Project Title:	Baltic+ BalticAIMS
WP:	WP2: Service Chain Specification
Document Title:	D2.2 Data and Platform Provisioning Plan
Version:	1.1 Final version
Author(s) and affiliation(s):	Mikko Kervinen, Sampsa Koponen, Jenni Attila SYKE Norman Fomferra, Alicja Balfanz, Tonio Fincke Gunnar Brandt, Martin Böttcher Carole Lebreton, Carsten Brockmann BC Petra Philipson, Susanne Thulin BG
Version history:	0.5 Version for PM2, 16.6.2021 1.0 Version for PDR, 16.9.2021 1.1. Final version, 5.10.2021
Distribution:	ESA, Project team, public

Contents

Abstract

The overall objective of WP2 is the specification of the BalticAIMS service chain. This data and platform provisioning plan describes of datasets, system structures, interfaces and deployment strategies required to implement the user service portfolio defined in D2.1. The comprehensive technical descriptions are available in D2.3.

Glossary

Chl a	Chlorophyll a
CMEMS	Copernicus Marine Environment Monitoring Service
EO	Earth Observation
HELCOM	Helsinki Commission
MSI	MultiSpectral Instrument
OLCI	Ocean and Land Color Imager
S2	Sentinel-2
S3	Sentinel-3

1 Introduction

1.1 Purpose and scope

This document collects data requirements and processing infrastructure requirements of the BalticAIMS data system and plans how they will be fulfilled.

Topics covered in this document are

- Satellite input data identification with sensor, product type, areas, time period
- System elements with focus to the different external interfaces
- ICT processing resources with VMs, cores, memory, disk space
- ICT storage space with cloud storage type and volume
- Planned temporal profile, the time resources are required
- Estimated ceiling price

The input data identification and the estimation of resources required is mainly derived from the use cases that come with certain data requirements, areas of interest, and temporal coverage. Other parameters are the applied processing and the way data is organized in data cubes.

The other documents generated in WP2 are:

- *D2.1: Service Portfolio Definition*: Contains descriptions of the showcases
- *D2.3 Service Delivery Chain Specification*: Provides a comprehensive technical description of the end-to-end data acquisition, processing, analysis, delivery and integration capabilities of the system.

1.2 Document overview

After this formal introduction

section 2 is an overview of the show cases, services, and functional system elements
section 3 identifies the raster datasets and vector datasets for BalticAIMS, both time series and static
section 4 describes the different interfaces of the BalticAIMS data provisioning system
section 5 defines the deployment of BalticAIMS services to a virtualized physical cloud infrastructure and estimates costs for the runtime of the project.

References are at the end of the document.

2 System overview

This section is an overview over showcases, services, and the functional system elements.

2.1 Showcases

The 5 show cases are:

- A: Provide EO based information to be used in user legacy systems for spatial planning
- B: Monitor the effects of nutrient flow from the drainage basin to the coastal waters
- C: Monitoring the impacts of coastal activities
- D: Combination of Coastal Zone mapping and CMEMS coastal water quality material
- E: Monitoring of temperature anomalies

Each show case is further elaborated in several user stories with concrete applications. Details are provided in D2.1.

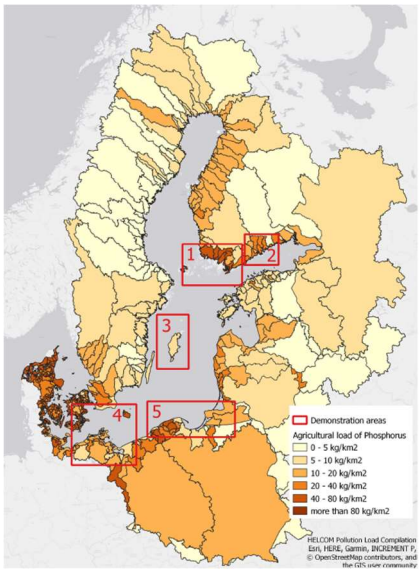


Figure 1: BalticAIMS showcase areas

There are 5 areas the show cases will be applied to in different combinations:

- 1. Archipelago Sea, Finland
- 2. Helsinki
- 3. Gotland
- 4. Mecklenburg-Vorpommern
- 5. Poland

2.2 Services

The BalticAIMS services used by the different show cases are:

- Showcase A makes available EO data for spatial planning using the user systems, mainly GIS. The GIS will read from BalticAIMS data interfaces, in particular WMS, WCS and WFS, to access data cubes and GeoDB. Simple static files are served from BalticAIMS geo file server as well.
- Showcase B provides data to analyse the effects of nutrient flow from drainage basins to coastal water. Main service used is TARKKA. The showcase may also use the OGC services provided by BalticAIMS.
- Showcase C provides data to analyse the impact of coastal activities. TARKKA demonstrates the service. GIS interfaces may be used as well.
- Showcase D maps coastal zones. It uses the BalticAIMS data service to include selected layers into the analysis.

