# Data services innovation at EUMETSAT: The EUMETSAT Products Customisation Toolbox

**Daniel Lee (presenter), Fernando Jose Pereda Garcimartin, Peter Miu, Rodrigo Romero, Oscar Perez Navarro**

EUMETSAT, Eumetsat-Allee 1, Darmstadt, Germany

## ABSTRACT

The EUMETSAT Products Customisation Toolbox is one of the EUMETSAT big data Pathfinder projects, intended to explore the possibilities of applying state of the art big data approaches and data processing strategies to EUMETSAT's products, and to lower barriers for working with these data. The purpose of the Toolbox is to allow users to customise EUMETSAT products in order to fit their workflows, by allowing them to aggregate, extract, resample, and reformat products. In addition to these functions, the Toolbox also exposes other capabilities to users that are useful in diverse environments.

The Toolbox initially targets users from the Numerical Weather Prediction community within EUMETSAT's Member States who are already familiar with EUMETSAT products. During the pathfinding phase, the Toolbox supports a limited set of EUMETSAT data with the intent of easing access to some of EUMETSAT's flagship products that we currently disseminate to users. In the future, the Toolbox should support a wider range of products, as well as additional customisations.

Some of the major developments foreseen for the Toolbox include support on more computing platforms and full support of all EUMETSAT products, past, present and future. Currently, the Toolbox can be used on Ubuntu Linux and is accessible via graphical user and command line interfaces, as well as via import as a Python package. This ensures maximal compatibility with common scientific data processing workflows. With the planned addition of an interface as a web service, the Toolbox will be an excellent utility for processing EUMETSAT ad-hoc, or in operational workflows, whether it is deployed on a desktop PC, inside an HPC cluster, or within as a cloud-native application in a microservice ecosystem, ensuring that users have easy access to EUMETSAT products in the format – and the environment – which best suits their needs.

## PURPOSE

EUMETSAT produces and disseminates many meteorological satellite products in their native format (e.g. NetCDF), native projection (e.g. in satellite projection) and native spatiotemporal coverage (e.g. PDU of 3 minutes for EPS).

Many users work with EUMESAT data using common GIS or remote sensing applications rather than applications built specifically for EUMETSAT data. In many cases, this makes it necessary for them to customise the products into a different projection (e.g. Mercator, Platte Carree), format (e.g. BUFR, GRIB, NetCDF, HDF4/5, GeoTiff) or

spatiotemporal coverage (e.g. a specific continent, a country, a region, a tile or region of interest) before importing the data into their standard workflows and environments.

