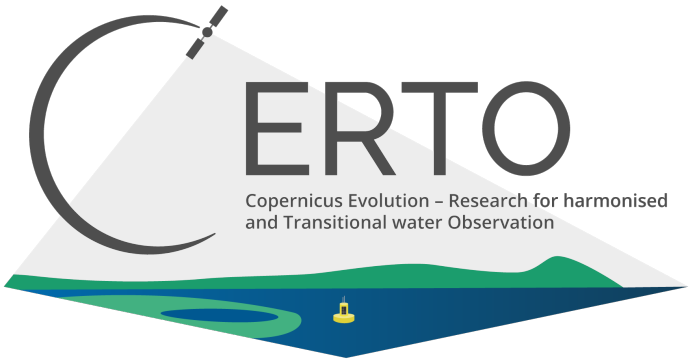

Horizon 2020

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

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

Project Number: 870349

Deliverable No: D7.1		Work Package: 7	
Date:	30-JUN-2020	Contract delivery due date	30-JUN-2020
Title:	Initial Report on DIAS & Cloud Provider Capabilities		
Lead Partner for Deliverable	PML Applications		
Author(s):	Amy Westlake, Oliver Clements, Kim Hockley, PML Applications; Steve Groom, PML		
Dissemination level (PU=public, RE=restricted, CO=confidential)			PU
Report Status (DR = Draft, FI = FINAL)			DR

Acknowledgements

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

Table of Contents

Glossary

API	Application Programming Interface
C3S	Copernicus Climate Change Services
CAMS	Copernicus Atmosphere Monitoring Service
CEMS	Copernicus Emergency Management Service
CLMS	Copernicus Land Monitoring Service
CMEMS	Copernicus Marine Environment Monitoring Service
DIAS	Data and Information Access Service
ENS	Elastic Node Server
ECMWF	European Centre for Medium-Range Weather Forecasts
EO	Earth Observation
GB	Gigabyte (unit of storage capacity)
GNSS	Global Navigation Satellite System
HTTP	HyperText Transfer Protocol
JSON	JavaScript Object Notation
NRT	Near real-time
PB	Petabyte (unit of storage capacity)
RAM	Random Access Memory
SSD	Solid State Drive
TB	Terabyte (unit of storage capacity)
VM	Virtual Machine
vCPU	virtual Central Processing Unit

1. Executive Summary

Water quality is a key worldwide issue relevant to human food consumption and production, industry, nature, and recreation. The European Copernicus programme includes satellite sensors designed to observe water quality and services to provide data and information to end-users in industry, policy, monitoring agencies, and science. However, water-quality data production is split across three services, Copernicus Marine, Copernicus Climate Change, and Copernicus Land, with different methods and approaches used and some areas, notably transitional waters, are not supported by any service.

The CERTO project aims to address this lack of harmonisation by undertaking research and development necessary to produce harmonised water-quality data from each service and extend Copernicus to the large number of stakeholders operating in transitional waters. The main output of the project will be a prototype system that can be “plugged into” the existing Copernicus services, developing Data and Information Access Services (DIAS), or popular open-source software used widely by the community (e.g. SNAP). This report is an initial assessment of the features and capabilities of the five main DIAS systems available at the time of writing (May 2020).

2. Introduction

This report compares the five main Copernicus DIAS systems (CREODIAS, WEkEO, Mundi, ONDA and Sobloo) against the requirements of the CERTO processing system.

The DIAS systems comparison features the technical services, pricing, technical offerings, and support options available, and how they meet the project requirements set out in the feature and requirement section. We compare the different pricing options available from the DIASs against a specification of virtual machines that would meet the base requirements for the CERTO system.

It is fully anticipated that over the lifetime of the CERTO project the DIAS services will evolve with user requirements, the availability of data will change with the launch of new Copernicus satellite missions, and the pricing will be updated to meet these changes. This document will be updated periodically to reflect the service and price evolutions.

3. DIAS Landscape

3.1. CREODIAS

Copernicus Services Available Locally
CREODIAS provides access to almost all Copernicus services; CAMS, CEMS, CLMS, CMEMS, and satellite data from; Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5P.