- Showcase E makes available EO data to analyse temperature anomalies. It mainly uses TARKKA to demonstrate the service.

2.3 System elements

BalticAIMS data services are based on TARKKA, XCube, and GeoDB. The involved elements are shown in Figure 2.

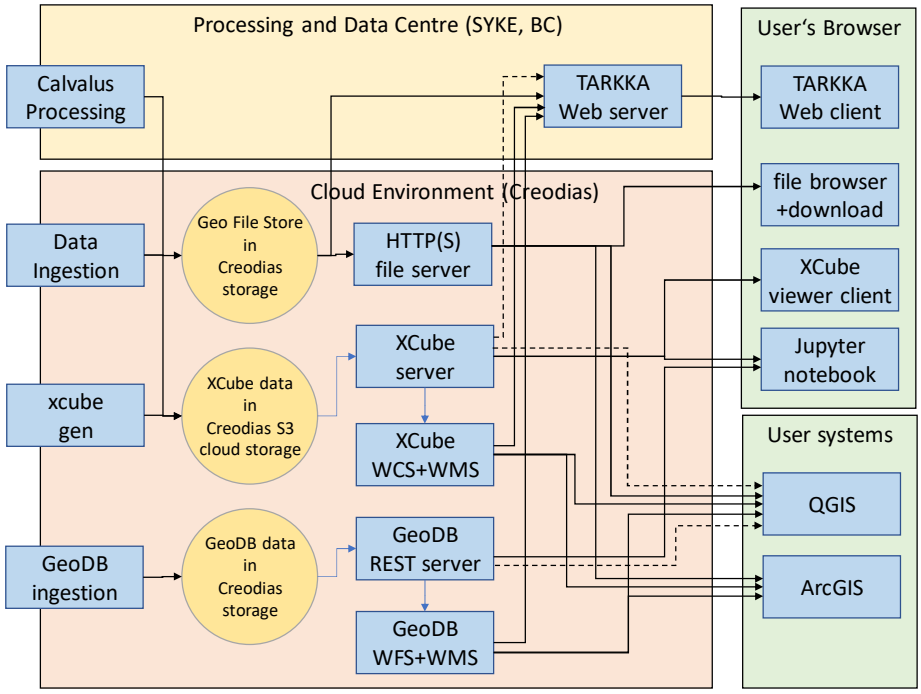


Figure 2: BalticAIMS system elements

- Raster time series data is either ingested and prepared with xcube gen, or it is processed by BalticAIMS and stored in XCube zarr format. This data is accessed via XCube server and viewer or via OGC WMS or WCS either by TARKKA or by user GIS applications.
- Feature data is inserted into GeoDB and served via OGC WFS either to TARKKA or to user GIS applications.
- Unstructured or simple file data can also be stored in a structured geo file store and served via HTTP for download or direct access by user GIS applications.

Because WCS is a rather slow and coarse-grained protocol we may switch over to plug-ins for QGIS and TARKKA to access the XCube REST API directly instead (dotted lines). This depends on whether the plug-ins can be developed in time.

3 Datasets

Main purpose of the BalticAIMS system is to serve datasets suitable for show cases in an analysis-ready form to expert users. The datasets are derived from user stories that each has its set of datasets required. The datasets per show case are collected in the Datasets table of [A2.1 Datasets Table]. While the dataset table is the reference this section provides an overview of the data and defines its representation in the system as data cube, geo-db, or files on the geo file server.

3.1 Data sources and other features of the input data

The Datasets table of [A2.1 Datasets Table] describes each dataset with 26 attributes. The attributes comprise:

- identifying information with **record name** (e.g. HROC WQ L3 inwater), **data source** (e.g. CMEMS), **dataset name** (e.g. HR-OC L3 BGC)
- **data type** (raster data or vector data), **file type** (e.g. NetCDF, GeoTIFF, shapefile), and **means of access** (e.g. FTP, WCS)
- **variables** included (e.g. CHL, SPM) and required
- projection
- spatial extent and spatial resolution provided and required
- **temporal coverage** and **stepping** provided and required
- **processing** requirements and data **conversion** requirements for cube generation
- **show cases** that use the dataset and **relevance** of the dataset for the show cases

The show cases and relevance help to define a sequence of integration.

