Horizon 2020

H2020 LC-SPACE-04-EO-2019-2020

Copernicus Evolution – Research for harmonised and Transitional-water Observation (CERTO)

Project Number: 870349

Deliverable No: D2.1		Work Package: 2	
Date:	30-DEC-2020	Contract delivery due date	30-NOV-2020
Title:	Technical requirements for the CERTO prototype		
Lead Partner for Deliverable	PML Applications		
Author(s):	Ben Calton, PMLA; Oliver Clements, PMLA; Amy Westlake, PMLA		
Dissemination level (PU=public, RE=restricted, CO=confidential)			PU
Report Status (DR = Draft, FI = FINAL)			FI

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme grant agreement N° 870349

Table of Contents

Reference Documents

RD1	Open Geospatial Consortium – Well Known Text representation of coordinate reference systems - http://www.opengis.net/doc/is/wkt-crs/2.0.6
RD2	Web Map Service interface standard - https://www.ogc.org/standards/wms
RD3	Web Coverage Service interface standard - https://www.ogc.org/standards/wcs

Glossary

API	Application Programming Interface
CMEMS	Copernicus Marine Environment Monitoring Service
EO	Earth observation
MSI	Multi Spectral Imager
netCDF	Network Common Data Format
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
WCS	Web Coverage Service
WMS	Web Map Service
WP*x*	Work Package *x*

1 Executive Summary

- The overall objectives of WP2 are to identify user needs, content and quality requirements, and concerns in terms of specific information required on any aspect of the state of transitional waters.
- Interactions between the project partners and case study partners resulted in a set of user and content requirements, captured in the deliverable D2.1 'Content and user requirements for the CERTO prototype' (section 4).
- This document builds on the requirements laid out in D2.1 to define the technical capabilities required of the prototype.

2 Introduction

The purpose of this document is to describe the technical requirements of the CERTO prototype system such that implementation of the described system in WP7 fully addresses the needs of the user community and case study partners. This document will define the case study regions, the variables which be produced, processing levels of available products, the timeliness of data, data formats, and delivery methods.

3 Regions of Interest

3.1 Case Study Areas

Prior to deployment in one or more Copernicus service after the completion of CERTO, the prototype, initially, will produce data for geographic regions across Europe as defined in this section in support of WP8 in year 3 of CERTO. It is envisaged that all parameters produced by the prototype will be available in all areas, calibrated and validated by means of in situ data; however, the spatial and temporal resolution may vary between parameters depending on the capabilities of the sensor and/or the algorithm used.

In each of the case study areas the spatial extent is defined by a Well Known Text polygon [RD1], and the data products provided will cover this extent as a minimum and in some cases coverage may extend beyond this.

3.1.1 Curonian Lagoon

Spatial extent:
POLYGON((20.52 54.89,21.29 54.89,21.29 55.73,20.52 55.73,20.52 54.89))

