Europe Aid/132517/D/SER/Multi

Monitoring for Environment and Security in Africa (MESA)

CONTRACT NO. EDF 2013/315-947

eStation 2.0

User manual

Draft 1.1

July 2015

Date: //
Ref: MESA/DOC/eStation/…

Abstract / Résumé
This document provides instructions for the exploitation of the eStation 2.0 system, including the visualization/analysis component.

	Name	Position
Prepared by	Marco Clerici	JRC-EC responsible for MESA
Contributions/Reviews by	Antoine Royer Jurriaan Van't Klooster	JRC-EC thematic Expert for MESA IT-GIS Specialist

	Details	Date
Version history	Draft 0.1	December 2014
	Draft 1.0	May 2015
	Draft 1.1	July 2015

Contents

List of Figures

ACRONYMS and DEFINITIONS

AMESD	African Monitoring of Environment for Sustainable Development
ACMAD	African Centre of Meteorological Applications for Development
AGRHYMET	Centre Régional de Formation et d'Application en Agrométéorologie et Hydrologie Opérationnelle
AU	African Union
BDMS	Botswana Department of Meteorological Services
CICOS	Commission Internationale du Bassin Congo-Oubagui-Sangha
CWG	The MESA Continentalisation Working Group
EO	Earth Observation
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EUMETCast	EUMETSAT's primary dissemination mechanism for the near real-time delivery of satellite data and products
FTP	File Transfer Protocol
GIS	Geographical Information System
IOC	Indian Ocean Commission
JRC	Joint Research Centre of the European Commission
MESA	Monitoring for Environment and Security in Africa
MOI	Mauritius Oceanography Institute
REC	Regional Economic Communities
RIC	Regional Implementation Centre
TA	Technical Assistance
TAT	Technical Assistance Team
THEMA	Regional and Continental Thematic Actions

1. INTRODUCTION

1.1 SCOPE OF THE DOCUMENT

This document describes the functionalities of eStation 2.0 application and explains how the Final User can benefit from its features. It is meant mainly for the thematic expert making use of the system, and describes both the 'processing' and 'visualization/analysis' components.

1.1 DOCUMENT ORGANIZATION

The present document is structured into the following chapters:

- Chapter 2: Overview of the eStation 2.0

This is the basic introduction to the eStation 2.0, which provides application's rational, the overall structure (hardware and software-wise), an overview of the GUI, some essential notions to understand the system functioning and an overview of the services defines. It is meant for the thematic User, but reference to this section is done also from other documents (e.g. the [RD-1] – Administration Manual).

- Chapter 3: User Guide

It describes all functionalities that can be controlled through the GUI by the thematic User, namely the Dashboard, Acquisition, Processing, Data Management, Analysis, System and Help panel. Note that the 'Analysis' panel contains the visualization tools that represented the 'EMMA' tool in eStation 1.0.

1.1 APPLICABLE AND REFERENCE DOCUMENTS

Id	Title	Date	Reference
AD-1			
AD-2			
AD-3			
AD-4			

Table 1: Applicable documents

Id	Title	Date	Reference
RD-1	eStation 2.0 Administration Manual		AdminMan
RD-2			
RD-3			

Table 2: Reference documents

2. OVERVIEW OF THE ESTATION 2.0

2.1 SYSTEM CONCEPT

The eStation 2.0 is the evolution of the system delivered to beneficiaries of the AMESD project, and is mainly intended to provide the MESA[1] Regional and Continental Implementation Centres (RICs/CIC) and National Focal Points (NFPs) with an instrument to receive, process and visualize Earth Observation data for environmental monitoring and climate services. The design of the eStation 2.0 has been based on the lessons learned from the AMESD project, and on feedbacks from Users; as a result, the main purposes of the system are:

- Ensuring continuous reception of EO data, and easy the collection of missing data
- Facilitating the post-processing and the link with other tools (e.g. QGIS, SPIRITS)
- Facilitating user interaction (UI rather than coding)
- Ensuring Hardware redundancy
- Proposing customized functions for data analysis.

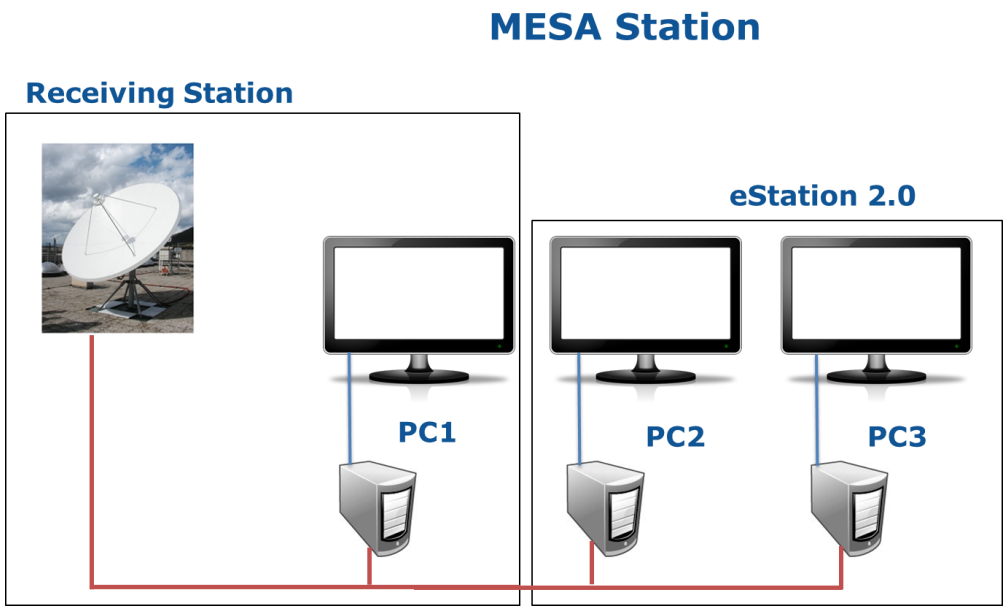


Figure 1: overview of the eStation 2.0 as part of the MESA station

In this respect, the eStation 2.0 is meant to be a light processing server for EO datasets, rather than a stand-alone GIS platform (like QGIS or similar commercial solutions). The functioning of the application is organized around a number of Services, as displayed in Figure 2, namely:

[1] See http://rea.au.int/mesa/

- Get Services: to systematically acquire from PC1 (Receiving Station) and from remote ftp/http servers EO data. There are therefore two 'get' services that can be configured and controlled independently:
    - Get EUMETCast (data from PC1)
    - Get Internet (data from remote ftp/http servers)

- Ingestion Service: to convert from the various 'native' format to the GTIFF reference format of the eStation 2.0. This operation also includes – optionally – re-projection and clipping to a specific region of interest.
- Processing Service: to derived from the input data additional products, like long term statistics, anomalies, and other added-value indicators.
- System Services: to manage the data-synchronization between PC2 and PC3, the data backup and all operations to be executed continuously in the background.

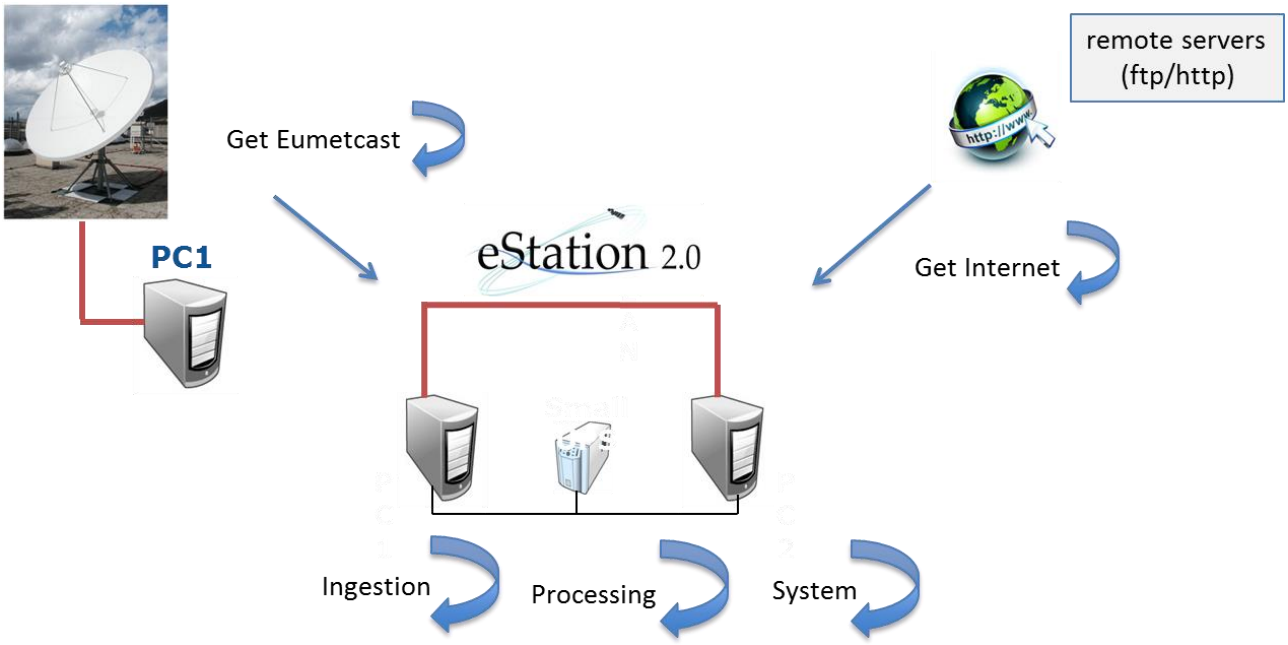


Figure 2: Overview of the Services running on the eStation 2.0

These Services are meant to run continuously and w/o User supervision on the eStation 2.0, once properly configured by the User. According to the type of installation (on a single computer or on PC2 and PC3 of a MESA station), the role of each computer (PC2/PC3) and mode each computer (Nominal or Recovery mode) the Services might be activated or de-activated. Also note that the Visualization functionalities, including the timeseries analysis, are performed on the fly, and there is no service associated to them.

2.2 System Structure and configurations

The eStation 2.0 is an application mainly developed in python, for the services and processing component, and in ExtJS 5.0 for the visualization and GUI component. A postgresql database stores the information for the EO products definition, for the remote sources for retrieving these data and all relevant User configurations (see Chapter 4 of Administration Manual for DB description). The application has a series of software dependencies, including GDAL library and its python wrapper for geo-processing, Mapserver for image rendering, and a list of python non-standard modules, including ruffus for the processing engine.

Within the MESA project the application is integrated into the Linux Ubuntu 12.04 OS, and all dependencies are part of the default installation. The integration in Windows OS is also considered, though it is not presented in this Guide.

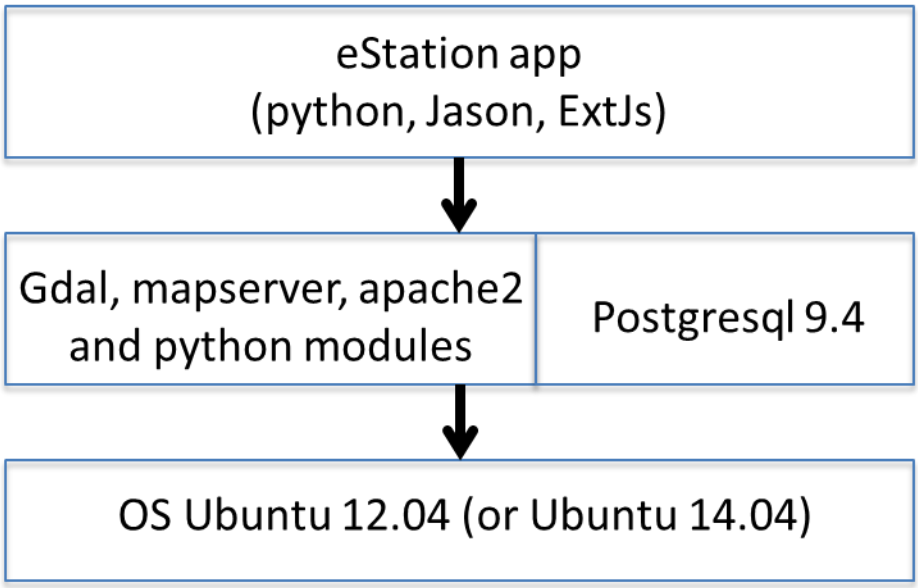


Figure 3: eStation 2.0 SW organization

The following options are foreseen for the installation and configuration of the System, according to the available HW.

2.2.1 Full MESA eStation (default case)

The MESA beneficiaries will receive a dedicated HW for the installation of the eStation, consisting in 3 computers (see Figure 1): PC1 (aka 'Receiving Station'), which is connected to the external antenna, and PC2 and PC3 for the installation of the eStation 2.0.

eStation 2.0 is therefore installed on 2 computers, and the various tasks are shared between the machines, in order to ensure load share. In this case, the OS of the machines is Ubuntu 12.04. The connection to the Internet – when available in the Institution - is enables, so that the system can access additional data sources (see get internet Service) and Remote User support is also feasible.

Note that an identical configuration can be reached by installing the new eStation 2.0 software on the PC2/3 of the AMESD station.

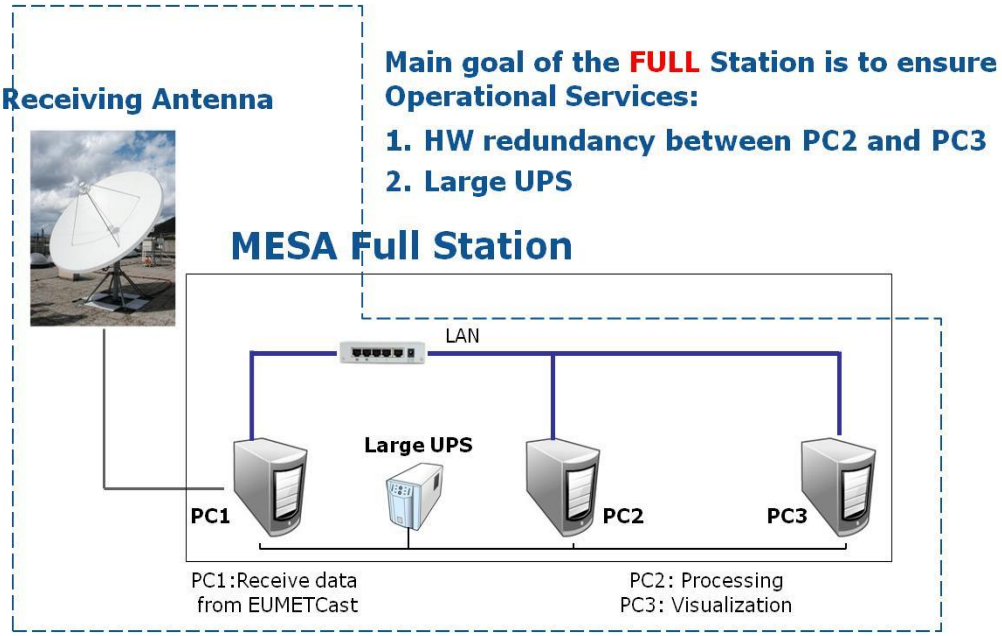


Figure 4: Overview of the MESA Full Station

2.2.2 Single Computer (or Light eStation)

The eStation 2.0 can also be installed on a single computer, and all processing/visualization functionalities can be activated, with the exception of the ones related to the hardware redundancy (i.e. the data backup to PC3 and the activation of the Recovery Mode).

This option is foreseen for all the Users who do not receive the 'Full MESA Station', and who intend to use a single computer for the eStation, limiting the HW investment. Connection to the EUMETCast Receiving Station and to the internet will be activated whenever possible; the resulting system is not recommended for implementing the MESA operation services, rather for associated Users like University, students, thematic experts.

2.2.3 Live USB key/HD

An external USB device (disk/key) can be used as a 'boot' device for an existing computer (either desktop or laptop) by modifying the BIOS settings, and without modifying original hosting machine. With this approach, which is somehow similar to temporarily replace a disk of the computer for a limited amount of time, the User can keep both the eStation application and datasets on a single external disk, and 'run' it from different machines, when necessary. It is therefore suggested for non-operational use of the eStation (demo, training, travelling), similarly to AMESD 'StandAlone' application, but with the benefit of having all features of the system available.

2.2.4 Virtual Machine

Various 'virtualization' environments exist (including VM-ware and Oracle Virtual Box) for providing the User with the option of running, e.g., an Ubuntu 'virtual machine' on a Windows computer. In such an environment, the eStation 2.0 can be installed and run, with some limitations of the performances and disk

space. Similarly to the Live USB solution, Virtual Machine installation is suggested for demonstration and testing purposes rather than normal operations.

Table 1 summarizes the options of the various above-described installations, in terms of number of computers, operating system, connection to EUMETCast and the Internet and expected usage.