Some limited customisations are available when ordering data from EUMETSAT, e.g. via the EUMETSAT Data Centre. However, these capabilities are not exposed to users of EUMETCast, thus requiring users to find or implement their own software if they want to apply the same customisations to data received in near real-time. This can be expensive, requires expert knowledge, and does not guarantee uniform outputs when performing the same customisations on archival data requested from the Data Centre.
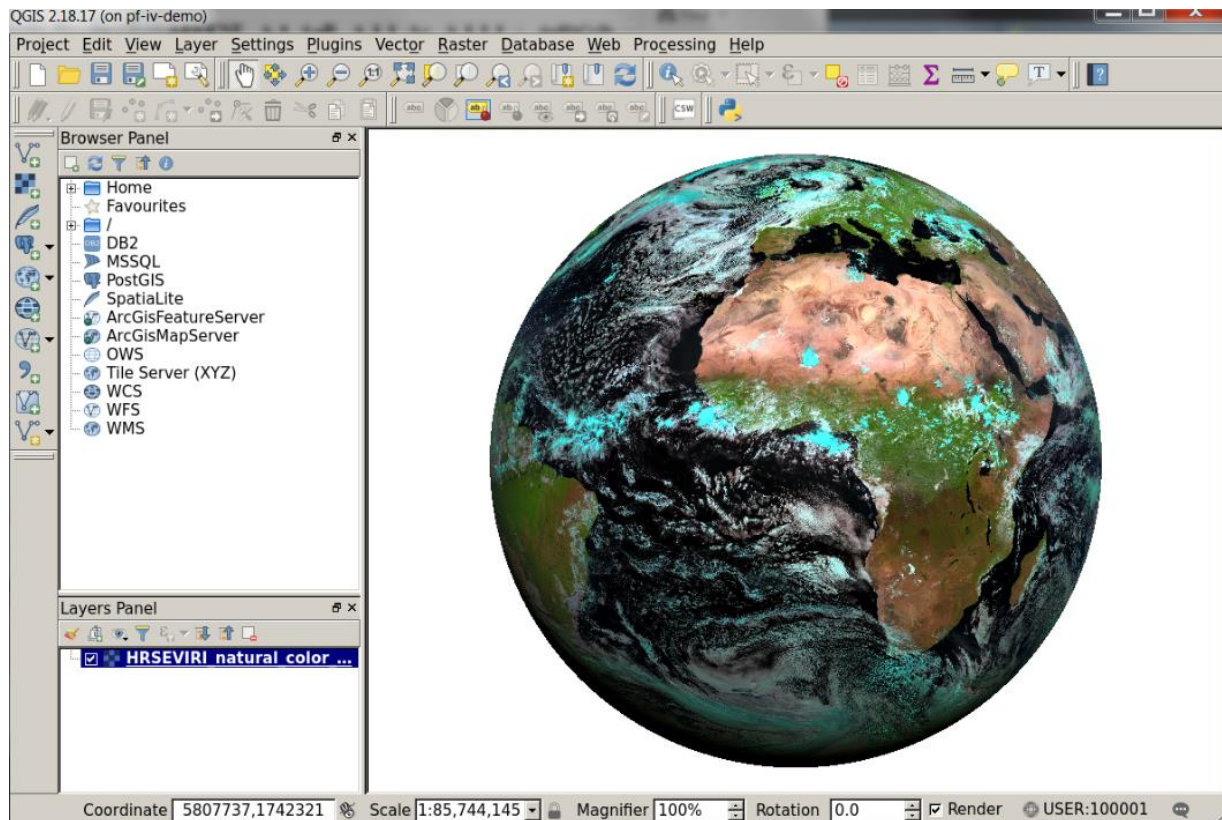
The EUMETSAT Products Customisation Toolbox ("Toolbox") offers users a simple, user-friendly and free set of capabilities to perform product customisations. The Toolbox is environment agnostic and portable, meaning that EUMETSAT can deploy it in its numerous Ground Segment systems. This is also advantageous for users who use the Toolbox in their own systems, as it will produce the same results as if the data were customised at and retrieved directly from EUMETSAT.

The goal of the Toolbox is to ease workflows for users by delivering our many complex products from disparate sources, customised for their individual use cases, such as processing or visualisation in standard GIS and Remote Sensing software. The Toolbox will extend the existing functionality in all EUMETSAT data access services, including in other EUMETSAT Pathfinders. It will be offered to users as a simple, user-friendly and free package, which will save product preparation time and allow users to focus on their core business rather than data preparation.

## KEY FUNCTIONS

The EUMETSAT Products Customisation Toolbox will provide the means to customize native data from EUMETSAT by filtering and aggregating, and by transforming it into the format, projection and Region of Interest (ROI) that interests the end-user of the data. This will ease the integration of EUMETSAT products into commonly used GIS software and stimulate the uptake of EUMETSAT data.

The Toolbox already supports many products from Meteosat Second Generation (MSG), EUMETSAT Polar Service (EPS) and the Ocean and Sea Ice Satellite Application Facility (OSI SAF) at this time. The final intent is to support all past and future products.

*Figure 1: The EUMETSAT Products Customisation Toolbox makes it simple to customise EUMETSAT native format products and prepare them for use in standard processing software, as shown in this example using QGIS.*

The main functions of the Toolbox are:

1.  **Aggregation**: users can aggregate e.g. EPS products into longer segments or full orbits, or OSI SAF global products to include multiple time steps;
2.  **Layer filtering**: users can extract individual bands from multi-band products;
3.  **Re-projection and re-sampling**: users can change the geographic projection and re-sample products into the desired resolution;
4.  **Region of interest (ROI) extraction**: users can extract regions of interest by using a customisable pre-set region or providing bounding-box coordinates;
5.  **Re-formatting**: users can read products in their native format and write them in various common output formats (including BUFR, GRIB, netCDF, HDF4/5, GeoTIFF, JPEG and PNG). This opens the possibility of consolidating re-formatting software, including in the EUMETSAT Ground Segment;
6.  **Quicklook** generation: users can create simple visual representations of products for manual granule selection, cataloguing, etc.;
7.  **Compression / decompression**: users can losslessly compress and decompress products for efficient transmission and storage.

These functions are explored in greater detail in the following sections.

## Aggregation

Near-Real Time products are distributed in small packets in order to facilitate timely product generation and data assimilation. Many analysis workflows, however, require time series data. It may also be relevant whether data is

contiguous across Product Dissemination Unit (PDU) boundaries. The Toolbox can "stitch" together various product input files and create a unified output.