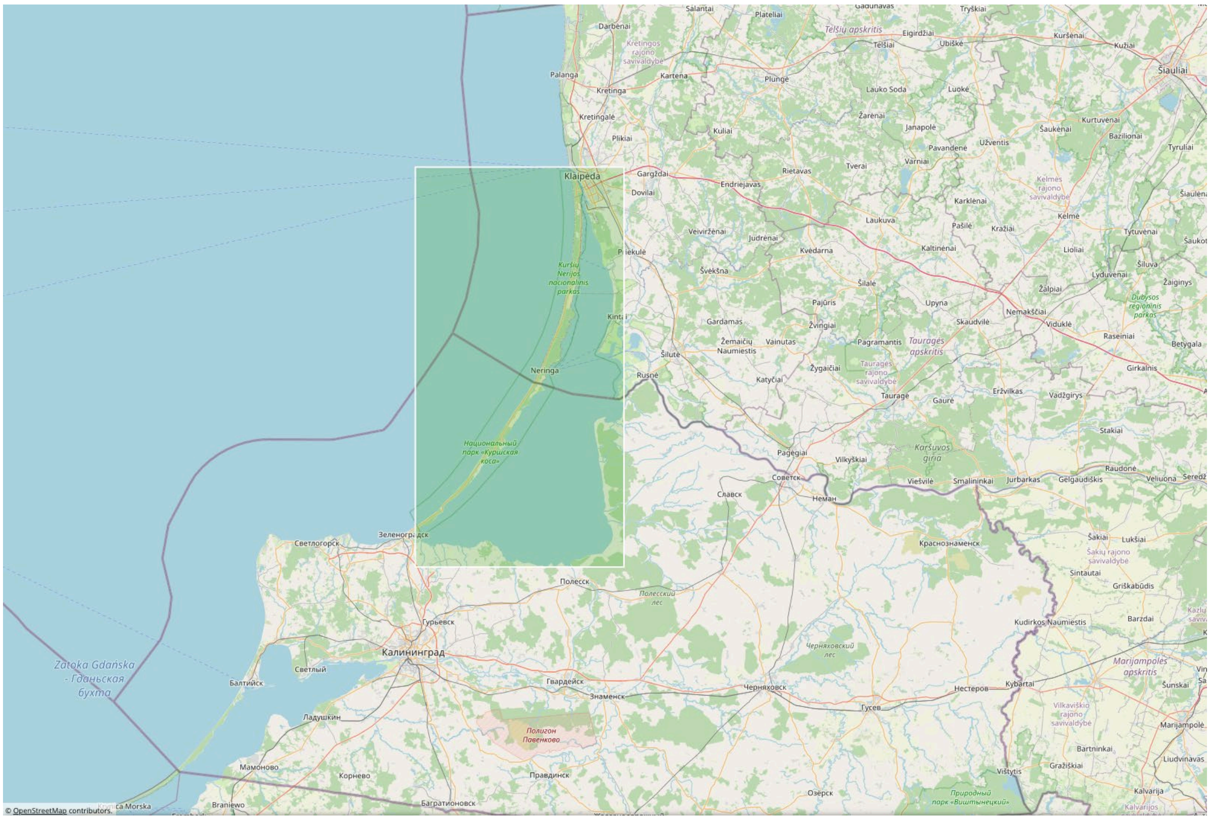


Figure 1 - Map displaying case study area for the Curonian Lagoon highlighted by a green box

3.1.2 Danube Delta

Spatial extent
POLYGON((28.274 44.4019, 28.274 45.4977, 29.8875 45.4977, 29.8875 44.4019, 28.274 44.4019))

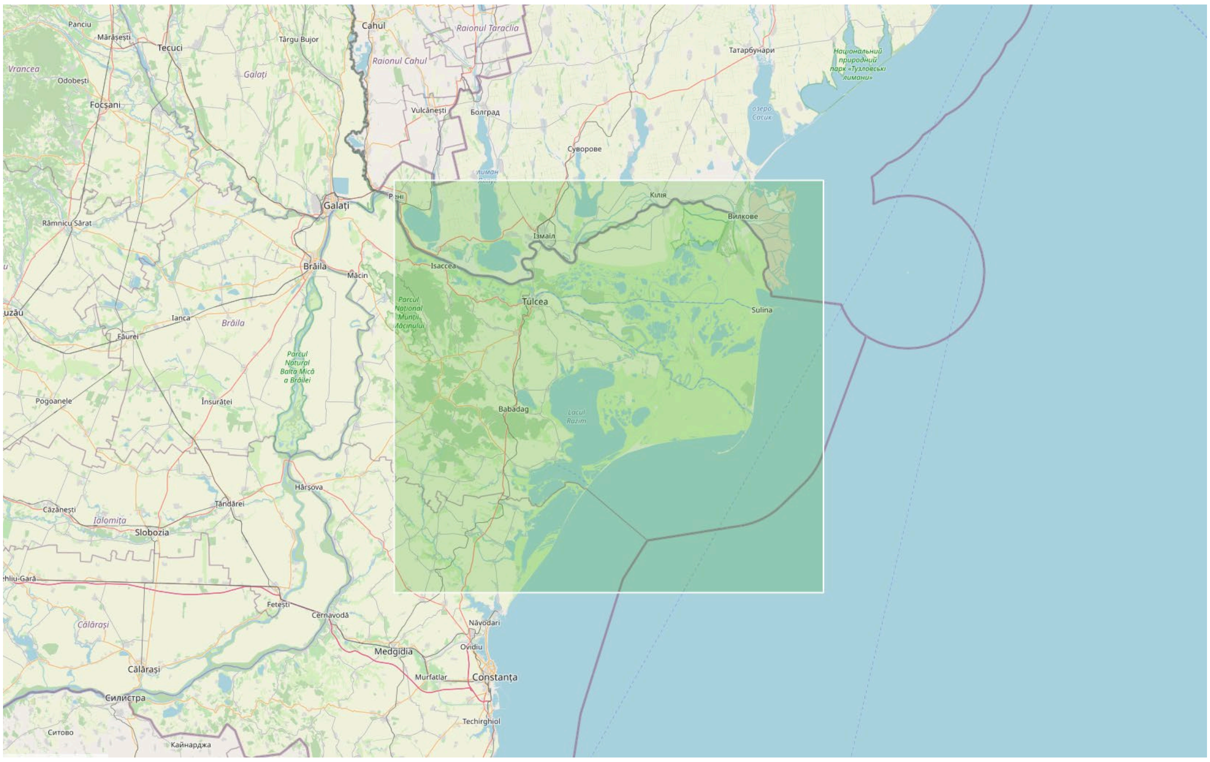


Figure 2 - Map displaying case study area for the Danube Delta highlighted by a green box

