EUMETSAT Contract No. EUM/CO18/4600002180/JML

evolutions studies
automated optical sensor registration monitoring tool

Algorithm Input / Output Data Definition Document

Deliverable D07

Prepared by

Stefan SCHEIBLAUER and Thomas Nagler
ENVEO IT, Innsbruck, Austria

Document Reference: ENVEO-OSMON_D07.IODD
Issue / Revision: 1/3
Date: 02.11.2020

Document controlled by Stefan Scheiblauer

EUMETSAT STUDY CONTRACT REPORT			
Contract No: EUM/CO18/4600002180/JML	Subject: Deliverable D07 Algorithm Input / Output Data Definition Document		Contractor: ENVEO
	Star Code:	No of volumes: 1 This is volume no: 1	Contractor's Ref: Deliverable D07

Abstract:

The goal of this the Input / Output Data Definition document (IODD) is to provide a precise description of the input and output files used and generated by the OSMON Core-Module and the OSMON Reporting and Analysis Module.

The work described in this report was done under EUMETSAT Contract. Responsibility for the contents resides in the authors or organization that prepared it.

Authors:
Stefan Scheiblauer and Thomas Nagler

EUMETSAT Study Manager: Johannes Müller	EUMETSAT BUDGET HEADING

This page is intentionally left blank.

DOCUMENT CHANGE LOG

Issue/ Revision	Date	Modification	Modified pages	Observations
1.0	21.06.2019	All new	All	NA
1/1	12.07.2019	Copyright inserted Updated Section 2.1; 2.2 added AVHRR product guide to referenced documents, added data volume estimates	All	Include EUMETSAT comments
1/2	24.07.2019	added definition, Updated Section 2.32 and 4	All	Include EUMETSAT comments
1.3	02.11.2020	Update section 2, Input /Output	1, 5-12	

This page is intentionally left blank.

TABLE OF CONTENT

APPLICAPLE AND REFERENCE DOCUMENTS

ID	Source
[RD-1]	Sentinel-2 User Handbook Issue 1 Rev 2, 24 July 2015, ESA
[RD-2]	LANDSAT 8 (L8) DATA USERS HANDBOOK, LSDS-1574 Version 4.0, April 2019, Department of the Interior U.S. Geological Survey
[RD-3]	Sentinel-3 OLCI Product Data Format Specification - Level 1 products, S3IPF.PDS.004.1 Issue 2.2, 09 October 2017, ACRI ST
[RD-4]	Sentinel-3 SLSTR Product Data Format Specification - Level 1 products, S3IPF.PDS.005.1 Issue 2.7, 06 February 2018, ACRI ST
[RD-5]	AVHRR Level 1b Product Guide, EUM/OPS-EPS/MAN/04/0029, Issue v3A, 21 January 2011, EUMETSAT
[RD-6]	ENVEO IT, (2019), Evolution Studies - Automated Optical Sensor Registration Monitoring Tool. Output data format and Reference data set Deliverable D09/D10/D11 ENVEO-OSMON_D09_D10_11.ODF_RDB_DDS_1/3

Note: If not provided, the reference applies to the latest released Issue/Revision/Version

ACRONYMS

AVHRR ..... Advanced Very High Resolution Radiometer
DW ..... Data Window
GCP ..... Ground Control Point
I/O ..... Input/Output
MB ..... Megabyte
MetOp ..... Meteorological Operational Satellite
MROI ..... Medium Resolution Optical Imager
MSI ..... Multispectral Instrument
NIR ..... Near Infrared
OLCI ..... Ocean and Land Colour Instrument
OLI ..... Operational Land Imager
PNG ..... Portable Network Graphics
QL ..... Quick Look
SLSTR ..... Sea and Land Surface Temperature Radiometer
SWIR ..... Shortwave Infrared
TIR ..... Thermal Infrared
TIRS ..... Thermal Infrared Sensor
VIS ..... Visible

DEFINITION OF TERMS

GCP	In the OSMON project the term Ground Control Point (GCP) refers to an image template covering a prominent feature on the Earth Surface, which can be identified in medium resolution optical images. GCP image templates are extracted from VIS, NIR, SWIR, TIR spectral bands of Landsat-8 OLI and TIRS data. The geographic location of a GCP is exactly specified by the geographic map projection, ellipsoid and local datum, corner coordinates and pixel spacing in Easting and Northing direction.
MROI	Medium Resolution Optical Imagers is used as a general name of medium resolution optical satellites such as Sentinel-3 SLSTR, Sentinel-3 OLCI and MetOp/AVHRR.
MROI-DW	This term is used for MROI-satellite data covering the same area as the corresponding GCP.MROI-DW are raster files in the same map projection, ellipsoid and local datum, corner coordinates and pixel spacing in Easting and Northing direction. To achieve the same pixel spacing as GCP, oversampling of the native resolution of the MROI data is applied.

