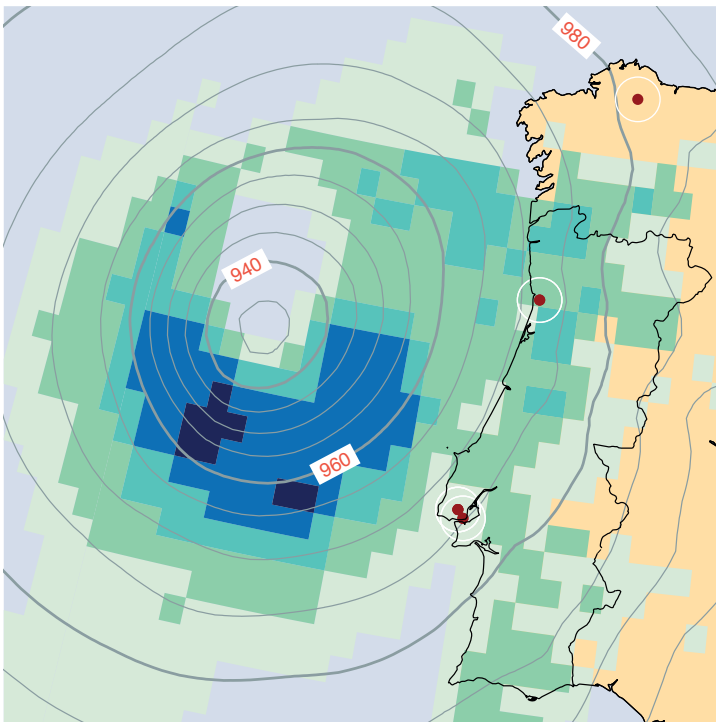




COMPUTING

Migration from GRIB1 to GRIB2: preparing ECMWF model output for the future



This article appeared in the Computing section of ECMWF Newsletter No. 175 – Spring 2023, pp. 32–37.

Migration from GRIB1 to GRIB2: preparing ECMWF model output for the future

Robert Osinski, Matthew Griffith, Sébastien Villaume

In 2022, ECMWF started a multi-year effort to migrate its daily operations data output from the file format GRIB edition 1 (GRIB1) to GRIB edition 2 (GRIB2). The project is partly a response to the call for global numerical weather prediction (NWP) at convection-permitting resolutions set out in ECMWF's ten-year Strategy 2021–2030. Such resolutions require GRIB2 rather than GRIB1 data because of the limitations of GRIB1 grid definitions. ECMWF has already produced Integrated Forecasting System (IFS) output on vertical model levels using the GRIB2 format for several years. Here we present an overview of where we are in the transition from GRIB1 to GRIB2 for all of our output.

Setting the scene

GRIB1 was created in 1985 and was commonplace by the early 1990s. It was not designed to accommodate horizontal grid resolutions needed to resolve convective-scale phenomena (1–4 km), which are called for in the ten-year Strategy and which will be used in the digital twins for the EU's Destination Earth initiative, in which ECMWF participates.

GRIB1 also has other limitations and disadvantages. The most important one is that GRIB1 was deprecated by the World Meteorological Organization (WMO) more than a decade ago in favour of GRIB2, and that it has not been referenced in the WMO Manual on Codes since 2016 (World Meteorological Organization, 2022). Other limitations of GRIB1 include:

- A maximum vertical resolution of 127 levels; this limitation was hit in 2011 and is the reason why the data on vertical model levels were migrated to GRIB2 during the implementation of Cycle 37r2 (137 vertical model levels).
- GRIB1 allows for the definition of only 128 different parameters and does not have the necessary metadata to describe modern NWP outputs, such as ensembles and probabilities.
- GRIB1 does not have an official built-in mechanism to extend its metadata. Over the years, ECMWF has extended the limited GRIB1 metadata through the permitted local section found in section 1 of the header. This is how ensemble members and ensemble size were introduced in GRIB1. The drawback of these extensions is that they are not endorsed by the WMO and thus not part of the official data format.

