2023

The Copernicus Sentinel Data Access Worldwide Benchmark

Courtesy of ESA EO Graphic Bureau

Service Design Document

Funded by the EU

TABLE OF CONTENTS

Funded by the EU
opernicus
esa European Space Agency

Funded by the EU
opernicus
esa European Space Agency

DOCUMENT APPROVAL

Title	Copernicus Sentinels Data Access Worldwide Benchmark Service Design Document
Issue/Version	Issue 2/Version 2

Document Signature Table		
Authors	Alessandra Giustiniani (bookcaptain), Nicola Cafaro and Marco de Vitis - Exprivia Emmanuel Mathot and Fabrizio Pacini - Terradue	Alessandra Giustiniani Firmato digitalmente da Alessandra Giustiniani Data: 2023.05.04 10:06:27 +02'00'
Approved	Antonio Vollono - Exprivia	

DOCUMENT CHANGE RECORD

Issue/Revision	Date	Reason for Change	Changed Pages/Parag.
2.2.1	31/03/2023	Metrics update for E6 scenario	• Sec. 4.6, 8.4, ANNEX E update
2.2	14/10/2022	Update for new E6 scenario	• Sec. 4.6, 8.4, ANNEX E update
2.1	30/09/2022	Update for Q9/Q10 computation	• Sec. 3.10, 8.4, ANNEX E update
2.0.1	29/10/2021	Updated thresholds for Q3 and Q53 indicators. Removed note on SP5 collection; clarification on otherData metric.	• Sec. 3.3, 2.5 update • Sec. 3.1, 3.9
2.0	05/06/2021	Updated for Q4 computation, E2E scenarios 3 and 4	• Sec. 3.4, 4.3, 4.5, 8.3, 8.4 update
1.6.2	15/01/2021	Update of Q1 computation	• Sec. 3.1 update
1.6.1	01/12/2020	Updated for E2E scenarios 2 and 5 + Q9	• Sec. 4 update • Annex C update • Annex F update
1.3	03/08/2020	Minor updates on QoE and preliminary new E2E scenarios 2 and 5	• Sec. 3 update • Sec. 4 update
1.2.1	03/08/2020	Updates on Q10, TS07, TS15 and relative TestCases; integration on TP06, TP07 and TP10	• Sec. 3 update • Annex C update
1.2	24/07/2020	Update version including implementation for CCN#2	• Sec. 4 update
1.1	18/05/2020	Update version including design for CCN#2	• Sec. 4 added • Included ESA comments
1.0	24/04/2020	Update version prior to May data campaign collection.	-
0.6	31/01/2020	Updated version for SORR review	-
0.6 Draft	02/12/2019	Updated version after November data campaign collection.	• Sec. 3.1 New Q1 definition • Sec. 3.5 Q5 defnition • Annex: upgrades according to above
0.5 Draft	05/08/2019	1st Version according to the new documents' structure. For ESA internal review	-
0.5	20/09/2019	Revised according to ESA comments	-

Funded by the EU
Copernicus esa European Space Agency

DOCUMENT DISTRIBUTION

This document is intended for open distribution.

Funded by the EU
Copernicus
esa European Space Agency

Acknowledgments

The Sentinel Data Access Worldwide Benchmark project is under an assignment from the European Space Agency funded by the European Union in the frame of the Copernicus Programme.
The views expressed herein can in no way be taken to reflect the official opinion of the European Space Agency or of the European Union.

Funded by the EU
Copernicus
esa European Space Agency

1. INTRODUCTION

1.1 DOCUMENT SCOPE AND PREPARATION

The *Copernicus Sentinels Data Access Worldwide Benchmarking* service aims at providing impartial and reliable information on the Sentinels data accessibility, providing a comprehensive and factual overview of the conditions experienced by users in Europe and beyond.

As graphically represented in Figure 1, the service results are captured in three different kinds of deliverables that are:

- The *Service Design Document* (this document), presenting the main characteristics of the service and providing key information that is needed to interpret the service results.
- The *Service Specific Reports*, presenting the service results for a given target site (e.g. the ESA's hubs or the DIAS); and
- The *Service Summary Report*, presenting an overview of the core benchmarking results to the Copernicus governance.

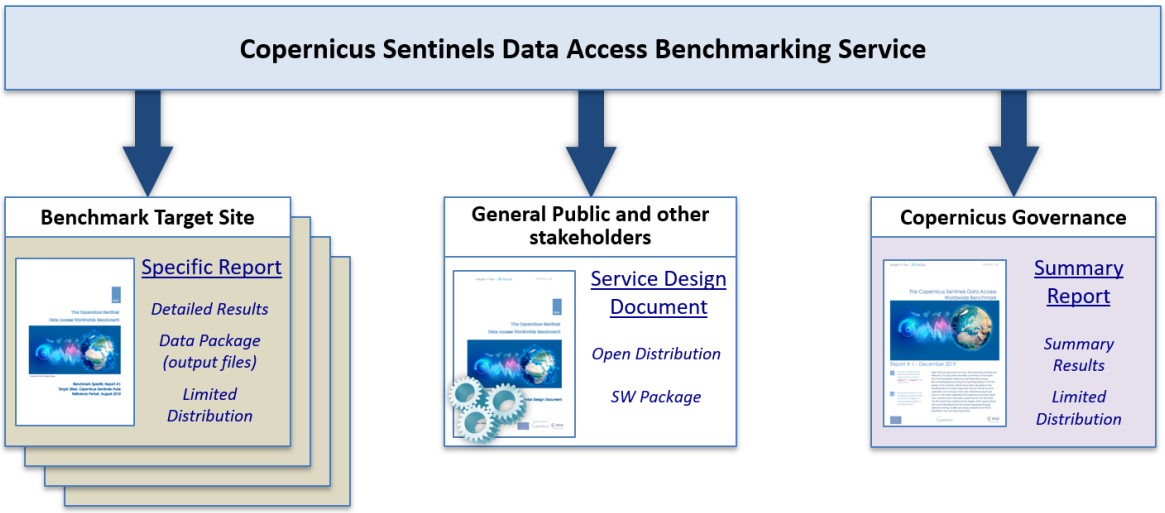


Figure 1: Graphical representation of the core deliverables from the Copernicus Sentinels Data Access Worldwide Benchmark.

While the service reports are issued following specific benchmarking campaigns and approximately every 6 months, the *Service Design Document* is updated as needed during the project lifetime to reflect improvements in the service, e.g. following the consolidation and extension of the available metrics as well as the finalization of the test sites network.

The attention is focused on the definition of a rigorous strategy that allows a comprehensive and easy-to-understand evaluation of the overall benchmarking results. The methodology is continuously calibrated and validated through extensive testing of the core target sites as well as through exchange with core stakeholders and experts via ESA.

Comments and suggestions for improvements of the methodology can be emailed to cdab@exprivia.com.

Funded by the EU
Copernicus esa European Space Agency

1.2 BACKGROUND AND OBJECTIVES

Copernicus Sentinel user products are made available by ESA to all users on a free, full and open basis via the Copernicus Open Access Hub. ESA has created also dedicated access points to serve specific user communities such as the Member States (the CollHub), the International Partners (the IntHub) and the Copernicus Services (the CopHub). In addition, the EC has launched an initiative to offer the data via highly scalable computing infrastructures called Copernicus Data and Information Access Service (or DIAS). 4 DIAS are managed by ESA and 1 by Eumetsat (see https://www.copernicus.eu/en/access-data/dias). Table 1 presents a prospect of the Sentinels data access services under ESA management. These constitute the core target sites for the benchmarking services.

Portal	URL	Description/Main Objective
Copernicus Sentinels Open Access Hub	https://scihub.copernicus.eu	The reference point for accessing Copernicus Sentinels data. It provides complete, free and open access to all available Sentinel user products.
Copernicus Sentinels Collaborative Data Hub	https://colhub.copernicus.eu	The Collaborative Data Hub provides Collaborative Ground Segments with access to a Rolling Archive of user products.
Copernicus Sentinels International Access Hub	https://inthub.copernicus.eu	The International Access Hub offers a Rolling Archive of 21 days of Copernicus Sentinels user products. The access to this Hub is restricted to International partners having established agreements with the European Commission.
Copernicus Services Data Hub	https://cophub.copernicus.eu	This Hub is dedicated to Copernicus Services (Atmosphere, Marine, Land, Climate Change, Security, Emergency).
Sobloo DIAS	https://sobloo.eu/	Cloud-based platforms providing centralised access to Copernicus Sentinels data and information products from Copernicus operational services. DIAS platforms also provide processing tools. Each of the five competitive platforms also provides access to additional commercial satellite or non-space data sets as well as premium offers in terms of support or priority
Mundi DIAS	https://mundiwebservices.com	
Onda DIAS	https://www.onda-dias.eu/	
Creo DIAS	https://creodias.eu	

Table 1: The Copernicus Sentinels Data Access services under ESA management.

Besides ESA's managed initiatives, Sentinels data are distributed also from partner entities (e.g. Eumetsat) as well as redistributed from a number of third parties, as fully allowed under the Copernicus open and free data policy. These are institutional (National), commercial and/or scientific redistributors of Copernicus Sentinel user products. Many of these sites have specialised their offer by hosting limited but targeted collections of interest for their community of reference. Therefore, they greatly differ in terms of e.g. the offered products types (e.g. Sentinel-2 data only), the geographical coverage (one country or one continent only), the time frames (e.g. recent data only), the latency of availability of the products and, for the commercial ones, the associated costs.

Portal	URL	Portals self-description
Austrian National Sentinel Data Hub access	https://data.sentinel.zamg.ac.at	The Sentinels National Mirror Austria provides free and open access to Sentinel-1 and Sentinel-2 user products.
Belgium National Sentinel Data Hub access	https://terrascope.be/en	User-friendly platform to access satellite data and derived information to the Belgian users.
Finland National Sentinel Data Hub access	https://finhub.nsdc.fmi.fi	Finnish Meteorological Institute's National satellite data centre provides easy access to latest satellite data from Baltic Sea region and Northern Hemisphere, dedicated data products unveiling the changes in Arctic environment and supports worldwide research activities of Arctic conditions.