Installation	Number of PCs (for eStation only)	OS	EUMETCast connection	Internet connection	Foreseen Usage
Full Station	2	Ubuntu (>12.04)	yes	whenever possible	This is the default case for the 'official' MESA beneficiary Institutions, included in the list of the Supply Contract - Lot 1
Single PC	1	Ubuntu (>12.04)	whenever possible	whenever possible	This case is envisaged for 'additional' Users of the MESA project, or external Users, willing to exploit a single PC for the application, e.g. partner Universities.
Live USB	1	any[2]	whenever possible	whenever possible	Demonstration of the tool, analysis by thematic User on additional PCs (e.g. while travelling).
VM	1	any	whenever possible	whenever possible	Demonstration of the tool, analysis by thematic User on additional PCs (e.g. while travelling), tests of the installations.

Table 3: summary of options for the various installations.

[2] The USB device (key/disk) will be connected to a computer to be re-booted from the device; therefore the OS of the PC is not relevant.

2.3 SYSTEM OVERVIEW

The main interface to the application in through a web interface that can be accessed at the address **localhost/esapp/.** The home page is the 'Dashboard' (see Figure 5), which presents the overall status of the MESA Station and allows the control of the Services.

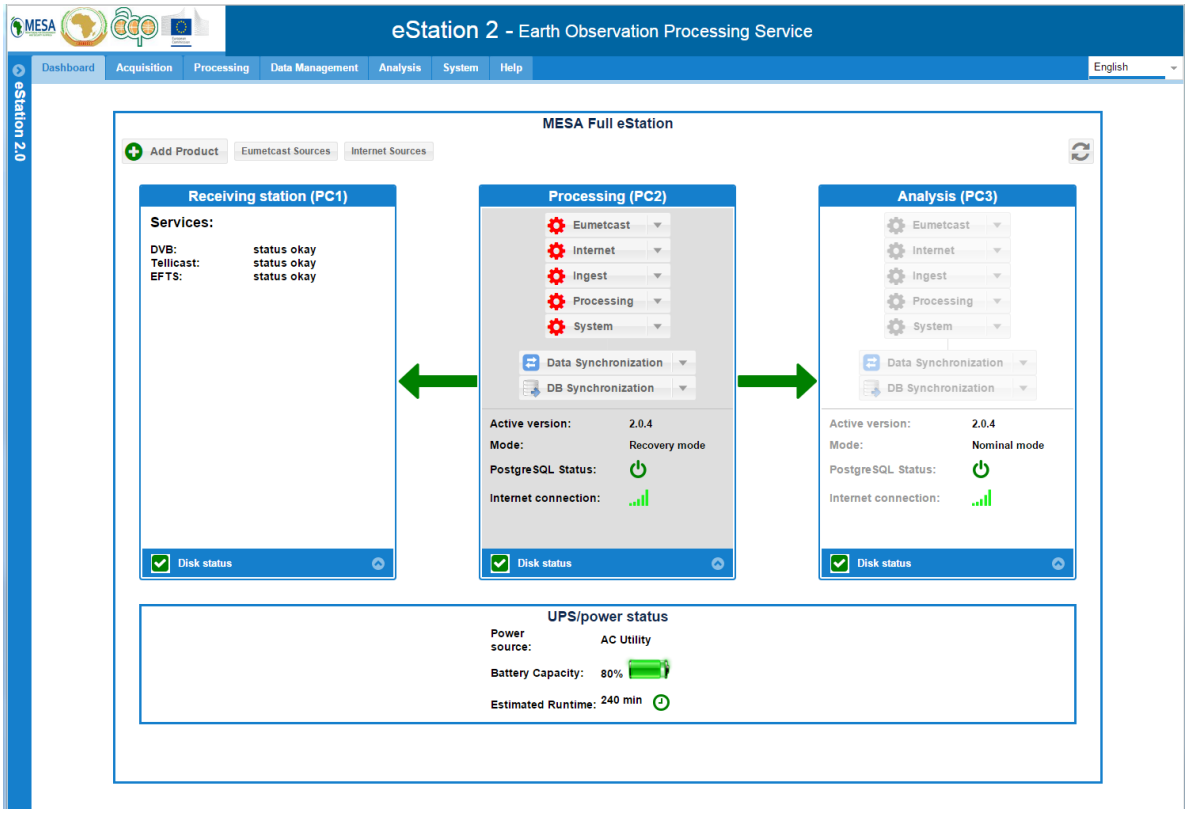


Figure 5: Overview of eStation 2.0 UI

The other tabs presented in the page are:

- **Acquisition**: to view/control the status of the services for retrieving and ingesting the EO data.
- **Processing**: to activate/de-activate the derivation of new products through the processing Service.
- **Data Management**: to view/control all datasets (i.e. GTiff files) existing on the system, and complete the incomplete timeseries.
- **Analysis**: to perform the data analysis and generate images for bulletins/reports. It replaces the former EMMA application of the version 1.0
- **System**: to control the status of the application, move to Recovery Mode (and back) and to perform basic diagnostic operations.
- **Help**: to access the Help resource locally or remotely available.

2.4 eSTATION 2.0 ESSENTIAL CONCEPTS

This section defines a list of concepts underpinning the conception and implementation of the eStation 2.0; whenever possible, the definitions make use of examples, and reference to international standards is made.

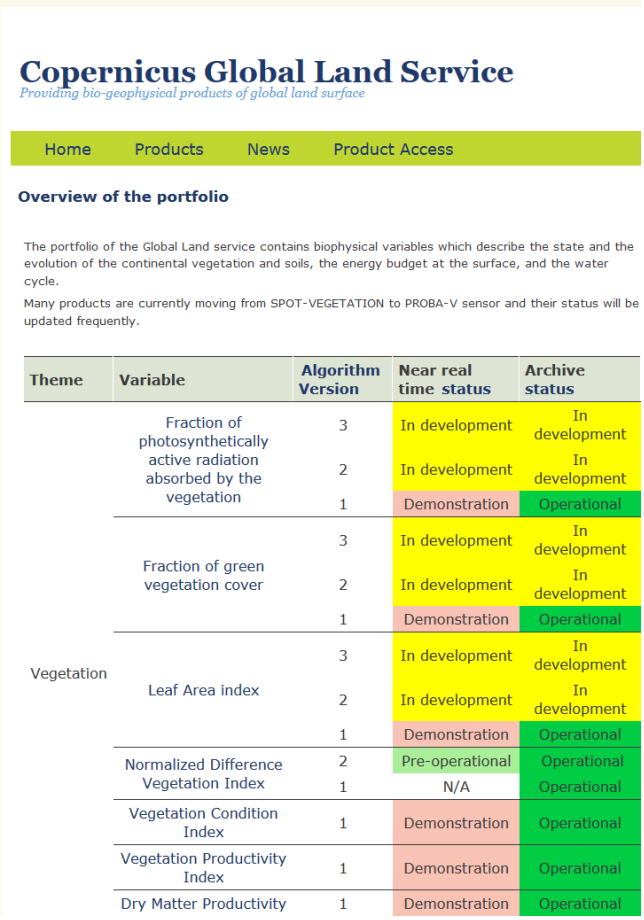
EO Products

An Earth Observation product is a bio-geophysical quantity describing the status of a component of the Earth (land surface, atmosphere or radiation budget) as derived by satellite observations and numerical modelling. The 'Burnt Area', 'Dry Matter Productivity', 'FAPAR' and Leaf Area Index are, among others, EO datasets made available by the Copernicus Global Land Service, as displayed in **Error! Reference source not ound.** A 'Product' is characterized by the geographic extension covered by the images (it might be global, regional or local), by the images geographic or projected co-ordinate system, by the temporal extension, which depends normally on the instruments it is based on for data collection. Additionally, when looking at the distribution of the product to the Users, it is also relevant the file format (e.g. GTIFF), the data policy under which it is distributed and the means of distribution to the Users (e.g. via satellite of through the internet).

Figure 6: EO datasets from the Copernicus website.

Version

A product's **version** identifies a specific collection of images, depending mainly on the algorithm used for the computation; the various versions of the Copernicus products are displayed in Figure 1Figure 7 under the column 'Algorithm version'. For the 'incoming' products on the eStation, the version is the one defined by the data provider, and might be set as 'undefined' if it is unique and there is no clear reference on the documentation or traceability about it in the documentation.



Copernicus Global Land Service

Providing bio-geophysical products of global land surface

Home | Products | News | Product Access

Overview of the portfolio

The portfolio of the Global Land service contains biophysical variables which describe the state and the evolution of the continental vegetation and soils, the energy budget at the surface, and the water cycle.

Many products are currently moving from SPOT-VEGETATION to PROBA-V sensor and their status will be updated frequently.

Theme	Variable	Algorithm Version	Near real time status	Archive status
Vegetation	Fraction of photosynthetically active radiation absorbed by the vegetation	3	In development	In development
		2	In development	In development
		1	Demonstration	Operational
	Fraction of green vegetation cover	3	In development	In development
		2	In development	In development
		1	Demonstration	Operational
	Leaf Area index	3	In development	In development
		2	In development	In development
		1	Demonstration	Operational
	Normalized Difference Vegetation Index	2	Pre-operational	Operational
		1	N/A	Operational
	Vegetation Condition Index	1	Demonstration	Operational
	Vegetation Productivity Index	1	Demonstration	Operational
	Dry Matter Productivity	1	Demonstration	Operational

Figure 7: Version of Copernicus GL products

Subproduct

An 'incoming' product might contain more than a single variable/layer, e.g. the ancillary information as the 'Status Map' (or Quality Flag) of the NDVI Spot-Vegetation product; furthermore, from the incoming product several added-value products and indicators can be generated on the eStation, like long term statistics and anomalies. In order to keep a clear reference to the 'original' (or 'native') product, it is decided to have a 'two-level' identification approach so that more than one **subproduct** can be associated to the same **product**. More specifically, for each product on the eStation it will exist:

- One 'native' product, which has exactly the same name of the product with the '_native' suffix, and refers to the 'incoming' EO product. This 'subproduct' cannot be visualized in the eStation, it is actually a 'reference' for the incoming dataset, i.e. to the files received on the eStation before their ingestion.
- One or more 'ingested' products, according to the number of layers that are extracted from the incoming dataset.
- Zero, one or more 'derived' products, computed on the eStation by the processing Service.

As an example, in Figure 8 the **subproducts** for 'fewsnet-rfe' product (version 2.0) are displayed: a single 'native' is defined - as it is always the case – named 'fewsnet-rfe_native'; a single 'ingested' product exists, as the incoming dataset as no any ancillary information (and the rainfall estimate), while several subproducts are generated on the system.

productcode [PK] character varyir	subproductcode [PK] character varying	version [PK] character varying	defined_by character vai	activated boolean	category_id character vai	product_type character varying
fewsnet-rfe	fewsnet-rfe_native	2.0	JRC	TRUE	rainfall	Native
fewsnet-rfe	10d	2.0	JRC	FALSE	rainfall	Ingest
fewsnet-rfe	1monmin	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	1monmax	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	1mondiff	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	10dmax	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	10dmin	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	1moncum	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	1monavg	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	10dnp	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	10davg	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	10dperc	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	1monperc	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	10ddiff	2.0	JRC	FALSE	rainfall	Derived
fewsnet-rfe	1monnp	2.0	JRC	FALSE	rainfall	Derived

Figure 8: sub-products existing for the 'fewsnet-rfe' product, 2.0 version.

In the case of the Vegetation VGT/PROBA, the situation is more articulated, as we consider four different version on the eStation 2.0: the Spot Vegetation-1 and 2 collection, plus the PROBA-V collection, version 2.0 (not operational) and 2.1. Therefore four 'native' subproducts are defined, one for each version. In ingestion, two sub-products are extracted from the Spot-V products (NDVI and SM, i.e. Status map), while only the NDVI is present in PROBAV versions, being the quality information embedded in the same byte.

productcode [PK] character varyir	subproductcode [PK] character varying	version [PK] character varying	defined_by character vai	activated boolean	category_id character vai	product_type character varying
vgt-ndvi	vgt-ndvi_native	spot-v2	JRC	FALSE	vegetation	Native
vgt-ndvi	vgt-ndvi_native	spot-v1	JRC	TRUE	vegetation	Native
vgt-ndvi	vgt-ndvi_native	proba-v2.1	JRC	TRUE	vegetation	Native
vgt-ndvi	vgt-ndvi_native	proba-v2.0	JRC	FALSE	vegetation	Native
vgt-ndvi	sm	spot-v2	JRC	FALSE	vegetation	Ingest
vgt-ndvi	ndv	spot-v2	JRC	TRUE	vegetation	Ingest
vgt-ndvi	sm	spot-v1	JRC	FALSE	vegetation	Ingest
vgt-ndvi	ndv	spot-v1	JRC	TRUE	vegetation	Ingest
vgt-ndvi	ndv	proba-v2.1	JRC	TRUE	vegetation	Ingest
vgt-ndvi	ndv	proba-v2.0	JRC	FALSE	vegetation	Ingest

Figure 9: sub-products existing for the Spot/PROBAV NDVI

Dataset

A ***dataset***, in eStation jargon, is an ensemble of images related to the same ***EO products,*** for a given geographic and temporal extensions. The concept of 'dataset' is close to the one of 'product', and it stresses the idea of having an ensemble of files generated for the same product for a given region and a given period.

Mapset

The '**mapset'** is the ensemble of information identifying the geolocation of a raster file. Therefore, it includes the following elements, which are not all independent from each other:

1. Spatial Reference System (SRS) - defined through the SRID[3] and referring, by default, to the EPSG authority. It logically includes:

1.1 Geographic Coordinate System (GCS), including Datum.
1.2. Map Projection (if any - name and parameters).

2. Pixel size (unit, value)
3. Boundary Box (ULx/y, LRx/y) or 'Extent' or 'Origin'
4. Raster size (Xsize, Ysize)

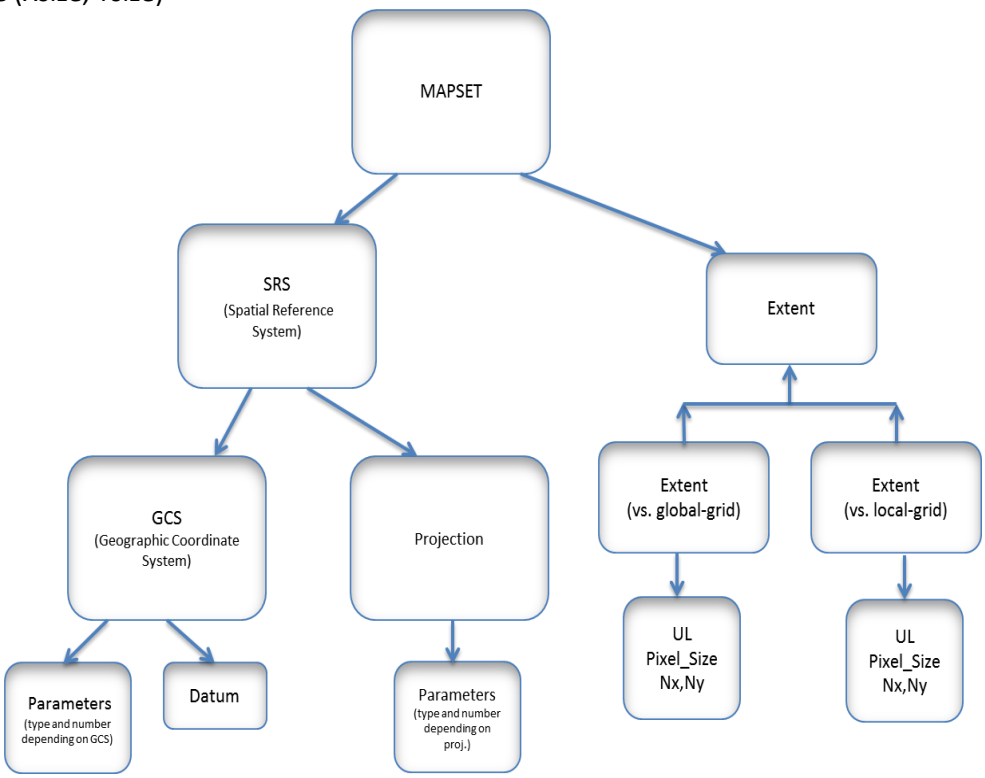


Figure 10: mapset object organization

The overall '**mapset**' object is defined in Figure 10. Note that there are several ways of defining the 'Extent' of a mapset, and the one of choice for the eStation is by defining a. Upper-Left Corner[4] (in the units defined by the SRS), b. the pixel size (in the units defined by the SRS) and c. the number of pixel and rows (Nx,Ny).
The '**mapset**' replaces the concept of 'ROI' (existing on eStation 1.0) and offers the possibility to have all geo-referencing information in a single object, which is stored in a single table of the database, and is convenient for direct re-projection of an image from an original to a target 'mapset'.

[3] see http://en.wikipedia.org/wiki/Spatial_reference_system

[4] Note that the co-ordinates of a pixel refer to the Upper-Left corner of the pixel, not to its centre.