### Layer filtering

EUMETSAT data contains a rich sampling of the spectrum – for example, the IASI instrument on board the Metop satellites views the Earth with 8461 channels. These channels are carefully chosen in order to cover the primary use cases for each instrument; however, for a given use case, only a limited number of channels might be required. The Toolbox allows users to easily extract only those channels that are of interest to them so that they can transmit and store smaller data volumes.

### Re-projection and re-sampling

Data fusion workflows require data to share a common input projection and grid spacing. The Toolbox utilizes the Geospatial Data Abstraction Library in order to reproject and resample data into all common projections.

### Region of interest extraction

A common data reduction tactic is to extract only the geographic area that is of interest for a given study or use case. The Toolbox makes it possible to extract all observations that fall within a given bounding box, eliminating irrelevant data from downstream processors.

### Reformatting

EUMETSAT produces products in several formats, (e.g. BUFR, GRIB, netCDF, HDF, HRIT, EPS, CAP, etc.), many of which cannot be imported into standard geodata processing workflows. Users from the Numerical Weather Prediction (NWP) community are already in possession of specialized software which can process these data, but users from other disciplines often require other formats, such as netCDF, HDF, GeoTIFF, JPEG, etc. The Toolbox is capable of re-formatting all supported products into several common file formats.

### Quicklook generation

The Toolbox can generate so-called "quicklooks" – visual representations of the product in question – so that users can quickly judge the relevance of a given product. This is especially useful for ingesting time series of data into local archives where users can discover, visually evaluate and access stored products before applying an entire processing chain to them.

### Compression and decompression

EUMETSAT plans to extend the functionality of the Toolbox to include compression and decompression. This will allow EUMETSAT to use one common software for reducing data volumes before storage and dissemination, and enable any user to decompress that data losslessly. Users will profit from the higher timeliness and lower transmission and storage costs of products that are compressed before dissemination.

## INTERFACES

The Toolbox currently supports three interfaces, with a fourth interface in planning:

- Graphical user interface (GUI),
- Command line interface (CLI),

- Application programming interface (API), and
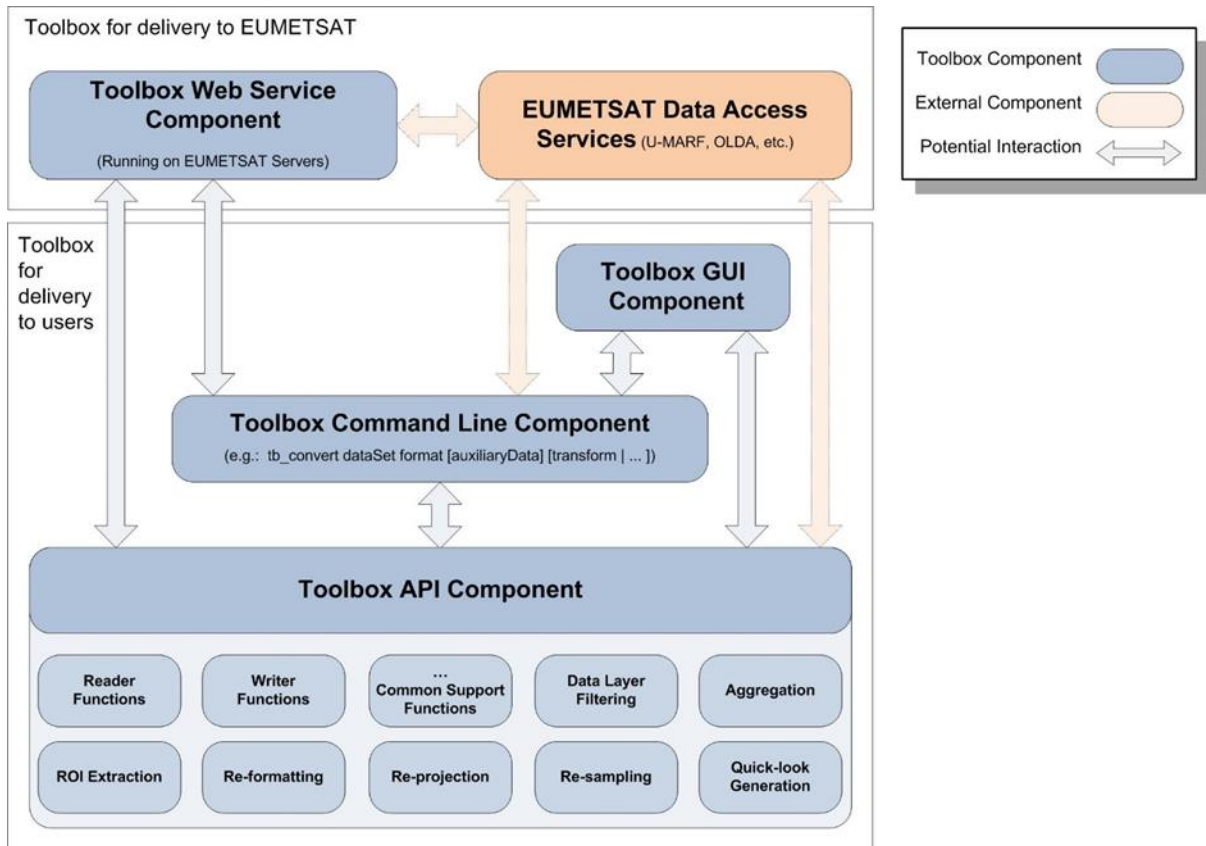- Web service interface (currently in planning).



