subgroup: 'Mappings'*)

Plenary presentation

Lead/Rapporteur: John Caron (UNIDATA)

File structure of NetCDF

- How to convert collection of GRIB messages → NetCDF file.
- Topics and notes (not meant to be definitive).

General

- Our intention is to create a shared document to identify issues, collect knowledge of how to create CF compliant files from GRIB, and add advice on "best practices", when not otherwise prescribed by CF. (Who should host this?)
- We will need to extend CF, then feedback working implementations into CF.
- CF should standardize on existing practices, so extension to CF must be tested and when successful, standardized. Similar to "local tables" in GRIB-2.
- CF may modify or not accept proposal. So then we have a choice to change possible already written files, or create an extended CF Conventions.
- CF is oriented towards making files easy for readers to use, not to make writers easy or efficient.

When should all information be put into one file, when split into multiple files? Best practices?

- Single horizontal coordinate system (GDS) per file.
- Single reference time per file (records from a single model run).
- Optionally can make separate files split by variable.
- Optionally can create time series for single variable and single level (e.g. what nco can do).
- When possible, should be dictated by storage needs (e.g. efficiency of the mass store). User should "see" logical collections, or create subsets as needed.

Which dimensions should be unlimited – time and level?

- Unlimited means that the variable can be appended to along that dimension, otherwise dimension lengths are fixed at file creation.
- Can have multiple unlimited dimensions in NetCDF-4 (only one in NetCDF-3).
- Unlimited has strong effect on NetCDF-3; it radically affects the storage layout, and thus the read performance. For NetCDF-4, layout is controlled by the chunking strategy.
- Very important to understand consequences of chunking in NetCDF-4.
- UNIDATA's nccopy can copy files and change chunking. We think it is mostly efficient but would like feedback.

Choosing variables

- How to define which GRIB records go into a variable?
- When do you use one variable with a vertical dimension versus multiple variables; e.g. "2m temperature" and "10 m temperature" versus temperature (level=2, y, x).
- Need to allow user configuration of this process.
- Is it possible to describe the GRIB messages to NetCDF variable mappings in a scientist-readable format for interoperability with multiple software tools?

Rectangular problem

- When creating nD NetCDF arrays (n>2), may not be "dense" (e.g. all levels for all times).
- Happens a lot, especially for collections of forecasts across multiple runs (Forecast Model Run Collections (FMRC) in the CDM).
- What to do when slicing and concatenating existing NetCDF files – gaps in data due to "non-dense" dimensions.
- Can set to missing data, but how to indicate to user where the data actually exists?

Dimensions

- How are multiple time dimensions handled?
- How are ensemble dimensions handled?
- What happens to dimension order for ensemble axis?
- Order of dimensions: CF prefers time, level, y, x (due to COARDS compatibility), but others are possible (performance/tools must be considered). Chunking can improve performance if used carefully.
- Need for extra dimensions like ensemble (realization) axis. What order to put that axes?
- What to do with dimensions of size 1?

Spherical harmonics

- How to store in NetCDF/CF. Currently no convention covering these (?)
- There is a CF convention for non-rectangular thin grids. But not well used, some consider it deprecated, because compression solves the size problem.
- Non-uniformly spaced coordinates (e.g. Gaussian) are OK.
- Need to convert to rectangular x, y?

Time coordinates

- Non-uniformly spaced time coordinates are OK.
- Different ways of defining time axes/forecast start-time and lead-time.
- Specify all times relative to, for example, 19000101 or relative to forecast start-time? The former requires two time coordinates. Maybe we should create a best practice document for archives/distribution?

Plenary discussion summary

- Need to create a shared document/wiki.
- Need to extend the CF Conventions.
- More discussions are needed on many of the issues raised.

Working Group: Mappings (*Subgroup: 'Parameter/standard names/units plus statistical processes/cell methods'*)

Plenary presentation

Lead/Rapporteur: Bruce Wright (Met Office)

- Need to map a group of keys to another group of keys.
- Should only map to CF Standard Names that follow the Standard Name grammar? May need to add new Standard Names.
- Discussions on handling of Statistical Processes – believe both standards support what is required from the other.
- Do we expose these concepts (as for GRIB-API)? Probably not, as there will be thousands, most with no common use.
- Metarelate exists in a development form and it maps one source to one target, but both source and target can be arbitrarily complex in terms of combinations of keys.
- It uses the common keys in GRIB2 (rather than looking at each template in GRIB2) to reduce the number of mappings and risks of ambiguity.
- It uses long random strings for the mapping names to avoid people using them as identifiers in their own right. So far, it only covers parameters, but could cover the other areas required.