Sources of datasets (or Data source)

Datasets of the eStation are retrieved mainly from the EUMETCast dataflow (i.e. from the Receiving Station, or PC1) and from remote servers (http or ftp), which are both defined as 'Datasets sources' or simply '**data sources**'; they can be associated to an EO product in order to make automatic the retrieval and ingestion of data in the system. 'Data sources' can be seen as a local (EUMETCast) or remote (internet) repository where the EO products are available, or are made continuously available by dissemination as soon as they are computed.
The 'EUMETCast sources are described in a database table ('***eumetcast_source***') that contains all metadata as existing in the EUMETSAT Product Navigator[5] , as in Figure 11. Few of these fields are actually relevant for the data processing on the eStation, and they are:

- eumetcast_id: to uniquely identify the source and associate it to a product.
- filter_expression_jrc: a regular expression to identify a 'dataset' among all the existing ones.

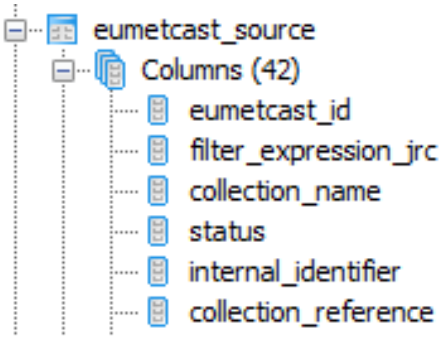


Figure 11: Eumetcast datasource (detail)

The 'internet' sources (see table '***internet_source***') includes some additional relevant elements necessary to access the remote repository, like the base url of the data provider, the credentials (username/password), the regular expression to filter the datasets, and the start and end data to select a temporal window.

[5] See http://navigator.eumetsat.int/discovery/Start/Explore/Quick.do;jsessionid=03AF8AD82E5C74242E359DCEE525A326

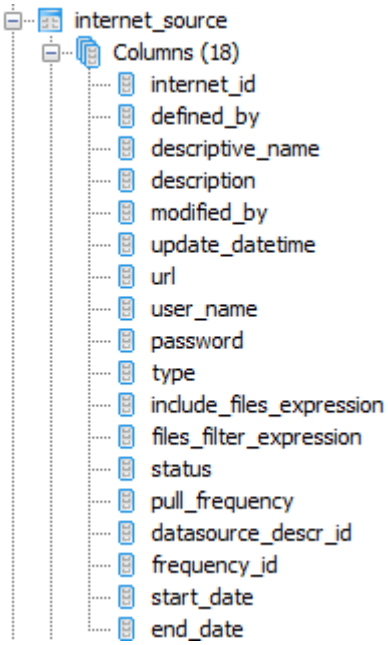


Figure 12: Internet datasources

Additional information for managing the datasource in defined in the '***datasource_description***' table (see Figure 13), which contains elements to describe the 'incoming' files in terms of:

- File naming rules and file extension, which allows the 'get' services' to identify and manage the incoming files, e.g. by extracting the date/time of the image.
- The geo-reference, i.e. geographic coverage (e.g. global) and native mapset; note that in some cases the information on the geo-reference is coded in the incoming file itself, and there is no need to fill this field.
- The pre-processing type to be applied during the ingestion, which varies according to the file format (HDF, GTiff, netcdf) and the tiles organization (see 2.5.2 for more detail)

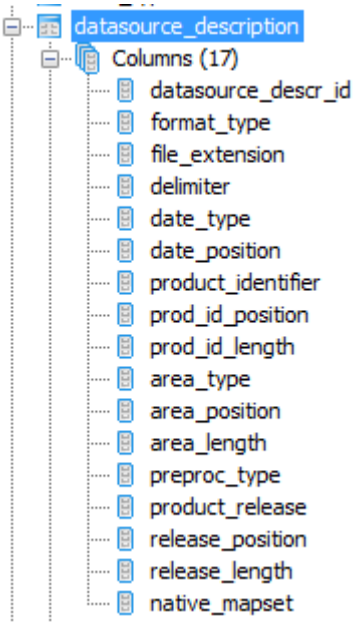


Figure 13: Datasource description table

Services

The eStation 2.0 '**Services**' are processes in charge of executing continuously a specific set of operations, according to the Users settings. The logical workflow for the User is therefore:

- Define/verify the settings on the eStation for each of the services (e.g. 'ingestion')
- Activate the Service itself and monitor its execution.
- Optionally modify the Settings (w/o need of re-starting the Service)
- Optionally stop/restart a Service in case of unforeseen circumstances.

The Services are execute on the eStation as 'daemons', e.g. as detached process that runs in the background, and will therefore continue their execution even if the parent process (e.g. the User Interface or the terminal) are stopped.

The following Services are implemented on the system:

- Get EUMETCast: to copy data from PC1 to eStation 2.0
- Get Internet: to copy data from remote servers eStation 2.0
- Ingestion: to convert the incoming products into GTiff format, and optionally re-project to the defined 'mapset'.
- Processing: to derive from the incoming products additional indicators.
- System: to manage all background operations, such as data/database synchronization between PC2/3, database dump, system diagnostic.

In nominal conditions of the MESA Station, the Services and the 'Analysis' option should be activated on PC2 and PC3 as in the following table:

Service	Status on PC2	Status on PC3
Get EUMETCast	ON	OFF
Get Internet	ON	OFF
Ingestion	ON	OFF
Processing	ON	OFF
System	ON	ON
Visualization	OFF	ON

eStation2 standard format

The files ingested on the eStation2, or generated by the processing service, present some common characteristics that are been defined to facilitate the exploitation of the Users, not only on the eStation but also in third-party software. There characteristics include a common file format (GTIFF), a unique convention for scaling the physical values in digital number and for encoding the 'no-data', and a list of metadata, written a 'tags' in the GTIFF file.

Data coding

(To be completed)

Nodata encoding

(To be completed)

List of metadata

(To be completed)

Data Archives

The Datasets generated on an eStation can be transferred to another machine, e.g. for completing the timeseries, by generating a 'Data Archive' or 'Archive'. This is a simple .tar compressed (.tgz) file which contains a series of GTIFF images, which are then copied in the target machine to the correct directory.

Define archive naming rules

Limits to archive size

Example of archives

Add a reference on how to create/import archives.

2.5 DATA PROCESSING

As explained in **Error! Reference source not found.** - **Error! Reference source not found.**, the eStation is mainly an EO data processor, and three main types of Services exist for processing data: the data retrieval (or 'get' services), the format conversion ('ingestion' service) and the generation of derived products ('processing' service). These layers are described in the current section, with reference to the service definition (what it does) and to its configuration (mainly, the description of database tables containing the settings).

As already mentioned in section **Error! Reference source not found.**, there are two distinct services for retrieving external datasets on the eStation, from the PC1 (Get EUMETCast) and from remote servers (Get Internet). They have some commonalities, being the latter slightly more complex for the need of accessing various and different directories structures.

2.5.1 Get EUMETCast Service

Overview (what it does)

The service copies files that are made available by Tellicast and EFTS services in a directory of PC1, which is mounted on eStation through 'samba' server. This is a 'pull' approach and it does not take care of the house-keeping of the files in the original directory (see Figure 14)

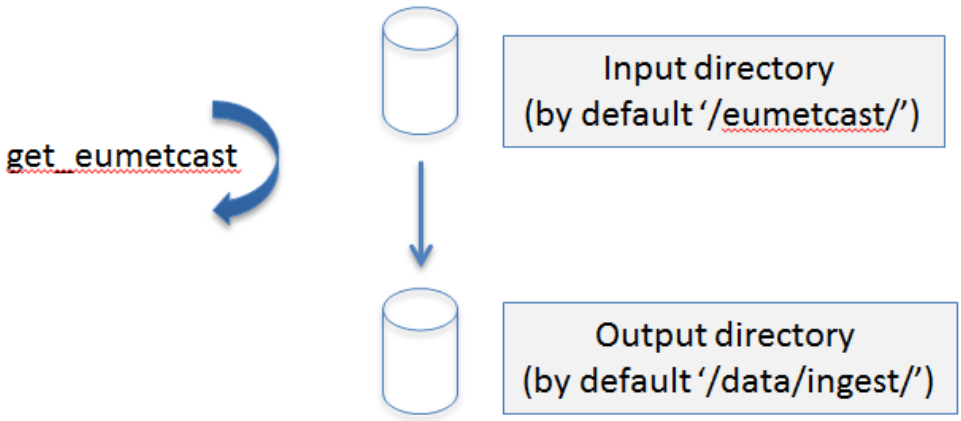


Figure 14: Get EUMETCast Service

As anticipated in paragraph 2.4, a location containing files belonging to the same EO product is defined as a 'data source'. The 'data sources' are processed independently from each other, and they can be activated/deactivated while the Service is running, so that they are taken into consideration at the next process cycle without need of a re-start.

Implementation (how it works)

The service should also ensure that a file in the input directory is copied to the output directory only once, and not continuously overwritten; therefore a list[6] of the already copied files is created and maintained for each of the data sources. Note that the files removed from the input directory are also removed from the

[6] This file is called 'get_eumetcast_processed_lists'.

list and, as a consequence, if a file is re-disseminated after having been deleted from input dir[7], it will be copied again by the Service.

The overall organization of the Get EUMETCast service can be represented as below:

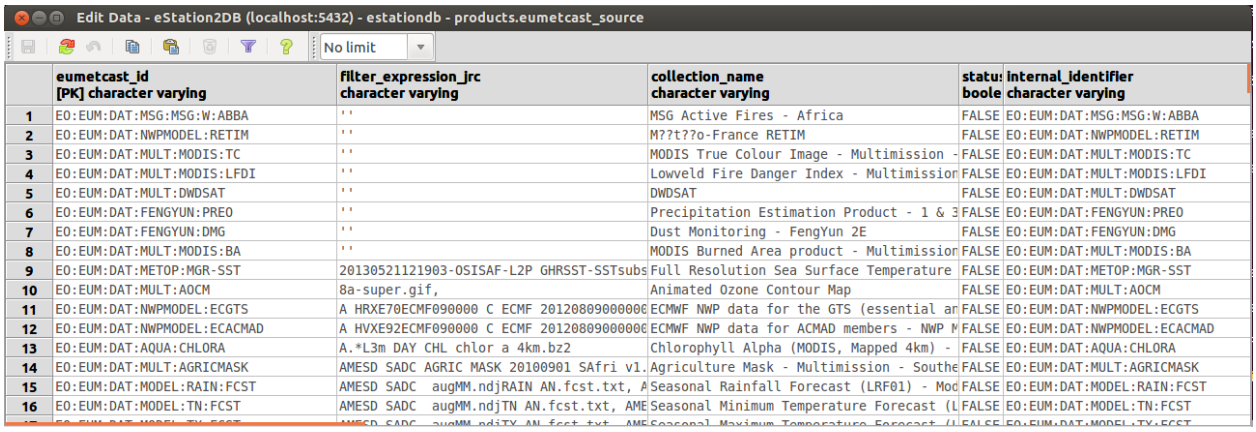
Loop over all active EUMETCast data sources and, for each of them:

- Create a list of files in input directory that match the EUMETCast source (i.e. a regular expression[8]).
- Generate list of files to be copied, i.e. the existing ones not yet copied.
- For each of the files to be copied:
  - Copy from 'input' to 'output' directory
  - Add the file to the list of copied files
- Check if all files in the list still exists in the filesystem (and clean the list accordingly)
- Save the list

Configuration

The Service is configured and controlled by two tables in the postgresql database (see also Chapter 4 of the Administration Manual).

The *eumetcast_source* table contains the description of all products disseminated by EUMETCast: Figure 15 displays a subset of the table, including the columns *eumetcast_id* (used to identify uniquely the source) and the *filter_expression_jrc*, used for associating the source the input files.



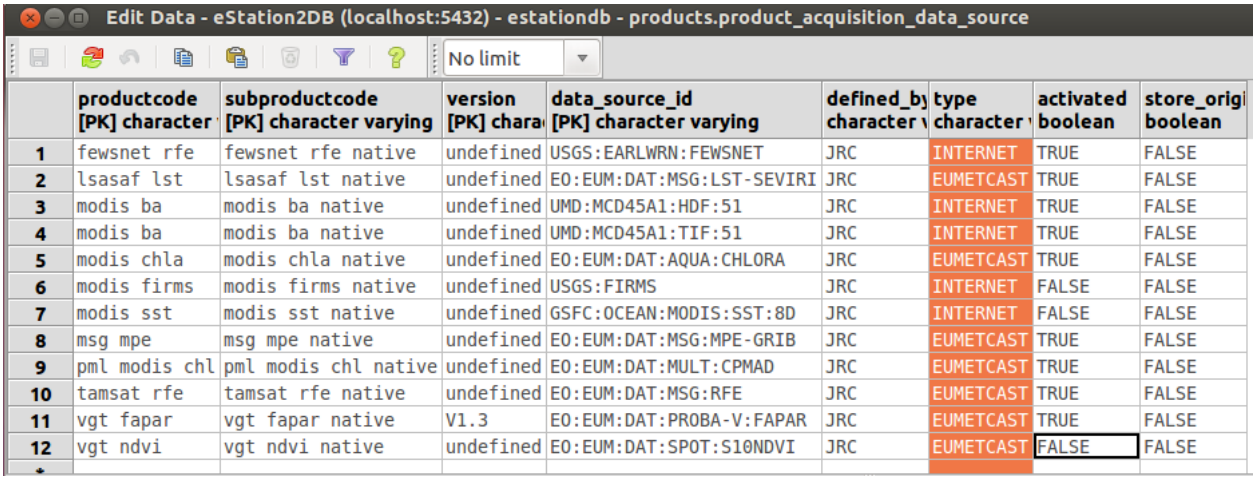
	eumetcast_id [PK] character varying	filter_expression_jrc character varying	collection_name character varying	statu: boole	internal_identifier character varying
1	EO:EUM:DAT:MSG:MSG:W:ABBA	''	MSG Active Fires - Africa	FALSE	EO:EUM:DAT:MSG:MSG:W:ABBA
2	EO:EUM:DAT:NWPMODEL:RETIM	''	M??t??o-France RETIM	FALSE	EO:EUM:DAT:NWPMODEL:RETIM
3	EO:EUM:DAT:MULT:MODIS:TC	''	MODIS True Colour Image - Multimission -	FALSE	EO:EUM:DAT:MULT:MODIS:TC
4	EO:EUM:DAT:MULT:MODIS:LFDI	''	Lowveld Fire Danger Index - Multimissior	FALSE	EO:EUM:DAT:MULT:MODIS:LFDI
5	EO:EUM:DAT:MULT:DWDSAT	''	DWDSAT	FALSE	EO:EUM:DAT:MULT:DWDSAT
6	EO:EUM:DAT:FENGYUN:PREO	''	Precipitation Estimation Product - 1 & 3	FALSE	EO:EUM:DAT:FENGYUN:PREO
7	EO:EUM:DAT:FENGYUN:DMG	''	Dust Monitoring - FengYun 2E	FALSE	EO:EUM:DAT:FENGYUN:DMG
8	EO:EUM:DAT:MULT:MODIS:BA	''	MODIS Burned Area product - Multimissior	FALSE	EO:EUM:DAT:MULT:MODIS:BA
9	EO:EUM:DAT:METOP:MGR-SST	20130521121903-OSISAF-L2P GHRSST-SSTsubs	Full Resolution Sea Surface Temperature	FALSE	EO:EUM:DAT:METOP:MGR-SST
10	EO:EUM:DAT:MULT:AOCM	8a-super.gif,	Animated Ozone Contour Map	FALSE	EO:EUM:DAT:MULT:AOCM
11	EO:EUM:DAT:NWPMODEL:ECGTS	A HRXE70ECMF090000 C ECMF 20120809000006	ECMWF NWP data for the GTS (essential ar	FALSE	EO:EUM:DAT:NWPMODEL:ECGTS
12	EO:EUM:DAT:NWPMODEL:ECACMAD	A HVXE92ECMF090000 C ECMF 20120809000006	ECMWF NWP data for ACMAD members - NWP M	FALSE	EO:EUM:DAT:NWPMODEL:ECACMAD
13	EO:EUM:DAT:AQUA:CHLORA	A.*L3m DAY CHL chlor a 4km.bz2	Chlorophyll Alpha (MODIS, Mapped 4km) -	FALSE	EO:EUM:DAT:AQUA:CHLORA
14	EO:EUM:DAT:MULT:AGRICMASK	AMESD SADC AGRIC MASK 20100901 SAfri v1.	Agriculture Mask - Multimission - Southe	FALSE	EO:EUM:DAT:MULT:AGRICMASK
15	EO:EUM:DAT:MODEL:RAIN:FCST	AMESD SADC augMM.ndjRAIN AN.fcst.txt, A	Seasonal Rainfall Forecast (LRF01) - Mod	FALSE	EO:EUM:DAT:MODEL:RAIN:FCST
16	EO:EUM:DAT:MODEL:TN:FCST	AMESD SADC augMM.ndjTN AN.fcst.txt, AME	Seasonal Minimum Temperature Forecast (L	FALSE	EO:EUM:DAT:MODEL:TN:FCST

Figure 15: *eumetcast_source* table.

A specific EUMETCast source is associated to an eStation product though the *product_acquisition_data_source*, which is displayed in Figure 16. In order to uniquely identify the native subproduct, the triplet *product/subproduct/version* is specified in the table. Note also that this table is common between the two 'get' services (EUMETCast and Internet).