Funded by the EU
Copernicus esa European Space Agency

French National Sentinel Data Hub access	https://peps.cnes.fr	PEPS - Plateforme d'Exploitation des Produits Sentinels - est développée et opérée par le CNES afin de répondre aux besoins nationaux en facilitant l'accès aux données et leur utilisation par les utilisateurs publics et privés pour stimuler l'innovation et développer le secteur de la valeur ajoutée, facteur clé du succès de Copernicus.
German National Sentinel Data Hub access	https://code-de.org/	The Copernicus Data and Exploitation Platform – Deutschland (CODE-DE) is the German entry point to the Sentinel Satellite Systems, their data products and the products of the Copernicus Services.
Hellenic National Sentinel Data Hub access	https://sentinels.space.noa.gr	Hellenic Mirror Site is the official Copernicus data access point for Greece publicly serving satellite data from the Sentinels over the region of South & Southeastern Europe, Middle East & North Africa, in a timely manner. It is part of the Copernicus Collaborative Ground Segment.
Italian National Sentinel Data Hub access	http://collaborative.mt.asi.it/	
Norwegian National Sentinel Data Hub access	https://colhub.met.no	
Portuguese National Sentinel Data Hub access	https://ipsentinel.ipma.pt/	The portal IPSentinel - "Portuguese infrastructure for storing and providing images of Sentinel satellites", allows free and open access to data from Sentinel satellites obtained for the Portuguese territory including the area of responsibility for search and rescue in the Atlantic.
Swedish National Sentinel Data Hub access	https://swea.rymdstyrelsen.se/portal/	Access from the Swedish National Space Agency
British National Sentinel Data Hub access	http://sedas.satapps.org/	The Sentinel Data Access Service (SEDAS) is a mechanism designed and developed to allow end users to search and download Sentinel data. The SEDAS portal will contain a 30 day rolling archive of Synthetic Aperture Radar (SAR) data from Sentinel 1A and 1B. (Product Types: GRD, SLC & OCN) as well as Optical data from Sentinel 2 (Product Type: L1C).
Czech Republic National Sentinel Data Hub access	https://dhr1.cesnet.cz/	
Geoscience Australia (GA)	http://www.copernicus.gov.au/	Copernicus Australasia is a regional hub supporting http://www.copernicus.eu/ and multifaceted Earth observation programme to date. We provide free and open access to data from Europe's Sentinel satellite missions for the South-East Asia and South Pacific region.
NASA Alaska Satellite Facility	https://vertex.daac.asf.alaska.edu	ASF maintains the NASA archive of synthetic aperture radar (SAR) data from a variety of satellites and aircrafts. ASF maintains a complete archive of processed Sentinel-1 data, available within three days of acquisition. Registration is required, and terms and conditions apply.
USGS Earth Resources Observation and Science (EROS) Center	https://www.usgs.gov/science-explorer-results?es=sentinel	The Sentinel-2 data were acquired, processed, and generated by the European Space Agency (ESA) and repackaged by USGS into tile-based bundles. The Sentinel2Look Viewer is a prototype tool that was developed to allow rapid online viewing and access to the USGS Sentinel-2 image archives.
Copernicus Online Data Access (CODA)	https://coda.eumetsat.int/	The Sentinel-3 Marine Products are available through the Copernicus Online Data Access (CODA) service. CODA is an online rolling archive with https access to Sentinel-3 Level 1 and Level 2 (Marine) global data in different latency modes
Google Cloud Sentinel-2 Data repository	https://cloud.google.com/storage/docs/public-datasets/sentinel-2	One way to query, visualize, and analyse the Sentinel-2 data is by using Google Earth Engine, where the data is available in the image collection with id COPERNICUS/S2. Sentinel-2 scenes are added daily as they become available. Data is typically available 1-2 days after publishing by Copernicus.

Funded by the EU

Amazon	http://sentinel-pds.s3-website.eu-central-1.amazonaws.com/image-browser/	Amazon S3 is a storage platform of Amazon Web Services (AWS). Level 1C scenes and metadata, in https://docs.aws.amazon.com/AmazonS3/latest/dev/RequesterPaysBuckets.html S3 bucket
Sinergise EO browser	https://sentinel-hub.com/explore/eobrowser	New Sentinel data are added regularly, usually within few hours after they are available on Copernicus OpenHub.
CloudFerro sp	http://www.cloudferro.com/en/eocloud/	EO Cloud Platform
Sentinel Australasia Regional Access (SARA)	https://copernicus.nci.org.au/sara.client/#/home	Free access to data from all Sentinel satellites for the South-East Asia and South Pacific region.
WEkEO DIAS	https://www.wekeo.eu	WEkEO is the EU Copernicus DIAS reference service for environmental data, virtual processing environments and skilled user support.
Luxembourg Space Agency Data Centre	https://www.collgs.lu	It has a massive archive including all S2 online
Microsoft Planetary Computer	https://planetarycomputer.microsoft.com/	Free access to Sentinel-1 GRD and Sentinel-2 L2A products

Table 2: Examples of other Copernicus Sentinels Data Access services. The explanation text is based on the declarations of the providers (i.e. cut and pasted from the portals).

As it can be seen, these sites have different characteristics and offer different services, depending on their core objectives and on their user community of reference. Users have therefore a number of possible access points available to discover, view, download and locally manipulate the Sentinels data. The choice of the preferred site/s will be guided by the user specific needs (e.g. the available collections, the cost, the timeliness) but also by the user experience. This can greatly vary, depending not only on the performances and the objectives of the sites themselves but also on additional factors related to the user specific environment and connectivity.

The Copernicus Sentinels Data Access Worldwide Benchmark service aims at providing impartial and reliable information on the Sentinels data accessibility from the access points under ESA management, providing **a comprehensive and objective overview of the conditions actually experienced by users**. The core ESA requirements are recalled as follows:

- The *benchmark* shall be **representative of the most common users' behaviours** when discovering, downloading and, where applicable, processing the Copernicus Sentinels data.
- The benchmark must be **reproducible and verifiable**. This implies that both the test strategy as well as the results are extensively documented and recorded so that tests can be consistently repeated and results compared. The software codes will also be published at due time during the project.
- The benchmarks must be **robust and objective**. Measures for the internal diagnostics must be put in place, to be able to automatically disentangle underperformances of the test network that could impact the measurements quality.
- The benchmark service must be **configurable and expandable**. This implies the capability to add and/or remove test sites, target sites and test cases without perturbing the on-going benchmark operations.
- The *benchmark* shall be **geographically representative worldwide** as relevant for Sentinels data access.
- The *benchmark* shall be as **representative** as possible of the most common technological solutions adopted by Sentinels users. This implies that different solutions must be chosen for the test platforms, including different service providers and cloud vs. bare metal solutions.

The original requirement for the benchmark to be "**representative of the most common user's behaviour**" it's very hard to set expecially for the DIAS target sites. In fact an enhancement of the benchmarking activity is foreseen in order to investigate "**typical user operation**" of specific target sites. This imply to address application-oriented end-to-end data flows such as: Rapid Mapping, NDVI mapping, Rapid mosaicking/Sea ice mapping, Trend mapping, Interferogramme computation.

Funded by the EU
Copernicus esa European Space Agency

Specific details on those new user's scenario will follow on next document deliveries as soon as they will be defined and implemented.

It is worth to highlight the emphasis that is put on the reliability and trustability of the exercise, to be achieved through the transparent communication of the tools and of the obtained results. The verifiability and reproducibility of the results, through an open discussion with ESA and the interested parties, supports the ultimate goal to achieve a robust benchmarking of the Sentinels data accessibility. In a longer term perspective, the production of standardized metrics and rigorous tools to be shared publicly in support to third party self-benchmarking is considered instrumental for streamlining the benchmarking strategies.

Computer-systems performance analysis often feels more like an art than a science. Indeed, different individuals can sometimes reach apparently contradictory conclusions when analyzing the same system or set of systems. While this type of ambiguity can be quite frustrating, it is often due to misunderstandings of what was actually being measured, or disagreements about how the data should be analyzed or interpreted. These differences further emphasize the need to clearly communicate all results and to completely specify the tools, techniques, and system parameters used to collect and understand the data.

From Lilja, David J., Measuring computer performance: a practitioner's guide, ISBN 0521641055

1.3 DOCUMENT OVERVIEW

This document is structured as follows:

- This introductory section
- The benchmarking strategy (Section 2)
- The benchmarking QoE indicators (Section 3)
- Known perspectives and possible ways forward (Section 4)
- ANNEX A – Recall of Benchmark Service Architecture
- ANNEX B – The current network of test sites
- ANNEX C – Description of Test Scenarii and Test Cases
- ANNEX D – Configuration parameters and datasets for benchmark operations
- ANNEX E – Detailed description of benchmark basic metrics
- ANNEX F - The benchmarking SW Suite (not yet present in this version of the document)

Funded by the EU
opernicus esa European Space Agency

1.4 ACRONYMS AND DEFINITIONS

Term	Acronym	Description
Application Index	APDEX	Open standard methodology used to measure the performances of SW applications
Benchmark Trusted Goldmarks		test sites characterised by known and guaranteed performance, used for the computation of the core indicators
Core target sites		the Sentinels Data Access sites under ESA responsibility, to be benchmarked as part of the activity
Data and Information Access Services	DIAS	Cloud-based platforms providing centralised access to Copernicus Sentinels data and information products from Copernicus operational services
Reference SW Suite	RSWS	The publicly available SW package allowing the benchmark operations towards core target sites
Quality of Service	QoS	Measurement of the overall performance of a service, in particular services provided on a network
Quality of Experience	QoE	Measure of the satisfaction of a customer with a service in all the aspects of the experience
Q1 Reference Collections		The Copernicus data collections used as a reference to compute the Q1 (see Section 3.1)
Q1 Reference Site		The site used as a reference for the definition and computation of the Q1 Reference Collections
Test Sites (or User Sites)		The HW platforms (physical or virtual) from which the tests for the quality assessment are launched

1.5 REFERENCES

Document Title	Identifier	Internal Reference
Copernicus Website	www.copernicus.eu	[RD1]
Sentinels OnLine	https://sentinels.copernicus.eu	[RD2]
Statistics of Sentinels Data Access Hubs	https://scihub.copernicus.eu/reportsandstats/	[RD3]
Sentinel data Legal Notice	https://sentinels.copernicus.eu/documents/247904/690755/Sentinel_Data_Legal_Notice	[RD4]
Copernicus Sentinel Data Access - 2018 Annual Report by SERCO	Available at: https://scihub.copernicus.eu/twiki/do/view/SciHubWebPortal/AnnualReport2018	[RD5]

Funded by the EU

2 THE BENCHMARKING STRATEGY

2.1 OVERVIEW

The benchmarking service aims at assessing the quality of the accessibility services provided by the Copernicus Sentinels data portals that constitute the **Benchmark Target Sites**. The target sites under ESA responsibility (i.e. the 4 ESA data Hubs and the 4 DIAS developed under ESA contracts) constitute the **Core Target Sites** of the service. A series of Third Party providers will be also assessed to support external benchmarking of the core target sites.