1. INTRODUCTION

The goal of this the Input / Output Data Definition document (IODD) is to provide a precise description of the input and output files used and generated by the OSMON Core-Module and the OSMON Reporting and Analysis Module.

2. CORE-MODULE

The Core Module (Figure 1) estimates the shift of a level-1 MRO-DW with respect to a GCP-DW. Additionally, it supports the use of auxiliary geospatial datasets to ensure the functionality of orthorectification and quality checks with vector datasets. The estimated shifts are written to a PostgreSQL tables. If necessary, the MROI-DW can be locally stored together with a quicklook of the MROI bands.

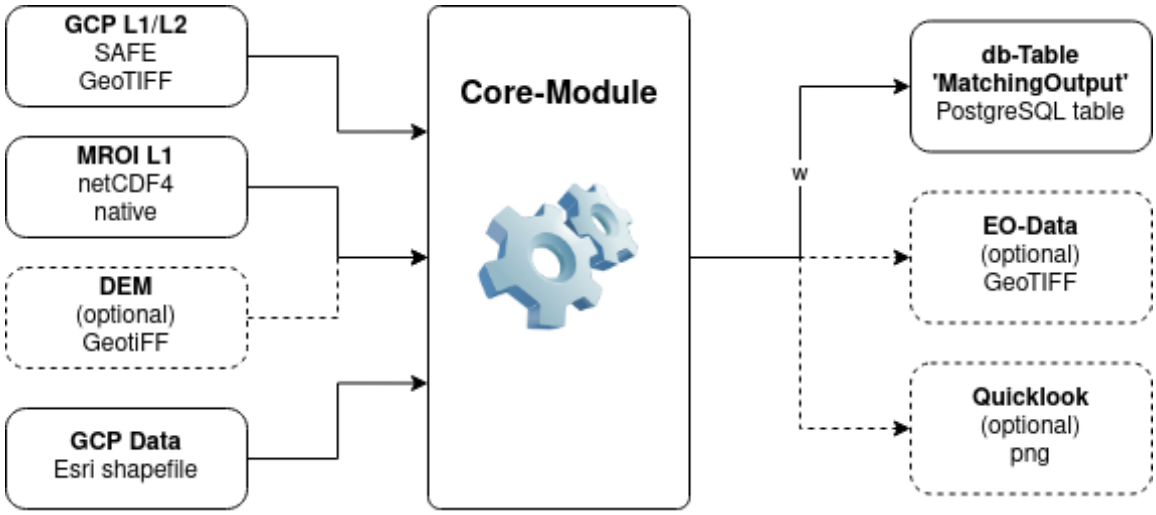

Figure 1: Schema of I/O files of OSMON Core-Module

2.1. Input data and file types

2.1.1. Satellite imagery

The high resolution satellite imagery for the reference datasets are Sentinel-2 Level-2A data in SENTINEL-SAFE format [RD-1] and Landsat-8 Level-1 data as GeoTIFF [RD-2]. The Core-Module supports all bands in its native resolution and reads in read only mode (Table 1). Metadata of the

Sentinel-2 product is available in the enclosed xfdumanifest.xml file. Landsat-8 comes with additional information stored in metadata text file (*MTL.txt). The meta data files are considered as mandatory and in absence of these files the software will return an ERROR message, ending the shift estimation.

Sentinel-3 OLCI [RD-3] and SLSTR [RD-4] data is provided as Level-1B data in the netCDF4 disseminated in 'Stripes' or 'Frames'. Metadata of the Sentinel-3 products are available in the enclosed xfdumanifest.xml file. The meta data files are considered as mandatory and in absence of these files the software will return an ERROR message, ending the shift estimation. Level-1 data of MetOPs AVHRR-3 [RD-6] is supported by the Core-Module in native format ([RD-6]). Inserted satellite data should follow the naming conventions specified by the distributing agency, the data maintain the folder and file structure defined in the documents listed in Table 2.