*Figure 2: The Toolbox interfaces and underlying functions.*

The following sections provide more detail on each interface.

**Graphical user interface (GUI)**

The Toolbox GUI has an attractive and intuitive front-end that allows users to use all Toolbox functions through graphical icons and visual indicators. Users can deploy the GUI natively in their local operating system, i.e. Windows, Linux and Mac. This interfaced is primarily aimed at novice and new users, general users who have very basic format transformation needs, or users who wish to interactively produce and test workflows which should be automated at a later point. Workflows generated via the GUI can be saved and re-used via the other interfaces.
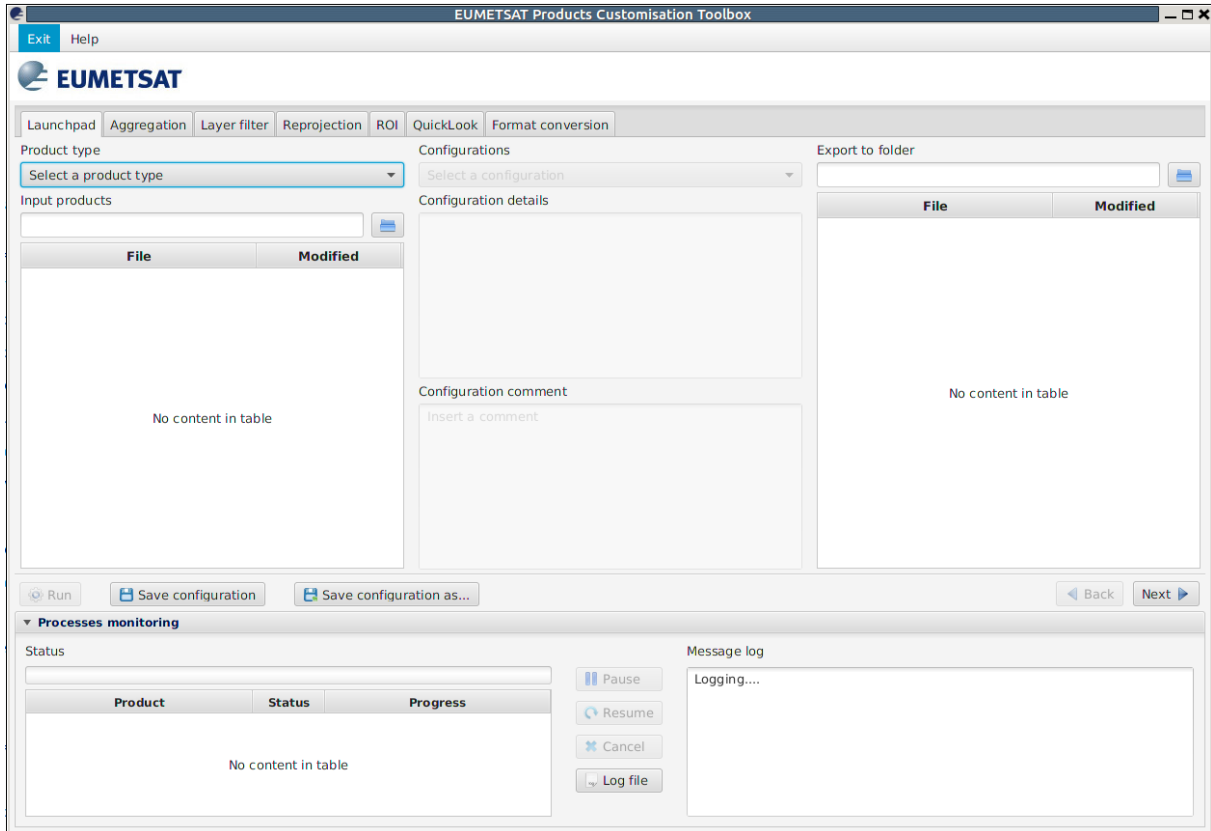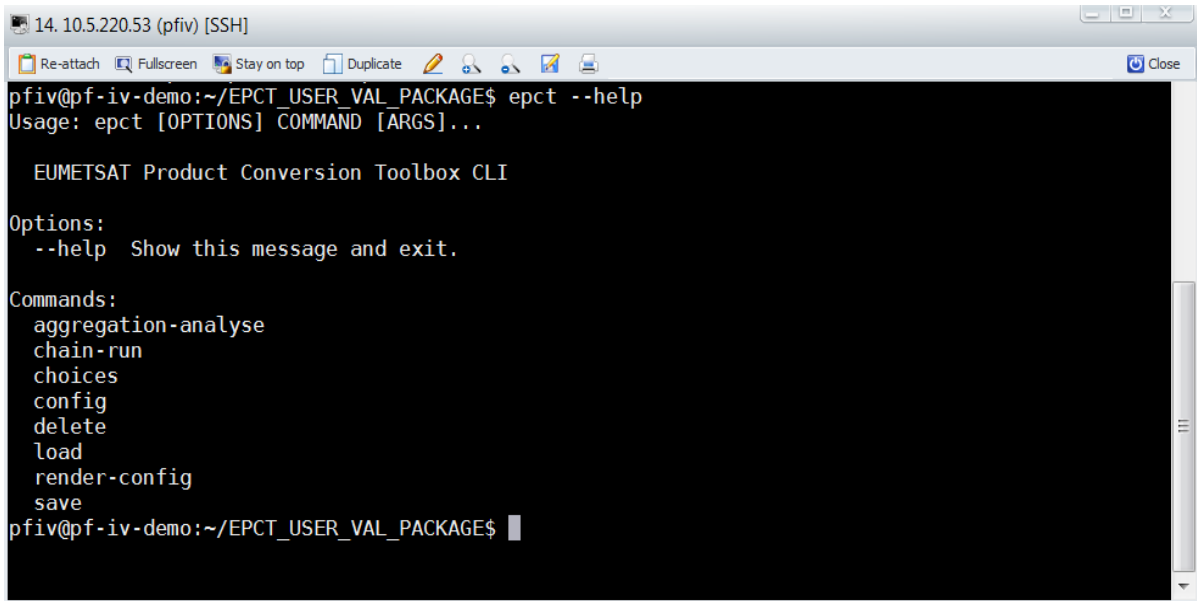
*Figure 3: The Toolbox GUI.*