3.1.3 Elbe Estuary and German Bight

Spatial extent:
POLYGON((6.014 53.119, 5.881 54.994, 7.434 55.041, 7.377 55.940, 9.157 55.940, 9.157 55.041, 10.719 55.032, 10.652 53.148, 6.014 53.119))

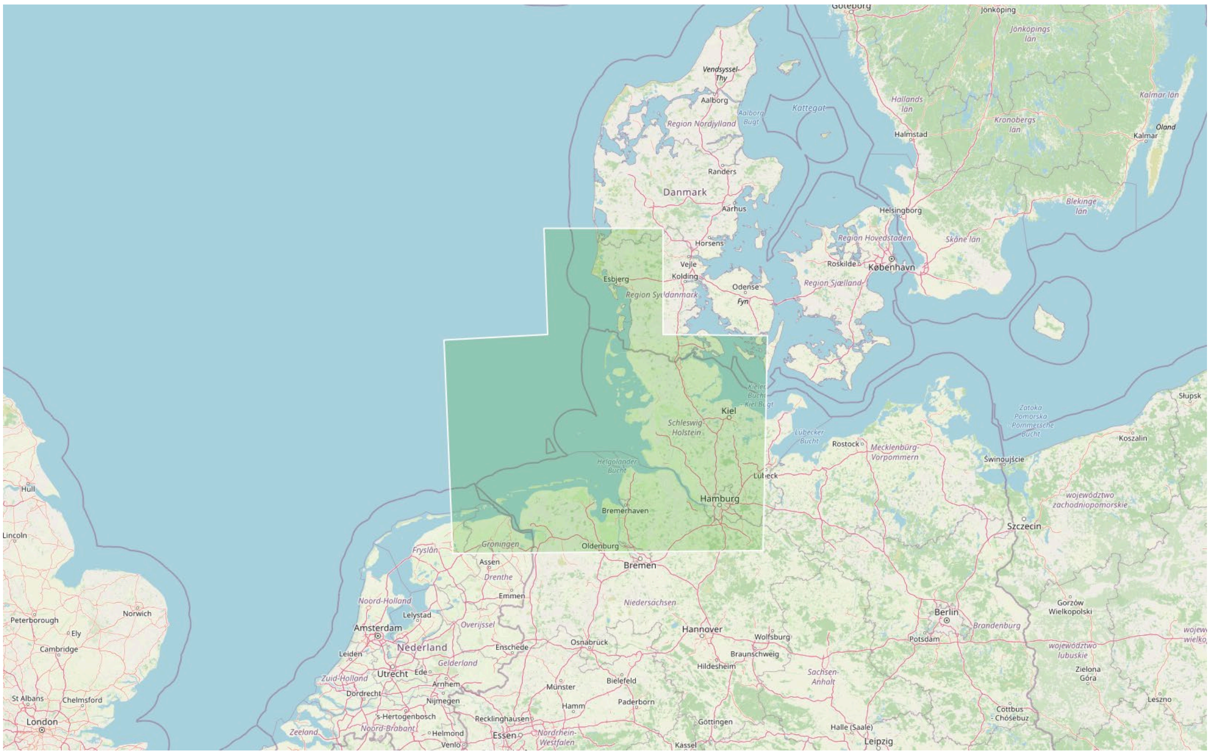


Figure 3 - Map displaying case study area for Elbe Estuary and German Bight highlighted by a green box

3.1.4 Tagus Estuary

Spatial extent:
POLYGON((-9.55 39.09, -8.53 39.09, -8.53 38.34, -9.55 38.34, -9.55 39.09))