We think it would be worth using 'METARELATE' to gain more experience with the mappings. Possibly go through the templates to look at the additional mappings required or take a small set of parameters (e.g. some of the standard CMIP5 parameters) to (a) see if they are in the database or (b) how we would add them.

Plenary discussion summary

- We need to clearly define what a 'parameter' is.
- Does it include the vertical co-ordinate?
- How does it relate to the standard_name?
- How does it relate to a variable name?
- May need to create many new CF standard names
- Should there be engagement with the CMIP conventions to see how they handle similar issues?
- The formation of a group consisting of GRIB and NetCDF experts is key.

Appendix 1: Workshop attendees

Note that the views expressed by the participants contained within this document are their own personal opinions and do not necessarily represent those of their organisations.

Pierre-Antoinne	Bretonnière	*IC3*
Paul	Berrisford	*ECMWF*
John	Caron	*UCAR/UNIDATA*
Michael	Claudon	*Meteo France*
Francisco	Doblas-Reyes	*WCRP;CMIP5; IC3*
Prashanth	Dwarakanath	*LiU/NSC*
Simon	Elliot	*WMO Chair of DRMM; EUMETSAT*
Enrico	Fucile	*Member of DRMM; ECMWF*
Manuel	Fuentes	*ECMWF*
David	Hassell	*Reading University*
Rosalyn	Hatcher	*Reading University*
Mark	Hedley	*Met Office*
Martin	Juckes	*STFC*
Heiko	Klein	*Norwegian Meteorological Institute*
Michael	Kolax	*SMHI*
Luis	Kornblueh	*MPI*
Sibylle	Krebber	*DWD*
Per	Lundqvist	*LiU/NSC*
Matthew	Manousakis	*ECMWF*
Kevin	Marsh	*ECMWF*
Chris	Little	*Met Office*
Kristian	Mogensen	*ECMWF*
Shahram	Najm	*ECMWF*
Kevin	O'Brien	*NOAA*
Alison	Pamment	*STFC*
Matthew	Peroutka	*WMO member MDRD; NOAA*
Paul	Poli	*ECMWF*
Tiago	Quintino	*ECMWF*
Baudouin	Raoult	*ECMWF*
Atsushi	Shimazaki	*WMO*
Stephan	Siemen	*ECMWF*
Hamish	Struthers	*LiU/NSC*
Jeremy	Tandy	*WMO Chair of MDRD; Met Office*
Eizi	Toyoda	*WMO Co-Chair of MDRD; member of DRMM; JMA*
Sébastien	Villaume	*SMHI*
Jörg	Wegner	*DKRZ*
Bruce	Wright	*Met Office*

Appendix 2: Workshop programme

Wednesday 24 September		
09.30-10.00	Registration and coffee	
10.00-10.10	Welcome/housekeeping	Florence Rabier (ECMWF)
10.10-10.40	GRIB and NetCDF: Setting the scene	Baudouin Raoult (ECMWF)
10.40-11.00	Parameter naming in GRIB and CF	Alison Pamment (STFC)
11.00-11.15	Applications of the CF data model	David Hassell (University of Reading)
11.15-11.35	***Coffee break***	
11.35-11.50	The role of WMO inter-programme expert team on data representation, maintenance and monitoring	Simon Elliott (EUMETSAT)
11.50-12.05	Observations and measurements as a basis for semantic reconciliation between GRIB and NetCDF	Jeremy Tandy (Met Office)
12.05-12.20	Don't solve problems, copy success: Leveraging standards and conventions to improve interoperability	Kevin O'Brien (NOAA)
12:20-12.35	Lesson from incompatible units of time-integrated GRIB parameters	Eizi Toyoda (JMA)
12.35-12.50	NetCDF metadata standards for climate model intercomparisons	Martin Juckes (STFC)
12.50-13.05	TBD	
13.05-13:45	***Lunch break***	
13.45-14.00	GRIB and NetCDF in a world of competing standards	Matthew Peroutka (NOAA)
14:00-14.15	GRIB to NetCDF/CF as part of Unidata's THREDDS project	John Caron (UCAR)
14.15-14.30	The challenges of using NetCDF and GRIB for managing forecast data at the Met Office	Bruce Wright (Met Office)
14.30-14.45	Software for managing and sharing metadata translation information	Mark Hedley (Met Office)
14.45-15.00	ECMWF data decoding tools for users and operations	Enrico Fucile (ECMWF)
15:00	**Working groups**	
15.30	***Coffee break (in meetings)***	
16:30-16.45	Climate Data Operators (CDO)	Luis Kornblueh (MPI)
16.45-17.00	The CF checker tool	Rosalyn Hatcher (University of Reading)
17:00	**Feedback from working groups/plenary discussion**	
18.00	***Close day 1***	
20.00	***Social event***	