In order to achieve its objectives, the benchmarking service encompasses a combination of measurements, interpretation and communication activities that include:

- The systematic execution of benchmarking operations, simulating typical **user scenarios** from different test sites located in Europe and around the world;
- The cleaning, analysing and interpretation of the results;
- The presentation of the results towards various actors;
- The monitoring of the results and of the benchmark quality over time; and
- The implementation of service evolutions as appropriate.

For the sake of the benchmarking service, "typical" operations performed by the Sentinels users' can be sketched according to the following scenarios:

- *Users performing simple data search and single downloads (TS01).* This is generally common for occasional users, especially for optical data such as the ones from Sentinel-2.
- *Users performing complex data search and bulk download (TS02).* This reproduces the most common feature typical of occasional users performing some specific or dedicated assessment (e.g. over given geographic areas and/or given time periods). This applies to many different kinds of users, especially common on the exploitation of Sentinel-1 and Sentinel-2 data. When backward in the past, this scenario can involve data that are available on an off-line archiving facility.
- *Users performing systematic periodic data search and related remote data download (TS03 and TS04).* This pertains to operational users and can be applied to local, regional as well as at global level. At global level, it is a typical scenario for scientists and/or policy users and it typically relates to Sentinel-3 and Sentinel-5P data although it can relate also to Sentinels 1 and 2.
- *Users explore the overall data offer from a given distribution site* (*TS05 and TS06*)
- *Users accessing and processing Sentinels data from within a cloud environment (TS11, TS12 and TS13).* This scenario is typical of the DIAS users, and can reproduce the data load approaches described in the scenarios described above but within a cloud environment and including the related processing and, optionally, download and/or publication of the resulting products.
- *Users performing simple search and view of data over a specific location (TS09).* This mostly relates to optical data such as the ones from Sentinel-2.

These scenarios are described in Annex 8.1 of the current document.

In order to be able to simulate more comprehensive user operations involving the generation of value-added information in support to specific applications, five end-to-end scenarios have been added as follows:

- *Users generate NDVI maps in support to e.g. forest monitoring or agriculture (E1).* This scenario concerns very simple processing or manipulation of same-product parts. This implies e.g. extraction of features from a given product and computation of a generic index (e.g. bands manipulation).
- *Users perform rapid mapping in support to e.g. emergency management (E2)* This concerns the rapid identification of data of relevance and their display following some specific filter (e.g. cloud free). Area sizes are relatively small (e.g. cities/ basins) and little or no processing is involved. Focus is on stringent timelines and identification of best pre vs. post event images.
- *Users perform rapid mosaicking in support to e.g. sea ice mapping (E3).* This concerns the computation of a generic index derived from mosaicking small amounts of Sentinels data of the same type over a

Funded by the EU
Copernicus esa European Space Agency

medium size area (e.g. the Arctic Sea) and within a given time window (e.g. daily composites). Focus is on timeliness and mosaicking (TBC). This scenario is typical of operational users having stringent timeliness constraints such as, for example, support to sea ice navigation in the Baltic Sea during wintertime or support to emergencies in the immediate aftermath of a crisis over a specific location.

- *Users analysis and map long-term trends (E4).* This concerns complex processing of multiple data of the same type over long time windows to build multi-temporal sequences derived from comparison of data over the same area. Concerned time spans and area sizes can be span from medium sizes to very large ones. When applied over very long time frames and at global level, this scenario may require powerful AI computing. Typical examples can be related to monitoring of deforestation or of air quality parameters. Focus is on orchestration and processing performances when huge quantities of data are involved. Options for storage and products exploitation/visualization may be at stake.
- *Users compute select Synthetic Aperture Radar interferometric pairs and compute interferogrammes (E5).* This involves the complex computation derived from multiple time series of the same data types. Focus is on tools facilitating the data search for the identification of interferometric pairs and for building multi-temporal stacks.

These scenarios are supposed to be executed within the cloud computing environments available from the target sites and are described in Annex 8.3 of the current document.

With the exception of the five end-to-end scenarios and all operations executed within the cloud environments (for which the user infrastructure is irrelevant), the above mentioned user operations are executed from a number of **Test Sites** distributed in Europe and in various locations around the world. The test sites network is composed of **30 nodes**, whose geographical distribution is as follows:

Continent	Number of Test Sites
Europe	**18**
North America	**3**
South America	**2**
Asia	**4**
Africa	**2**
Oceania	**1**

Table 3: Worldwide Distribution of the Test Sites Network

This list must not be considered static: it evolves during the project lifetime according to the measured performances, the obtained benchmarking results and cost/benefits considerations.

A limited set (5) of top performing sites in Europe are chosen to represent the benchmark reference with respect to the European hubs: they are used to compute the various quality indicators, as it is explained in the following Sections.

The Test Sites are managed remotely as part of a broad architecture designed to provide a regular and automated monitoring of the target sites in an automated way. A complex orchestration of operations from the various test sites to the various Target Sites is operated, whereby each test site issues one or more requests towards one or more target sites, according to a pre-defined test-to-target matrix. The results are systematically stored, analysed and reported. Details can be found in ANNEX A - Service Architecture.

2.2 FROM QUALITY-OF-SERVICE TO QUALITY-OF-EXPERIENCE

Funded by the EU
opernicus esa European Space Agency

The benchmarking service must assess the quality of the data access services as experienced by the users. An evolution is thus required from the concept of **Quality of Service (QoS)** (that is measured at the data access points from the service providers) to the wider idea of **Quality of Experience (QoE)**, where the key factor is not only to measure the data access performance parameters, but the perception that users have of the provided service.

The key questions to be addressed for evaluating the user's perception can be summarised as follows:

Q1. How large is the Copernicus Sentinels collection made available by the target site? How rich is the actual data offer?
Q2. How reactive is the target site in responding to the user's requests?
Q3. How quickly can users find (and visualize) the desired Sentinels products?
Q4. How quickly can users download the identified Sentinels product/s of interest?
Q5. Where applicable, how efficiently can users access and process Sentinels data in a cloud environment? How costly is this service?
Q6. How stable are the above measured performances along time?
Q7. How stable are the above measured performances with respect to the users' load?
Q8. How variable are the above measured performances with respect to the user location in Europe and around the world?
Q9. How rich is the offer of tools and services and complementary datasets to the users?
Q10. How effective is the support provided to the users to allow a full exploitation of the services?

In order to answer these questions in a simple, yet representative, objective and quantifiable way, ten QoE indicators were defined that effectively synthesize the measured performances. Each target site can then be evaluated according to its set of QoEs and the derived classification.

The first indicators (Q1 to Q8) are automatically computed: each functionality, dedicated requests are automatically executed tens or even thousands of times during the reporting period and from different test sites in Europe and around the world. How can all these results be synthesized in an effective way? A hierarchical computation was developed that progressively refines and combines the obtained results, based on the following principles:

- For each metric MX, we derive a statistic representation of the percentage of users that are satisfied, tolerating or dissatisfied with the service. We do so though a so-called **application performance index** (or **APDEX**) that is closer to the user perception with respect to the engineering result.
- The APDEX, which was defined after the APDEX Alliance[1], leverages on the distinction of three performance zones (i.e. *Satisfied*, *Tolerating* and *Frustrated*) and weighs the number of times the user experience falls in one of these zones (with weights of 1, 0.5 and 0 respectively) over the totality of the relevant test counts according to the formula below:

$$APDEX(MX) = \frac{N\ Satisfied\ Counts + 0.5 \times (N\ Tolerating\ Counts)}{Total\ Counts}$$

In an automatic performance evaluation system, like ours, the Tolerating Counts are the values of the Metric that are considered acceptable, whereas the Satisfied Counts are the values of the Metric that are considered more than acceptable. The APDEX classification of a Metric can be represented as follows:

[1] The method was derived from the APDEX Alliance, a community of enterprises and vendors striving to understand and improve application performance using the Apdex methodology which links application performance to business objectives https://www.apdex.org

Funded by the EU

Figure 2: APDEX zones.

- By definition, the APDEX values fall between 0 and 1 and are proportional to the user satisfaction levels (i.e. APDEX=0 means that no user is satisfied while APDEX=1 indicates that all user samples were in the "satisfied" zone).
- In order to have a stable reference for the quality assessment of the target sites from the European perspective, the APDEXES relevant to the QoEs Q2, Q3, Q4, Q5, Q6 are computed only from a limited set of test sites located in Europe that constitute Europe's Reference Test Sites (see Annex B to this document) and over the reporting period. Longer time frames are evaluated separately as part of the time stability criterion (see Sub Section 3.6) while geographic variability is assessed as part of criterion Q8 (see Sub Section 3.8).
- For the assessment of the Geographic Variability (Q8) we make use of all the test sites spread all over the world. In this case we repeat the measurement of all the other criteria and assess the variability of the results with respect to the geographic location.
- Quality criteria and overall scores of target sites can be easily compared.

For each target site, the whole process develops as follows:

- Step1: Results from the benchmarking operations are recorded in a number of elementary metrics that are computed from each test site (the full list of the metrics can be found in ANNEX E to this document)
- Step2: The elementary metrics are filtered to obtain corresponding application indexes (or APDEXs)
- Step3: The APDEXs relevant to a given functionality (i.e. related to the questions Q1 to Q7) are weighted and combined as appropriate to obtain the corresponding QoE indicator
- Step4: The target site is classified according to the QoE indicator values for each of the specific questions.