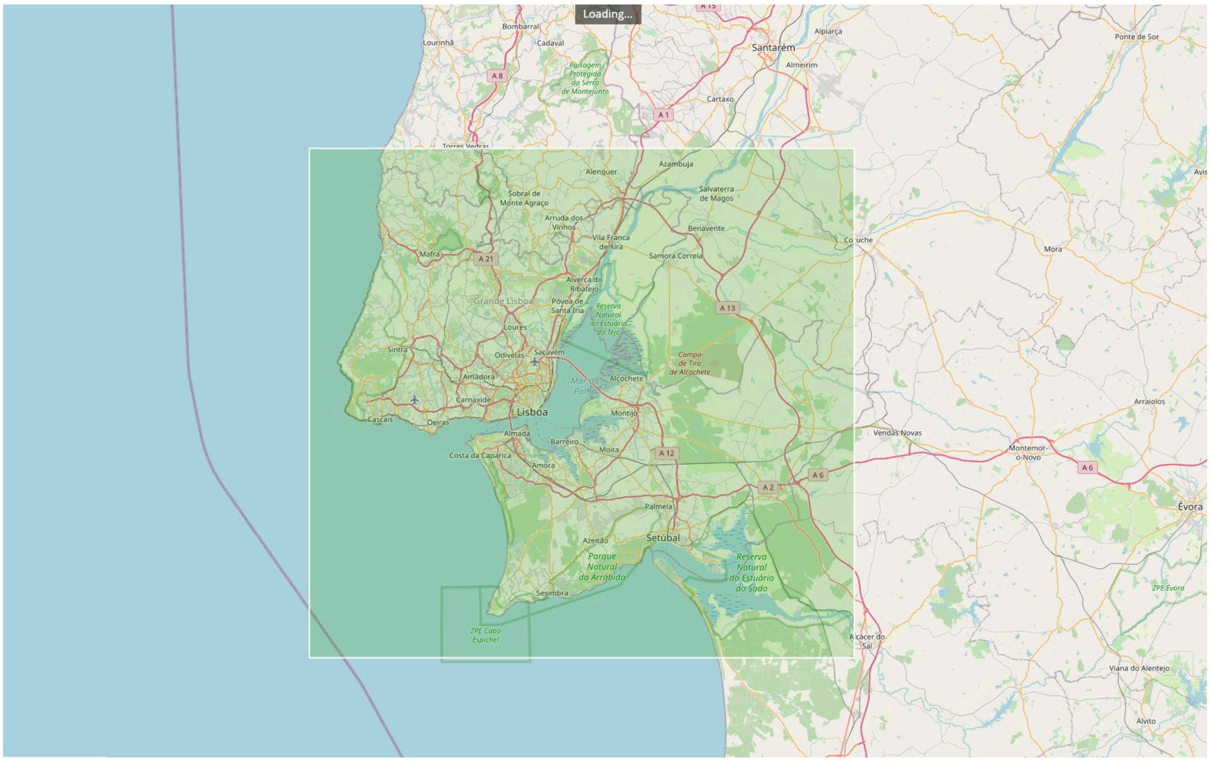


Figure 4 - Map displaying case study area for the Tagus Estuary highlighted by a green box

3.1.5 Tamar Estuary

Spatial extent:
POLYGON((-4.41 50.53,-3.87 50.53,-3.87 50.14,-4.41 50.14,-4.41 50.53))