3.2 Raster datasets

EO time series raster datasets suitable for the different show cases extracted from the Dataset table are listed in Table 1. A subset of them (up to all of them, following showcase priorities) will be converted into data cubes to demonstrate the showcases.

Table 1: Time series EO raster datasets for BalticAIMS

Variables and source	Representation	Extent, resolution	Showcases
HR-OC TUR, CHL, SPM	data cube daily, monthly averages	Baltic Sea 100m	A1, B, C, D2, E2
HR-OC RGB	data cube daily, monthly	Baltic Sea 100m	A, D
SYKE HR+MR WQ (tur, cdom, sdt, algae)	data cube temp. aggreg. TBD	Northern Baltic Sea 60m, 300m	A, B, C
SYKE HR+MR SST	data cube daily	Baltic Sea 100m, 1km	A1, C, E1, E2
SYKE data fusion gap-filled SST	data cube daily	Finland 100m	E
CMEMS SST	date cube daily	Baltic Sea 2km	A, E
SentinelHub RGB (MSI, OLI, OLCI)	WMS on-the-fly retrieval used with TARKKA	global 10m, 15m(?), 300m	A, B, C, D, E
Cyanoalert (chl, cyanobacteria, ys)	data cube temp. aggreg. TBD	Baltic Sea 300m	A

Time series raster datasets will be added as cubes into XCube. Each dataset is configured as a cube on its own. The name of the cube will be the “record name” of the dataset in the Dataset table (with updates of the table to harmonise names where useful). The names of the variables will be those of column “variables required”, again with harmonisation where useful).

Additional static raster datasets suitable for showcases are listed in Table 2.

Table 2: Static raster datasets for BalticAIMS

Variables and source	Representation	Extent	Showcases
Corine LC	raster with 44 LC classes (?)	Europe 100m	B1, C
SYKE HR LC	raster with 44 LC classes	Finland 20m	B, C
Corine LC backbone 2018 (not yet available)	raster with 12 classes	Europe	B
CLMS Global LC	raster, 23 classes annual since 2015	global 100m	B
CLMS HR-VPP (NDVI, PPI, LAI, FAPAR)	raster, seasonal trajectories of PPI (?)	Europe 10m	B
SYKE river impact areas (mean tur)	yearly 2003-2011	Coast of Finland, 60m	B
Upwelling Areas (not yet available)	raster	Gulf of Finland 300m	E
SEPA HR land cover	raster	Gotland 10m	B
Risk areas (nitrogen, phosphorus, health)		Gotland 10m	B
Symphony background (nitrogen, phosphorous)	raster	Sweden 250m	B
HELCOM MADS indices (BSII, BSPI, physical disturbance, loss, heat input, nitrogen, phosphorous)	raster	Baltic Sea 1km	C, E, B

The static datasets can either be provided in GeoTIFF on the geo file server (baseline). Or it can be provided as layer in XCube (during fine-tuning of showcases). In particular for large datasets with higher resolution the XCube solution has the advantage of variable subsets that can be retrieved. Datasets configured as XCube layers will follow the same naming rules as for the previous table. Datasets served by the HTTP geo file server will be organised by datasets as defined in section 4.6 below.

3.3 Feature Datasets

There are vector and point datasets with temporal extent (Table 3) and with rather one-time data (Table 4) suitable for the different showcases. Again, a subset of them will be implemented as datasets showcase by showcase following priorities.

Table 3: Time series feature datasets for BalticAIMS

Variables and source	Representation	Extent	Show cases
FMI Sea ice	Polygons and attributes	Baltic Sea	C
SYKE leisure boat data	TBD	demo areas	C
HELCOM agriculture loads (nitrogen, phosphorus)	Polygons and attributes		B
In-situ stream WQ			B
FMI in-situ SST	time series	Coast of Finland, point-wise observations	E
City of Helsinki in-situ SST	time series	Coast of Helsinki	
Agricultural EU support blocks *)	polygons+attributes yearly	Gotland	B

*) if data license allows

Table 4: static feature datasets for BalticAIMS

Variables and source	Representation	Extent	Show cases
Agriculture production (crop type)	Polygons and attributes (aggregated by drainage basin) or 1km gridded data	Finland 2015, 2019	B
HELCOM marine protected areas	polygons+attributes	Baltic Sea	D
Corine Land Cover	polygons+44 LC classes (?)	Europe	B, C
CLM coastal LC/LU	polygons, attributes	coastal area	D
Corine LC backbone 2018 (not yet available)	polygons+18 classes	Europe	B
HELCOM dredging sites	points, lines, polygons	Baltic Sea	C
Badestellen MV	points	MV, Germany	C, D
Protected areas MV	polygons, attributes	MV, Germany	C, D
BSH (bathymetry, water transport network, marine use, vessel traffic)	polygons, attributes	German coast	C, D
SMHI Drainage basins	polygons, attributes	Gotland	B
Swedish Board of Agriculture prod. places of animal husbandry (cattle, pigs, sheep, poultry, goats)	polygons, points	Gotland	B
HELCOM areas (MSPA, rivers, hotspots)	polygons, attributes	Baltic Sea	B, C