Technical Services
CREODIAS offers cloud-based virtual machines running CentOS, Ubuntu or Windows. The resources can be increased based on the users' needs, ranging between 1 GB and 128 GB RAM, and 8 GB and 384 GB SSD. At an extra cost, dedicated server virtual machines are also available ranging from 48 GB to 498 GB RAM. In terms of containers, two types are available; Docker Swarm and Kubernetes. An extensive REST API is available with access to OpenStack (cloud system), Finder (data viewer) and billing data. The API can export data in various formats; GeoTIFF, JP2, JPX, PNG, JPG. Integrated Jupyter Hub & Notebooks support is available free of charge for every registered CREODIAS user.

The CREODIAS cloud offering is called Hybrid Cloud and is provided by CloudFerro, it consists of a private cloud, located in Reading, UK and access to the public EO Cloud localised in Warsaw, Poland. Both solutions are based on the open source cloud system called OpenStack with access to 6PB of EO data and 150PB of climate data.

Earth Observation Applications
Two Earth observation applications are available; EO Finder and EO Browser.

- EO Finder is a search engine, designed for selecting products based on sets of specific variables - such as time, place, collection, processing level. The data can be accessed with JSON-based API.
- EO Browser is a front-end web application for querying and browsing all imagery datasets from various missions. Data can be searched by location name or GNSS coordinates, with a smart calendar option for quick overview of available dates. Results are returned in tabular form or a spatial overlay, with export options of the current view to a JPG or full resolution GeoTiff.

Pricing
CREODIAS have two main payment methods available: a flexible prepay per hour based on usage and a fixed, long term contract where access is booked for a certain length of time. For long-term usage, the fixed contract works out to be more cost effective; however, where usage may vary and management is required, the prepay per hour is more suitable.

There are multiple discounts available for the platform, including for contracts longer than 1 year or for large cloud environment orders, and a "special discount" of 18.5% for scientific institutions. CREODIAS offer free credits worth 150 EUR for new EO data users to test and evaluate the platform, plus free of charge EO data access from CREODIAS cloud environment.

Support Options
CREODIAS has multiple freely-available support options, including an extensive FAQ section with detailed tutorials and procedures, a searchable knowledge-base, and an accessible support email address for queries.

3.2. WEkEO

Copernicus Services Available Locally
WEkEO, implemented by EUMETSAT, ECMWF and Mercator Ocean International, provides access to partial set of the data produced by Copernicus services; CMEMS, CAMS and C3S, together with Sentinel-1, Sentinel-2, and Sentinel-3 data.

Technical Services
WEkEO has limited information about their services online compared to the other DIASs. Their virtual processing environments are available in various sizes, ranging from 16GB RAM up to 4TB with either CentOS or Ubuntu (both Linux) operating systems. Windows mini and full are also supported, if the client provides their own licence.

They have a basic API called Harmonised Data Access (HDA) API which allows access to the whole data catalogue, with filtering, searching and downloading functionalities. This API allows datasets usage with Jupyter Notebooks and WEkEO virtual machines.

Pricing
WEkEO provides a minimal free service including datasets, data downloads, Jupyter Notebooks, and support. Extended services start at 66 EUR/month which includes the free services plus Virtual Machines, processing tools, and networking. Virtual machines start from 124.68 EUR/month up to 3,304.72 EUR/month, customised plans are available for larger virtual machines than offered.

Support Options
WEkEO has a number of support options available, although these are limited. They provide a small list of FAQs, limited documentation, and a contact form.

3.3. Mundi

Copernicus Services Available Locally
Mundi provides access to the following Copernicus services: CEMS and CLMS, with Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5P data. It also has Landsat data.

Technical Services
Mundi's cloud service, called Open Telekom Cloud and hosted in Germany, offers virtual computing servers with scalability and load balancing. For virtual machine offerings users can choose between 1 and 60 vCPUs as well as 1 to 940 GB RAM with unlimited scalability. It runs on a "pay as you use" system where users only pay for monthly usage of storage (starting from 5 GB which is free).

Jupyter Notebooks integration is available and has pre-installed tools to enhance the user experience. Examples are available in a shared environment with the possibility to install complementary tools to fit your requirements.

Pricing
Mundi has a customised pricing structure, with a selection of free tools including data, sandbox, and basic support but with the option of paying for extended services. For Mundi

Cloud you pay depending on the configuration chosen but extended support begins at 750 EUR a day. Advanced tools come at a custom price whilst Copernicus and Landsat data are free.