This is illustrated in the figure below:

Funded by the EU
opernicus
esa European Space Agency

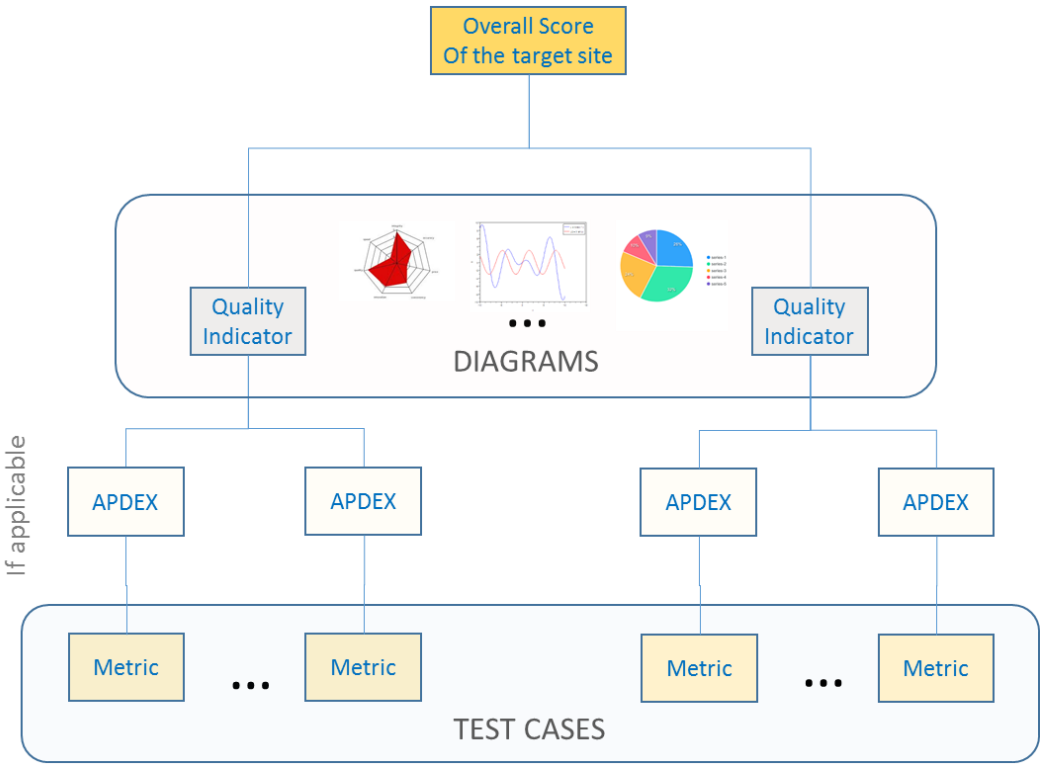


Figure 3: Schematic representation of the hierarchical calculation of the QoE indicators, leveraging on the elementary metrics.

Correspondingly, radar diagrams can be created to represent in a graphical form the quality of the Target Sites. The following drawing illustrates this concept:

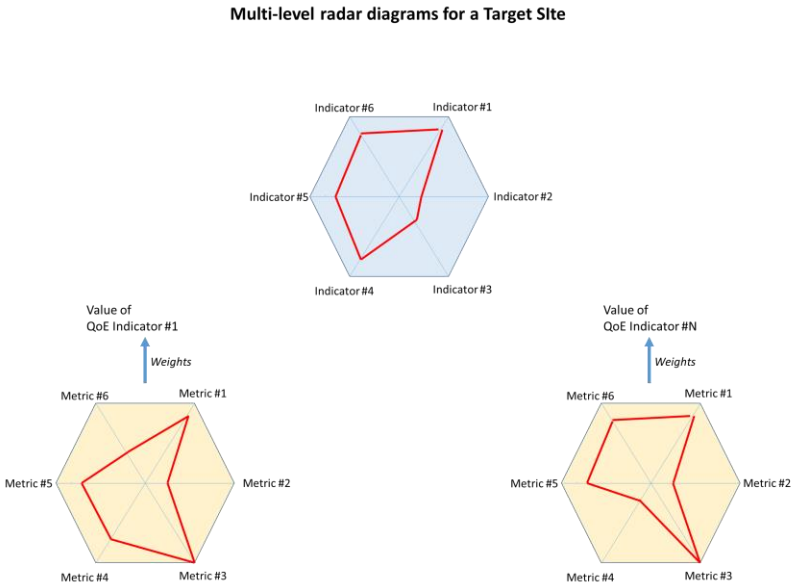


Figure 4: Hierarchical diagrams.

A fundamental aspect to consider when benchmarking the EO data providers is the evolution of results along time. The values of elementary metrics are progressively accumulated and their analysis provide an interesting perspective on service quality.
The following figure illustrates the above concept:

Funded by the EU
opernicus
esa
European Space Agency

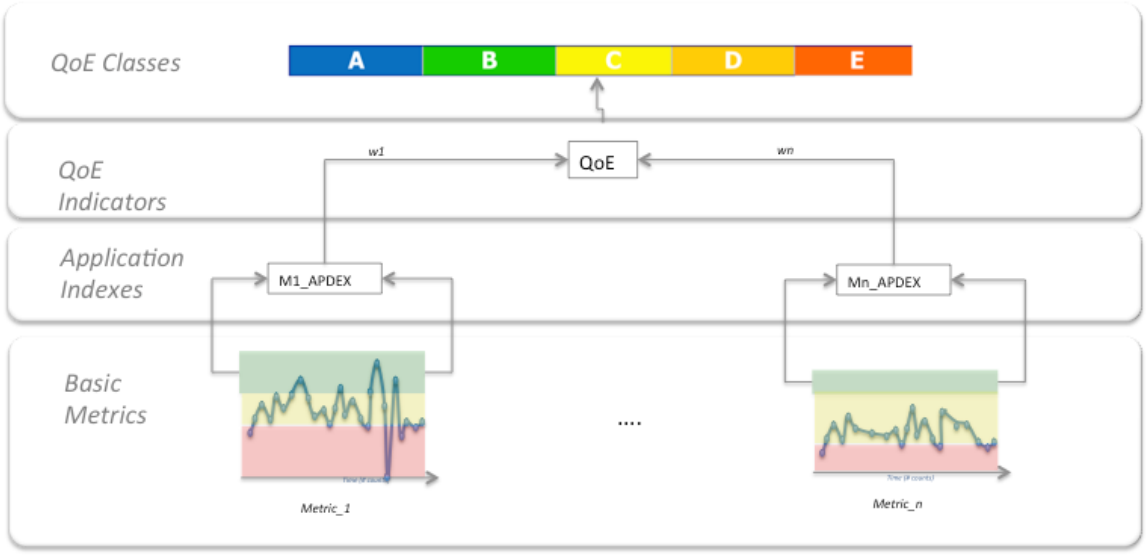


Figure 5: Hierarchical diagrams showing the QoE classification based on the time sequences of the basic metrics.

The colour codes and the letters (A-E) give an immediate impression on the Quality of Experience indicators.

For the indicators Q9 and Q10, the difficulties to measure these features in an objective, repeatable and verifiable manner has been carefully considered and a mix of manual and automated operations is performed with evaluations performed against pre-defined criteria.

Finally, it is worth mentioning that there are other aspects that have not been considered here but that are equally important for the users: a "catchy" graphical layout, a well-planned easy-to-navigate and consistent information content, the availability of richly contributed marketplaces and hosted third-party applications, where users can easily and quickly navigate through value-added available resources can be winning assets. However, it seems not possible to express a quantifiable and repeatable evaluation thereof and they are not taken into account in the current version of the benchmarking.

2.3 FROM CROSS-CUTTING TO END-TO-END APPLICATION-SPECIFIC EVALUATIONS

The quality of experience described in the previous Section is computed across different Sentinels product types, data volumes and scenarios and is therefore not specific to any application or user type. Nevertheless, experienced users who elaborate value-added information from Sentinels data would be more interested about evaluations targeted to their specific application. In order to assess these usage-specific performances, the benchmark service adopts five end-to-end scenarios and evaluates them "vertically" from an end-to-end standpoint.

End-to-end scenarios are built through the execution of flows that include both engineering activities and automated operations. Five different steps can be distinguished in a generic end-to-end scenario as follows:

- **Step #1 Registration & Offer**. This is the starting point for our user that wants to use advanced services from a provider. Indeed, the use cases from this template needs to provision a development and a production environment. Those resources are not available

Funded by the EU
Copernicus esa European Space Agency

for guest user. We therefore assume the most of the time, user requires a user account registration with the opening of a commercial account or a subscription to use paying resources. Since commercial operations such as using it's credit card or being charged directly from it's bank account is not as trivial as using free services from an free account, we also assume our user needs some support from the help-desk to set-up its commercial account. At the end of this step, we should have a registered user ready for provisioning all the resources necessary for the rest of the scenario.

- **Step #2 Application Integration**. This is a large step including a lot of manual procedures that may vary a lot according to the use case. The main objective is to have an application "integrated" in the provider platform. This means that the user's application is ready for a scalable production in their hosted processing service offer. It also includes the user data management via the usage of the storage service of the data provider. The step contains several actions implying a cost that must be accounted for reporting as a metric. At the end of the step, the user should have a container (e.g. docker), preferably hosted on provider platform ready to be executed on provider resources (Data processing step).

- **Step #3 Data Access**. This is the step dedicated to the data access of the scenario. It consists mainly in assessing the data query and access service first in terms of functionalities in the context of the use case. Then on the second hand, the benchmark shall be performed with data queries and download tests.

- **Step #4 Data Processing**. This is the step dedicated to the cloud processing of the scenario. It consists mainly in assessing the processing performance in the context of this use case. To achieve the objective of this step, the operator shall execute Test scenario #15 of the cdab benchmark suite with the parameters relative to NDVI processing (Processing Scenario [ps] in cdab test suite)

- **Step #5 Data Exploitation**. This is the step dedicated to the data exploitation of the scenario. It consists mainly in assessing the way of exploiting processing results on the provider platform. To achieve the objective of this step, the user will perform a manual evaluation of the adequacy of the tool for data exploitation and some automated tests for catalogue query for the data results if available.

The activity diagram below shows the implementation of the end-to-end flow.