The feature datasets can either be provided as files on the geo file server. Or they can be provided as collection in GeoDB. On the geo file server the structure follows the same archiving rule as for the raster data:
/vectordata/<topic>/<record name>/[<region>/][<temporal coverage>]/<filename>.<extension>
The collection name in GeoDB shall be the record name from the Datasets table (with harmonisations where useful).

4 External interfaces

This section describes the different interfaces of the BalticAIMS data provisioning system.

4.1 TARKKA

TARKKA is a public service of the Finnish Environment Institute, where anyone can browse and view SYKE's open satellite data. The map interface of the service is designed to present both high-resolution (10 m - 60 m) and medium-resolution (300 m - 1 km) satellite data. Current service can be found at syke.fi/TARKKA.

Tarkka implements many easy to-use features for exploring and analysing the data layers:

- Map navigation
- Selection of layers
- Date selection for temporal layers
- Free-test search functionality for feature datasets
- Comparison of layers with swipe-functionality
- Export of the visualization to image file
- Extraction and visualization of point-wise and regional time series from raster data

In BalticAIMS, TARKKA service is customized to serve and visualize datasets relevant to selected showcases and required new functionality will be added to the service. TARKKA will be used as a frontend for showcases B, C, D and E.

Figure 3: TARKKA web application for data visualisation and analysis

4.2 XCube viewer

The XCube viewer is a web client to visualise time series of raster datasets served by XCube servers.

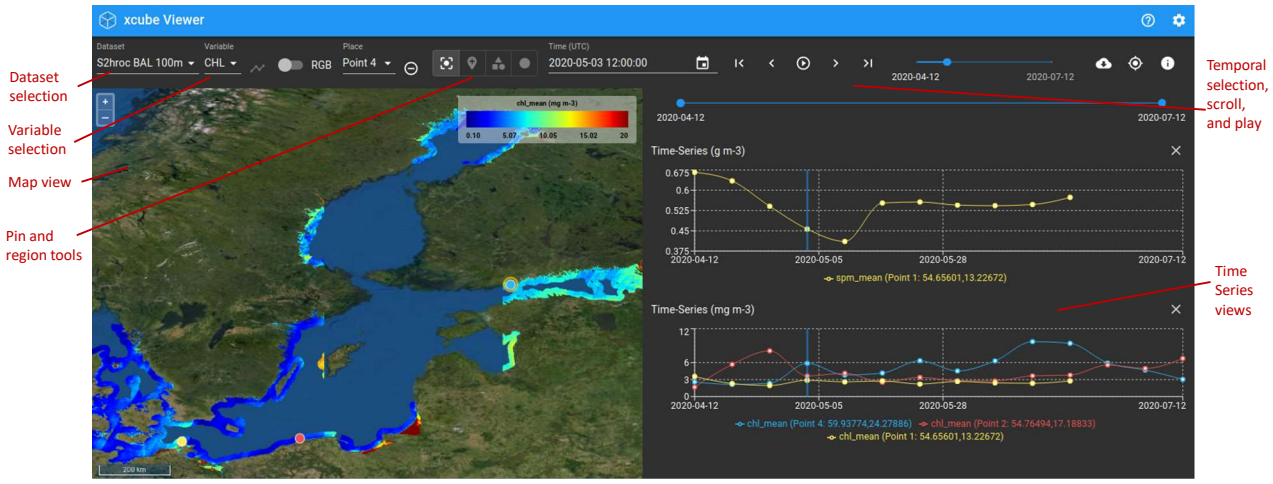


Figure 4: XCube Web viewer for inspection of time series of raster datasets

The viewer shows datasets served by the BalticAIMS XCube server and their variables. The map view displays the data at a certain time step and can play the time series as a sequence. The time series view shows graphs of variables at selected points or regions.

4.3 OGC Web Coverage Service interface

The OGC WCS for BalticAIMS is a standard interface to serve raster data. It will serve the same data also available in XCube viewer, but as data layers suitable for integration into user GIS applications.

- GetCapabilities lists the BalticAIMS raster datasets as coverages. CoverageIds are the dataset record names.
- DescribeCoverage lists the spatial and temporal extent of a coverage and the variables as DataRecord field names.
- GetCoverage requests specify the spatio-temporal and field subset requested and returns a file in GeoTIFF or NetCDF format containing the respective data.