[7] Retention time of input directory is normally 1 week.

[8] A 'regular expression' (see https://en.wikipedia.org/wiki/Regular_expression) is a string containing characters that can match several strings, e.g. several similar filenames.



Edit Data - eStation2DB (localhost:5432) - estationdb - products.product_acquisition_data_source

No limit

	productcode [PK] character	subproductcode [PK] character varying	version [PK] chara	data_source_id [PK] character varying	defined_by character	type character	activated boolean	store_origi boolean
1	fewsnet rfe	fewsnet rfe native	undefined	USGS:EARLWRN:FEWSNET	JRC	INTERNET	TRUE	FALSE
2	lsasaf lst	lsasaf lst native	undefined	EO:EUM:DAT:MSG:LST-SEVIRI	JRC	EUMETCAST	TRUE	FALSE
3	modis ba	modis ba native	undefined	UMD:MCD45A1:HDF:51	JRC	INTERNET	TRUE	FALSE
4	modis ba	modis ba native	undefined	UMD:MCD45A1:TIF:51	JRC	INTERNET	TRUE	FALSE
5	modis chla	modis chla native	undefined	EO:EUM:DAT:AQUA:CHLORA	JRC	EUMETCAST	TRUE	FALSE
6	modis firms	modis firms native	undefined	USGS:FIRMS	JRC	INTERNET	FALSE	FALSE
7	modis sst	modis sst native	undefined	GSFC:OCEAN:MODIS:SST:8D	JRC	INTERNET	FALSE	FALSE
8	msg mpe	msg mpe native	undefined	EO:EUM:DAT:MSG:MPE-GRIB	JRC	EUMETCAST	TRUE	FALSE
9	pml modis chl	pml modis chl native	undefined	EO:EUM:DAT:MULT:CPMAD	JRC	EUMETCAST	TRUE	FALSE
10	tamsat rfe	tamsat rfe native	undefined	EO:EUM:DAT:MSG:RFE	JRC	EUMETCAST	TRUE	FALSE
11	vgt fapar	vgt fapar native	V1.3	EO:EUM:DAT:PROBA-V:FAPAR	JRC	EUMETCAST	TRUE	FALSE
12	vgt ndvi	vgt ndvi native	undefined	EO:EUM:DAT:SPOT:S10NDVI	JRC	EUMETCAST	FALSE	FALSE

Figure 16: product acquisition datasource table

The configuration of the Get EUMETCast Service consists in two main steps:

1. Defining (modifying/adding) the sources in *eumetcast_source table*.
2. Associating the source to the products in the *product_acquisition_data_source* table.
3. Activating/Deactivating the single source in the *product_acquisition_data_source* table.

The first two steps are pre-set in the eStation 2.0 by JRC for all the products proposed to the Users, and need to be done only for additional products (Advanced Users). The activation/deactivation might be modified by the User according to its thematic needs and some specific operational constrains. The procedure for performing these operations though the GUI is described in Section 3.

References (Advanced Users)

Table 4 contains all the elements in the filesystem, both files and directories, that are relevant for the Service implementation, and is meant as a Reference for advanced Users.

Element	Directory	File	Example/Default	Description
Input Dir	/eumetcast/	-	-	Input directory of the service.
Output Dir	/data/ingest/	-	-	Output directory of the service.
Process	<BASE_DIR[9]>/eStation2/apps/acquisition/	get_eumetcast_py	-	Python module that implements the Service.
Process pid file	/tmp/eStation2/services/	get-eumetcast.pid	-	Stores the pid[10] of the service.
Processed list	/eStation2/get_lists/get_eumetcast/	get_eum_processed_list_<source_id>.list	get_eum_processed_list_EO:EUM:DAT:MSG:MPE-GRIB.list	List of the files already copied for a specific source.

[9] The base directory of the eStation installation.

[10] Process Identifier, a unique integer number associated to the Linux process.

Ancillary Info	/eStation2/get_lists/get_eumetcast/	get_eum_processed_list_<source_id>.info	get_eum_processed_list_EO:EUM:DAT:MSG:MPE-GRIB.info	Information on the execution of the Service for the specific source (displayed in the GUI).

Table 4: Filesystem elements relevant for Get EUMETCast service

2.5.1.1 Get Internet

Overview (what it does)

The service copies files available on remote ftp/http serves to the local machine, into a directory that is by default the input directory of the ingestion service.

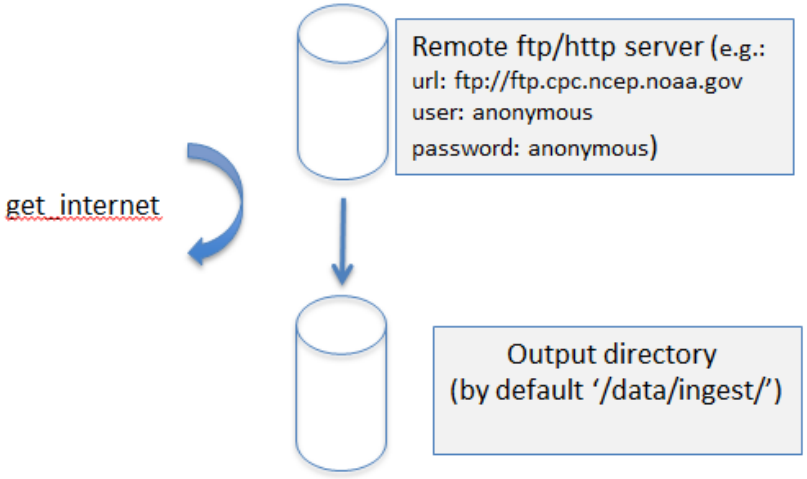


Figure 17: Get Internet Service

This service is similar to Get EUMETCast, for the principle of having different 'data sources' for the various datasets, and for keeping a list of the already copied files. Some differences exist for the need of specifying more elements to access the remote server, and take account of the directory and file naming and organization.

Implementation (how it works)

The main differences in the mechanism of the Get Internet Service with respect to the Get Internet are:

1. In addition to the location of the remote files (a URL address), user credentials have to be provided (username and password).
2. On the remote server a complex directory structure can exist (see examples below).
3. Unlike for the NRT[11] dissemination of EUMETCast, a full collection of data normally exists on the remote server, which potentially covers a large time windows; therefore time sub-setting has normally to be considered.

When a file is removed from the remote source, its name is not removed from the list of the processed file. As a consequence, files removed and re-inserted on the remote server are not downloaded a second time: the only option for re-downloading files is to manually clean the processed list.

The overall process for the Get Internet service is described below:

Loop over active Internet sources

- Create a list of files on the remote server that match the internet source definition (either type 1 or 2).

[11] Near Real Time

- Compute list of files to be copied, i.e. the ones available but not yet copied
- For each file to be copied:
    - Download the file to the local target dir.
    - Add the file to the list
- Save the list

Configuration

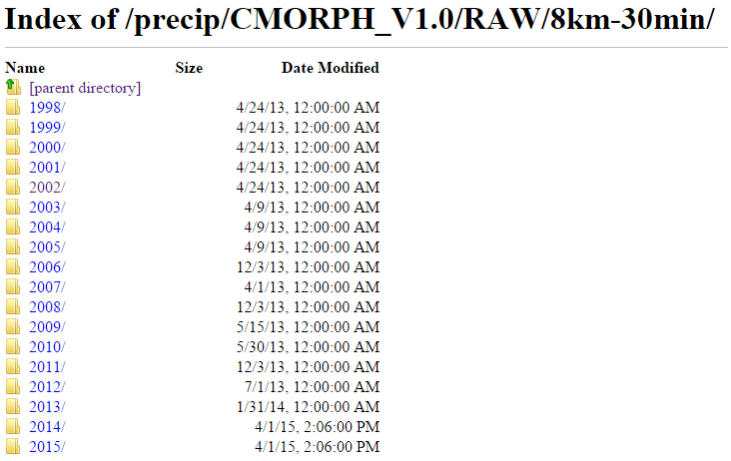
The Get Internet service takes into account the existence of both ftp servers, whose directory tree can be 'navigated' to search for the requested files, and http servers, where the exact location and naming of the files has to be known in advance. The two cases are described separately hereafter.

Type 1: ftp servers

Let's start from an example and consider the NOAA ftp server at the address ftp://ftp.cpc.ncep.noaa.gov, which is possibly a very rich and articulated site we access for retrieving, e.g., the CMORPH V 1.0 dataset. The specific dataset we are interested in (8 km resolution, 30 minute repeat cycle raw data) is located under the directory:

ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/RAW/8km-30min/

This address is therefore the starting point of our search in the server: as displayed in Figure 18, data are organized in subdirectories named after the year. The filename is like CMORPH_V1.0_8km-30min_201103.tar.



Index of /precip/CMORPH_V1.0/RAW/8km-30min/

Name	Size	Date Modified
[parent directory]		
1998/		4/24/13, 12:00:00 AM
1999/		4/24/13, 12:00:00 AM
2000/		4/24/13, 12:00:00 AM
2001/		4/24/13, 12:00:00 AM
2002/		4/24/13, 12:00:00 AM
2003/		4/9/13, 12:00:00 AM
2004/		4/9/13, 12:00:00 AM
2005/		4/9/13, 12:00:00 AM
2006/		12/3/13, 12:00:00 AM
2007/		4/1/13, 12:00:00 AM
2008/		12/3/13, 12:00:00 AM
2009/		5/15/13, 12:00:00 AM
2010/		5/30/13, 12:00:00 AM
2011/		12/3/13, 12:00:00 AM
2012/		7/1/13, 12:00:00 AM
2013/		1/31/14, 12:00:00 AM
2014/		4/1/15, 2:06:00 PM
2015/		4/1/15, 2:06:00 PM

Figure 18: example of ftp server (CMORPH dataset)

The idea is to identify the files to be downloaded through a regular expression, composed by two parts:

- A fixed prefix that represent the starting point of the search. It is called 'url' and in our example is:

  url = ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/RAW/8km-30min/

- A variable part for identifying all subdirectories and filenames we are interested in. In our example is:

  include_files_expression = [12][0-9][0-9][0-9]/CMORPH_V1.0.*

Note that the part '[12][0-9][0-9][0-9]/' corresponds to the 'year' subdirectory, and CMORPH_V1.0.* matches all files, regardless to their date.

These two variables are sufficient to define the 'internet source' for the ftp servers.

Type 2: http servers

Unlike for the ftp servers, on the http servers there might be restrictions in reading the contents of a directory, so that it is possible to access and download a given file, but not to 'walk' the directory tree down to that file. Consider as an example the Ocean Colour datasets distributed by GSFC-NASA at the address http://oceandata.sci.gsfc.nasa.gov/cgi/getfile/: entering in a browser this address the remote directory is not displayed, while it is possible to download the files by indicating the full name, e.g.:

http://oceandata.sci.gsfc.nasa.gov/cgi/getfile/A2015048.L3m_DAY_CHL_chlor_a_4km.bz2

As a consequence, we cannot read the contents of the remote directories and match it with some regular expression; the full filename has instead to be known in advance, including the variable part related to the observation date. The approach adopted for the http servers is therefore slightly more complicated than for the ftp servers, and is based on 3 elements, as described below.

- A fixed 'url' is provided as the initial part of the URL address, e.g.:

  url = http://oceandata.sci.gsfc.nasa.gov/cgi/getfile/

- A 'template' is provided to define the remaining part of the path (subdirectories and filename), which depends on the date. This part is still called 'include_files_expression', as for ftp servers, but its definition is different. It is not anymore a 'regular expression', rather a template containing %*type* elements that represents part of a date field (e.g. year, month, day of the month), and for the CHL source of our example will be:

  include_files_expression = A%Y%j.L3m_DAY_CHL_chlor_a_4km.bz2

  Note that the same notation as in UNIX *date* function is adopted[12] : in the following table we reproduce the format fields most commonly used:

Format	Description	Example
%Y	4-digit year	2015
%m	2-digit month	12
%d	2-digit day of	01
%H	2-digit hour	23
%M	2-digit minute	59
%j	3-digit day-of-year (from 1 to 366)	121
%{dkm}[13]	1 digit dekad of month	1,2 or 3

[12] See e.g. http://www.cyberciti.biz/faq/linux-unix-formatting-dates-for-display/

[13] This format is a specific implementation on the eStation, and represent the 'decade' in the range 1..36.

Table 5: date formats

- Three fields are defined to define all dates in a given period, namely:
  - start_date, end_date of the period
  - frequency, i.e. the repeat cycle of the product (every day, every 30 minute)

On the basis of the 'start_date', 'end_date' and frequency, all possible dates are computed and for each of them the corresponding filename is derived by using the expression defined through the *include_file_expression*. This filename is added to the initial path specified at point a, in order to have the full path.

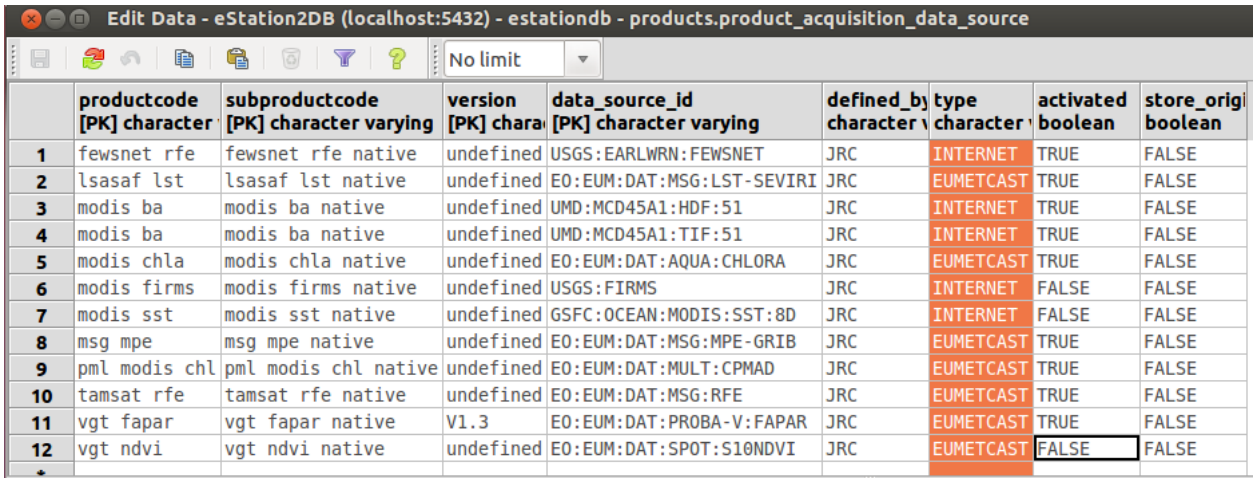
Configuration

The Service is configured and controlled by two tables in the postgresql database (see also chapter 4 of the Administrator Manual). The 'internet_source' table contains the information necessary to identify the remote source and the files to be retrieved from it; in Table 6 a description of the columns relevant for the current discussion is provided.

Column	Description	Example/List
internet_id	Unique Identifier, user defined	GSFC:CGI:MODIS:CHLA:1D
defined by	Who has defined the field (either JRC or the user)	JRC
descriptive Name	A descriptive name to identify the source	MODIS 4km Chla Daily
description	A (possibly more detailed) description	MODIS 4km Chla Daily
url	Url address of the ftp or http server. It includes the 'fixed' part of the full path.	http://oceandata.sci.gsfc.nasa.gov/cgi/getfile/ ftp://ftp.cpc.ncep.noaa.gov
username	User name for server access	anonymous
password	Password	anonymous
type	Type of server to be accessed (ftp or http)	ftp: type 1 above http_tmpl: type 2
include_filter_expression	Expression to match the variable part of the full path. It is a 'regular expression' for ftp servers and a 'template' for http ones.	A%Y%j.L3m_DAY_CHL_chlor_a_4km.bz2
files_filter_expression	Expression for matching the downloaded files in the ingestion phase. It is always a regular expression (also for the http) and refers only to the filename (not subdirectories).	.*.L3m_DAY_CHL_chlor_a_4km.bz2
status	Status of activation of the source: it should always be on, unless the source is obsolete or still under test. Note that the activation of the get for the source is done in pads table.	True
pull_frequency		
frequency	The string identifying the repeat cycle (or frequency) of the dataset to be retrieved. It applies only to http server.	e1dekad (i.e. every 'dekad) e1month (i.e. every month)
start_date	Start date of the period to be considered, in format YYYYMMDD. It applies only to http server.	20150101
end_date	End date of the period to be considered, in format YYYYMMDD. It applies only to http server.	20150631

Table 6: Contents of Internet source table (partial)

One, or more, specific sources can be associated to a 'native' subproduct though the 'product acquisition data table', which is displayed in Figure 16. Note that the table is common between the two 'get' services.



	productcode [PK] character	subproductcode [PK] character varying	version [PK] chara	data_source_id [PK] character varying	defined_by character	type character	activated boolean	store_origi boolean
1	fewsnet rfe	fewsnet rfe native	undefined	USGS:EARLWRN:FEWSNET	JRC	INTERNET	TRUE	FALSE
2	lsasaf lst	lsasaf lst native	undefined	EO:EUM:DAT:MSG:LST-SEVIRI	JRC	EUMETCAST	TRUE	FALSE
3	modis ba	modis ba native	undefined	UMD:MCD45A1:HDF:51	JRC	INTERNET	TRUE	FALSE
4	modis ba	modis ba native	undefined	UMD:MCD45A1:TIF:51	JRC	INTERNET	TRUE	FALSE
5	modis chla	modis chla native	undefined	EO:EUM:DAT:AQUA:CHLORA	JRC	EUMETCAST	TRUE	FALSE
6	modis firms	modis firms native	undefined	USGS:FIRMS	JRC	INTERNET	FALSE	FALSE
7	modis sst	modis sst native	undefined	GSFC:OCEAN:MODIS:SST:8D	JRC	INTERNET	FALSE	FALSE
8	msg mpe	msg mpe native	undefined	EO:EUM:DAT:MSG:MPE-GRIB	JRC	EUMETCAST	TRUE	FALSE
9	pml modis chl	pml modis chl native	undefined	EO:EUM:DAT:MULT:CPMAD	JRC	EUMETCAST	TRUE	FALSE
10	tamsat rfe	tamsat rfe native	undefined	EO:EUM:DAT:MSG:RFE	JRC	EUMETCAST	TRUE	FALSE
11	vgt fapar	vgt fapar native	V1.3	EO:EUM:DAT:PROBA-V:FAPAR	JRC	EUMETCAST	TRUE	FALSE
12	vgt ndvi	vgt ndvi native	undefined	EO:EUM:DAT:SPOT:S10NDVI	JRC	EUMETCAST	FALSE	FALSE

Figure 19: product acquisition datasource table

References (Advanced Users)

Table 7 contains all the elements in the filesystem, both files and directories, that are relevant for the Service implementation, and is meant as a Reference for advanced Users.