Funded by the EU
opernicus
esa
European Space Agency

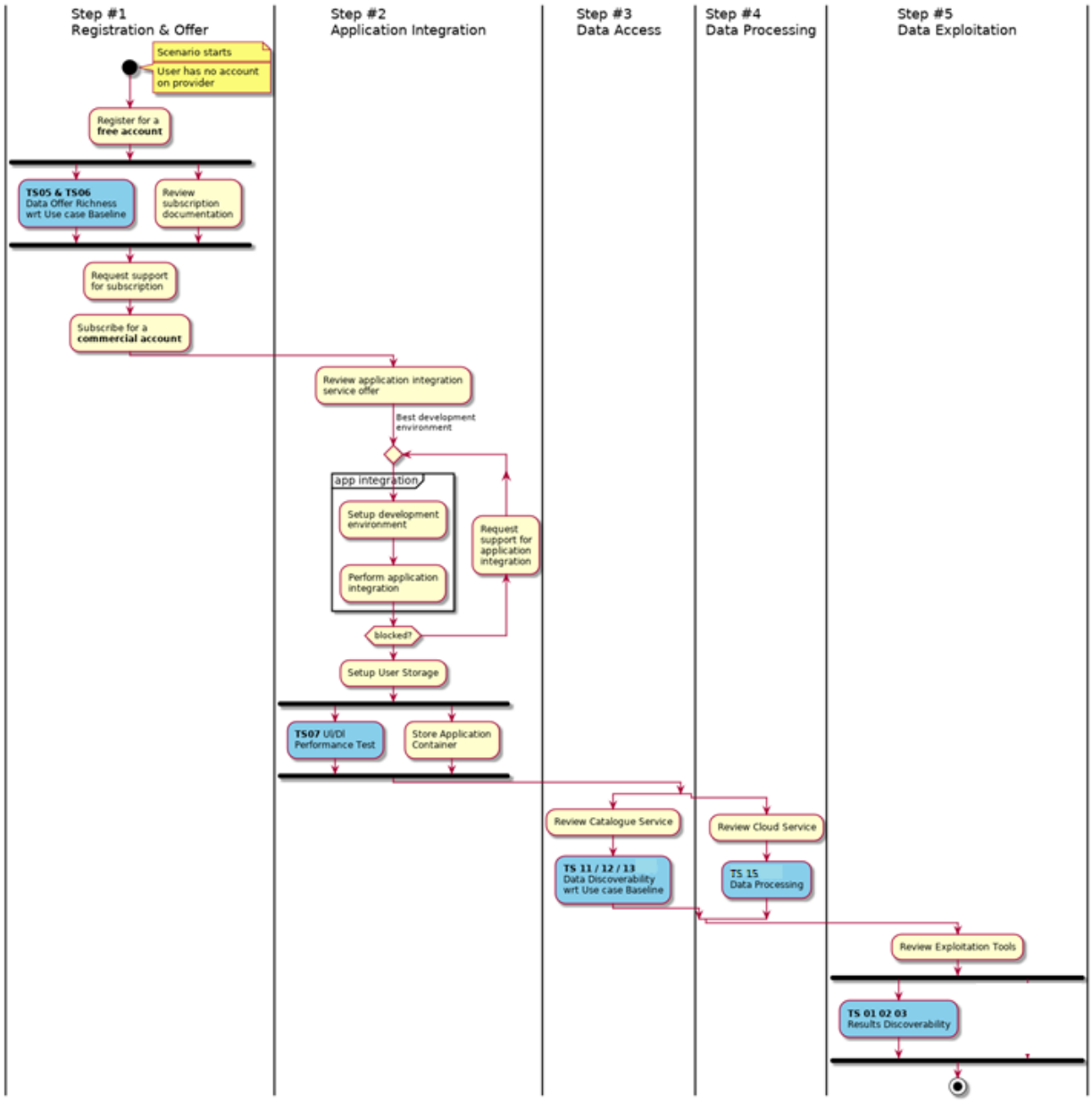


Figure 6 Activity diagram for E2E Test Scenario. The single blocks refer to simple steps either to be performed in an automated way (in blueish) or manually (in yellow). Test cases and procedures are detailed in Annex C to this document.

The figure above shows the sequential execution of the different procedures or test cases. For each of these, one or more metrics can be measured/noted that represent an evaluation of the outcomes. These metrics may be the same as the metrics already used to benchmark the cross cutting services (as described in the previous Section) or a customization of them that takes into account the specificity of the target end-to-end scenario (e.g. in terms of the data set of interest or its relative weight).

In the same manner as the short term for the "Quality-of-Experience (QoE) indicator" was chosen as "Q", the End-to-End (E2E) indicator short term is "E". The attached sequence number range from 1 to 5, where 1 is the NDVI mapping scenario, so at the end E1 means the "E2E quality indicator for NDVI scenario".

Throughout the flow, metrics are computed for each step and then aggregated to derive a single end-to-end total score as follows:

Funded by the EU
Copernicus
esa European Space Agency

$$E[1..5] = \sum_{i=1}^{5} Weigth_i * E[1..5][i]\_score$$

Where [1..5] is the E2E indicator for each scenario (E1, E2, …), i is the step of Use Case scenario as depicted in the figure above. The different steps are generically taken to have relative weights as indicated in the Table below.

Step	E[1..5] term	Weight
#1 Registration & Offer	E[1..5]1_score	0.30
#2 Application Integration	E[1..5]2_score	0.20
#3 Data Access	E[1..5]3_score	0.20
#4 Data Processing	E[1..5]4_score	0.20
#5 Data Exploitation	E[1..5]5_score	0.10

Table 4: Indicative E2E score terms and relative weights for each step of the End-to-End scenarios.

More in details, each score will be computed as the combination of different metrics retaled to that steps as shown in the figure above. However, since the single metrics used in each step for each scenario can vary, they are explained in details in the following Sections. Details on the user scenarios/procedures and the metrics extracted from the automated and manual tests are specified in Annexes C and E to this document.

The E2E E6 (High Demanding, Massive Processing With S-2 Data) have the steps #2 and #5 not applicable, so the score terms and relative weights will be the following:

Step	E6 term	Weight
#1 Offer	E61_score	0.20
#3 Data Access	E63_score	0.40
#4 Data Processing	E64_score	0.40

Table 5: Indicative E2E score terms and relative weights for each step of the E6 End-to-End scenario.

2.4 CALIBRATING AND VALIDATING THE BENCHMARKING SERVICE

A benchmarking service in charge of evaluating third parties performances must be validated and robust. The service key challenge is to gather multiple and complex information and to present it in a synthetic yet exhaustive manner. This relies on the adequacy of some key elements that are:

- The reliability, representativeness and effectiveness of the service operations
- The parameters used in the computation of the QoE classification and
- The performances of the test site network and the choice of the ones to be used as trusted goldmarks.

Dedicated procedures have been put in place to calibrate and validate these, as it is explained in the following sub-sections.

Funded by the EU
opernicus esa European Space Agency

2.4.1 The randomization approach and the optimization of the service operations

The user operations via API are performed in an automated way. A set of "typical queries" are launched from each test site on the catalogue API of each target site. The queries are generated randomly picking up filters from a pre-defined Mission Configuration Dictionary. Some examples of simple and complex queries that are randomly generated according to this method are:

- "Mission Sentinel-1 Track [16 TO 169]",
- "Mission Sentinel-2 Count 86 Level-1C",
- "Mission Sentinel-1 Count 62 Stripmap",
- "Mission Sentinel-2 A Level-2A",
- "Mission Sentinel-3 B Near Real Time",
- "Mission Sentinel-3 Near Real Time Track [330 TO 355]",
- "Mission Sentinel-1 Count 47 Level-2 Ocean (OCN)",
- "Mission Sentinel-3 Count 27 A"
- "Mission Sentinel-2 Level-2A From Sunday, October 22, 2017 To Thursday, June 13, 2019 intersecting United Republic of Tanzania",
- "Mission Sentinel-1 Count 13 From Sunday, January 22, 2017 To Tuesday, June 27, 2017 intersecting Cook Islands".

Details are provided in Appendix D to this document.

The randomization process aims at ensuring that the discovery and download performances of a target site are properly benchmarked (i.e. avoid caching by launching each time a new query) and that the functionality is properly exercised over a wide spectrum of filtering capacity and results accuracy (as opposed to repeating a limited set of fixed pre-defined operations). If the service is run over a sufficiently large period and in combination with a fairly significant load factor, the randomization process ensures that these objectives are met whilst maintaining a fair exercise of all target sites. Should this not be the case, there could be biases in the trial requests sent to the different target sites and this would jeopardize the objectivity of the benchmarks.

It is therefore crucial that the representativeness and the fairness of the approach be validated during the reporting period, at least during an initial ramp-up phase. This is achieved via the (a posteriori) analysis of the statistics of the performed queries and of the subsequent downloads (e.g. number of returned items from the catalogue, size of the data products that were downloaded). This analysis also aims at establishing the minimum time interval that is necessary to ensure that the objectives of the randomization process are met and that biases are very unlikely.

Optimisation of service operations implies also the need to limit errors when accessing target sites.
In this light, another element that needs monitoring to validate the benchmark results relates to the internally-generated interference errors. In fact, most target sites require the generation of accounts and in order to limit the overall number of accounts, sometimes the same account is shared between multiple test sites. This implies that requests launched at the same time could interfere the one with the other and impact on the results. Although the orchestration has been optimised in order to minimise (and, where occurring, purge) this effect, there might still be possible anomalies that must be detected especially for tests that are performed very frequently.

2.4.2 The QoE thresholding

As mentioned in the previous Section, the benchmarking service is based on the computation of simple, comprehensive indicators that are taken to represent the quality of experience of a user. These QoE indicators are derived through statistical computations based on the following key parameters:

- the frustration and the satisfaction thresholds used to compute the single APDEXes for a given metric and

Funded by the EU
opernicus esa European Space Agency

- the weights used to combine the APDEXes into a single indicator.

The setting of these parameters heavily affects the scores and so deserves careful analysis.

The "performances zones" as well as the weights depend on the metric and on the use case. In practice, they address the following questions: when a user performs a given operation (e.g. trying to connect to a site, performing a query, downloading products) what are the values that he/she considers satisfying? At what point he/she would drop the activity and feel frustrated?