4.4 OGC Web Map Service interface

The OGC WMS for BalticAIMS is a standard interface to serve images of raster data. It will serve the same data also available in XCube viewer, but as data layers suitable for integration into TARKKA.

- GetCapabilities lists the BalticAIMS raster datasets as layers. Layer names are the dataset record names.
- GetMap requests specify a spatio-temporal subset, resolution, projection, and returns a file in PNG or JPEG format containing the image.

WMS can also serve images of feature layers in the GeoDB.

4.5 OGC Web Feature Service interface

‘The OGC WFS for BalticAIMS is a standard interface to serve vector data with attributes. It will serve the data available in the BalticAIMS GeoDB in a form suitable for the integration into user GIS applications.

- GetCapabilities lists the BalticAIMS vector datatsets as FeatureType names together with their respective spatial and temporal extent.
- DescribeFeatureType returns the attributes available for a dataset.
- GetFeature inquires all or selected attributes (called properties in WFS) of all or selected vectors of a dataset specifying a bounding box, a filter expression, or a list of identifiers, and optionally an order attribute. The response is returned as XML document (application/gml+xml; version=3.2).

4.6 Integration into user systems

GIS applications QGIS or ArcGIS support the integration of data as layers. These layers may be raster data layers or vector data layers. In addition to local files layers can be retrieved from services supporting WMS, WCS, or WFS.

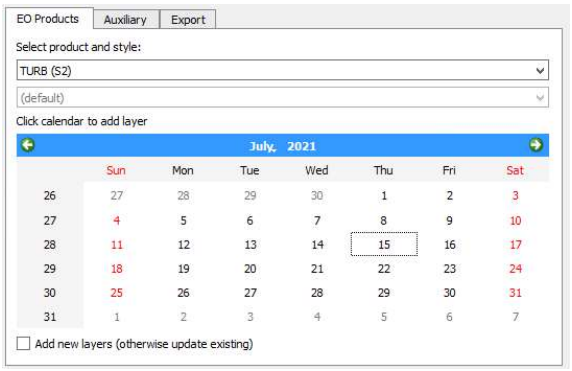
The integration of BalticAIMS into QGIS comprises

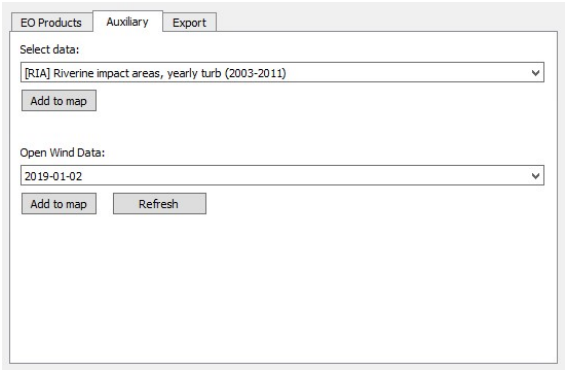
- The selection of the BalticAIMS WCS server or WFS server as data source
- The specification of authentication information
- The selection from the offered datasets and variables or feature collections
- The addition as layer

Optionally, as alternative for WCS, a plug-in for XCube can be developed to integrate XCube data layers into QGIS. Advantage of the WCS solution is that any GIS can integrate BalticAIMS data sources easily. Advantage of the plug-in solution is that it can directly access the XCube REST interface without conversion into a file, with better performance.

4.6.1 QGIS plugin

For non-expert users, point-and-click tool for importing the data layers from OGC interfaces to QGIS is provided. However, use of the plugin requires that the user can install new components in their GIS system/workstations. This plugin can then be used to list the available products and select the dates to import from temporal datasets.





4.7 Jupyter notebooks

Jupyter notebooks provide direct access to the GeoDB Python API and the XCube Python API. Examples of notebooks show how these interfaces can be used. Jupyter notebooks play a minor role in BalticAIMS because the focus is on data services and the integration into user systems and existing work practices.

4.8 Geo file server interface

Files of static data that is neither converted into a cube nor into a collection of GeoDB can be served by a HTTP(S) server of BalticAIMS. Datasets provided in the file server will be structured in directories following the archiving rule
/rasterdata/<topic>/<record name>/[<region>/][<temporal coverage>]/<filename>.tif
with

- topic: one of land, water, (extended if helpful)
- record name: entry from Dataset table
- region: optional, if there are separate datasets for different regions, e.g Gotland
- temporal coverage: optional, if there are separate datasets for different periods, e.g. 2018
- filename: either the original file name as provided by the data source, or a suitable name to identify the file

5 Deployment

Main parts of the BalticAIMS system will be deployed in a cloud service to allow access by users from anywhere. The cloud service selected is CreoDIAS. CreoDIAS hosts complete time series of Sentinel input datasets and provides openstack for infrastructure setup and management and S3 for cloud storage. Other parts relevant for BalticAIMS are hosted by SYKE or as part of the Finnish Collaborative Ground Segment, in particular TARKKA and CalFIN.