Table 1: High resolution satellite data input used as GCP

Satellite	Sensor	Processing level	Format	Metadata file
Sentinel-2 A/B	MSI	Level-2A	SAFE	*_Metadata.xsd
Landsat-8	OLI	Level-1	GeoTIFF	*MTL.txt

Table 2: Medium resolution satellite data input

Satellite	Sensor	Processing level	Format	Metadata file
Sentinel-3 A/B	OLCI	Level-1B	netCDF4	xfdumanifest.xml
Sentinel-3 A/B	SLSTR	Level-1B	netCDF4	xfdumanifest.xml
MetOp A/B/C	AVHRR/3	Level-1	native	

Table 3: Naming convention of supported datasets

Satellite	Sensor	Reference document
Sentinel-2	MSI	[RD-1]
Sentinel-3	OLCI	[RD-3]
Sentinel-3	SLSTR	[RD-4]
Landsat-8	OLI	[RD-2]
MetOp	AVHRR-3	[RD-6]

2.1.2. Vector datasets

In addition to the satellite imagery vector data characterising the GCPs is required. This ESRI shapefile holds the shape of the depicted feature (lake, island, …) as polygon in WGS84 coordinates. The attribute table of the dataset contains the following entries:

Attribute	**Description**
cpid	Unique ID of the GCP
name	Name of the depicted feature. If unavailable, the entry is marked with 'NA'
country	Name of the country where the GCP is located
continent	Name of the continent where the GCP is located
featureclass	Type of depicted feature. Implemented are coastline segment, lake segment, island, lake, reservoir, and geological feature.
elevation	Geodetic height of the feature derived at its centre location.
tmin_jan	Climatological monthly minimum temperature in January derived at GCP centre
tmin_feb	Climatological monthly minimum temperature in February derived at GCP centre
tmin_mar	Climatological monthly minimum temperature in March derived at GCP centre
tmin_apr	Climatological monthly minimum temperature in April derived at GCP centre
tmin_may	Climatological monthly minimum temperature in May derived at GCP centre
tmin_jun	Climatological monthly minimum temperature in June derived at GCP centre
tmin_jul	Climatological monthly minimum temperature in July derived at GCP centre
tmin_aug	Climatological monthly minimum temperature in August derived at GCP centre
tmin_sep	Climatological monthly minimum temperature in September derived at GCP centre
tmin_oct	Climatological monthly minimum temperature in October derived at GCP centre
tmin_nov	Climatological monthly minimum temperature in November derived at GCP centre
tmin_dec	Climatological monthly minimum temperature in December derived at GCP centre

A yaml file provides the relationship between the vector data and reference satellite imagery where the cpid is used as key and identifies a GCP. The product name of the reference dataset is listed as values for the corresponding GCP (Figure 2).

```
S001133E009480:
- LC08_L1TP_186061_20160310_20170328_01_T1
S001786E136798:
- LC08_L1TP_103061_20160812_20170322_01_T1
S001874E030773:
- LC08_L1TP_172061_20160628_20180526_01_T1
S002697E100079:
- LC08_L1TP_127062_20170519_20170525_01_T1
S002766E118879:
- LC08_L1TP_115062_20150408_20170410_01_T1
S002768E121515:
- LC08_L1TP_113062_20130420_20170505_01_T1
S003244E126084:
- LC08_L1TP_110062_20161016_20170319_01_T1
S004182W063418:
- LC08_L1TP_233063_20170728_20170810_01_T1
```

Figure 2 Auxiliary yaml file to link GCPs with reference datasets

2.2. Auxiliary inputs data file types

If for the pre-processing of the input satellite data a digital elevation model is required, a projected DEM can be handed over to the Core-Module. The supported file format is GeoTIFF. In the visualization of the results a water body mask provided as Esri shapefile can be superimposed to the MROI image chip. Both data sets are optional (Table 4: Auxiliary input data).

Table 4: Auxiliary input data

Dataset	(Spatial) Data Type	Format
DEM	Raster	GeoTIFF
Water Body Data	Vector	Esri shapefile

2.3. Output

During the processing database tables are continuously read and filled. These tables are described in in [RD-6]. Here we provide details on the PostgreSQL table holing the actual matching results.

The Core-Module estimates the shift between the GCP-DW and the MROI-DW. Each shift estimate and corresponding information are stored as one row in the database table “shifts” described in 2.3.1. A quicklook of the resampled MROI-DW is saved as PNG (or JPG) to a hierarchical directory following the conventions described in section 2.3.2. Quicklooks are stored only for the one VIS-band and one TIR band, resulting in 2 quicklook-files.