The answer to these questions is not straightforward and is potentially prone to a wide range of interpretations. For many sectors and functionalities (e.g. news reading, e-commerce, music streaming…), typical values for the performances zones have been defined through dedicated activities and surveys. However, they cannot be straightforwardly applied to the Sentinels benchmarking. The exercise is not trivial: on one hand, one wants to use "absolute" values that provide an absolute reference. On the other side, in order to be representative and useful, they should ensure a reasonable representation of the Sentinels data access "ecosystem". For instance, they should be set in a way that they allow a meaningful scale of values avoiding saturation (i.e. all target sites considered satisfying/unsatisfying).

The initial values have therefore been set starting from the literature but they were successively refined through dedicated sensitivity tests over the preliminary results obtained for the Copernicus Sentinels Hubs. Frustration threshold values have been set that capture, roughly, the 95th percentile of the benchmarking results.

The current values are included in the QoE descriptions in Section 3. However, they are expected to be fine-tuned after possible users feedbacks (e.g. in case the values are found unrealistic or not representative of the target user community), the availability of more robust results from an extended period of operations, the inclusion of more test sites and, especially, the inclusion of more target sites. The following procedure is therefore followed to calibrate the parameters (thresholds and weights) for a given QoE during the ramp up phase, following the addition of new test sites or target sites:

- Step 1: Calculate the QoEs (2, 3, 4 or 5) for all *test-to-target* connections - including the new ones - using the current values for a sufficiently significant period (i.e. at least one week TBD)
- Step 2: Compare the QoEs: do they make sense? Might there be any saturation effect?
- Step 3: Evaluate readjusting one or both the threshold values by inspecting the basic metrics statistics for the most representative sites. Can a potentially more appropriate value be identified?
- Step 4: If YES, implement new value and go back to step 1.
- Step 5: If NO, evaluate readjusting the weights and go back to step 1

It is worth highlighting that the choice of these parameters touches the core of the user quality of experience. This is made more complex by the fact that the benchmarking service, likewise the ESA Sentinels dissemination systems, does not target a specific user category.
Scientists perform climate change analyses using historical data over long-term time series, forecasters download freshly published data as they are available, civil protection officers look for the most recent images in the aftermath of a disaster. Other categories explore data with no particular time constraint: farmers, public authorities, start-ups, policy makers, ONGs, media, students or even amateurs.
All of the above categories can be considered Copernicus Sentinels users. However, they have very different needs and, very likely, report very different "perceptions" or "user experiences". In addition, the specific situation of the single users could play a role, including their own HW/SW situation (or their perception of it), their psychological profiles, situational factors such as e.g. the urgency of the data need (e.g. for emergency or for finalising a given simulation) and, most important of all, **the availability of alternative (and better!) portals and data sources at hand.**

2.4.3 The test sites network calibration

The quality and the stability of the test sites network is a critical component of the quality of the benchmarks. Constant monitoring is therefore necessary in order to ensure the intrinsic stability of the

Funded by the EU

measurements and that oscillations that are measured over time are not due to variations of the test sites but only reflect performance variations of the target sites. The exercise is made complex by the fact that performances are also affected by additional intrinsic factors such as the geographical distance from the target sites' connectivity along the line. The objective is twofold: on the one side, to guarantee an absolute (as much as feasible) calibration of the Test Sites network that is independent from the target sites to be benchmarked and, on the other side, to qualify the specific connection over which the target site is being benchmarked.

In order to accomplish this, a standard off-the-shelf service (speedtest) is used that measures the hardware and (almost) local network capabilities of a client server through its connectivity with respect to nearby high performance target servers.

This calibration process is used to confirm the reliability and consistency of the test sites network and especially of the ones that are chosen as the European Reference Set. This is achieved by computing a Calibration Quality Indicator associated to each test site (see details in ANNEX B - The Worldwide Test Sites Network). Whenever a test site significantly deviates from mean optimal values, the following actions are triggered:

- Inspect the machine for issues related to the calibration test suite, or contact the provider for assistance
- Fix issues (if any) and restart the calibration test suite
- If the results are still poor, discard the site from the evaluation of the results
- If the results are poor over repeated periods, consider the replacement of the test site.

The test sites calibration is executed at regular intervals alongside the benchmarking operations. Calibration test campaigns are run every month to gather reliable statistics. If these campaigns trigger the change of reference test sites this is explicitly reported in the specific reports.

2.5 SERVICE EVOLUTION

The current service scope is defined to cover the core metrics related to the basic data dissemination operations. This originally included automated "flat" user scenarios and the completion of the test sites network (V1), but was later expanded to include a set of application-specific scenarios that allow to benchmark the performances of the target sites from a tailored, end-to-end standpoint (V2).

During the service set-up, much focus is put on the calibration and validation of the benchmarking parameters (e.g. configuration parameters, thresholds and weights). This is enriched through time by gaining more information from a diverse set of test sites and of target sites as well as through exchange with core stakeholders and experts via ESA.

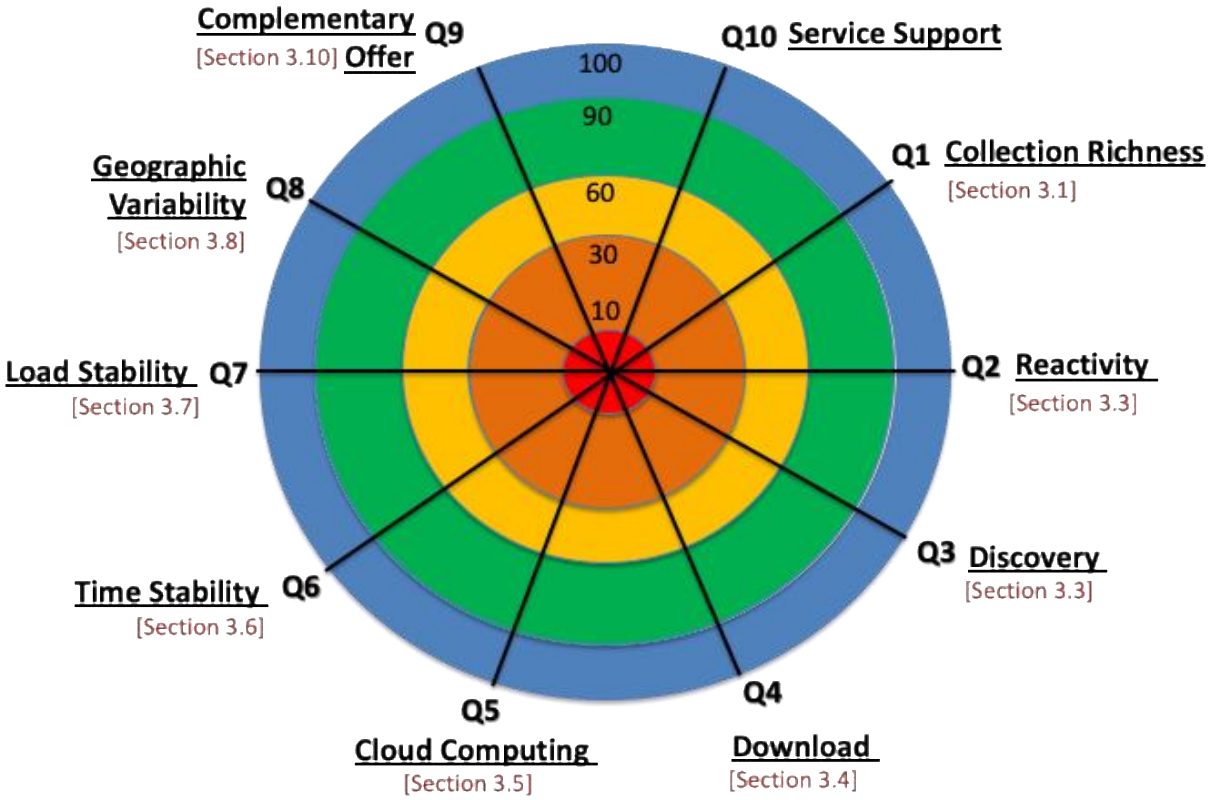
The benchmarking service is open to evolutions in case additional needs are spotted e.g. through coordination loops with core stakeholders and/or by ESA. Possible evolutions include:

- new user test sites, to ensure a wider representation of the user ecosystem and/or to satisfy specific needs from ESA in specific countries or with specific providers;
- new third party target sites, following e.g. third parties which might be interested to be "benchmarked" according to the established framework;
- new or adjusted user scenarios, whereby new functionalities could be introduced following user interests or emerging trends and technological developments.

Funded by the EU
Copernicus esa European Space Agency

3 THE QUALITY-OF-EXPERIENCE INDICATORS

In the following sub-sections, the QoE criteria are described in details, together with the approach used to compute them.



3.1 Q1: RICHNESS OF COPERNICUS DATA COLLECTIONS

Question: How large is the Copernicus collection made available by the target site? How does it compare with respect to its reference data source (e.g. OpenHub)?

This question addresses the richness of given Copernicus data collections (taken as the "Q1 reference collections") offered by a given target site. This is assessed not only in terms of the data products but also in terms of their metadata as well as their latency of availability.

This indicator is computed through comparisons with the most complete collections of the same type available at a pre-defined source, which is taken as the "Q1 reference site". This site drives the composition of the reference collections, which might evolve dynamically with the reference quality. This means that, if the Q1 reference site unilaterally deletes "obsolete" products and/or replace them following e.g. reprocessing campaigns, only the up-to-date products would be considered relevant for the Q1 reference collections. The OpenHub maintains an API advertising the deleted products[2]. As an example, in April 2018 the OpenHub started a campaign to repackage Sentinel-2 products into single-tile ones[3]: the obsolete "multi-tile" products might still be made available from a target site however they would not enter in the computation of the Q1. In this sense, the *Q1 indicator pursues a measurement of the alignment of the target site with respect to the reference source*.

[2] https://scihub.copernicus.eu/twiki/do/view/SciHubUserGuide/ODataAPI"
[3] https://scihub.copernicus.eu/news/News00310

Funded by the EU
Copernicus esa European Space Agency

Currently, the OpenHub is taken as the Q1 reference site[4] and the Q1 reference collections are described in the Table below. In the future, the analysis will be extended to other Copernicus products and information (e.g. the Copernicus Services products and the Sentinel-3 L2 Marine data) and the corresponding data sources will be taken as reference sites for these collections.