Element	Directory	File	Example/Default	Description
Input Dir	/eumetcast/	-	-	Input directory of the service
Output Dir	/data/ingest/	-	-	Output directory of the service
Process	<BASE_DIR>/eStation2/apps/acquisition/	get_eumetcast_py	/srv/www/eStation2/apps/acquisition/get_eumetcast.py	Python module in charge of the service
Process pid file	/tmp/eStation2/services/	get-eumetcast.pid		Stores the pid[14] of the service.
Processed_list	/eStation2/get_lists/get_eumetcast/	get_eum_processed_list_<source_id>.list	get_eum_processed_list_EO:EUM:DAT:MSG:MPE-GRIB.list	
Ancillary Info	/eStation2/get_lists/get_eumetcast/	get_eum_processed_list_<source_id>.info	get_eum_processed_list_EO:EUM:DAT:MSG:MPE-GRIB.info	

Table 7: Filesystem elements relevant for get_eumetcast service

[14] Process Identifier, a unique integer number associated to the Linux process.

2.5.2 Ingestion Service

Overview (what it does)

The main goal of the ingestion Service is to extract from the retrieved files the *subproducts* needed by the thematic User, for the specific *mapsets* he has defined. These *subproducts* are therefore stored in the standard eStation format (GTIFF containing specific tags), and ready on the system for visualization and further processing.

The complexity of the ingestion service, with respect to the *Get* services, relies, a part from the geo-processing of various formats, on the fact that several *subproducts* can be extracted from the same files, and for more than one *mapset*. Furthermore, the same *product* might have been retrieved from difference *sources* (e.g. EUMETCast and Internet, or different internet servers), having each source a different file naming and format.

Implementation (how it works)

To deal with the above described complexity, the overall service is organized in two main steps:

- Identifying the files from a *source* to be processed for a specific *product.*
- Process the files to extract the *subproducts* for the defined *mapset (*or *mapsets).*

Step 1: select the files for a product/source

The overall mechanism of the ingestion loop is described in Figure 20. Its role is to select a list of files existing in the input directory to be passed to a specific routine that extracts from them the defined *subproducts* for the active *mapsets*.

As first action, the list of all *products* whose ingestion is active (see also <screenshot of Acquisition tab>) is created. For each of them, the *sources* it has been retrieved from are identifying, and each *source* is treated separately, as the files coming from the various sources of the some product might be in a different format, or at least have a different filename.

Once a *source* is selected, it is possible to identify in the input directory all the files retrieved from that source, and to group them by date[15]. These files can subsequently be treated in order to extract from them one or more *subproducts*, for one or more *mapsets.*

[15] For each date, one or more file can exist, depending on the policy adopted by the data provider. In same case, namely the low resolution products, a single file exist, having continental or global coverage. With higher spatial resolutions, the information is stored in various files, also known as tiles or regions.

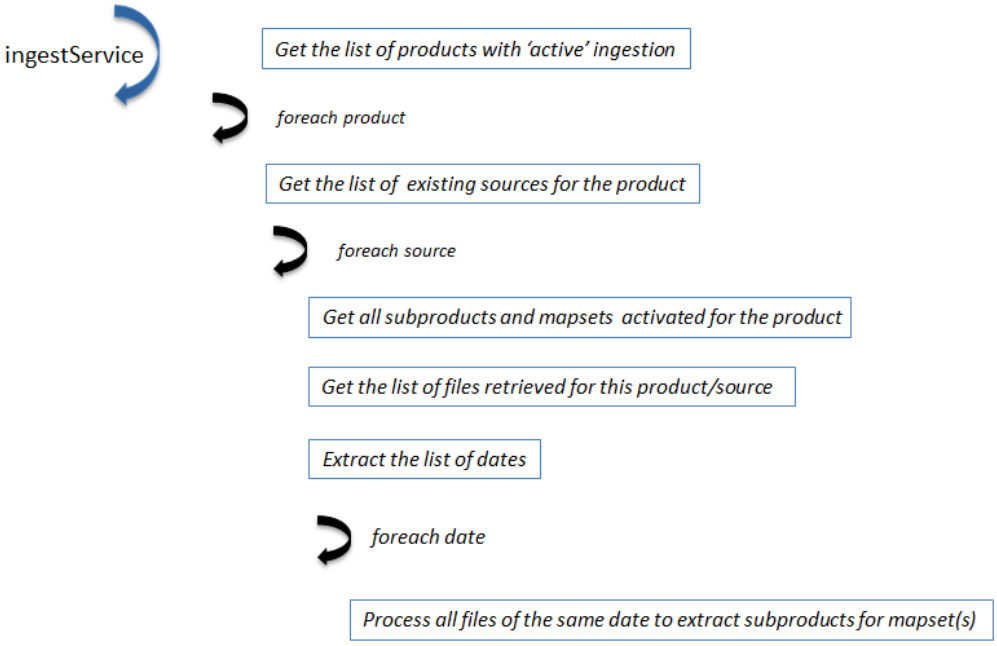


Figure 20: Ingestion Service structure

Step 2: process the files to extract subproducts

A single file, or a list of files covering adjacent geographic areas, are passed to a routine to process them and extract the numerical value to be converted into physical values and stored in eStation2 format. Each of the file can be in various file format (e.g. GTIFF, HDF4, HDF5, netcdf, HRIT, grib) and can contain one or more layers. The operations performed in this step are listed in Figure 21.

A *pre-processing* is applied in order to have, as intermediate step, a set of GTIFF and geo-referenced files containing a single *subproduct.* According to the *native* format, the series of performed operation is different. In the most general case, the *pre-processing* does the following:

- Unzip the files (from .gzip, .tgz, .bz2)
- Extract the physical values for each subproduct, and do mosaicking
- Write the values in a GTIFF format
- Geo-reference the file.

This pre-processing highly depends on the nature of the input files: a set of pre-processing routines are defined in order to deal with the most common cases.

Once the intermediate files are generate in a temporary working space, the generation of the eStation2 standard files is performed by:

- Converting the digital numbers to physical values, and convert back to digital number with a standard convention (see 2.4).
- Apply the geographic clipping/re-projection to generate the output with the defined *mapset* (i.e. for a specific boundary box and projection)
- Write to the files the eStation2 metadata (see 2.4)

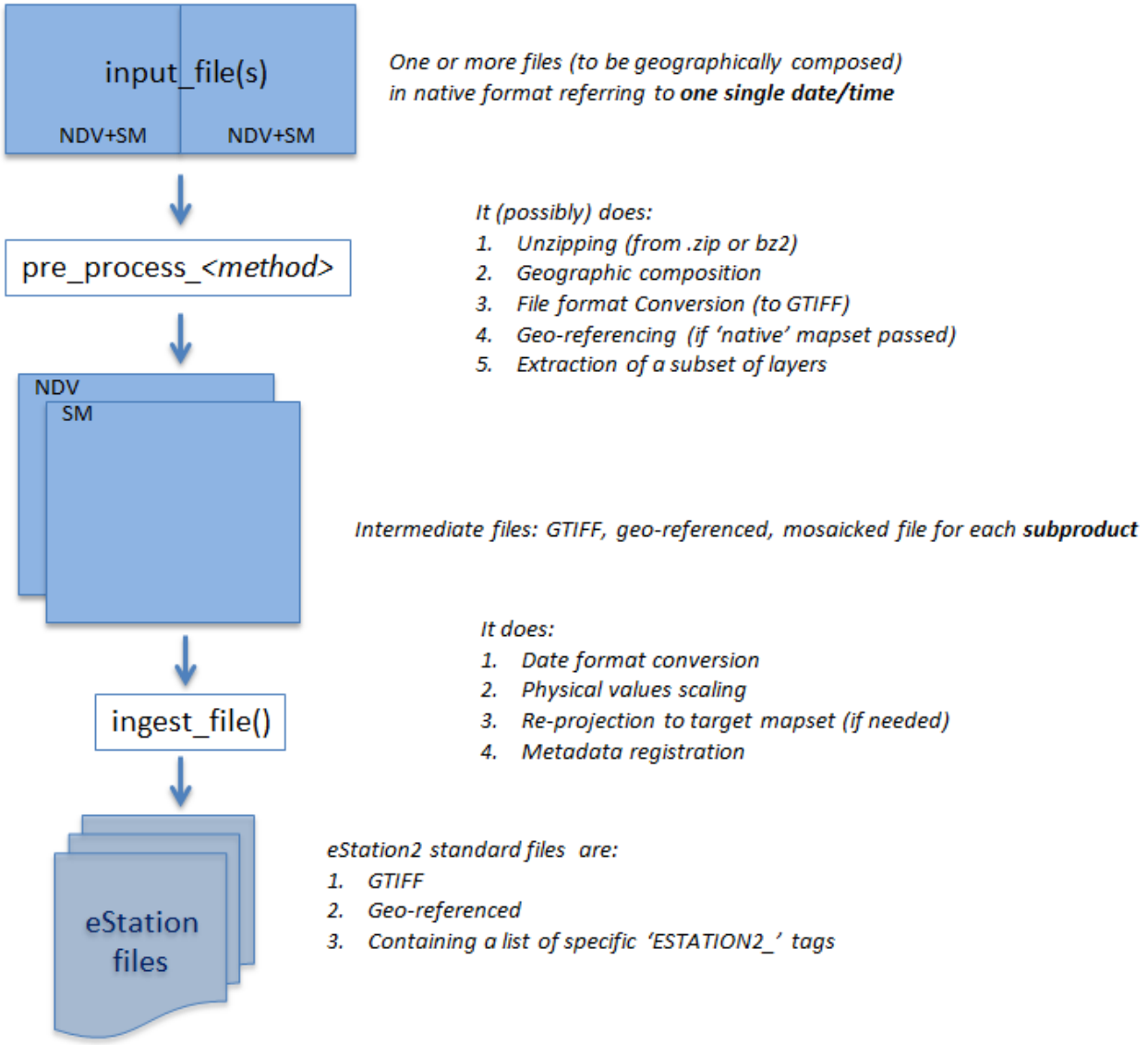


Figure 21: Ingestion data flow

Configuration

Several tables are involved in the configuration of the ingestion mechanism, and their relationship is represented in figure, in a simplified manner. The tables used for the Get services are also involved because, as already specific, the format and naming of files containing the some product can vary according to the source.

The Product table is the pivot table for the eStation2: therein all retrieved/processed and visualized products have to be defined. The Ingestion table establish the relationship between a *product/subproduct*[16]and the mapset we will apply in the ingestion. The product_acquisition_data_source table associate a *product* and a source, which can be of 'EUMETCast' or 'Internet' type. This table has been

[16] Always identified by the *product/version/subproduct* triplet.

already described in previous paragraphs, as well as the Datasource_Description (see xxxx), which mainly defines the rule adopted for the file naming. The Sub_Datasource_Description table contains the description of the contents of the input files (i.e. how many layers are present, what are the scale factor and offset, nodata coding). This table is directly linked to the Product table, to establish a relationship between the various layers in the files, and the associated *subproducts.*

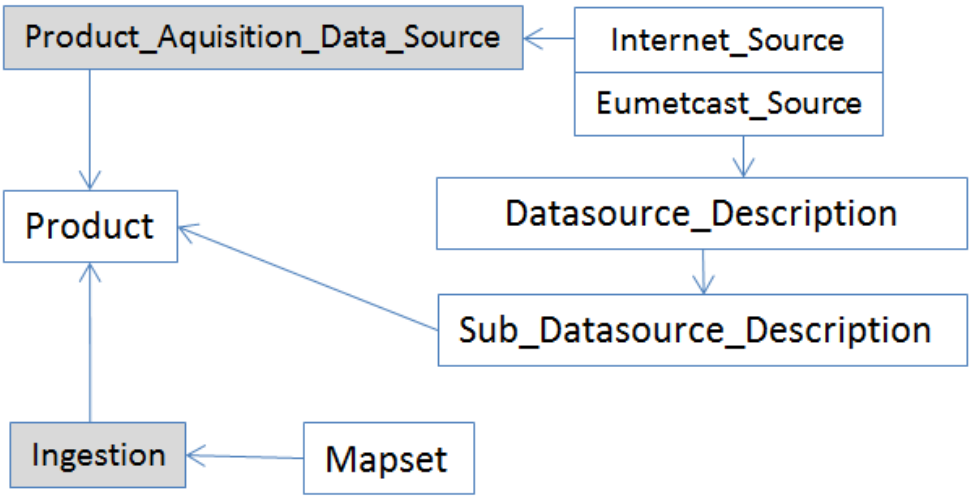


Figure 22: DB tables for ingestion

(To be completed)

References (Advanced Users)

(To be completed)

2.5.3 Processing Service

Overview (what it does)

The 'processing' service is devoted to compute EO products and indicators from the ones already existing in the system, i.e. to implement some algorithms and put them in operations. These algorithms can include re-projection functionalities, computation of temporal composition (e.g. from 10d to 1 month precipitations), computation of long term statistics and anomalies, or more complicated operations.

Implementation (how it works)

(To be completed)

Configuration

(To be completed)

References (Advanced Users)

(To be completed)

2.5.4 System Service

Overview (what it does)

(To be completed)

Implementation (how it works)

(To be completed)

Organization (where settings are stored in DB)

(To be completed)

References (Advanced Users)

(To be completed)

3. USER'S GUIDE

In the eStation 2.0 system most of functionalities are accessed through the user interface and there is no integration in the Ubuntu menu has been implemented. The current chapter describes all the operations that can be performed from the User Interface.

3.1 ACCESSING THE USER INTERFACE

- Access on the local machine [Open a browser (Mozilla Firefox) and go to the address localhost:/
- How to access the user interface on a different machine on the network
- What to do if the page is not displayed -> refer to trouble-shouting

Dashboard

The dashboard presents an overview of a MESA Full eStation or MESA Light eStation and the status of activation of the various services. When opening the GUI of the eStation in a browser, the Dashboard is the predefined entry page, and it is meant to be accessed for the basic verification of functionality, and initial diagnostic operations.

