9th October 2015

EUMETSAT Secretariat

Dear Secretariat and OPS-WG representatives,

We would like to raise the attention of the OPS-WG and the EUMETSAT Secretariat to what we see as critical issues concerning the suggested file format and compression methods for the MTG FCI level-1 data. After looking more into the specifications and test data provided this summer, we have become concerned that the proposed format and data compression will be unnecessarily sub-optimal for the end users.

Our main concerns regard:
- The suggested compression method and strategy
- The apparent lack of interaction with the CF community and what appears to be rather vague efforts made to strive at a CF compatible file format

Data compression and decoding on the user side

The FCI format specification was presented and discussed at the 37th STG Operations Working Group meeting (Document=EUM-STG-OPSWG-37-15-DOC-04). Citing from the minutes:

*The **Secretariat** advised that the presentation of the format to STG-OPSWG and STG-SWG was part of the consolidation process to allow comments from the Delegates to be taken into consideration.*

***Sweden**, noting that the FCI level 1c data will be provided in netCDF-4, queried whether the standard netCDF libraries would be able to read the data. The **Secretariat** responded positively, but explained that the issue is to first put together all the data chunks back into a reference grid. In addition to that, the compression and the IR 3.8 channel radiance encoding*

are non-netCDF standard, which means that add-in codes will be required. An open source decompression software will be made available..

So, the essence of the answer given to Sweden above by the Secretariat is that standard netCDF libraries will <u>not</u> be able to decode and read the data.

Further in [RF 2] it is read:
2.1.5 No Special Internal Compression
No special internal compression is currently applied to the radiance arrays. In order to increase compression ratios above those obtained from using the default netCDF compression, the disseminated radiances are expected to be compressed using a purpose-built charLS (lossless Jpeg) compression module. An additional software module is required for the HDF libraries at the receiver's end in order to invisibly decompress this data. Chunks retrieved from the archive are not currently expected to have a special compression
applied.

Providing an external module to well established library like netcdf4 and hdf5 will generate non negligeable problems for operational users, where the standard libraries are installed system-wide and used by other operational software. In addition we can imagine many disappointed users learning that a netCDF file from EUMETSAT cannot be decoded with standard netCDF libraries.

We fully acknowledge the need to keep bandwidth usage to a minimum, saving costs and minimising latency, but introducing non-standard encoding techniques that will require significant customisation on the user side must be very well motivated. Especially, when there are very efficient standard compression algorithms available in the HDF5 library. Even though these compression techniques are not available in the netCDF API today, we anticipate chances are higher that they may become available in the future, rather than the suggested charLS.

So, the questions that need to be answered are

- Has Eumetsat considered other compression algorithms already available in the HDF5 library, but not (yet) in the netCDF4 API, and how do they compare to *charLS*, for instance *SZIP?*
    - How much bandwidth is saved using the proposed method compared to the gzip compression available today in netCDF?
    - How much bandwidth is saved compared to other HDF5 built in compression algorithms?
- What is the plan for Eumetsat in the long term? Can it be expected that the modification to the library will be integrated upstream, so that it will be available in standard ways in the future?
- Considering that compression efficiency is so crucial, has EUMETSAT considered using the HDF5 file format, rather than netCDF, and thus having direct access to several efficient compression methods?

We are not convinced the suggested netCDF compression solution is sufficiently motivated, and would like the secretariat to answer the above questions and to make a serious effort to strive for a compression solution that both consider bandwidth and usability. Any customisation on the user side is regarded as an important limitation in usability.

Format specification and CF conformance:
In [RF 2] it is read:
2.1.3 CF Convention Conformance
The CF 1.6 and forthcoming CF 1.7 conventions do not cover the enhanced netCDF-4 constructs that are used in the MTG products such as groups, enumerated data types and unsigned data types. This means that the MTG products cannot currently conform to existing CF conventions. It is hoped that the creation of a CF 2.0 that is compatible with netCDF-4 will allow the products to be made CF compatible.

How is Eumetsat interacting with the CF community? Has the current format proposals been submitted for review to the cf-metadata mailing list as should be done with any dataset aspiring to CF compliance? The current proposal looks actually quite far away from CF conventions and more specifically from the work done by the cf-satellite group:
http://cf-trac.llnl.gov/trac/wiki/SatelliteData

What is Eumetsat's engagement in the creation of CF 2.0?

Concerning geolocation the FCI level-1c familiarisation document [RF 2] says:
3.5 How can I geolocate the data?
The FCI L1C radiance data is registered to a reference grid with fixed latitude and longitude positions according to the spatial resolution of the data. In order to reduce the size of the dataset, the radiance and quality flag variables are not geolocated but instead have pixel positions within the grid. The formulae for creating the reference grid and linking pixel position to latitude/longitude position are given in the FCI L1C Dataset User Guide [FCIL1DUG]. Normally the required variables are encoded into the dataset, but for this release, you will need to retrieve the values from [FCIL1DUG].

Why is this document not mentioning the grid mapping in the CF conventions (http://cfconventions.org/Data/cf-conventions/cf-conventions-1.6/build/cf-conventions.html#grid-mappings-and-projections) since the projection described here is the "geos"/"geostationary" projection available in proj.4 and the cf conventions 1.7?

References:
[RF 1]
EUM-STG-OPSWG-37-15-DOC-04_MTG_FCI_Level_1_Format_Specification_EUMEDS_10607.pdf
[RF 2]
ftp://ftp.eumetsat.int/pub/OPS/out/test-data/FCI_L1C_Format_Familiarisation

Sincerely,

Trygve Aspenes, Janne Kotro, Adam Dybbroe and Lorenzo Clementi

STG OPS-WG representatives for Norway, Finland, Sweden and Switzerland