GRIB2 resolves these limitations and brings critical new features:

- **Support for vertical resolutions with more than 127 levels:** as above, this was a critical feature when the model was upgraded to use 137 vertical model levels.
- **Support for horizontal resolutions at sub-kilometre scales:** planned resolution increases in ECMWF's ten-year Strategy are well within this scope.
- **Support for millions of different parameters:** NWP parameters can now be encoded with much more freedom.
- **Support for ensemble, re-forecast and post-processed products:** this addresses the most common limitations of metadata to describe products with context.
- **Support for a wide range of compression methods:** this is critical with increasing data volumes at higher model resolutions.
- **Support for rich metadata:** this enables more prescriptive parameter descriptions and improves discoverability and indexing.
- **Introduction of templating:** this allows the continuous integration of new templates when additional metadata is required.

These headlines, some of which are shown in Figure 1, describe some of the fundamental design changes which have been put in place within the GRIB2 format. We discuss these in more detail below.

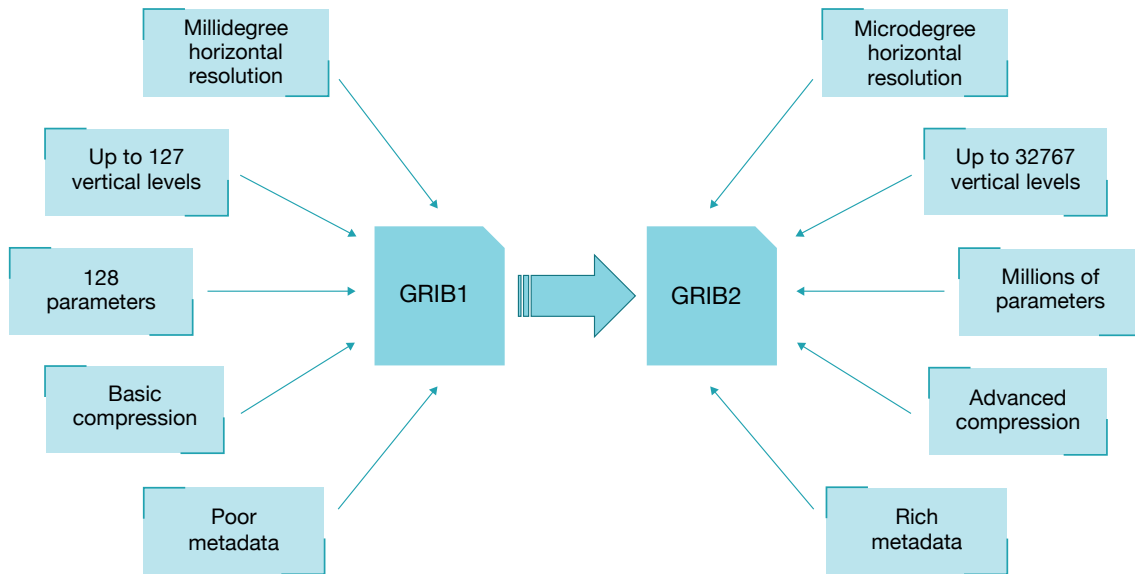


Figure 1 Some of the differences between the GRIB1 and GRIB2 file formats.

GRIB2 design philosophy

The new features and improved design of GRIB2 allow for a much more self-descriptive data format with an improved user experience. Expanding on the above points, GRIB2 encompasses the following:

- **Horizontal resolution:** previously limited to millidegree precision, this can be encoded up to the precision of a microdegree, allowing resolutions below the kilometre scale. This is sufficient to encode data according to the resolution increases planned at ECMWF for the next decade.
- **Parameters:** the total number of parameters that can be defined is in practice unlimited. In GRIB2, a parameter is no longer simply represented by a single entry in a code table but by a combination of entries in various code tables. Thus, the minimum number of metadata keys is now a triplet: discipline, parameter category and parameter number. The top level of the hierarchy gives the discipline within which the parameter is defined, such as meteorology, hydrology or oceanography. Then, within each discipline, the parameters are organised into categories. For instance, in the discipline ‘meteorology’, one can find the categories ‘momentum’, ‘temperatures’ or ‘short-wave radiation’. Finally, one selects a parameter within the category. For complex parameters, additional keys are required. The GRIB2 section 2, called the local section, is reserved for encoding centre-specific, local metadata for a parameter. This is used at ECMWF for Meteorological Archival and Retrieval System (MARS) keys, e.g. class and stream. It has the advantage that the message itself must conform to the standards of the GRIB2 data format.
- **Compression:** the data representation section offers a wide range of compression methods. This helps reduce the size of archived data, which will increase significantly with higher model resolutions. Recently, a fast and lossless compression algorithm with a high compression ratio, developed by the Consultative Committee for Space Data Systems (CCSDS), has been implemented for GRIB2 in ECMWF’s IFS, and this feature will be activated in the implementation of IFS Cycle 48r1 later in 2023 (Betke et al., 2022).
- **Rich metadata:** The metadata set describing a given parameter is rich, enabling better parameter descriptions and improved discoverability and indexing in accordance with FAIR data principles. In the context of the migration-to-GRIB2 project, we have developed and will develop new templates to extend the metadata to enable the encoding of all of our products.

- **Templates:** The grid section, product section and data representation section of a GRIB2 message can now be templated, allowing the continuous integration of new templates when additional metadata is needed. Template extensions can be requested from the WMO twice a year through an amendment procedure of the manual on codes called the ‘Fast Track procedure’. The request is then processed by the WMO over a period of around six months, after which the amendments are published by the WMO and are ready to be used operationally.

To aid the understanding of the approach behind GRIB2, it is good to look at an example. We shall compare the metadata between a GRIB1 and GRIB2 parameter encoding for “Maximum temperature at 2 metres in the last 24 hours”. This is presented in Table 1.

Maximum temperature at 2 metres in the last 24 hours	
GRIB1	GRIB2
Table2Version = 128 (bears no specific meaning)	discipline = 0 (meteorology)
IndicatorOfParameter = 51 (Maximum temperature at 2 m in the last 24 hours)	parameterCategory = 0 (temperatures)
IndicatorOfTypeOfLevel = 1 (surface)	parameterNumber = 0 (temperature)
	typeOfFirstFixedSurface = 103 (height above ground in metres)
	scaleFactorOfFirstFixedSurface = 0
	scaledValueOfFirstFixedSurface = 2
	typeOfStatisticalProcessing = 2 (maximum)
	lengthOfTimeRange = 24
	indicatorOfUnitForTimeRange = 1 (hour)

Table 1 Comparison of metadata describing a meteorological parameter in GRIB1 and GRIB2. The names of the keys correspond to those used in ecCodes, an ECMWF package for decoding and encoding messages in WMO formats. When the value taken by a key is followed by an explanation in brackets, it is because it references an entry in a table. For example, entry 103 in the table representing “Fixed surface types and units” corresponds to “height above ground in metres”.

GRIB2 metadata follows a strategy that can be summarised by the ‘what, where, when, and how’ approach. This is illustrated in Table 1, using a colour-coded key. It can be thought of as follows:

- **‘What** is being encoded?’ This is the base or core of the parameter and is always defined by the keys discipline, parameterCategory and parameterNumber (in red). For this example, we have discipline 0 (meteorology), parameter category 0 (temperatures) and parameter number 0 (temperature). If the parameter requires no more metadata, we could stop here. However, in the vast majority of cases we then use additional keys to extend the scope of the parameter:
- **‘Where** is the parameter defined?’ This refers to the vertical spatial range or spatial position for which my parameter is valid (in purple). Here, it is at a specific fixed level – at 2 metres above the surface.
- **‘When** is the parameter defined?’ This refers to the time range or time point for which my parameter is valid or is processed (in green). Here, we can see that we perform processing over a 24-hour time period. This combines with the ‘how’ key to indicate the kind of processing we perform.
- **‘How** is the parameter processed in time?’ This key tells us how we want to statistically process the parameter in time (in orange). Here, we can see it is ‘maximum’. This combines with the ‘when’ key to give us a maximum in the last 24 hours.