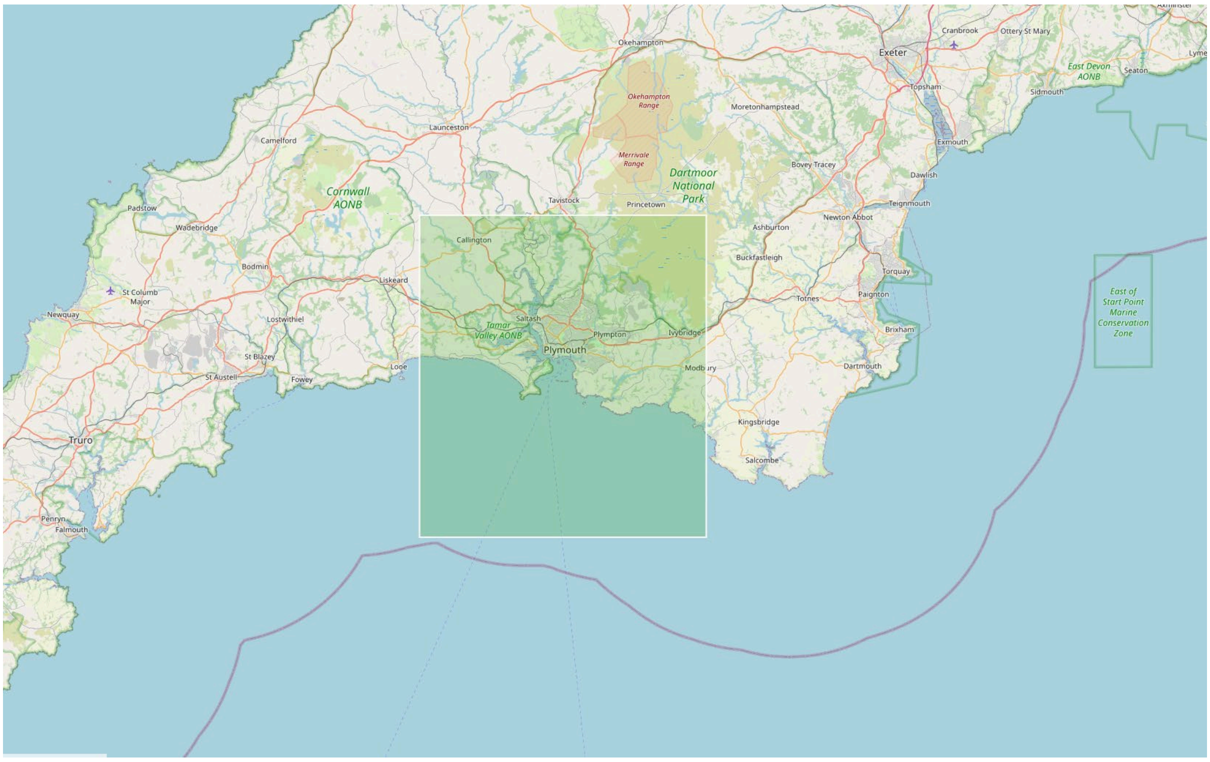


Figure 5 - Map displaying case study area for Plymouth Sound highlighted by a green box

3.1.6 Venice Lagoon and North Adriatic Sea

Spatial extent:
POLYGON((13.45 45.80, 13.45 44.30, 12.00 44.30, 12.00 45.80, 13.45 45.80))

Figure 6 - Map displaying case study area for the Venice Lagoon and North Adriatic Sea highlighted by a green box

3.2 Copernicus Marine Regional Areas

The Copernicus Marine Environment Monitoring Service (CMEMS) is one of six services of the EU's Copernicus Programme. CMEMS provides access to a vast catalogue of marine data where EO and model products are tailored to one of seven geographic areas; the spatial extent of each area is demonstrated in Figure 7. EO products are provided by entities known as Thematic Assembly Centres (TAC), whilst model products are provided by Monitoring and Forecasting Centres (MFC). Specifically, the Ocean Colour TAC (OCTAC), which is led by CNR with involvement from PML Applications, provides ocean colour products for all regions in near-real time mode plus a reprocessed product covering the entire timeseries from 1997 to present.

Ocean colour products are currently provided at 1km resolution for each CMEMS areas. Commencing in Q2 2021, CMEMS will deliver Chlorophyll-a concentration and remote sensing reflectance at 300m resolution using data from Sentinel 3A and 3B OLCI.

The CERTO prototype will be capable of producing the full suite of CERTO variables as defined in this document (section 4 on page 11) for the coastal zone of each of the CMEMS regional areas with the exception of the Global area; in the CMEMS context the coastal zone is defined as 25km from the shoreline.

Figure 7 - CMEMS geographical regions (Credits: Copernicus Marine Environment Monitoring Service)

3.3 User-specified areas

As a result of interactions with user groups as part of dissemination and exploitation activities to be carried out in WP9, it is possible that there will be requests for products for specific areas. All requests will be recorded and collated, and where reasonable, we will also deliver CERTO data products specifically tailored to the user-specified area.

4 Variables

During the case study user engagement surveys a list of variables which users have stated a need or desire for was collated; a summary of these surveys can be found in “D2.1 Content and use requirements for the CERTO prototype”. The variables are grouped into three categories; those which we are able to provide at the requested resolution, those which we can provide but at a different resolution to that requested, and those which we are unable to provide.

Table 1 provides a list of target variables that the prototype will be capable of producing, provided that a suitably tuned algorithm is available.

Table 1 - Variables provided at requested resolution

Variable	**CF Standard Name**	**Units**
Chlorophyll-a concentration (CHL)	mass_concentration_of_chlorophyll_a_in_sea_water	kg m-3
Total Suspended Matter (TSM)	mass_concentration_of_suspended_matter_in_sea_water	kg m-3
Turbidity	sea_water_turbidity	1
Remote Sensing Reflectance (Rrs)	*<no standard name available>*	sr-1

Table 2 provides a list of variables which can be delivered by the CERTO prototype, however, not as the high resolution products requested by the users. For example, sea surface temperature is available from SLSTR only at 1km resolution.

Table 2 - Variables available at a lower resolution than requested

Variable	**CF Standard Name**	**Units**
Sea Surface Temperature (SST)	sea_surface_temperature	K
pH	*<no standard name available>*	
Sea Surface Height (SSH)	sea_surface_height_above_geoid	M

Finally, in Table 3 are variables which were requested by users which cannot be provided as a result of Earth observation from satellite.