Thursday 25 September		
09.00-09.15	Data reference syntax – governing standards within climate research data archived in the earth system grid federation	Michael Kolax (SMHI)
09.15-09.30	SPECS NetCDF convention: Dealing with climate predictions on the ESGF	Pierre-Antoine Bretonnière and Francisco Doblas-Reyes (IC3)
09.30-09.45	A short history of GRIB	Chris Little (Met Office)
09.45-10.00	On-the-fly GRIB to netCDF conversion within the MARS-ESGF integration	Sebastien Villaume (SMHI)
10.00-10.15	Data standardization at DKRZ	Jörg Wegner (DKRZ)
10.15-10.30	Gridded data from many sources – a data-user's perspective	Heiko Klein (NMI)
10.30-10.50	***Coffee break***	
10.50-11.05	Management of GRIB by WMO	Atsushi Shimazaki (WMO)
11.05-11.20	GRIB/NetCDF file compression	John Caron (UCAR/UNIDATA)
11.20-13.00	**Working groups**	
13.00-13.40	***Lunch break***	
13.40.14.25	**Working groups (including summing up)**	
14.25-15.30	**Feedback from working groups/plenary discussion**	
15.30	***Coffee break and close of workshop***	

Appendix 3: Suggested Questions for working groups to address

The sections below present the questions which were collected from participants as part of the pre-workshop questionnaire, and used as the basis for the working group discussions.

"Governance" Questions

- What is expected of GRIB in the future?
- What are the challenges that remain to unifying GRIB and NetCDF?
- What similar efforts are taking place around the world?
- How can we best manage these efforts?
- What are the barriers to modernizing data formats addressed?
- What are the issues? Which are solved (or solvable) and which not?
- How could we develop and authoritative mapping between GRIB2 and NetCDF?
- Who is going to be responsible for creating the initial mappings and maintaining them in the future as GRIB and NetCDF standards evolve?
- Will GRIB become too complex? Will NetCDF governance be regulated?
- Assuming that a mapping table for the metadata is to be implemented - which role will be played by the WMO?
- Can we outline of a clear path to moving to common standards for weather and climate communities?
- Can we develop a plan for making GRIB and NetCDF formats interoperable that meets the needs of both Met Office and WMO?
- Will the WMO include NetCDF/CF as an approved encoding?
- When CF Conventions change, will they be backward compatible?
- Can we gain some insight into the governance of GRIB?
- Can translation information is managed and shared in a trusted manner?
- Can we agree on a time plan for creating working groups and setting up a list of tasks to enable clear GRIB <-> NetCDF conversion?
- Do we need a Governance steering group?
- Are complex data modelling and mapping may the right way forward?
- Can we get a general agreement on how to map GRIB to NetCDF (parameters, units, metadata, file structures…?)
- Can we make sure we are "doing it right"?
- Can we get a general agreement on how we deal with future requirements (new grids, new parameters …)?
- Can we have a tighter collaboration between WMO and the CF community?
- Should/Will there be future workshops like this one? Can we commit to a plan for a way forward?

“Mappings” Questions

- How can we best support GRIB1/2-NetCDF mappings?
- Can GRIB and NetCDF ever be used interchangeably/reliably without information 'loss'?
- Can we develop and authoritative mapping between GRIB2 and NetCDF?
- How should individual GRIB records be stored in NetCDF (1 variable, all times/levels)?
- What are the limitations of CF-NetCDF?
- What are the current recommended tools/mappings?
- Can we keep the mappings more useful than confusing or harmful?
- Can we develop a functional and automated inter-format mapping convention?
- GRIB and NetCDF/CF communities to develop a formal mapping. Include this into the NetCDF-Java/CDM library.
- How seriously are the leading operational NWP centres adopting (or forced to handle) open standards/NetCDF?
- What is the impact of these mappings on other projects/technologies (MARS, ESGF, etc.)?
- What about Missing/wrong standard names for certain GRIB-parameters?
- Can we agree on standard variable names?
- Which semantics underpin information exchange in GRIB and BUFR? How one can losslessly transform from GRIB to NetCDF and vice versa? How are these semantics encoded in each format?
- Are there significant incompatibles between the metadata structures of GRIB and NetCDF?
- Are there incompatible structures in the protocols for encoding information (i.e. encoding of data in GRIB vs. NetCDF CF convention)?
- Can translation information be managed and shared in a trusted manner?
- Can we agree on a time plan for creating working groups and setting up a list of tasks to enable clear GRIB ⟷ NetCDF conversion?
- Can we develop a plan for making GRIB and NetCDF formats interoperable that meets the needs of both Met Office and WMO?
- Where are master/local tables published?
- Where will the mappings be officially documented? WMO website? CF website, modelling centres?
- What if there is no mapping for a parameter?
- Do we need a “Mapping” steering group?