The deployment defines a (virtualized) physical system architecture. On the basis of this physical architecture the deployment defines which service is deployed on which machine, and which data is stored in which storage.

5.1 Physical system layout

The "physical" system on CreoDIAS comprises

- a CreoDIAS account with credentials for openstack and credentials for the BalticAIMS buckets
- virtual machines
- a private network with a router towards other networks (internet) and an external IP, firewall rules
- disk volumes (SSD or HDD) associated with a VM each
- cloud storage buckets

Two layouts are foreseen at different stages, a reduced system for development (from T0+6 onwards), and an extended system for operations and demonstration (T0+12 onwards). The reduced system in Figure 5 deploys all continuous services on a single VM and allocates additional VMs for cube generation on demand. The extended system in Figure 6 scales to more cubes and more concurrent users and deploys different services on different machines, mainly one for GeoDB and one or two for XCube servers.

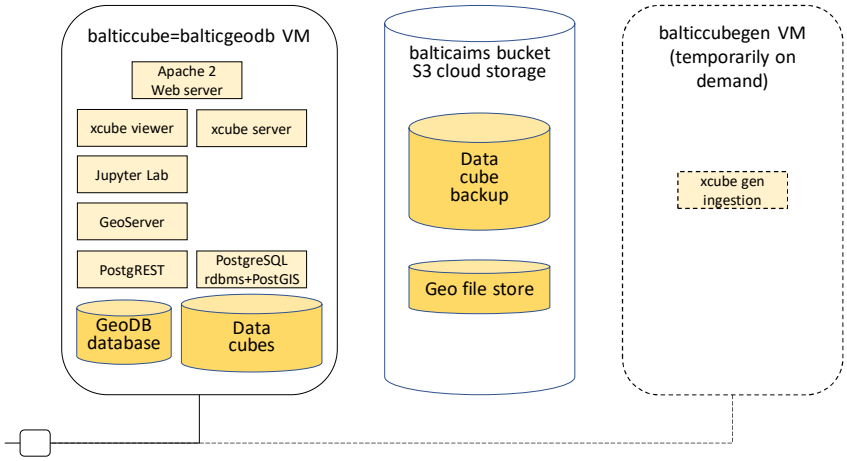


Figure 5: Development system layout and service allocation

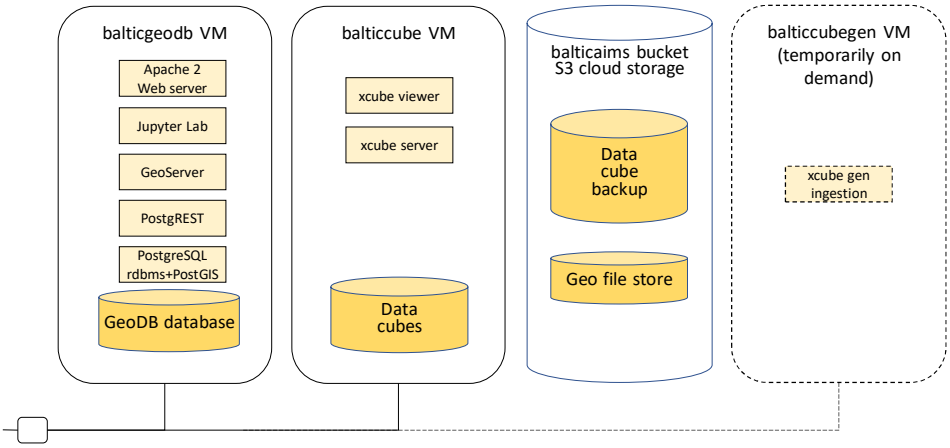


Figure 6: Operational system layout and service allocation

5.2 Data cube services

To deploy XCube services

- Create a cloud storage bucket balticaims for the data, either a single one for all cubes or separate ones for different regions. Set up the directory structure according to rule

s3://balticaims/xcube/<region>/<dataset name>/

- Create a volume for the life data cubes and mirror the structure from the bucket to the volume.
- Install the XCube server software and the XCube viewer software on the balticcube VM. Install the xcube configuration for BalticAIMS. Define ports for both services.
- Ensure that network access from internet to XCube server and to XCube viewer ports are allowed by firewall rules.
- Install a test cube into cube storage. Start viewer and server and test it.