Table 3 - Variables which cannot be provided

Variable	CF Standard Name	Units
Dissolved Oxygen	mole_concentration_of_dissolved_molecular_oxygen_in_sea_water	mol m-3
Nutrients	mole_concentration_of_dissolved_molecular_nitrogen_in_sea_water mole_concentration_of_dissolved_inorganic_phosphorus_in_sea_water	mol m-3

5 Processing Levels

EO data from satellite borne sensors are available at multiple processing levels depending on the needs of the user. Processing from raw L0 data to Level 1 (L1) involves various radiometric correction processes and geometric correction, and in the case of Sentinel 2 MSI, conversion to pre-defined intersecting tiles in JPEG2000 format, and during Level 2 (L2) processing atmospheric correction is performed. Processing to Level 3 produces a data product that is ortho-rectified and mapped to a standard regular grid for a single time slice; it is then possible to draw together multiple L3 scenes acquired over a week, a month, or longer to produce a L3 composite product (in CMEMS terminology L3 composite products are referred to as Level 4 products).

Within CERTO, data at L1 will be the primary input, where the prototype will apply atmospheric correction techniques improved by the efforts of WP5 to produce Level 2 products. These L2 products will be subject to the optimal algorithm based on water class membership from WP4 to deliver mapped L3 products.

Whilst data at all processing levels will be available to project partners for further analysis, it is anticipated that case study partners (and any potential external partners) are provided only with daily products. Should the case study partners require a composite product, for example to provide a weekly average view for a cloud affected area, a suitable temporal composite product will also be produced.

6 Timeliness

The prototype will be an ‘on demand’ system, meaning that requests for data products covering a specific region over defined time range will need to be requested by some means; data will not routinely be produced, for example, as they would be as part of a Copernicus service in quasi near-real time. WP7 will provide an interface where users can select one of the pre-defined case study areas, or define the spatial extent of the data requirement, the variables required for that region, and the temporal range that they are required for, plus any other configuration options that may be applicable. Upon completion of this request it will be assessed, initially manually, by the prototype development team in WP7 for completeness;

the requesting partner will be given an estimate of how long it will take to produce the data, and the processing will be started.

Once the data have been produced, they will be published using the methods described in Section 9 and the originator of the request will be informed of their availability.

7 Data Format

Data products will be delivered in netCDF (or Network Common Data Form). It is a community standard for sharing array-oriented scientific data in a self-describing, highly scalable, and portable binary format with interfaces in a wide range of common programming languages. It is broadly agreed that netCDF is the *de facto* format for EO data, and during the case study partner interviews it was specifically requested in 5 of the 8 interviews. In the interviews that did not specifically request this format, netCDF is the underlying storage format which would deliver the requested method.

8 Code Considerations for the Prototype

The construction of the prototype is a primary activity of WP7 and a more complete view of the technical architecture of the prototype is defined in D7.2 Solution Architecture Design Document. However, there are several overarching requirements which ensure community engagement and user uptake.

The prototype must be constructed using a commonly used and well documented programming language which offers a full suite of functionality, can be used to interface with other existing systems where required, and offers long term support. Python is a well-known and widely used language in the scientific community and it has a range of well-developed libraries for interacting with netCDF files, the primary data format of the project. Several contributing elements of the prototype are written in Python (for example, the Polymer atmospheric correction) which ensures seamless integration; however, Python can easily be used to interface with software in other languages by means of an extension or wrapper. The reverse is also true in that Python scripts can be executed by software written in other languages where integration is required, for example, when integrating the CERTO prototype with the open source user-focussed SNAP processing package, which will provide the means of accessing the CERTO prototype by end-users who wish to undertake their own processing.

Having clarity of which version of software is being utilised is key to successful user engagement. Within WP7 it was agreed that a three-number versioning system would be adopted, with each number separated by dots, e.g. 1.3.23. The three numbers together would refer to a single release, where the:

- First number = Major version number – drives a release or a breaking change
- Second number = Minor version number – iterative developments, planned updates within a development phase
- Third number = Incremental changes – updated each time a commit to the master code base occurs

The full version number will be associated with a git commit identifier for clarity. The CERTO prototype will ultimately be a collection of contributed software potentially each with their own version control system; all contributed software will keep its own version number which will be aligned to a CERTO release version ID.

9 Data Access and Delivery Methods

The data products that will be produced by the CERTO prototype will be made available via several methods and protocols to suit the various needs of the user community with differing levels of technical skill requirement.