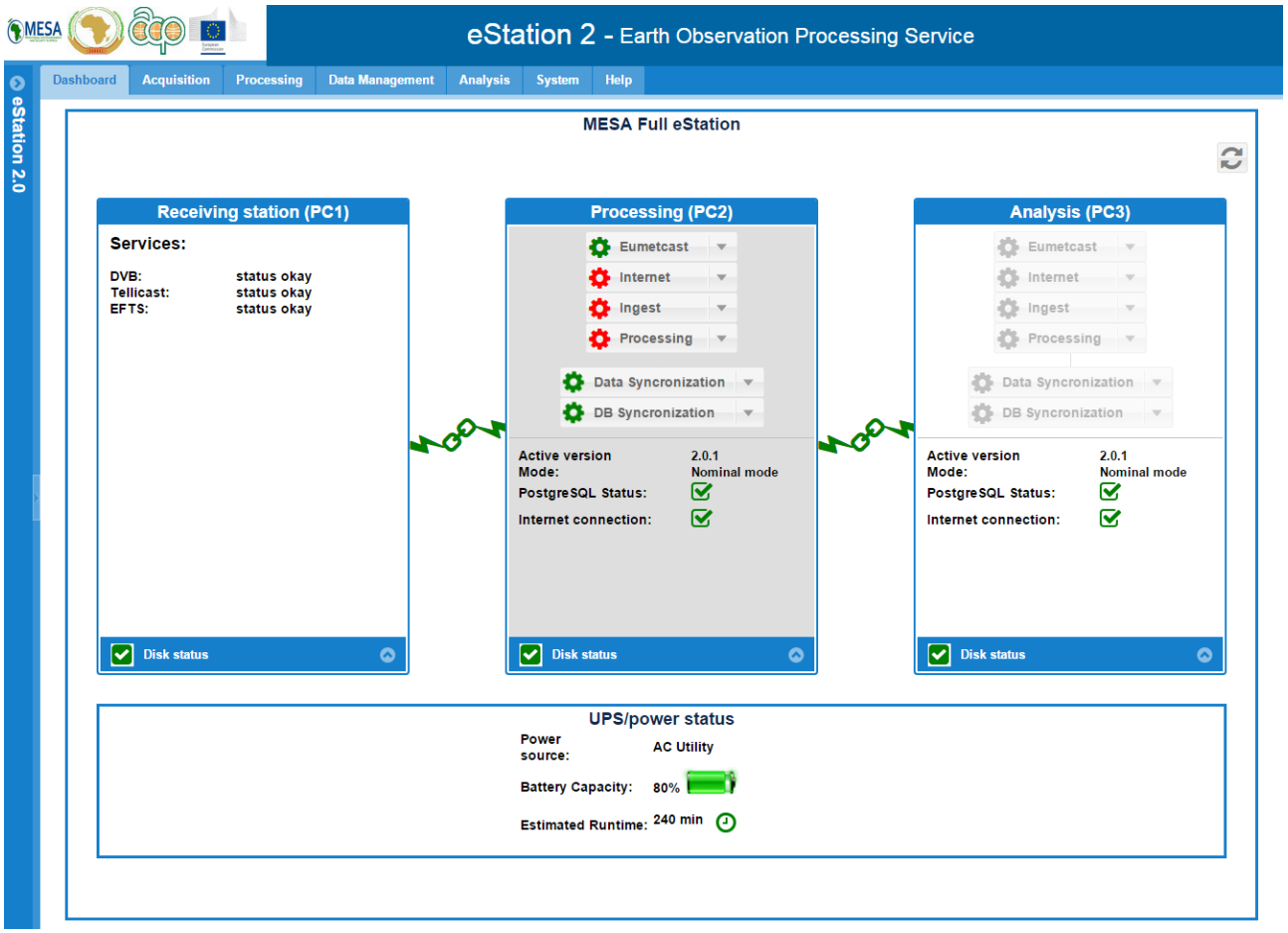


Figure 23: Dashboard of a MESA Full eStation

A panel representing the status of the three PCs is displayed; with on the bottom part the status of the external UPS. The grey background of a PC makes evident the machine the User is connected to (PC2, in case of Figure 2).

The GUI can be opened from PC2 or PC3, each with its own IP address or hostname (or localhost). In nominal mode all services are running on PC2 and can be controlled only from the GUI of PC2. In this case the services on PC3 are disabled. In degradation mode PC3 will take over the services when PC2 is down.

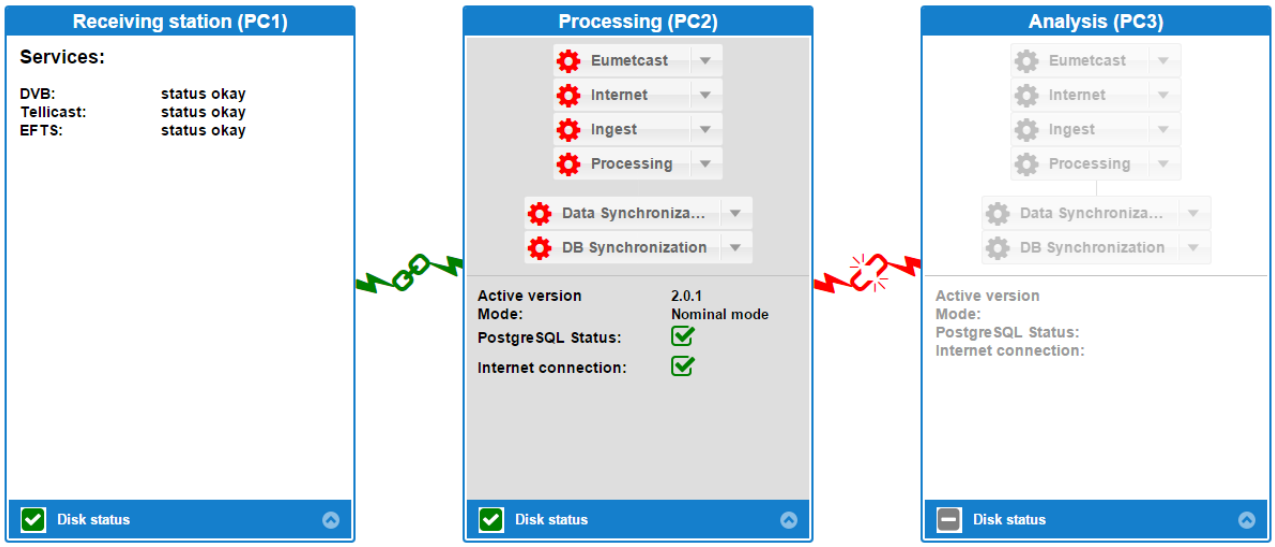
To control the services from the Dashboard, open the GUI from the PC where the services are running, in this case PC2.

To refresh the Dashboard, click on the  button.

Connections between the three PCs

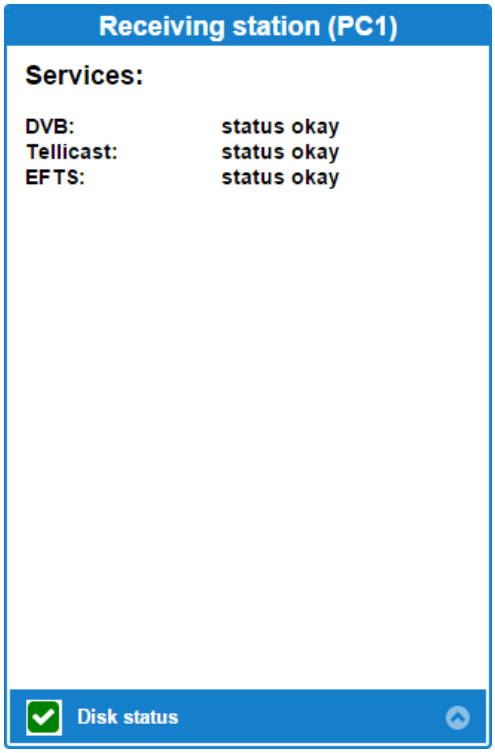
When two PCs can establish a connection between them, you will see a green chained connection  and a red unchained connection sign  when there isn't.

In the following situation, PC2 is connected to PC1 but not connected with PC3. Because P2 cannot connect to PC3, you will see all the information in the PC3 panel grey.



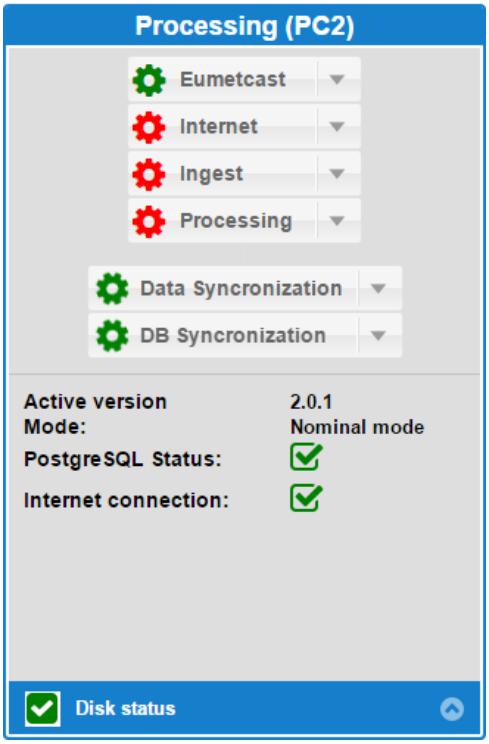
Description of PC1

Describe the functionality and tools available on the receiving station (PC1).



Description of PC2

PC2 has the role as Processing server and runs in nominal mode all services.

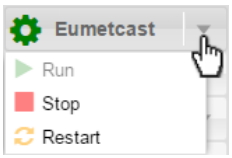


There are 6 services:

Eumetcast	Get data from the receiving station (PC1) for all activated products that have one or more activated Eumetcast data sources defined.
Internet	Get data from internet sources (FTP or Http) for all activated products that have one or more activated Internet data sources defined.
Ingest	This service is the pre-processing that ingests/converts the incoming raw data for a product to the pivot format, and subsets (clips) them to a specific 'mapset'.
Processing	This service runs all defined processing chains that are activated.
Data Synchronization	Synchronizes all the data on PC2 with the data on PC3.
DB Synchronization	Synchronizes the database on PC2 with the database on PC3.

The services Eumetcast, Internet and Ingest can individually be started, stopped or restarted.

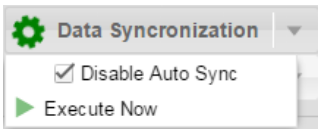
By clicking on the arrow next to the title, a menu will drop down with the items Run, Stop and Restart.



If the service is running, then the cog icon is green  and red  when the service is not running.

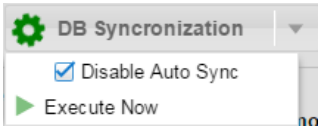
You can refresh the current status of all services by clicking on the title of a service menu button.

The **Data Synchronization** service is by default enabled to automatically synchronize the data every X minutes. The Auto Sync can be disabled by unchecking the checkbox in the Data Synchronization menu.



Manually execute the Data Synchronization by clicking on the "Execute now" menu item.

The **DB Synchronization** service is by default enabled to automatically synchronize the database. The Auto Sync can be disabled by unchecking the checkbox in the DB Synchronization menu.



Manually execute the DB Synchronization by clicking on the "Execute now" menu item.

For information only, in the PC2 panel under the services menu buttons you see the following info:

Active version:	The version of the installed eStation on PC2.
Mode:	Nominal mode or Degradation mode
PostgreSQL Status:	when the PostgreSQL database on PC2 is running. when the PostgreSQL database on PC2 is down.
Internet connection:	if there is an Internet connection from PC2. if there is NO Internet connection from PC2.

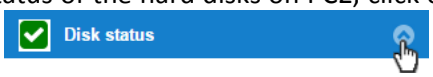
Disk status

On the bottom of the PC2 panel you see a quick look of the status of the hard disks on PC2.

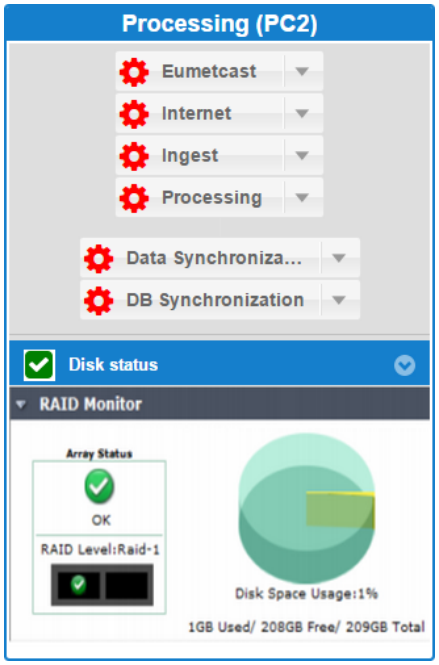
The icon means the hard disks are working well.

The icon means that there is a hard disk failure.

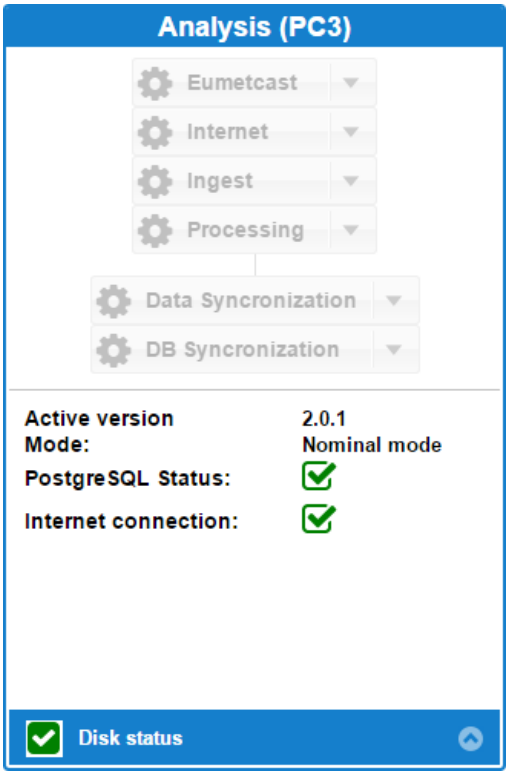
To see the full status of the hard disks on PC2, click on the up arrow in the circle on the bottom of the PC2 panel.



The hard disk status information will slide up.



Description of PC3



Description of UPS status

Acquisition

Under ‘Acquisition’ two main services are merged: the retrieval of data from the Receiving Station and the internet (get_eumetcast and get_internet services) and the ingestion, i.e. the pre-processing that converts incoming data to the pivot format, and subsets them to a specific ‘mapset’ (see …)

The acquisition page shows all activated products for each product category and allows the user to check and control the status of the services Eumestcast, Internet and Ingest, to check the status of completeness of the ingested datasets, activate or deactivate each individual Get and Ingest defined for a product and to display the log files associated to each individual Get and Ingest defined for a product.

The acquisition page also gives the user the possibility to activate or deactivate a product, to manage a products definition and its assigned Get and Ingest definitions, and to add a new product and assign an existing or new created Get and Ingest definition (all forms still to be implemented).

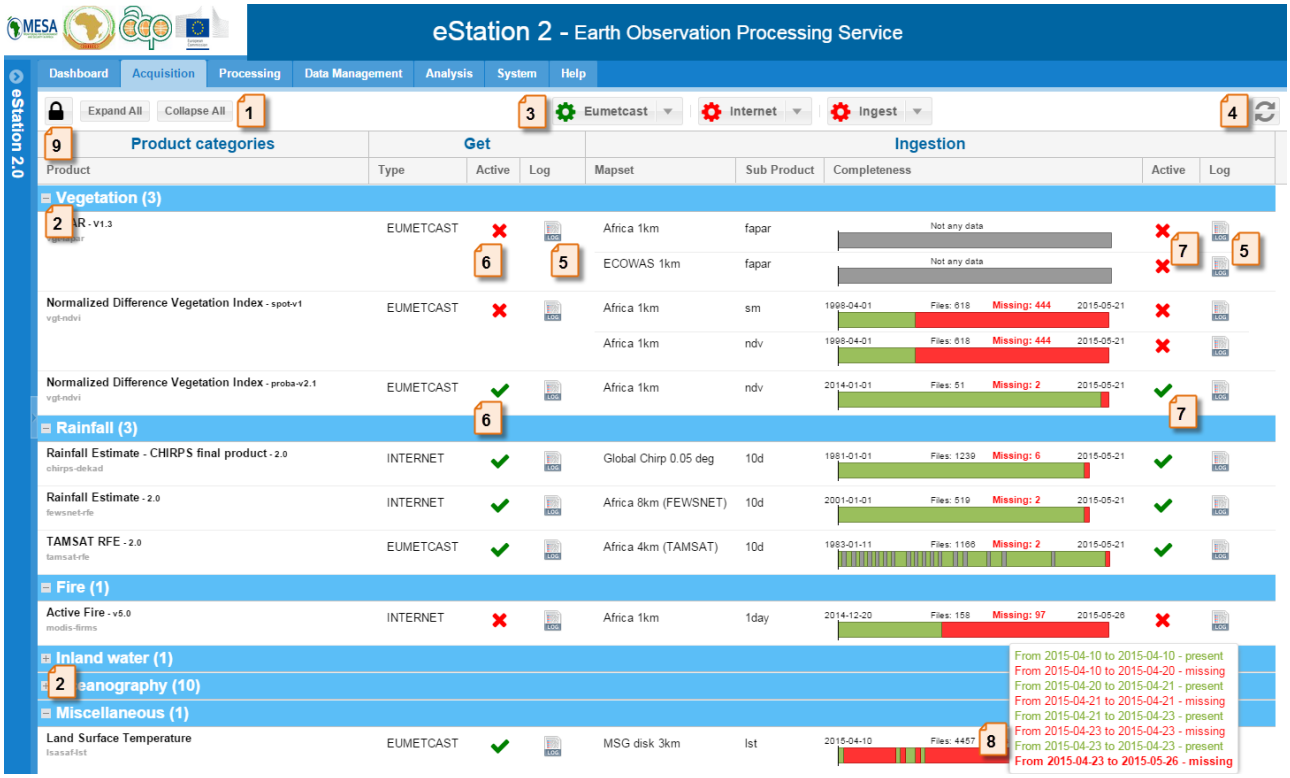
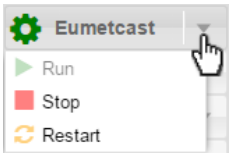


Figure 24: Overview of the Acquisition tab

Corresponding to the number in the above figure, the User can perform the following operations:

1. Expand all product categories to see all their activated products by clicking on the “Expand all” button and to collapse all product categories by clicking on the “Collapse All” button.
2. Expand or collapse each product category individually by respectively clicking on the + or – sign on the left of the title of a product category.
3. The services Eumetcast, Internet and Ingest can individually be started, stopped or restarted.