Mission	Product Types (always NTC unless stated otherwise)	Geographic Coverage	Time coverage	Size of collection (at CDR)	Q1 Reference Site
Sentinel-1	• Raw • L1 SLC • L1 GRD • L2 OCN	Global	since start of mission	~5,95 Millions	**OpenHub**
Sentinel-2	• L1C • L2A (single-tile only)	Global	since start of mission [L2A are available only since March 2017 onwards for EEA39 and December 2018 for the world]	~23,268 Millions	**OpenHub**
Sentinel-3 Land	• OLCI L1 • SLSTR L1 • STM L2 Land • OLCI L2 Land • SLSTR L2 Land • SYN L2	Global	since start of mission	~4,071 Millions	**OpenHub**
Sentinel-5P	• TROPOMI L2 O3 • TROPOMI L2 NO2 • TROPOMI L2 HCHO • TROPOMI L2 CO • TROPOMI L2 CH4 • TROPOMI L2 SO2 • TROPOMI L2 Aerosol Index • TROPOMI L2 Cloud	Global	since start of mission	~313.000	**OpenHub (currently S5P preOps)**

Table 6: Current Q1 Reference Collections. Note that only Non-Time-Critical (NTC) consolidated products (delivered on average in 24 hours after sensing) are considered here.

The Q1 indicator also checks the reliability of the Copernicus data offer that is publicly advertised on the target site's web portal. As introduced in Sect. 1.2, many sites do not provide full Copernicus data collections but rather offer targeted limited collections. These are generally defined as subsets of the Q1 reference collections (e.g. only Sentinel-2 L1C data products, only over Europe, only over Belgium, only the last 12 months etc etc). The consistency of the actually hosted data collections with respect to the public data offer is measured as part of the Q1 indicator.

[4]ESA's OpenHub offers the most complete open collection of Copernicus Sentinels data. Sentinel products are made systematically available on this portal after downlink and processing from the missions ground segments. Other sites (e.g. CollHub) provide quicker access but limited to restricted communities. The OpenHub is therefore taken as a convenient publicly available reference source for Non Time Critical data products.

Funded by the EU
Copernicus esa European Space Agency

It is important to notice that, with over 30Mi of products, a product-by-product comparison is not feasible and the metrics are computed by sending remote requests to the catalogues. The capability to selectively define the scope of the requests filtering out spurious products is crucial to ensure a fair and meaningful comparison. However this crucially relies on the availability of appropriate filters in the target site's APIs implementation.

The following parameters are taken into account for the computation of the Q1 indicator:

- The *data coverage* (%), providing a measure of the completeness of the Q1 reference data collections hosted on a given target site. This is computed as the ratio between the number of "Q1 reference collections" relevant data products[5] that are **available at the target site[6]** and the corresponding number at the reference site;
- The *catalogue coverage* (%), providing a measure of the completeness of the Q1 reference collections indexed in the catalogue of the target site. This is computed as the ratio between the number of the "Q1 reference collections" relevant items in the catalogue of the target site and corresponding number at the reference site.
- The *data availability latency (hours)*, providing a measure of how quickly the "Q1 reference collections" relevant data are available at the target site with respect to the reference site. This is computed as the average latency of the availability as compared to the date of publication on the reference site; and
- The *consistency with the public data offer* (%), which is a measure of the extent to which the public advertisement on "Q1 data collections" made by the target site is actually honoured. This is measured as the dataCoverage parameter above but limited to the subsets of the Q1 reference collections that are equivalent to the ones of the target site's public data offer (i.e. same timespan, same geographic coverage and same product types).

The calculations are performed at the level of the single product type and then averaged at mission level. The final scores are computed as the averaged of the values obtained from the single missions. This allows to remove any bias caused by the non-uniform distribution of collections over different mission (e.g. Sentinel-3 hosts many more collections that Sentinel-2) and ensures that each mission is tested individually and contributes equally to the final result. Sometimes, coverages over 100% for specific collections are returned by the benchmark tests. This could be caused by extra products hosted by the provider outside the scope of the reference Open Hub or by obsolete products not yet deleted from the catalogue of the target site. It is not always possible to filter away these products and this can artificially inflate the Q1 computation; to prevent this, the values of the metrics dataCoverage, catalogueCoverage and dataOfferConsistency are always capped to 100%.

To obtain a single score for the **Copernicus Collections Richness (Q1)** the above parameters are combined as follows:

$$\mathbf{Q1} = \mathbf{wcc1} * (\#\mathbf{M015}) + \mathbf{wcc2} * (\#\mathbf{M023}) + \mathbf{wcc3} * (\mathbf{APDEXQ1M013}) + \mathbf{wcc4} * (\mathbf{APDEXQ1M024})$$

Which is computed by using the elementary metrics, the tolerance zone parameters and the weights contained in Table 7.

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weights
#M023 dataCoverage	Not Applicable	Not Applicable	0.50
#M015 catalogueCoverage	Not Applicable	Not Applicable	0.15
#M013 avgDataAvailabilityLatency	Less than 6 hours	More than 24 hours	0.15

[5] When comparing data collections with the reference site only relevant data products are considered. This implies that items not relevant for the comparison (e.g. NRT, obsolete products that have been removed from the OpenHub, etc.) must be filtered out.

[6] Data products that are available locally at the target site or anyway under the full control of the target site provider i.e. with no need for downloading from third party portals or entities.

Funded by the EU
Copernicus esa European Space Agency

#M024 dataOfferConsistency	More than 99%	Less than 90%	0.20

Table 7: Index computation parameter used for the Copernicus Collections Richness QoE. The elementary metrics are described in Annex E.

The **Copernicus Collections Richness (Q1)** score ranges between 0 and 100. A value of 100 would correspond to complete Q1 reference collections (fully aligned with the ones hosted on the reference site), with products always being published within 6 hours from their publication on the reference site. Lower values might indicate longer latencies or focused collections not fully aligned with the Q1 reference ones. It is important to understand that this is not necessarily an indication of poor performances: for instance, target sites implementing rolling policies (e.g. CollHub) or offering national coverage (e.g. ISPPortugal) would fall in these categories. Lower Q1 values could also be due to a poor alignment with respect to the publicly declared offer.

According to this indicator, target sites are classified as follows:

Collections Richness Classes				
A	**B**	**C**	**D**	**E**
90< x	60< x ≤90	40< x ≤60	10< x ≤40	0< x ≤10
The site hosts a remarkably complete Copernicus collection, with minimum delay with respect to the products availability. This is in line with what it publicly advertises.	The site hosts a remarkable fraction of the Q1 reference collections, with good latency and alignment wrt its public data offer	The site hosts a fraction of the Q1 reference collection, or has some noticeable latency wrt its publication on the reference site	The site hosts a small fraction of the Q1 reference collections	The site offers a minimum fraction of the overall Q1 reference collection. This might indicate a very focused collection (e.g. only a given product type, only one single country, or a strict rolling policy).

3.2 Q2: SITE REACTIVITY

Question: How reactive is the Sentinels data access portal in responding to user requests?

This question relates to the overall responsiveness of the portal: when a user tries to connect, how quickly does the site respond? How frequently it is found unavailable (apart from known downtimes)? These performances can go unperceived if they keep within reasonable limits, but become extremely disturbing when low levels are hit.

The target site general reachability is evaluated by performing multiple concurrent remote HTTP web requests to the front endpoint of the target site that can be executed multiple times per day (see TC#101). The following parameters are thereafter measured:

- The time needed for the site to respond to the user request (both average and peak times are taken into account); and
- The associated error rates (i.e. when the site is not responding at all).

The QoE **Reactivity (Q2)** indicator is then computed by processing the above metrics according to the following formula:

$$\mathbf{Q2} = \mathbf{wr1} * (\mathbf{APDEXQ2M001}) + \mathbf{wr2} * (\mathbf{APDEXQ2M002}) + \mathbf{wr3} * (\mathbf{APDEXQ2M003})$$

Funded by the EU

Where the APDEXES are computed using the results obtained from Europe's Reference Test Sites during the reporting period and using the elementary metrics, the tolerance zone parameters and the weights contained in Table 8.

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weights
#M001 Average Response Time	Less than 0.1 s	More than 2.0 s	0.4
#M002 Peak Response Time	Less than 1.0 s	More than 4.0 s	0.2
#M003 Error Rate	Less than 1%	More than 10%	0.4

Table 8: APDEX computation parameters used for the Reactivity QoE. The elementary metrics are described in Annex E.

The value of the QoE indicator drives the classification of the target site, which is established according to the scheme represented in the figure below.

Site Reactivity Classes				
A	**B**	**C**	**D**	**E**
90< *QoE* ≤100	60< *QoE* ≤90	40< *QoE* ≤60	10< *QoE* ≤40	0< *QoE* ≤20
The site is extremely reactive to the users request, providing very quick responses and negligible error rates	The site has very good reactivity and very low error rates	The site has good reactivity and low error rates	The site has low reactivity and high error rates	The site has very low reactivity to the users requests. It presents large response times that may be coupled with unacceptable error rates

3.3 Q3: COLLECTIONS DISCOVERABILITY

Question: How efficiently can users find (and visualize) the desired Sentinels products?

The question is of utmost importance because the result of a query can strongly impact the user experience and in some cases drive the overall judgment on the site, e.g. when stringent timeliness requirements are at stake.

Discoverability depends on different parameters related e.g. to the degree of expertise of the users (i.e. how far the user has a precise knowledge about the searched item? Is it performing an open query or a very constraining one?), the search technique (i.e. is the search performed automatically via script or are we talking about a user navigating through the database looking for a specific acquisition?). The size of the searchable database and of the searched items can also play a role, with performances being much quicker for a reduced dataset.

The site performances in this respect are evaluated through multiple concurrent remote HTTP web requests to the front (opensearch) API and GUI of the target site using different queries, from basic ones (e.g. by mission, by product type…) to more complex ones (e.g. by geographic locations, time intervals, ingestion date…) (see TCs# 201, 202, 203 and 204, 209). Queries are executed with random combinations of options over pre-defined databases of e.g. geographic areas, product types and time intervals. Details can be found in Annex C to this document.