9.1 OPeNDAP

OPeNDAP is commonly used software to provide access to spatio-temporal data. It provides access at a specific URL where users can interact with data via command line, Internet browser, or a custom UI. It is also possible to use other netCDF compliant tools such as Matlab, R, IDL, and Python to access data. OPeNDAP provides methods to access and download spatial and/or temporal subsets of data without having to download the entire dataset, so is useful for particularly large datasets or low bandwidth scenarios.

9.2 Open Geospatial Consortium (OGC) Standards

The OGC have defined a number of applicable standards. A Web Map Service (WMS) produces maps of spatially referenced data dynamically from geographic information. This International Standard defines a “map” to be a portrayal of geographic information as a digital image file suitable for display on a computer screen as described in full in WMS specification [RD2].

The OGC Web Coverage Service (WCS) specification document [RD3] defines the service which supports electronic retrieval of geospatial data as “coverages.” Coverages are digital geospatial information representing space/time-varying phenomena, specifically spatio-temporal regular or irregular grids.

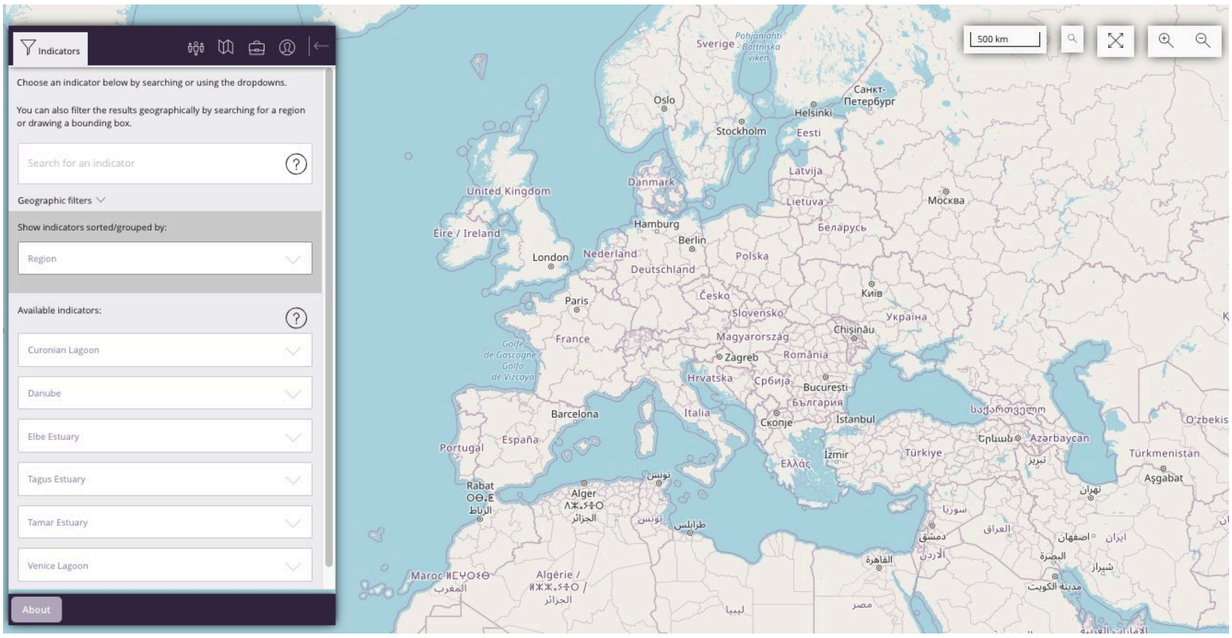
Both WMS and WCS are clearly defined standards that have a reliable and robust implementation in the THREDDS Data Server software. An instance of THREDDS offers anonymous user access to the CERTO data products via WMS and WCS; this interface is available at https://engage.certo-project.org/thredds. The CERTO data products will be available at a queryable WMS and WCS endpoints, meaning that users can request information about available variables, spatial and temporal range, available scale colour palettes and styles before making either a map request or a coverage request.