This key-value type design is very powerful and flexible and allows for a direct and intuitive mapping to keywords used in MARS.

Challenges

The migration to GRIB2 poses several challenges. The GRIB format has been tightly coupled to the Centre’s dataflow and many of our tools are designed to take advantage of the GRIB data format. While most tools use GRIB data in a transient manner and will only require migrating once, the MARS archive must continue to handle GRIB1 data properly for decades to come. At the time of writing, MARS has more than 200 PB of data in GRIB1 stored on tapes. Converting this data to GRIB2 to completely deprecate GRIB1 at ECMWF is not realistic as it would require a significant amount of time and resources.

Instead, we are planning to keep serving this data as is but will offer a tool to convert on-the-fly to GRIB2.

Another challenging aspect of the migration will be the implementation of the migration in operations. This will require preparation upstream and should be implemented in the form of a technical cycle (although this has not been decided yet). Test data in GRIB2 will be released more than six months ahead of implementation to enable our Member and Co-operating States and other users to adapt their workflows accordingly.

The biggest challenge of this migration will be to handle the legacy data formatting standards accumulated over the life of GRIB1. The MARS language and GRIB1 have been around since the early 1990s. Throughout the years, both have been extended to accommodate new types of data that could not have been foreseen and planned for during their design phases. These are types of data that are commonplace for ECMWF now, such as ensembles, seasonal forecasts, hindcasts, probabilities, waves, oceanography, hydrology, and land surface modelling. Understandably, this has created technical debt which has accumulated over the years, making certain aspects of the migration very tricky.

In some cases, a direct migration will be impossible and will require some redesign. For instance, GRIB2 prescribes the units of the parameters to specific SI units and does not allow for alternative, equivalent units. A good example is the precipitation parameters produced by the IFS in units “metres of water”, while GRIB2 expects the parameters to be expressed in kg m^{-2} . If we were to switch to producing our precipitation data in WMO standard units, this would create a discontinuity in the archive: a user trying to retrieve data spanning over the transition would receive part of the data in the old units and part in the new units. To solve this issue, we must define two sets of precipitation parameters, one set using standard WMO units and a second set with legacy units. The downside is that this second set is defined locally and not endorsed by the WMO. We could then either produce and archive both sets for convenience or we could only archive our local parameters and offer a conversion on-the-fly to the WMO parameters.

Timeline

A roadmap for the migration has been drafted (see Figure 2). The amount of work and the scale of the changes will not allow everything to be migrated at once. Several factors have been considered to set priorities and derive a workplan:

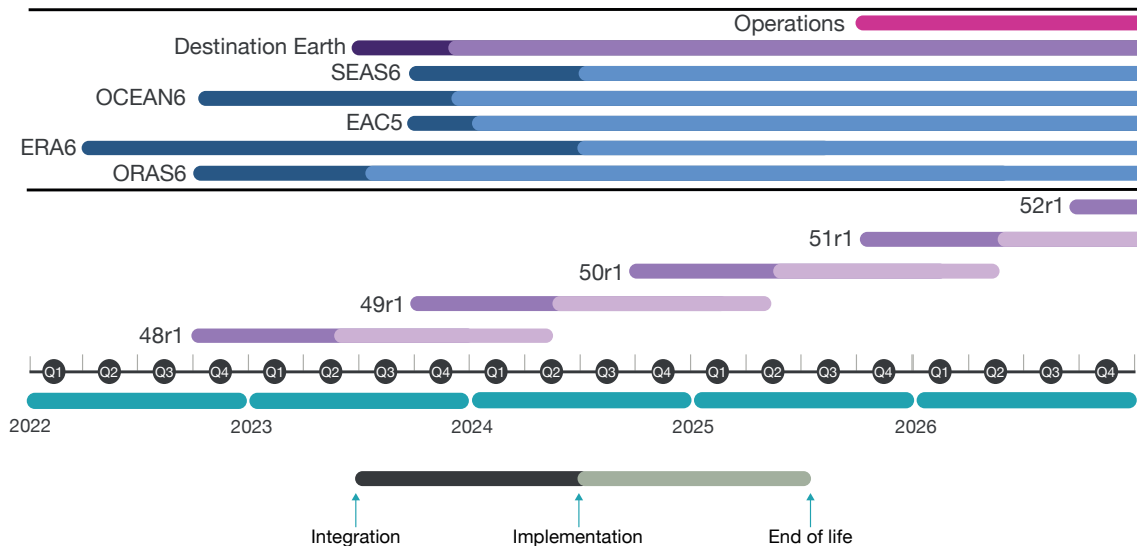


Figure 2 This migration roadmap indicates when GRIB2 is to be introduced in all of ECMWF’s weather forecasting operations (from IFS Cycle 51r1) and when various other services that will use GRIB2 will become operational.

- Any new dataset with a new type of data (not existing in GRIB1) shall be produced entirely in GRIB2. This will be the case for the ocean reanalysis ORAS6 and the real-time OCEAN6. Between 2018 and 2022, the European Flood Awareness System (EFAS), the Global Flood Awareness System (GloFAS) and the Fire Copernicus Emergency Management System (CEMS-Fire) were all released as GRIB2 only datasets following this principle.

- Any new dataset replacing an existing dataset (produced in GRIB1) shall also be produced entirely in GRIB2. By this, we mean any datasets with a well-defined beginning and end. The atmospheric composition reanalysis EAC5 (replacing EAC4), the next global reanalysis ERA6 (replacing ERA5), and the new seasonal forecasting system SEAS6 (replacing SEAS5) fall into this category.
- Any new parameter shall be defined only in GRIB2. This has already been common practice for the past five years. It is the main reason why some surface parameters or non-model-level parameters, in addition to those on vertical model levels, are encoded in GRIB2. This is acceptable because the parameters are new, and therefore they do not introduce a change of behaviour or discontinuity in workflows or in the MARS archive. An example of recently added parameters are new thermal comfort indices, such as the UTCI (Universal Thermal Climate Index).
- Our existing IFS GRIB1 parameters, produced by the operational suite, will be the last to migrate to GRIB2 with an extended period of testing prior to implementation.

Ongoing migration progress

The next dataset in the scope of this work to be released in GRIB2 is the ocean reanalysis ORAS6. The work for this dataset and for OCEAN6 started several years ago independently of this migration project. This is due to the use of unstructured ocean grids (ORCA grids) by the model, which cannot be represented in GRIB1. This makes this dataset a natural candidate for the GRIB2 data format. The ocean grids have now been implemented in GRIB2 and are fully supported by ecCodes 2.20.0 and higher and ECMWF's Meteorological Interpolation and Regridding (MIR) software package. ORAS6 and OCEAN6 are scheduled for production during the second half of 2023.

The next major milestone is concerned with the datasets which will be based on IFS Cycle 49r1, namely ERA6, SEAS6 and EAC5. Early last year, we conducted an exhaustive inventory of all the parameters and concepts needed in GRIB2 for these projects. For EAC5, several hundreds of new parameters would be required, due to two main factors:

- The introduction of many new chemical species and aerosols. For each species, we would need a complete set of physical observables: wet deposition of <species>, dry deposition of <species>, mass mixing ratio of <species>, etc.
- The emissions are now to be resolved by emission sector, leading to yet more parameters: emission of <species> from <sector>. Typical sectors include agriculture, industry, road and volcanoes.

Fortunately, we can use the rich metadata in GRIB2 and the flexibility to extend this metadata through new templates. It is now possible to specify the chemical species or aerosol and the source of emissions through separate metadata keys. The implementation in IFS-COMPO (IFS composition) of this new scheme is well under way.