## Command line interface (CLI)

All product customisations are available as binaries with text interfaces. This means that they can be used either in interactive shell sessions, e.g. in order to develop workflows or in headless environments. They can also be integrated into shell scripts or any program which makes system calls. An example use case for the CLI would be to embed the Toolbox in a EUMETCast Receiving Station and apply product customisations automatically as data is received.

*Figure 4: The Toolbox CLI.*

**Application programming interface (API)**

The toolbox API allows software developers to integrate EUMETSAT Toolbox functionality into their applications. Currently the API is exposed as a Python package. This means that developers writing programs in Python can use all functionalities of the Toolbox natively within their application and integrate its use with other popular scientific processing packages, such as the SciPy stack, xarray, Dask, and PyTroll. Customisations may of course also be applied within reproducible research contexts such as Jupyter Notebooks or a JupyterLab.

**Web service interface**

A planned evolution of the Toolbox will expose its functionalities via a REST API. This will make it possible to use the Toolbox as the backend format customisation system for all EUMETSAT Data Accessing Service, e.g. the UMARF ordering system. This will not only streamline formatting initiatives within EUMETSAT, but also ensure that customised products are identical, regardless of whether the customisation takes place at EUMETSAT or on users' premises. Furthermore, users will be able to deploy the Toolbox as a microservice if they are building cloud-native infrastructure.

## NEXT STEPS

The EUMETSAT Products Customisation Toolbox is approaching its close-out phase and will soon be ready for operationalisation. This will entail solidification and baselining of all interfaces, as well as extension to support new products formats and interfaces. The integration of the Toolbox with other EUMETSAT Pathfinder products synergetically adds value; the next development phase will build upon this foundation to provide streamlined product customisations to users in the format and environment of their choice, tailored to the needs of their applications.