2.3.1. Estimated Shifts

The shift estimates and relevant metadata (Table 5: Data-Table Attribute) to uniquely identify the MROI-DW form a structured data item in a table. This row is appended in the PostgreSQL table "shifts" depicted in Table 5 with predefined data types.

Table 5: Content and data type of output table

Data-Table Attribute	Data type	Description
eo_raster	Foreign key	EO raster the matching was applied to
ref_raster	Foreign key	Reference raster utilized to derive output
matching_band_combination	int	Characterizes the bands used in the matching routine
shift_x	real	Shift in east/west direction in meters [m]
shift_y	real	Shift in north/south direction in meters [m]
shift_along_track	real	Shift in along track direction in MROI pixel [pix]
shift_cross_track	real	Shift in cross track direction in MROI pixel [pix]
snr	real	Signal to noise ratio
cor_coeff	real	Correlation coefficient
error_flag	int	Matching error code
osmon_version	varchar	Version of the OSMON package
last_insert	varchar	Timestamp (UTC) of database insertion

2.3.2. Quicklook of Resampled MROI

By default, only a PNG quicklook of the MROI-DW in is stored. The resampled and projected MROI-DW can be saved optionally to local drive in GeoTIFF format (Float32). The file naming convention for MROI-DW and the containing directory structure is identified by the sequence of fields described here:

Directory structure:

Root/version/platform/instrument/NnnnnnEeeeeee /product/

Where:

- Version indicates the Version of the OSMON package as given in the software package (e.g. v1.0)
- The platform directory derived from the platform name and can hold several instrument sub directories.
    - LANDSAT_8 = NASA/ U.S. Geological Survey Landsat
    - METOP-0[12] = polar orbiting meteorological satellites operated by EUMETSAT
    - Sentinel-3[AB] = operated by ESA and EUMETSAT to monitor sea and land surfaces
- Instrument indicates the name of the sensor which collected the data.
    - OLCI = Ocean and Land Colour Instrument (OLCI) only with S3A/B/ …

  - SLSTR = Sea and Land Surface Temperature Radiometer (SLSTR) only with S3A/B/ …
  - AVHRR3 = Advanced Very High-Resolution Radiometer (AVHRR/3) only with

- NnnnnnEeeeeee is the GCP ID identifying the centroid of the GCP in latitude and longitude:
  - N = Northern hemisphere, S = Southern hemisphere
  - nnnnn = 5 digits, zero padded latitude with three decimal places ignoring the decimal separator
  - E = Eastern hemisphere, W = Western hemisphere
  - eeeeee= 6 digits, zero padded longitude with three decimal places ignoring the decimal separator
- Product filename as disseminated by the providing agency (Table 3).

The file base name BAND is the band/channel name according to product data format specifications. The filename extension ext is png for the quicklook and tif for GeoTIFF.

BAND.ext

- OLCI [RD-3] : Oa<b> e.g.: Oa17where:

Variable	Placeholder	Possible Values
Band	<b>	01-21

- SLSTR [RD-4] : band/grid/view e.g.: S3an where:

Variable	Placeholder	Possible Values
Band	<b>	S1-S9
Grid	<g>	‘i’ – 1km Thermal Infra-Red grid ‘a’ – 500m visible and SWIR “A stripe” grid ‘b’ – 500m visible and SWIR “B stripe” grid
View	<v>	‘n’ – nadir view ‘o’ – oblique view

- AVHRR3 [RD-6]: Ch <b> e.g.: Ch3a where:

Variable	Placeholder	Possible Values
Band	<b>	1, 2,3a 3b, 4, 5

3. REPORTING AND ANALYSIS-MODULE

The Reporting and Analysis Module (Figure 3) accesses the OSMON database in read only mode and performs the statistical analysis of the estimated shifts. The sensor, GCP and time span to be analysed can be defined by the user via a web-interface.

The analysis is then performed in an automatic or interactive mode. In the interactive mode the results are displayed on the web-browser, with the option to generate and download a pdf-file, in the automatic mode a pdf-file is generated and made available on the web.