Support Options
Mundi has multiple freely-available support options including a “getting started” guide, FAQs, full documentation, and a help contact form. Mundi is the only DIAS system to provide a full-service monitoring page showing the current uptime and status of Mundi services. There is also the option to pay to extend support and access their helpdesk if required.

3.4. ONDA

Copernicus Services Available Locally
ONDA provides access to CMEMS, CLMS and CAMS, and Sentinel-1, Sentinel-2, Sentinel-3, and Sentinel-5P data. ONDA is operated by a Serco Italia S.p.A. led consortium.

Technical Services
ONDA can supply a range of virtual servers based on a project's requirements, ranging from general purpose servers with balanced CPU and memory resources up to memory intensive or computing intensive servers for larger or faster processing power. They also offer the option to scale and configure custom environments or add extra storage to existing servers.

ONDA provides two different API’s; a basic Open Search API and an advanced API. The Open Search API simply allows for data discovery from the ONDA catalogue through HTTP requests. The second is an advanced API service that is free of charge for users who order a Virtual Server, it is run on an Elastic Node Server (ENS) which is Open Source providing a front-end to data storages for use with client Computing Instances.

Jupyter notebooks integration is available as part of a custom environment service which provides pre-configured virtual environments for all users who make use of Onda’s virtual servers.

Pricing
ONDA has a number of services available for free including data discovery, an advanced API, access to a WMS to view the data, and ability to download the data for free. Virtual servers, support, and data storage are charged extras. Four types of virtual servers are offered; General Purpose, Computing Intensive, Memory Intensive and simple storage volumes with pricing ranging from 7GB starting at 22 EUR/month up to 120GB RAM at 325 EUR/month excluding tax.

Support Options
ONDA has a number of support options available, including a list of 45 FAQs, documentation, multiple user guides, and a contact form. They also have a searchable glossary of technical terms and their definitions.

3.5. Sobloo

Copernicus Services Available Locally

Sobloo provides access to five Copernicus services: CMEMS, CLMS, CAMS, CEMS and C3S. For sentinel data Sobloo provides Sentinel-5p, Sentinel-3 (A), Sentinel-2 (A and B), and Sentinel-1 (A and B). Sobloo is operated by a consortium led by Airbus, Orange and Capgemini.

Technical Services

Sobloo's cloud service is based on OpenStack technology and provides open APIs with the ability to manage multiple regions on a pay-per-use model without commitment. For managed services they can offer management of either a Microsoft or Linux OS, taking care of backups, recovery and restore of single files or a complete system, and performance monitoring with monthly reviews. They offer a wide range of databases including PostgreSQL, MySQL, Microsoft SQL Server, Oracle, MongoDB, Cassandra, MariaDB, Couchbase, Redis, Neo4j, and ElasticSearch. Sobloo are the only DIAS to state that their managed services meet a 99.90% uptime as defined in their agreement.

Pricing

Sobloo has extensive price lists for various virtual machines and hosted services requirements.

Support Options

Sobloo has a number of support options available, including a list of FAQs, documentation, a getting started guide, and a contact form.

3.6. DIAS Service Summary Matrix

Table 1 displays a summary of the information captured in the previous sections for ease of comparison

Table 1 - Service summary matrix

Service/DIAS	CREODIAS	WEkEO	Mundi	ONDA	Sobloo
Data Availability					
C3S	✓	✓			✓
CAMS	✓	✓	✓	✓	✓
CEMS	✓		✓		✓
CLMS				✓	✓
CMEMS	✓	✓		✓	✓
Sentinel 1	✓	✓	✓	✓	✓

Service/DIAS	CREODIAS	WEkEO	Mundi	ONDA	Sobloo
Sentinel 2	✓	✓	✓	✓	✓
Sentinel 3	✓	✓	✓	✓	✓
Sentinel 5P	✓		✓	✓	✓
Pricing Options					
On demand pricing	✓	✓	✓	✓	✓
Reserved compute/long term pricing	✓	✓	✓		
Incentive free credit	€200	Offers free tier			
Support Options					
Online documentation /user guides	✓	✓	✓	✓	✓
Frequently Asked Questions	✓	✓	✓	✓	✓
Working Examples, e.g. Jupyter notebooks	✓	✓	✓		
Helpdesk support via email	✓	✓	✓	✓	✓
Telephone support			€750 /day		

4. Feature & Requirements Matrix

The primary requirements for the CERTO processing system are twofold.