ERA6 will also require several new implementations to support its release in the GRIB2 format. It will be the first dataset to have wave parameters in GRIB2 including 2D wave spectra (directions and frequencies). These spectra cannot be represented in GRIB2 with existing templates. For the wave spectra and wave parameters, we submitted six new templates to the WMO in November 2022. Additionally, ERA6 will also offer many new parameters, such as new water and energy budget parameters. These parameters, together with the parameters already produced in ERA5, were reviewed, mapped and requested through the WMO approval process when required. The templates and the parameters for ERA6 have just been accepted and will be published in the WMO Manual on Codes in May 2023.

We are also working on other aspects of metadata modelling for the migration. Recent developments, which will also be used in the Destination Earth initiative, include the new snow, soil and sea-ice multilayer schemes. The multilayer snow scheme and corresponding multilayer GRIB2 output will already be introduced in IFS Cycle 48r1. We also looked at how to encode metadata for the Extreme Forecast Index (EFI), the Shift of Tails (SOT), and anomalies based on climate distributions. Four new templates were created to encode these. Finally, we worked on a way to encode the metadata for optical parameters which are wavelength dependent, and we proposed four new templates to achieve this. These templates will be useful for parameters related to simulated satellite images and the radiation parameters used to produce the popular 'space view' images for IFS output.

Future developments

We are currently working on the design of templates to enable the encoding of tile-based parameters. Recent developments in land-surface modelling make use of the partitioning of the grid box into ‘tiles’ or ‘patches’ with their own properties and modelled physical processes. Typical tile classes include high vegetation, low vegetation, oceans, lakes, urban land and bare land. However, tile schemes can also be much more detailed in the granularity of the chosen tiles, including more than 20 different types of tiles accounting for different vegetation types across the globe. This kind of partitioning is particularly useful for parameters like 2-metre temperature, as this has large variations depending on the surface over which it is measured. For example, the effect of urban areas, lakes, oceans, and a forest could all be taken into account by encoding the temperature on the tile it is associated with. We are actively working with modelling teams from various European meteorological services to draft templates that could be used by all major European land surface models.

Finally, a template to encode a new type of horizontal grid, the HEALPix (Hierarchical Equal Area isoLatitude Pixelization) grid, is also in the pipeline. This grid, originally designed for cosmological applications, has recently gained much attention due to its attractive and versatile properties. This template, together with the tile templates, will be submitted to the WMO for validation during the next WMO Fast Track procedure.

Expected user impact

The migration from GRIB1 to GRIB2 is comparable to the migration from the Python 2 to the Python 3 programming language: it will require changes in workflows, from the modification of scripts to changes in existing practices. Certain features or parameters will need to be deprecated, too. However, importantly, this will not require a complete rewrite of applications and tools. As in the case of Python 2 and Python 3, we are expecting both ecosystems of dataflow to co-exist for several years. IFS Cycle 49r1, which is due to be implemented next year, is probably the most relevant example of this. This IFS cycle should be able to run in legacy mode in operations, but it should also be able to run in GRIB2 only mode, for ERA6 and Destination Earth. We are actively working on a technical solution allowing this switch rather than maintaining parallel releases of ecCodes.

The user interaction with GRIB2 messages via ecCodes will be the same as in GRIB1, i.e. mostly by setting/accessing edition-independent keys: `dataDate`, `dataTime`, `paramId`, `typeOfLevel`, etc. However, for certain parameters there will be changes in the representation used and in the method of access in GRIB2. The following are the most common examples of such changes:

- Some parameters will obtain a different `paramId` in GRIB2. The ‘soil temperatures level 1/2/3/4’ are a good example of this. In GRIB1, these are represented by four separate `paramIds` all on a unique level called ‘surface’. In GRIB2, they will all be represented by a unique `paramId` on four discrete soil levels/layers.
- A `paramId` in GRIB2 may need to be complemented by additional keys, for example a `wavelength` for optical parameters or a `chemId` to specify a chemical species/aerosol.
- A pre-existing GRIB2 representation of a parameter may become deprecated. This can happen when the representation was erroneous, incomplete or for other technical reasons. In this case, it will still be possible to read and decode such a parameter, but ecCodes will use the new representation when writing the parameter to a file.