5.3 GeoDB

To deploy GeoDB services

- define database locations on disk volume mounted to balticgeodb VM (identical to balticcube VM for development phase):

/balticdata/geodb/

- Install PostgreSQL server and PostGIS and PostgREST on balticgeodb.
- Install GeoServer software on balticgeodb.
- Install geodb software package
- create database and start postgresql server.
- Start Jupyter notebook and ingest test dataset into GeoDB. Test access from Jupyter notebook.

5.4 Jupyter lab

Jupyter lab is installed together with an Anaconda environment with installed xcube and geodb client packages. One example notebook for geodb and one for xcube are provided.

5.5 Geo file server

The BalticAIMS geo file server is implemented by an Apache Web server that serves static files organised in the directory tree as defined in section 4.6. The directory tree is hosted on S3 cloud storage of CreoDIAS in the balticaims bucket. The bucket is mounted as a S3FS fuse file system. The Apache configuration uses the path to the root of the directory tree as the root of the static content with a DocumentRoot configuration:

DocumentRoot /balticaims/geofiles

5.6 CalFIN services

The purpose of CalFIN in BalticAIMS is to provide the computing power for satellite data processing. SYKE has access to a massive parallel computing system called Calvalus, which is located at the Finnish National Satellite Data Center (NSDC). The NSDC is a Copernicus Collaborative Ground Station and as such has direct access to the Copernicus Core Ground Stations. This guarantees fast connections to Sentinel data. The Calvalus system currently has 31 nodes, 248 processors cores, and 674 terabytes of disk storage. This system is used for operational EO service provision by SYKE and is capable of processing all satellite data required by the project.

As an example, SYKE's Baltic Water quality processor combines use of Case2Regional-processor, cloud masking with Idepix-processor and in-house algorithms to produce turbidity data layer for each input tile. Based on this existing processing chain we can estimate that the cpu-time consumed in generation of one variable for single S2

tile is 45 to 50 minutes. This corresponds to a processing effort of 80 core-hours / data month / variable for the Archipelago region.

The use of this system is offered free of charge to the project and thus does not generate ICT costs for ESA.

The Calvalus system will be extended and configured to directly generate data cube zarr files and to store them in CreoDIAS cloud storage. This saves the cube generation step for data processed for BalticAIMS.

5.7 System resource estimation

The BalticAIMS data will be converted into a data cube for each region that is continuously extended over time with newly processed data. This section estimates the processing and storage effort required to create, host and provide access to this data cube.

According to the project schedule integration shall start after 6 months (T0+6 to T0+12) and operations shall start 6 months later until the end of the project (T0+12 to T0+24).

- T0+6 to T0+12 we need a reduced platform for development and the generation of the first two use case cubes/regions. Operations at T0+12 starts with data in the cubes instead of empty cubes.
- T0+12 to T0+20 is the operations phase where the cubes are systematically extended, and cubes for two more use cases/regions are created. For this period we need the complete infrastructure.
- T0+20 to T0+24 the cubes are still hosted to be available for assessment.

The estimation has a few uncertainties that will be removed while setting up cubes in the development phase:

- We currently assume that the data cubes have 15 variables. If a use case asks for more then more storage space is required.
- We currently assume that the standard inquiries are for time points (image pyramids) and for time series (time chunking). If there are other types of frequent inquiries the data cubes may need additional optimisations which require additional computing and storage.
- We currently assume that there are few concurrent users (around 4 at a time for showcase demonstrations, there may be more over time). If there are more concurrent users we may have to launch additional servers to provide the required performance in data access and viewing.
- If we decide for some reason to store all original observations instead of one observation per day per pixel (best pixel selection or averaging) this requires more space. The cubes we have generated so far do store the original observations and pre-calculate the daily means. We propose to restrict it to the daily means for the BalticAIMS cubes for the moment.
- If we add additional use cases with additional regions this will also increase the computing and storage.

To mitigate the uncertainties and to allow some flexibility e.g. in the number of variables or the resolution of land areas we use the area of the largest of the regions (Archipelago) as basis for the estimation.
We estimate the infrastructure costs for hosting a data cube service on CREODIAS as follows. Table 5 calculates the data cube storage volume for one region.