- Local access to relevant Copernicus datasets
- Ability to install and run containerised code

The required Copernicus data will differ from process to process but the availability of Sentinel-3 OLCI and Sentinel-2 MSI datasets is a minimum requirement. The time-series length is also an important factor considered in this evaluation. Alongside this some small factors need to

be considered, e.g. the file formats that are available, the mechanisms for accessing data, and the future addition of data for NRT activities.

Throughout the CERTO project we will be developing a processing system that utilises Containerisation. Containerisation in this case meaning isolated code that is engineered in such a way as it can be executed on a variety of systems, provided they support the appropriate container manager. There are several software options for containerisation, but the current industry standards are Docker and Singularity. The ability to run containerised code is essential as it allows the project outputs to be developed in environments specific to the processing tasks required. These can include various programming languages and tools

4.1. Software requirements matrix

The software requirements at this stage of the project are minimal due to the processing chain design approach taken. All the processing chain code will be delivered in a container and, therefore, the hosting infrastructure needs to provide support managing and operating containers.

Table 2 - software requirements matrix

Requirement	WEkEO	Mundi	ONDA	CREODIAS	Sobloo
Basic container support	✓	✓	✓	✓	✓
Advanced container management	✓	✓	✓	✓	✓

4.2. Data requirements matrix

Table 3 - data requirements matrix

	WEkEO	Mundi	ONDA	CREODIAS	Sobloo
Sentinel-2 A & B MSI	✓	✓	✓	✓	Only previous 9 months on local, full archive available via API
Sentinel-3 A & B OLCI	✓	✓	Partial archive - data start May 2018	✓	Only previous 9 months on local, full archive available via API

5. Pricing

For the pricing comparison we defined two virtual machines: a basic system, which may be used for testing and development purposes, and a more powerful option, which may be a production level specification. The specifications for the virtual machines are as follows.

Basic VM

- CPU: 4 cores
- RAM: 16 GB
- Storage: 100 GB local storage

Advanced VM

- CPU: 8 cores
- RAM: 64 GB
- Storage: 250 GB local storage

To assess the options provided by the DIAS providers we sent a request to their respective support systems for prices for the machines specified above. The requests were sent on 27th May and providers were given a week to reply. The DIAS providers mostly offer a tiered pricing based on whether you are renting for a short period or committing to rent for a longer period. These differences are reflected in the pricing.

Table 4 - pricing comparison – all prices quote are Euros per month excluding VAT

	Basic VM – short term	**Basic VM – long term (1 year contract)**	**Advanced VM – short term**	**Advanced VM – long term (1 year contract)**
WEkEO	104.75	76,40	280.18	174.96
Mundi *	149.66	98.64	339.89	249.00
ONDA	55.00	Not provided	215.00	Not provided
CREODIAS	104.75	76.40	280.18	174.96
Sobloo**	141.91	No response	378.04	No response

* Mundi DIAS prices are calculated using the open telekom cloud calculator advertised on their website as the preferred way to calculate prices. The calculator can be found at https://open-telekom-cloud.com/en/prices/price-calculator

** Sobloo DIAS did not respond to our specific request so the information about their pricing has been taken from a price list available on the DIAS website[1]

[1] https://sobloo.eu/sites/default/files/Cloud%20Services%20-%20sobloo%28Rev.%200226%29.pdf

6. Conclusion

The DIAS all provide the data required by CERTO in order to meet the objectives of the project. The majority have a full archive of data for relevant sensors, Sentinel-2 MSI and Sentinel-3 OLCI. Sobloo are only able to offer 9 months of relevant data on local storage with the rest of the archive accessible via an API; however, for the purposes of CERTO, this is a minor issue.

Price seems to play the largest role in differentiating the DIAS: WEkEO and CREODIAS use the same infrastructure provider so their prices are identical. However, the prices offered by the providers have the potential to change and will be updated in future releases of this document.

Support response is another area of categorisation; during the process of creating this deliverable we have sent numerous emails to the relevant support emails. In general, the response time for an enquiry was always within 24hrs, and it is worth bearing in mind that these were responses to general members of the public as we aren't paid users yet.

Overall the facilities, data, and costs associated with the DIAS are very similar across the board, with minor variance in costs.