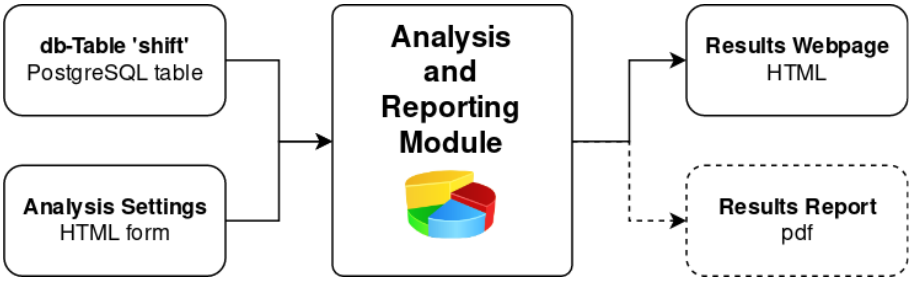

Figure 3: Schema of I/O files of OSMON Analysis and Reporting Module

3.1. Input data and file types

The user selects via the web-interface (HTML form) the data to be analysed. The for the analysis required options are:

- Instrument (single choice)
- Band (single choice)
- Satellite (multiple choices)
- Track
- GCP (single GCP, list of GCPs)
- Time span (UTC)
- Analysis options such as trends, warnings, thresholds, …
- Reporting options (automatic mode, interactive mode)

With this request the Reporting and Analysis Module queries the PostgreSQL table “shifts” described in section 2.1. The returned database entries are analysed and prepared for web display and/or provided as report described in section 3.2.

3.2. Output

The outcome of the analysis is displayed on the results HTML-webpage in the form of interactive timeseries and tables containing:

- Instrument
- Band with information on
  - processing software
  - processing hardware/environment
  - processing facility
- Satellite
- Track
- GCP
- Time span analysed (UTC)
- Timeseries of the estimated shifts as line plot in
  - Easting and Northing including linear trend line (meters)
  - Along-tack and Across-track direction including linear trend line (number of GCP-pixel)
- Status report including
  - Pixel spacing in meters
  - Number of marked outliers
  - Warnings

This result webpage can be wrapped on user request to a report in pdf-format and downloaded. In automatic mode, only the PDF file is generated and stored.

The file naming convention described here:

INS_BN_MMM_NnnnnnEeeeeee_GSD_L_YYYYMMDDTHHMMSS_report.pdf

- INS is the Instrument name:
  - OLC = Ocean and Land Colour Instrument (OLCI) only with S3A/B/ …
  - SLS = Sea and Land Surface Temperature Radiometer (SLSTR) only with S3A/B/ …
  - AV3 = Advanced Very High-Resolution Radiometer (AVHRR/3) only with MOA/B/C …
- BN is the zero padded band number
- MMM is the mission ID:
  - S3A = Sentinel-3A
  - S3B = Sentinel-3B
  - MOA = MetOp A

  - MOB = MetOp B
  - MOC = MetOp C

- NnnnnnEeeeeee is the GCP ID identifying the centroid of the GCP in latitude longitude:
  - N = Northern hemisphere, S = Southern hemisphere
  - nnnnn = zero padded latitude with three decimal places ignoring the decimal separator
  - E = Eastern hemisphere, W = Western hemisphere
  - eeeeee= zero padded longitude with three decimal places ignoring the decimal separator
- GSD is the Ground sampling distance zero padded in meters
- L is the processing level of the original EO-data
- YYYYMMDDTHHMMSS analysis timestamp in UTC without time zone.

4. DATA VOLUME ESTIMATE

The data volume occupied by GCP-DWs depends on the number of GPSs and scales linearly with the number of bands of the reference data and the size of GCP-DW. Each band is stored in a separate file. In the current GCP reference dataset the mean size of a GCP is 6 MB. Table 6 estimates the data volume occupied by the Landsat-8 GCP-DW.

Table 6: Estimated Data Budget for one band and eight bands of the Landsat-8 GCP-DW

	Number of GCPs	Estimated data volume for raster file of one band [GB]	Estimated data volume for raster files of 8 bands [GB]
GCP	1	0.006	0.048
All GCPs	500	3.000	24.000
All GCPs +20% margin	500	3.600	28.800

The data volume of MROI scenes with almost full global coverage, downloaded per day from the data archive into a temporary directory, is estimated assuming 14.4 orbits per day in Table 7. After extracting the MROI-DW covering the GCPs the Level-1B data sets are removed.

Table 7: Estimated Data volume for Sentinel-3 SLSTR, OLCI and METOP AVHRR Level-1B data per day with global coverage.