To date, development has focused on changes with limited impact for users. However, we are now entering a phase of the project where we are tackling the more visible changes with higher user impacts. These will be clearly announced in ecCodes release notes and through other appropriate channels of communications (see below). It is therefore recommended to use the latest version of ecCodes at all times to avoid issues in the migration of workflows and to benefit from the latest features.

MARS requests for operational data will also be affected by the migration to GRIB2. This is because the migration will inevitably create a discontinuity at the time of implementation. A user who wants to retrieve data overlapping the transition may need to use separate requests for the retrieval depending on the parameters of interest.

We are still working on the best approach to handle this transition. The options include (a) a minimal impact on the user side at the cost of adding more technical debt on the application side; (b) a balanced approach consisting of compromises on both sides; (c) a disruptive approach, clearing as much technical debt as possible and preparing MARS for the long term. A couple of concrete examples of MARS requests are presented in Table 2.

	Before migration	After migration
Soil temperatures in soil layers 1 to 4	param = st1/stl2/ stl3/stl4, levtype = sfc	param = st, levtype = sol, levelist = 1/2/3/4
Mass mixing ratio, total column mass density and emission mass flux for ozone and carbon dioxide	param = o3/tco3/e_ o3/co2/tcco2/e_co2	param = mmmr/tcmd/ emm chem = o3/co2

Table 2 A few examples of MARS requests before and after the migration to GRIB2.

Stay informed

Users are encouraged to follow the progress of the migration to GRIB2. For this purpose, a mailing list has been set up: mtg2@lists.ecmwf.int. This list is intended to be used to inform users about progress and changes in the migration. Users are also invited to continue to check ecCodes release notes for a general understanding of changes and bug fixes, as many may not be related to the migration. To subscribe, send an email to sympa@lists.ecmwf.int with the subject 'SUBSCRIBE mtg2@lists.ecmwf.int'. To report a problem related to the migration to GRIB2 or if you have a question about it, please follow the normal procedure and contact user support via the Service Desk.

The information which will be distributed via mailing list, as well as more details, data and code examples, can also be found on the MTG2 Confluence web page (<https://confluence.ecmwf.int/display/MTG2US/Migration+to+Grib+2+--+User+Space+Home>). You can get informed of any updates and changes by clicking the watch button on that page.

Conclusion

The migration to GRIB2 is an essential step in reaching the goals set out in ECMWF's ten-year Strategy. Naturally the migration will require adaptation, both for users and internally in various ECMWF workflows. However, it is imperative we make this change to support the future data requirements of ECMWF. In addition, the migration will bring numerous advantages, such as more detailed metadata, a more efficient compression of the data, and a more consistent encoding of parameters. Users are invited to stay informed on the migration to GRIB2 via the emailing list and web page mentioned in the previous section.

Further reading

Betke, E., T. Quintino, S. Smart & T. Wilhelmsson, 2022: Impact of GRIB compression on weather forecast data and data-handling applications. *ECMWF Technical Memorandum, No. 900*. <https://www.ecmwf.int/en/elibrary/81320-impact-grib-compression-weather-forecast-data-and-data-handling-applications>.

World Meteorological Organization (WMO), 2022: Manual on Codes – International Codes, Volume I.2, Annex II to the WMO Technical Regulations: Part B – Binary Codes, Part C – Common Features to Binary and Alphanumeric Codes, **WMO No. 306**. https://library.wmo.int/?vl=notice_display&id=10684#.

© Copyright 2023

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

The content of this document, excluding images representing individuals, is available for use under a Creative Commons Attribution 4.0 International Public License. See the terms at <https://creativecommons.org/licenses/by/4.0/>. To request permission to use images representing individuals, please contact pressoffice@ecmwf.int.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.