   By clicking on the arrow next to the title, a menu will drop down with the items Run, Stop and Restart.

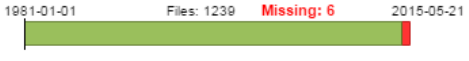


If the service is running, then the cog icon is green and red when the service is not running.

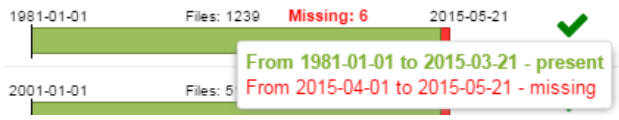
You can refresh the current status of all services by clicking on the title of a service menu button.

4. To refresh/reload the Acquisition page click on the refresh button .
5. View the log file for an individual 'Get' or 'Ingest' defined for a product, click on the LOG icon next to a Get or Ingest. A new window will open with the contents of the log file.
6. Activate/deactivate a single 'Get' source.

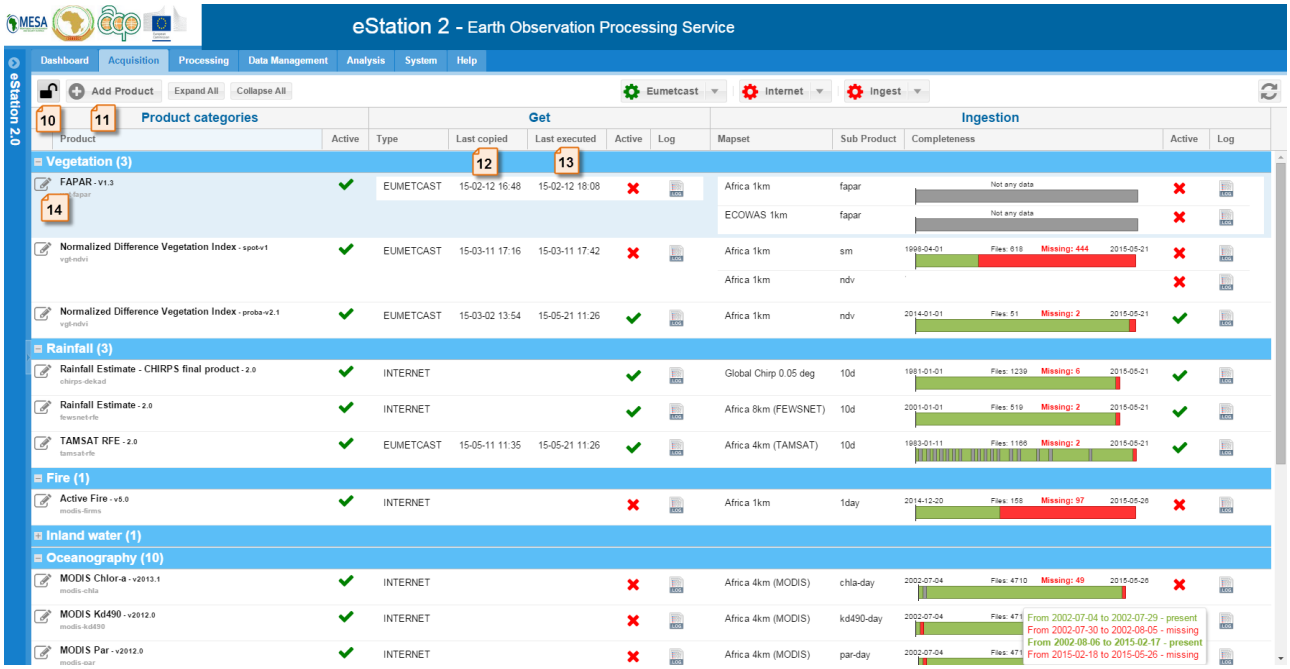
   The ✔ icon means the 'Get' source is activated. Click on the ✔ icon to deactivate the 'Get' source. The ✖ icon means the 'Get' source is deactivated. Click on the ✖ icon to activate the 'Get' source.
7. Activate/deactivate a single 'Ingestion'. An ingestion refers specifically to the source/and subproduct.

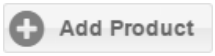
   The ✔ icon means the 'Get' source is activated. Click on the ✔ icon to deactivate the 'Get' source. The ✖ icon means the 'Get' source is deactivated. Click on the ✖ icon to activate the 'Get' source.
8. Check the status of completeness of ingested datasets. For each 'Ingestion' the Acquisition page shows a completeness chart 

, indicating the first date and the expected last date of the dataset, the total expected files for the dataset and the total of missing files. By going over a dataset completeness chart with the mouse pointer, a list of all the periods of present, missing and permanent missing files pops up:

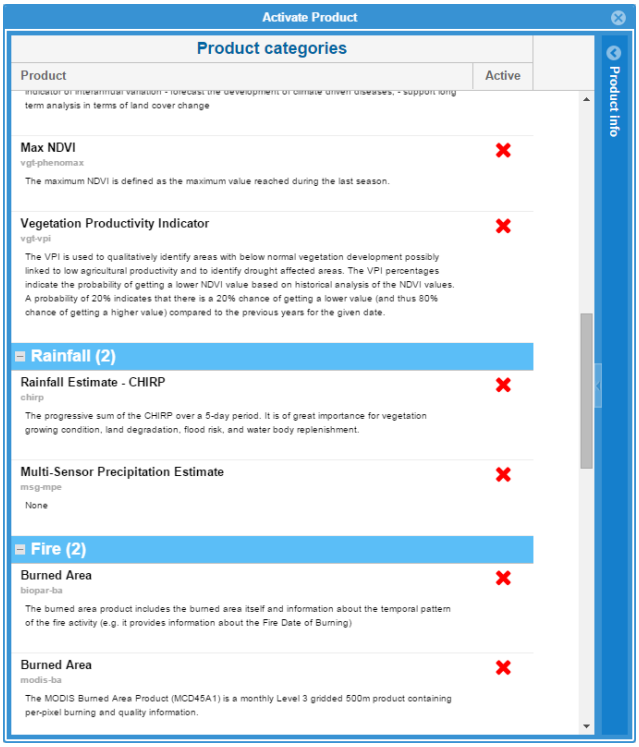


9. Lock/unlock the Acquisition page. Certain functionality and information is hidden when the Acquisition page is locked. The icon indicates that the Acquisition page is locked. Click on the icon to unlock the page. Hidden functionality and information will be shown and the lock icon becomes an unlocked icon . The following figure is the Acquisition page unlocked. The numbers indicate the added information and functionality.



10. Acquisition page unlocked. Click on the  icon to lock the Acquisition page.
11. All products listed in the Acquisition page are 'Activated' products, already defined in the system with assigned data sources and ingestions. For most products, especially those broadcasted through Eumetcast, the JRC will define their definition, their GET (data source) and Ingestion definition and assignment. To activate an already defined deactivated product, click on the 

button. A window will be shown with a categorized list of deactivated products.



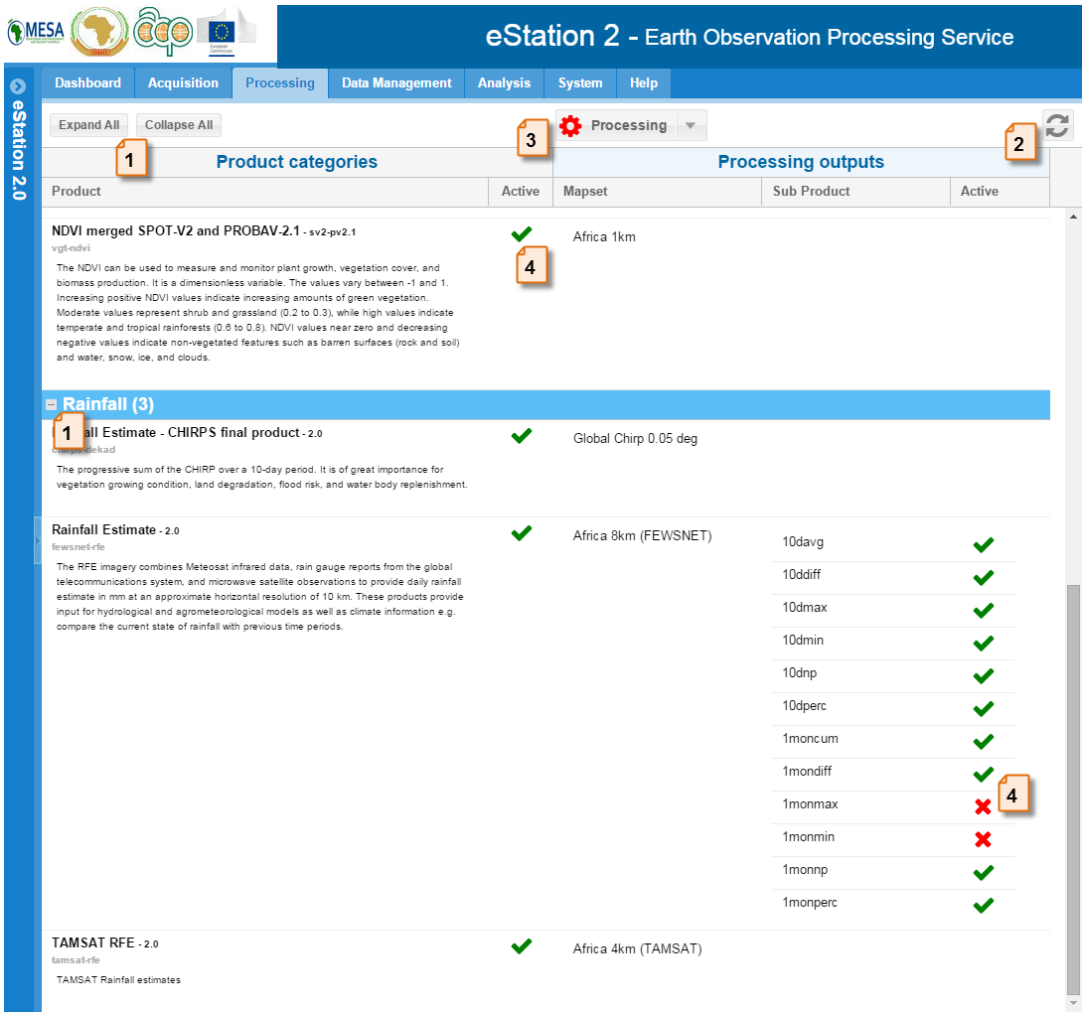
The  icon means the product is deactivated. Click on the  icon to activate the product. The product will be removed from the list and added to the list of Active products in the Acquisition page, when closing the Activate Product window.

12. The last copied column under 'Get' gives the date and time that for the 'Get' in question, a file has been copied.
13. The last executed column under 'Get' gives the date and time that the 'Get' in question has been executed (was running).
14. Editing a product. Click on the  icon to open the Edit product window.
    a. Edit general product data (Category/Version/Provider/Description/Descriptive Name).
    b. Add/Edit a specific source for getting data.
    c. Add/Edit a mapset (Ingest) for a product/subproducts.

Processing

The processing page is the interface to the 'processing' service, e.g. to the generation of EO products derived from the 'ingested' datasets.

The processing page allows the User to control the status of the processing service, and to activate/deactivate a single processing 'chain'.

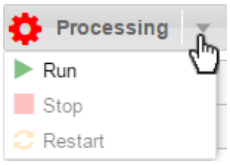


1. Expand all product categories to see all the products with one or more processing chains defined, by clicking on the Expand All button and to collapse all product categories by clicking on the Collapse All button.

   Expand or collapse each product category individually by respectively clicking on the + or – sign on the left of the title of a product category.

2. To refresh/reload the Acquisition page click on the refresh button.
3. The processing service can be started, stopped or restarted.

   By clicking on the arrow next to the title, a menu will drop down with the items Run, Stop and Restart.



If the service is running, then the cog icon is green and red when the service is not running.

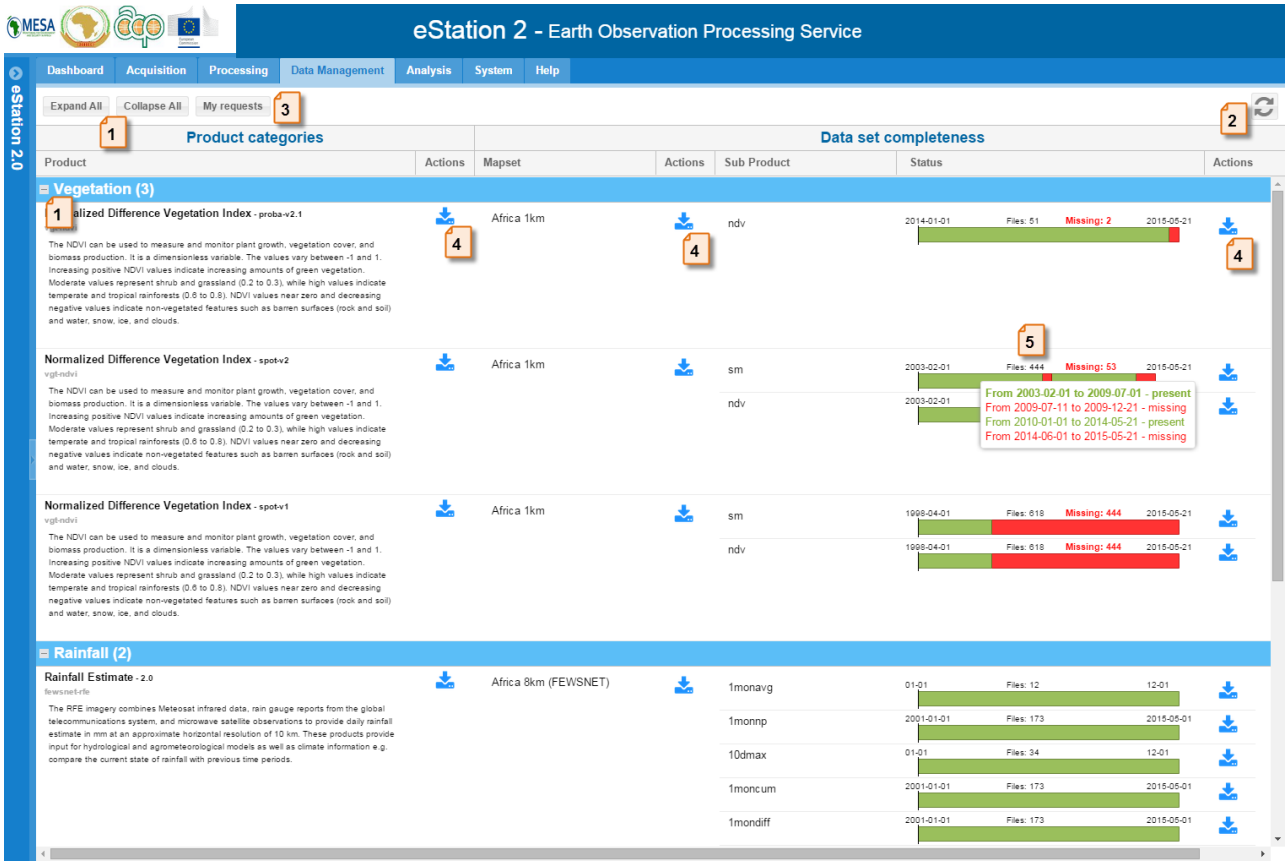
You can refresh the current status of the Processing services by clicking on the title of the service menu button.

4. Activate/deactivate all the products processing chains or a single processing chain.

The icon means the processing chain(s) for the product or the single processing chain is activated. Click on the icon to deactivate the all the products processing chains or the single processing chain. The icon means all the products processing chains or the single processing chain is/are deactivated. Click on the icon to activate the processing chain.

Data Management

The data management page is intended for having an overview and control over the completeness of the datasets existing on the eStation, and gives the possibility to send requests to another (remote) eStation to complete one or more datasets with missing files and for importing/exporting EO datasets.



1. Expand all product categories to see all the products with one or more processing chains defined, by clicking on the Expand All button and to collapse all product categories by clicking on the Collapse All button.

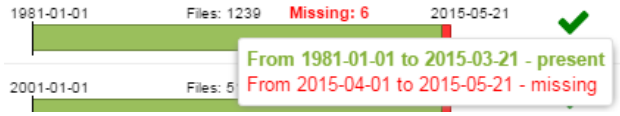
   Expand or collapse each product category individually by respectively clicking on the + or – sign on the left of the title of a product category.

2. To refresh/reload the Acquisition page click on the refresh button.

3. View the user's requests.
4. Send a request to complete all datasets of a product, all datasets of a mapset of a product or a single dataset.