	Sentinel-3 SLSTR	Sentinel-3 OLCI	METOP AVHRR
Estimated data volume of 1 orbit, L1B file	45 GB	30 GB	0.95 GB
Estimated data volume of 1 day, with global coverage	468 GB	432 GB	13.7 GB

MROI-DW used in the Core-Module are bilinear resampled to the same pixel spacing as the GCP-DW. Consequently, we assume the size of one MROI-DW with 6 MB per band. The input data volume scales linearly with the number of processed GCPs. The number of processed GCPs per relative orbit is

estimated with 34. Assuming all GCPs are covered once per day, 500 MROI-DW are processed per band and day. Table 8 gives an overview on the data volume per sensor as a function of the number of processed GCPs.

Table 8: Estimated data amount of MROI-DW (after extraction from MROI-scenes), processed in the Core-Module

	Number of processed GCPs	Data volume for 1 band [GB]	OLCI 6 bands [GB]	SLSTR 8 bands [GB]	AVHRR 5 bands [GB]
Per GCP	1	0.006	0.036	0.048	0.030
Per rel. orbit	34	0.204	1.224	1.632	1.020
Per day (global coverage, all GCPs)	500	3.0	18	24	15

Processed MROI-DW are not stored but, quicklook files of one VIS band and one TIR band with an approximate size of 600 KB are kept. The QL of the MROI-DW are stored to local drive and the link will be saved in the eo-table of the OSMON database. Table 9 summarizes the required storage in MB estimated for different numbers of processed GCPs (first column). The second column gives the estimated data volume for one VIS MROI-DW, the third for a VIS and a TIR MROI-DW of one instrument. Column four estimates the diskspace for one VIS and one TIR (where applicable) band of three instruments (OLCI, SLSTR and AVHRR3). The last column shows the expected data volume for one VIS and TIR (where applicable) quicklooks of OLCI and SLSTR MROI-DW derived from Sentinel-3 A/B/C and AVHRR/3 of MetOP A/B/C.

Table 9:Estimation of data volume for quicklooks of MROI-DW

	Number of GCPs	QL of VIS MROI-DW [MB]	QL of VIS & TIR MROI-DW [MB]	QL of VIS & TIR of 3 instruments MROI-DW [MB]	QL of VIS & TIR of 3 platforms and 3 instruments MB]
GCP	1	0.59	1.2	3.5	10.6
Per rel. orbit	34	19.9	39.8	119.5	358.6
Per day	500	292.9	585.9	1757.8	5273.4

The memory usage of the OSMON database with its 11 tables is dominated by the four major tables gcp-table, cor_contour, eo_raster-table and shift-table where the latter three increase with the number of processed GCPs. We consider the static part (Table 10) of the database separate from the increasing tables. Therefore, we estimate the total size of each table row in MB and scaled it with the estimated number of rows. The estimates are computed based on the number of bytes per table row and do not consider data compression by the database. For the tables with increasing number of rows we assumed Sentinel-3 A/B and MetOP A/B generating 19000 entries a day (Table 11)

The number of entries was estimated under the assumption of 2 operating platforms for Sentinel-3 and METOP, and 6 bands for Sentinel-3 OLCI, 8 bands for Sentinel-3 SLSTR and 5 bands for METOP AVHRR, each of them observing 500 GCPs per day.

Table 10: Estimated data volume of static OSMON-database tables (db-table)

Table	**Estimated size of one db-table row [MB]**	**Estimated number of rows in db-table**	**Estimated Size [MB]**
method	0.000092	1	0.000092
platform	0.000092	3	0.000275
sensor	0.000092	4	0.000366
spacecraft	0.000092	4	0.000366
featureclass	0.000073	5	0.000366
band	0.000092	17	0.001556
gcp	0.011728	510	5.981232
orbit	0.000038	1530	0.057781
		Total [MB]	6.04
		Total incl. 20% margin [MB]	7.25

Table 11: Estimated data volume increase per day of the OSMON-database tables (db-table). Assuming 6 OLCI bands, 8 SLSTR bands and 5 AVHRR bands derived from Sentinel-3 A/B and MetOP A/B

Table	**Estimated size of one db-table row [MB]**	**Estimated number of rows in db-table**	**Estimated Size [MB]**
cor_contour	0.000032	19000	0.608
eo_raster	0.001163	19000	22.097
shift	0.000212	19000	4.028
		Total [MB]	26.7

	Total incl. 20% margin [MB]	32.1

In case the reports are generated automatically an additional data volume increase of roughly 200 KB per report can be expected (Table 12). The size of the report depends on the design of the report, where a tabular arrangement consumes less disk space than a graphical representation.

Table 12: Estimated data volume of by automatically generated reports

Number of reports	**Estimated size of reports [MB]**
1	0.195