9.3 Visualisation Portal

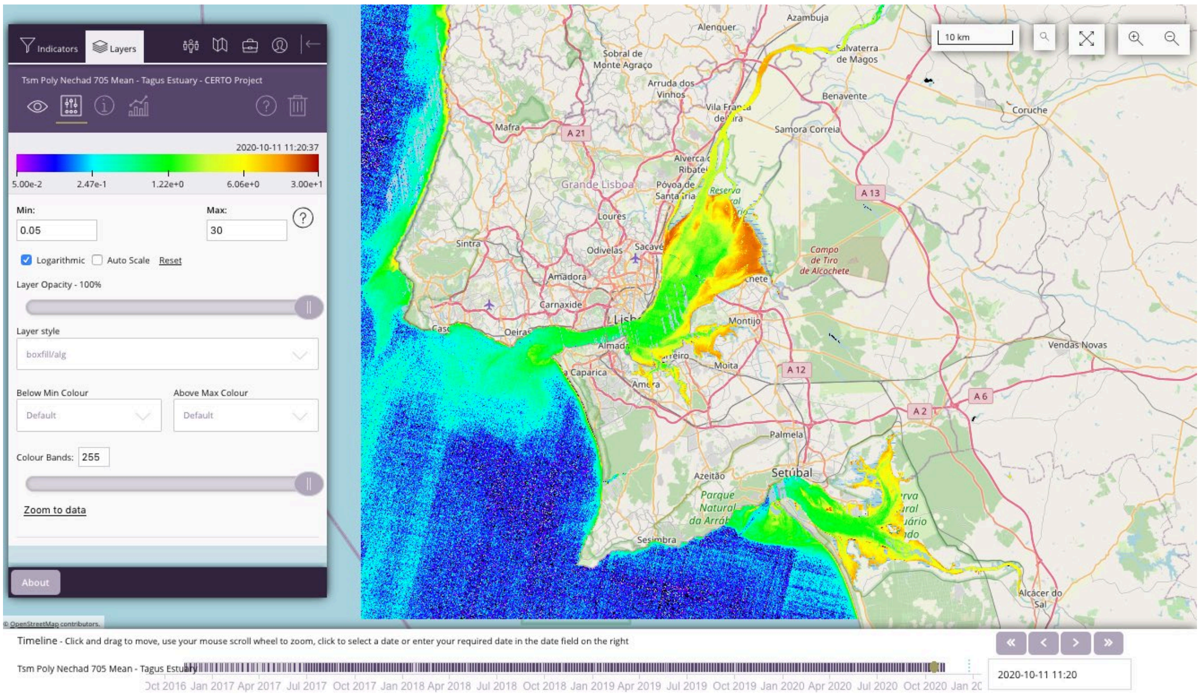
The GISportal is a data visualisation tool developed by PML Applications in previous FP6 (OpEc – Operational Ecology) and FP7 (eartH2Observe) projects; it provides an easy method for non-technical users to view and analyse data, as the only requirement is an internet browser, but is powerful enough to cater for more advanced user needs. An instance of this visualisation tool will be accessible via the main project website and will offer a predefined set of indicators or variables, grouped by type, provider, observation interval or source, which can be displayed on a map. A prototype portal is available at https://engage.certo-project.org/data/ and provides complete datasets of Sentinel 2 data for each of the six case study areas as defined above with water products *using current methods.* These data are being used by WP6 to develop Indicators and will also provide a baseline to quantify the improvements that will result from the developments undertaken in CERTO

With an indicator selected, the map and the timeline can then be used to explore the data spatially and temporally, as well as view any metadata associated with the indicator. Intuitive controls, such as click and drag to move, and double-click or scroll wheel to zoom in and out, allow users to manipulate the map. Moving a 'lozenge' on the timeline changes the date, or the user can select a specific date using the calendar or by typing it.

It is possible to change the scale settings to help highlight certain features in the data, for example, the user can define the minimum and maximum values, whether to use a logarithmic scale, and which scale bar palette to use to display the data. A suite of analysis tools allow users to produce timeseries plots, scatter plots or animations for a user defined area.



a



b

Figure 8 - Screenshots of the visualisation portal showing: a) the six user case study areas; b) total suspended matter (TSM) data in the Tagus estuary region.

9.4 Geographic Information System (GIS) Software

Some members of the user community or case study partners will already their own GIS software setup to meet the pre-existing needs. These systems range from open-source or freely available tools such as QGIS or Google Earth which are installed as client applications on workstations, through to propriety or custom-built systems that provide highly configurable access to multiple, centrally stored datasets delivered as client/server applications. The vast majority of these systems will be capable of interacting with OGC services, allowing users to access data via WMS without having to understand the WMS interface standard. Maps of CERTO data products can be combined with users' other data sources for comparison and analysis.

9.5 Direct File Download (S3)

In some cases, users will want to download data in their native format, perhaps for integration with internal systems, for analysis using specialist tools, or to host their own data access services. To accommodate this, all data produced by the CERTO prototype will be available for download via an API. Simple Storage Service, or S3, is a common standard used by all of the selected cloud providers which offers access to download netCDF data files.