5. Check the status of completeness of datasets. For each dataset the Data management page shows a completeness chart 

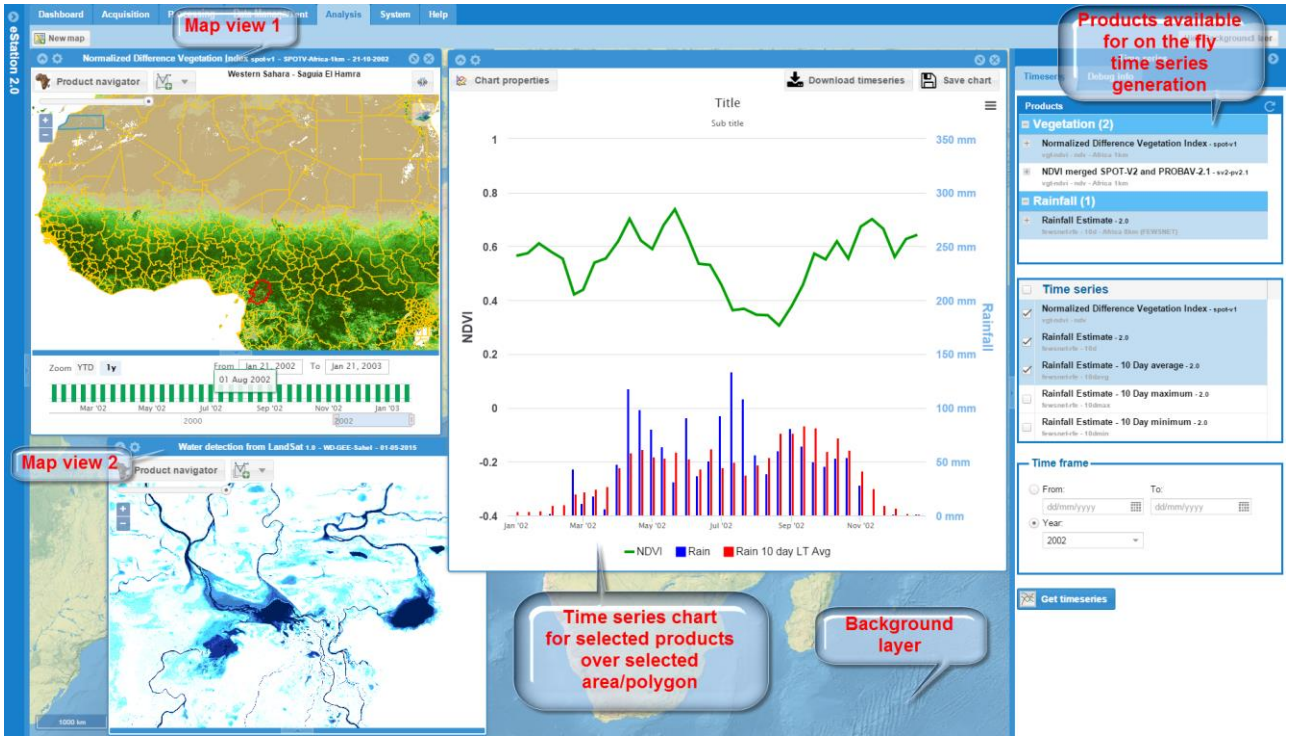
, indicating the first date and the expected last date of the dataset, the total expected files for the dataset and the total of missing files. By going over a dataset completeness chart with the mouse pointer, a list of all the periods of present, missing and permanent missing files pops up:



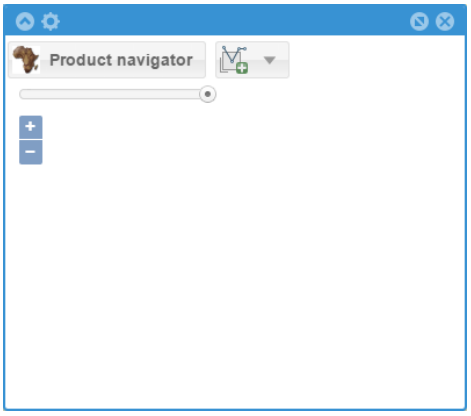
Analysis

The Analysis tool is the entry point for the eStation data visualization and analysis. It replaces the 'EMMA' web-viewer that was proposed on the eStation 1.0.

New in the analysis tool is the possibility to view more than one product, each in their own map view window and more importantly, time series are generated on the fly.



Open a new map view window by clicking on the "New map" button. A new empty map view window will be opened.

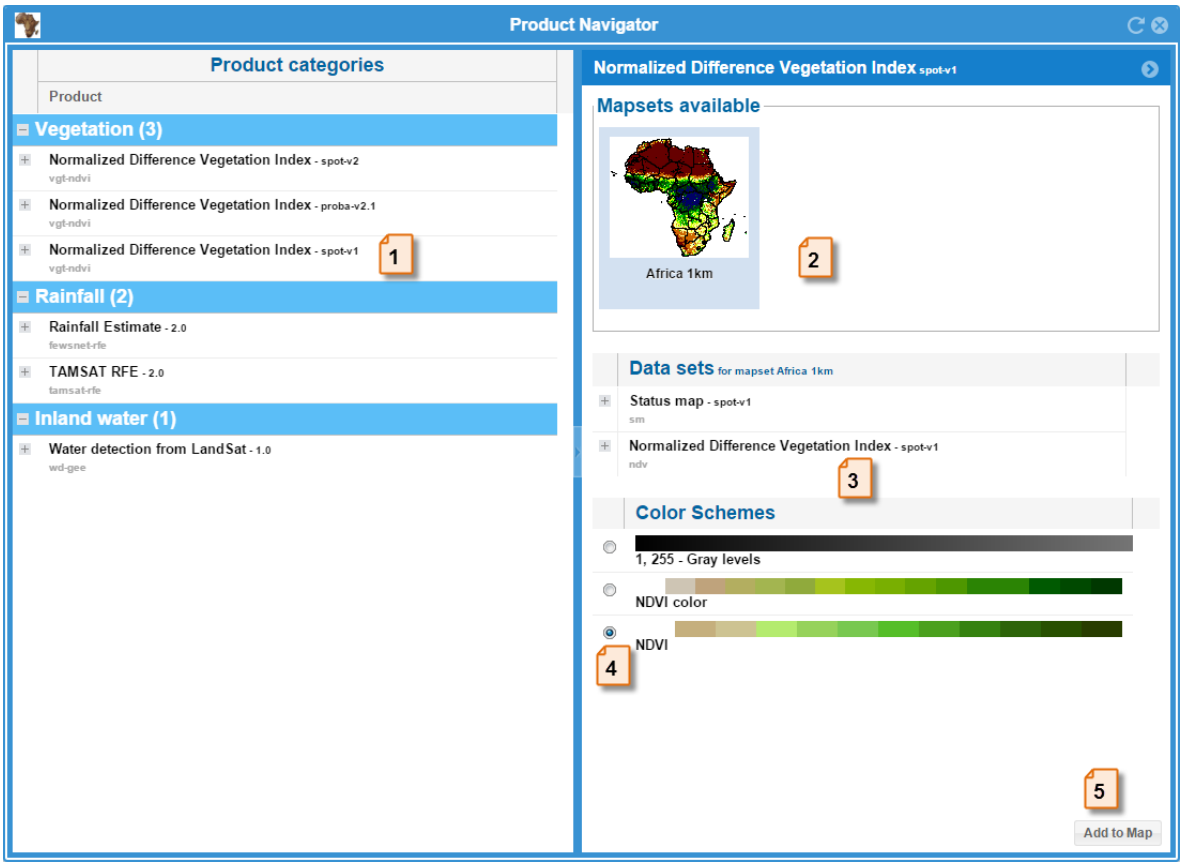


Map view functionality

Add product layer and its available timeline.

Product navigator

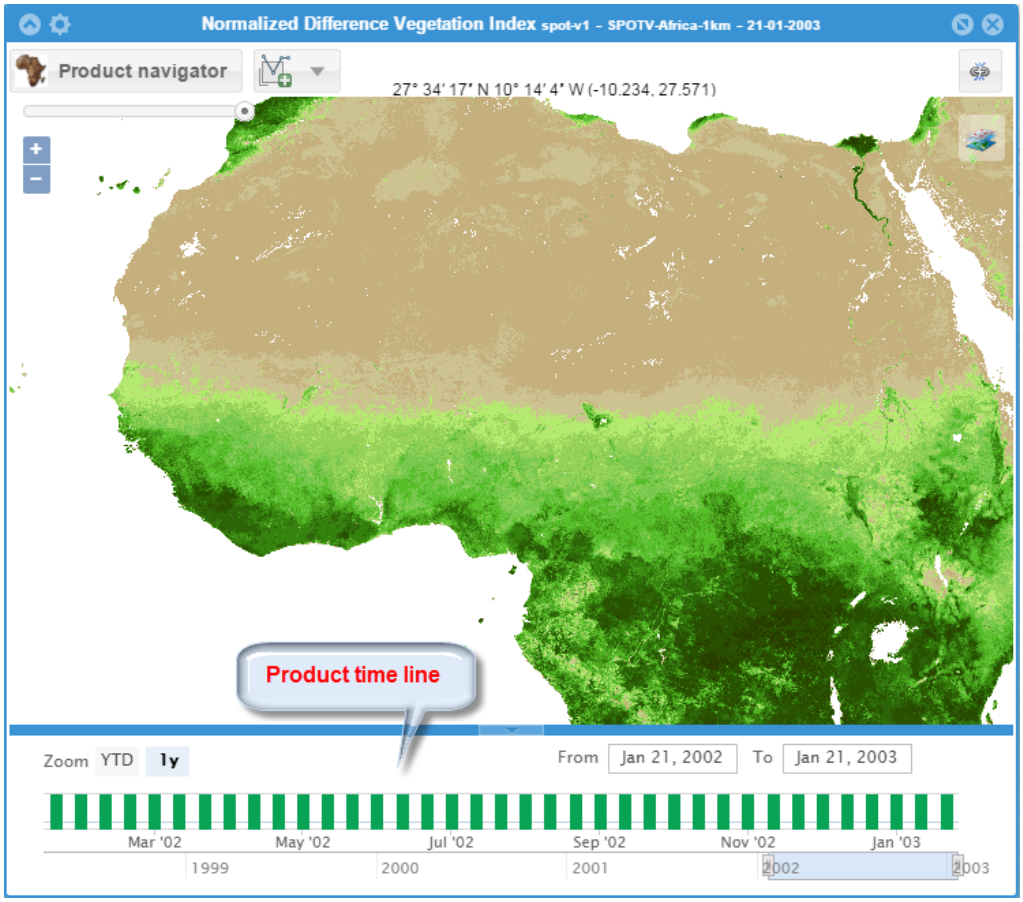
To add a product click on the “Product navigator” button. A product navigator window will be opened.



Steps to follow:

1. Select a product
2. Select a mapset
3. Select available data set for the selected map set
4. Select a Color scheme
5. Add to map

The last available date of the selected product dataset will be shown in the map view window.



Product time line

Add vector layer

Supplied vector layers (admin level 0 and 1)

Import vector layer (to be implemented)

Draw a polygon or line (to be implemented)

Map view functionality

Link/unlink map view window from background layer

Navigation: zoom in/out and panning

Layer switcher

Opacity slider

Show/hide tool bar

Collapse map view window

Full screen

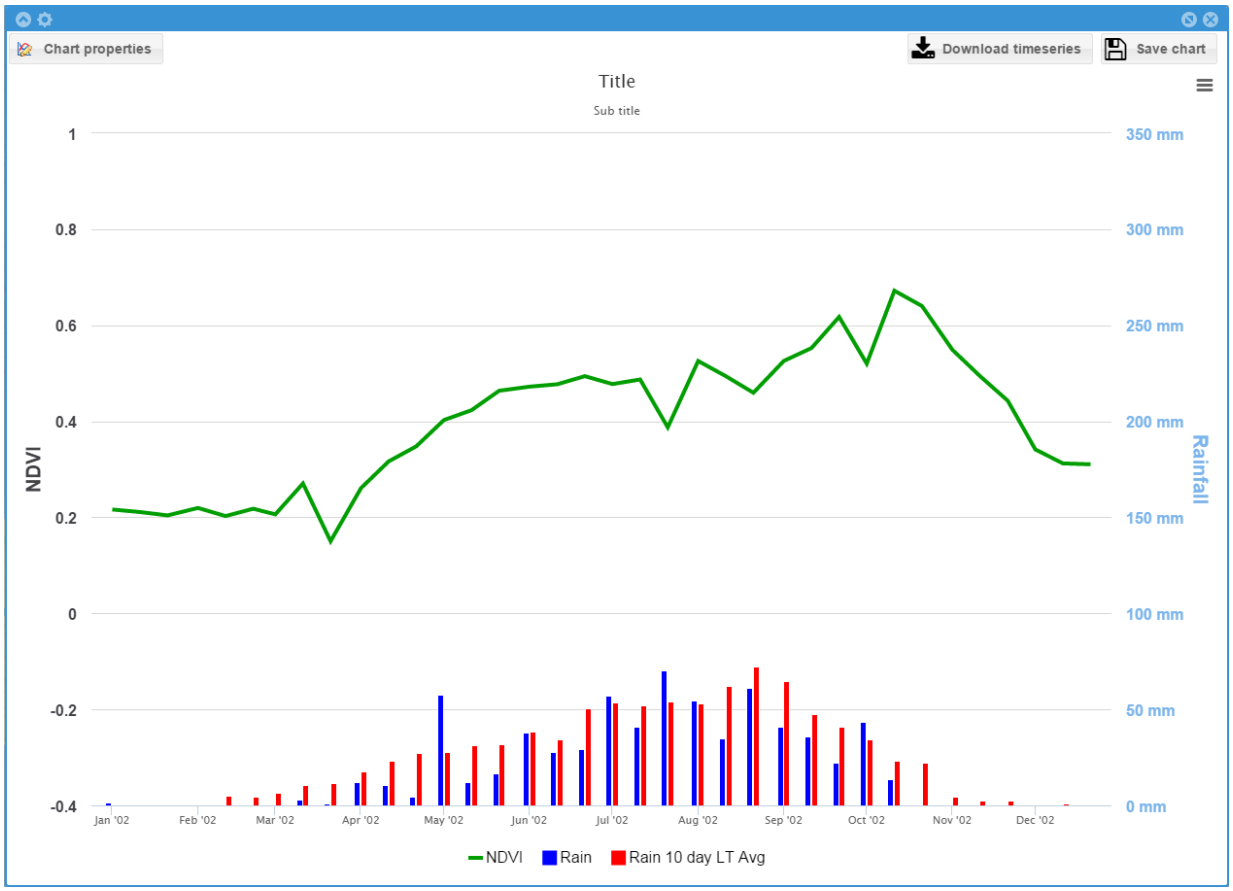
Time series chart

To generate time series, visualized in a chart, the user will have to select a polygon in one of the open map views. The selected polygon will be highlighted in red.

In the time series selection area on the left of the analysis tool, the following steps have to be taken:

1. Select one or more products
2. Select one or more product data sets
3. Select a time frame
4. Click on the “Get timeseries” button

A new time series chart window will be opened, showing the chart with the time series of the selected data sets, over the selected time frame.



System

The 'System' page hosts a series of operations that belongs either to the 'System Settings' categories or to some 'Diagnostic' tasks.

- *Definition of System Variables (directories)*
- *Machine configuration: role (PC2-PC3-single)*
- *Recovery/nominal mode*
- *Changing the IPs*
- *Logging files level .. delete the logging feels ?*
- *Diagnostic (here, or under help?)*

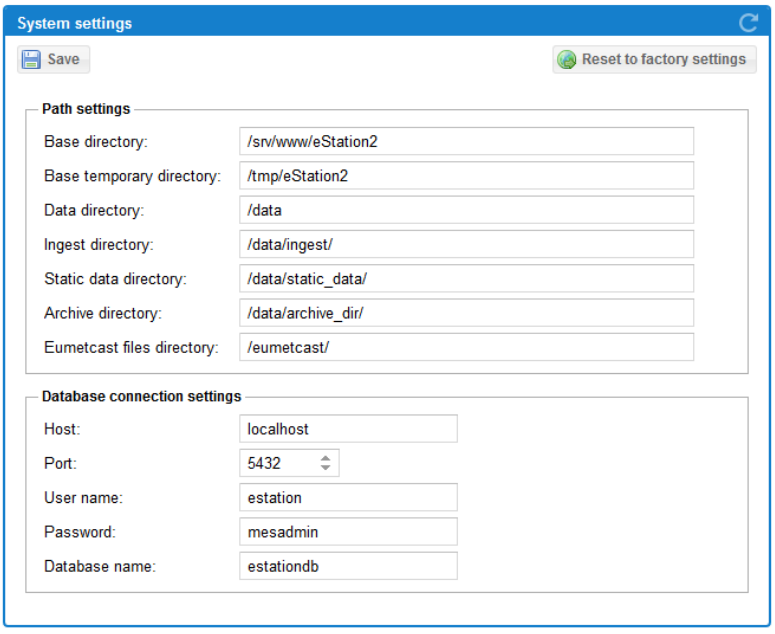


Figure 25: Settings page overview

Help