Table 5: Estimated data cube storage volumes

Item	Parameters	Size	Comment
Archipelago area, 60m pixel resolution	200 km * 150 km / (60m * 60m) =	9 Megapixels	
15 variables + RGB + lat, lon, time	15 * 4 Bytes + 4 Bytes (RGB) + 3*8 Bytes (lat, lon, time) =	88 Bytes / pixel	
Data covering project runtime, best pixel selection per day	3 years of data =	1095 days	
Raw data volume for 3 years and one area	9 Mega * 88 Bytes * 1095	870 GB	
Cube data volume with Pyramids and time chunking	Pyramids duplicate, time chunking adds raw volume again, overhead and compression may balance each other	2.6 TB	

Table 6 identifies infrastructure items, in particular VMs, block storage (SSD), and cloud storage. Minor costs like local disks, IP address, or network costs are included in the VM prices.

Table 6: Item costs for infrastructure elements

Infrastructure item	Item price	Duration of use	Comment
xcube server (4 cores, 32 GB RAM)	€ 160 / month	18 months	
2.6 TB SSD for cube data storage	€ 260 / month	18 months	
2.6 TB cloud storage for cube backup and viewer web page	€ 50 / month	18 months	
cube generation and extension VM (8 cores, 64 GB), required 150 hours initially, later ~1 hour/day	€ 20 / month (for temporal use)	14 months	

This results in estimated costs per region as listed in Table 7.

Table 7: Estimated Costs staged with additional regions

Region	Monthly costs	Duration	Estimated costs
Estimation for first region, one of each item, 50% margin	€ (160+260+50+20) * 1.5 = € 735	18 months	€ 13230 (€ 8820 / year)
Estimation for the second region, only last 3 items, assuming use of same xcube server	€ (260+50+20) * 1.5 = € 495	18 months	€ 8910 (€ 5940 / year)
Estimation for 3rd region, shorter time period due to later start assumed	€ (260+50+20) * 1.5 = € 495	12 months	€ 5940 (€ 5940 / year)
Estimation for 4th region, shorter time period due to later start assumed	€ (260+50+20) * 1.5 = € 495	12 months	€ 5940 (€ 5940 / year)
Second cube server to allow operations independent from integration from second cube onwards, during the end support for more concurrent users	€160 / month * 1.5 = € 240	12 months	€ 2880
Sum			€ 36900

The costs scale linearly with an offset for additional regions and for additional variables. Duplicating the region will duplicate storage and generation in the first place, but at some point we may need a second server as well. The costs can be lower than the estimated costs if XCube development allows to serve

cubes from S3 cloud storage directly. Then, SSD costs can be lowered considerably. SSDs then serve as cache only.

We also need the infrastructure for the development and testing period of the project. If we assume 18 months (SoW), 4 use case regions, some of them smaller, some of them as large as Archipelago, some interesting additional variables we detect during the project (that fully compensate for the smaller regions), and two of the regions start serious operations in the second year, we could estimate the budget for infrastructure costs for data cube hosting of:

€ 36,900.-

While the main part of the processing is done at the existing Calvalus system of SYKE, we will test the processing also at CREODIAS in order to assess the costs and potential of a full cloud system.

To test operational processing on CREODIAS we propose to do 3 months of processing of for one use case in the cloud. If we estimate 800 core hours per data month this will be about 2400 core hours, or 600 hours on a 4 core compute node. One compute node will be sufficient for our common needs. But we also want to test scalability that requires many nodes for a limited time. We need a master node to control processing. And we assume 2 month to set up and test in addition to 3 months common production and a bulk production test in the same order (equivalent to 3 month, but with more VMs in a shorter time):

Table 8: Estimated costs for a BalticAIMS cloud processing experiment

Item	Cost per month	Amount	Cost
compute VM (4 cores, 32 GB, local disk)	€ 160 / month	2 months preparation 3 months processing 3 months equivalents bulk processing (=30 VMs for 3 days)	€ 1280
master VM (2 cores, 4 GB, IP address)	€ 40 / month	2 months preparation 3 months processing 1 month bulk processing	€ 240
Cloud storage (1 TB)	€ 20 / month	8 months	€ 160
Margin		Factor 1.5 (allows to repeat half of the test after some experience)	€ 760
Estimate			€ 2440

5.8 Cost estimation summary

The estimated ICT costs for the cloud processing demonstrator are
€2,440.-

Adding the estimated ICT costs for the data cube of
€ 36,900.-

we get as estimated costs for the BalticAIMS ICT on CREODIAS
€ 39,340.-

We propose
€ 40,000.-
as ceiling price to be used for the NoR request.

6 References

[D2.1]
D2.1 Service Portfolio Definition

[D2.2]
D2.2 Data and Platform Provisioning Plan

[A2.1 Datasets Table]
BalticAIMS Dataset Attributes tables, Excel tables, version 2.0 (in progress, retrieved 16.09.2021)