The above operations are executed with different frequencies, from hourly (for the basic queries) to daily (for the complex ones). The following parameters are thereafter measured:

Funded by the EU
Copernicus esa European Space Agency

- The time needed for the site to respond to the user request (both average and peak times are taken into account);
- The error rates (on the server response and on the correctness of results);

The QoE **Discoverability (Q3)** indicator is then computed by processing the above metrics according to the following formula:

$$\mathbf{Q3} = \mathbf{wd1} * (\mathbf{APDEXQ3M001}) + \mathbf{wd2} * (\mathbf{APDEXQ3M002}) + \mathbf{wd3} * (\mathbf{APDEXQ3M003}) + \mathbf{wd4} * (\mathbf{APDEXQ3M012})$$

Where the APDEXES are computed using the results obtained from Europe's Reference Test Sites during the reporting period and using the metrics, the tolerance zone parameters and the weights contained in Table 9.

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weight
#M001 Average Response Time	Less than 2,5 s	More than 10 s	0.3
#M002 Peak Response Time	Less than 5 s	More than 20 s	0.1
#M003 Error Rate	Less than 1%	More than 5%	0.3
#M012 Results Error Rate	Less than 1%	More than 15%	0.3

Table 9: APDEX computation parameters used for the Discoverability QoE indicator. The elementary metrics are described in Annex E.

The value of the QoE indicator drives the classification of the target site, which is established according to the scheme represented in the figure below.

Discoverability Classes				
A	B	C	D	E
90< *QoE* ≤100	60< *QoE* ≤90	40< *QoE* ≤60	10< *QoE* ≤40	0< *QoE* ≤10
The site is extremely quick in finding and visualising the desired data	The site has a very good speed in finding and visualising the desired data	The site has average performances in finding and visualising the desired data	The site is slow in finding and visualising the desired data	The site is extremely slow in finding and visualising the desired data

3.4 Q4: DATA DOWNLOAD

Question: How quickly can users download the identified Sentinels product/s of interest?

This question is core to the data accessibility for any user wishing to use the data on their local infrastructure. We define download times as the elapsed time between the moment the request is submitted on the system and the one when the data is available at the user infrastructure. This quality criterion strongly depends on the network characteristics and on the performances of the user machine (i.e. writing speed on the local disk). Performances related to the target site, on the contrary, depend on the storage situation of the product of interest (e.g. is it available on-line or off-line?). The size of the requested products can also play a role, in that bandwidth saturation could be reached, prompting in some cases the interruption of the service.

The download performances are evaluated through single or multiple concurrent remote HTTP web requests to retrieve several product files from a set of selected URLs, where the URLs for download are

Funded by the EU
Copernicus esa European Space Agency

obtained from the results of the search scenarios (see TC#301). Also bulk downloads are executed, over a pre-defined set of selected URLs (see TCs# 302 and 303). When available, HTTP capability to download chunks or to recover failed downloads are tested to improve the download. Note that, wherever a given target site allows parallel downloads to be performed using a single account, this is fully exploited. Also products that are not available on-line are downloaded (see TC#304). A single download via GUI interface (see TC#309) is executed after the run of a simple search. Details can be found in Annex C to this document.

The above operations are executed on average from hourly (for downloading a single product) to daily (for bulk downloads and off-line downloads). The following parameters are thereafter measured taking into account all the results of the various downloads:

- The time needed for the site to respond to the user request (both average and peak times are taken into account);
- The rate of errors in the downloaded files (e.g. for instance based on checksum)
- The throughput, i.e. the speed at which data are being downloaded, computed as the ratio between the time it takes to download a file and its size; and
- The Q41_score as described below.

The QoE **Download (Q4)** indicator is then computed by processing the above metrics according to the following formula:

$$\mathbf{Q4} = \mathbf{wd1} * (\mathbf{APDEXQ4M001}) + \mathbf{wd2} * (\mathbf{APDEXQ4M002}) + \mathbf{wd3} * (\mathbf{APDEXQ4M003}) + \mathbf{wd4} * (\mathbf{APDEXQ4M004}) + \boldsymbol{wd5} * (\boldsymbol{Q41\_score})$$

Where the APDEXES are computed using the results obtained from Europe's Reference Test Sites during the reporting period and using the metrics, the tolerance zone parameters and the weights contained in Table 10.

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weights
#M001 Average Response Time	Less than 0.5 s	More than 2.0 s	0.10
#M002 Peak Response Time	Less than 1.0 s	More than 4.0 s	0.05
#M003 Error Rate	Less than 1%	More than 10%	0.20
#M005 Throughput	More than 8 MB/s	Less than 1 MB/s	0.50
Q41_score	See below	See below	0.15

Table 10: APDEX computation parameters used for the Download QoE indicator. The elementary metrics are described in Annex E.

The value of the QoE indicator drives the classification of the target site, which is established according to the following scheme:

Download Classes				
A	B	C	D	E
90< *Download* ≤100	60< *Download* ≤90	40< *Download* ≤60	10< *Download* ≤40	0< *Download* ≤10
Excellent download capabilities. Products are downloaded with highest throughput, with no or very limited errors	Very good download capabilities. Products are downloaded with high throughput, with limited errors and limited latency	Fair download capabilities. Products are downloaded with average throughput, limited errors and limited latency	Poor download capabilities. Products are downloaded with low throughput, many errors and/or long latency	Download capabilities are definitely not acceptable

Funded by the EU
opernicus esa European Space Agency

and minimum latency				

The **Q41_score** used for the Q4 Download Indicator is computed using the following metrics:

- The additional #M017 offline availability latency, due to the fact that a catalogued product is not immediately available online
- The data coverage online (%), providing a measure of the significance of the online collections with respect to the total in the catalogue

and according to the following formula:

$$\mathbf{Q41\_score} = \mathbf{APDEXQ41M017} + (\mathbf{1} - \mathbf{APDEXQ41M017}) * \mathbf{M023}^{*}$$

The #M023* is the #M023 dataCoverage metric extracted using test case custom configuration only for online data. The APDEX for M017 is computed using the following satisfaction and tolerance zone parameters:

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone
#M017 offlineDataAvailabilityLatency	Less than 1 hour	More than 1 day

The Q41 is meant to account for the access latency for collections that are not immediately accessible online. For target sites having all data online, it scores 15%. This value is then decreased depending on the amount of the data that are not online and the latency to access them.

3.5 Q5: CLOUD PROCESSING

Question: Where applicable, how efficiently can users access and process Sentinels data in a cloud environment? How costly is this service?

This question is typically referred to the Data Access Information Service (DIAS) platforms, providing cloud environments to exploit Copernicus collections. The use scenario is complex, because it encompasses several functionalities that range from the allocation of VMs (for a dynamic one) to the upload of user data to the processing.

The cloud computing performances must therefore be evaluated through the combination of several user operations. In order to account for different segments of use, both quick, basic scenarios as well as bulk processing must be accounted for. This implies that different categories of virtual machines must be considered from small ones (to account for one-off processes) to large ones (that usually accommodate bulk processing). The calculation of the associated latency must then take into account these differences through specific thresholds. In addition, since these environments are commercially provided, the related costs are also of relevance.

The Test Scenarios providing the metrics used for the Cloud Processing indicator are the TS11 and TS12 (see ANNEX C – Benchmark User Scenarii for the details**).**
Those scenario mainly regard the Cloud Service (Simple and Complex) local data search and download (single and multiple) on provisioned virtual machine (single and multiple).
The Test Cases used are the same of the Test Scenario TS01 and TS02 but in the context of cloud service. This means that the same metrics collected for the building of the Q3 Discoverability and Q4 Download indicators can be used to build two new scores (Q53 and Q54) with the same technique, but different

Funded by the EU

thresholds due to the different context. Moreover plain metrics such as the average provisioning latency is currently taken into account and in the next deliveries also the cost will be taken into account[7].

The effectiveness of user service (both commercial and technical) will also be evaluated and reported in the Specific Reports.

The QoE **Cloud Processing (Q5)** indicator is then computed by processing the above metrics according to the following formula:

$$\mathbf{Q5} = \mathbf{wd1} * (\mathbf{APDEXQ5M101}) + \mathbf{wd2} * (\mathbf{cost\_score}) + \mathbf{wd3} * (\mathbf{Q53}) + \mathbf{wd4} * (\mathbf{Q54})$$

Where the APDEXES are computed using the results obtained from Europe's Reference Test Sites during the reporting period and using the metrics, the tolerance zone parameters and the weights contained in following table:

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weight
#M101 Average Provisioning Latency	Less than 60 s	More than 90 s	0.1
cost_score	See below	See below	0.1
Q53	See below	See below	0.4
Q54	See below	See below	0.4

Table 11: APDEX computation parameters used for the Cloud Processing QoE indicator. The elementary metrics are described in Annex E.

The value of the QoE indicator drives the classification of the target site, which is established according to the scheme represented in the figure below.

Cloud Processing Classes				
A	B	C	D	E
90< *QoE* ≤100	60< *QoE* ≤90	40< *QoE* ≤60	10< *QoE* ≤40	0< *QoE* ≤10
The site is extremely good in discovery and download the desired data from the provisioned VM	The site is good speed in in discovery and download the desired data from the provisioned VM	The site has average performances in in discovery and download the desired data from the provisioned VM	The site is slow in discovery and download the desired data from the provisioned VM	The site is extremely slow in discovery and download the desired data from the provisioned VM

The **cost_score** used for the Q5 Cloud Processing indicator is computed according to the following formulas. The test scenarios TS11, TS12 and TS13 provide info in order to compute the absolute cost (M116 process duration and M115 overall duration) of the VM parallel runs on each TestCase (TC) run on the DIAS cloud or third-party provider according to the published hourly cost of each VM flavour and based on the process duration run:

$$\mathbf{TC_{cost}} = \sum_{VMRun} costHour(flavourName) * durationHour$$

In the reference period under consideration the max(TC_{cost}) and min(TC_{cost}) values greater than zero are computed among all the providers and TC, then for each provider:

[7] Cloud computing services may be provided under subscription or under a pay-per-use formula. The related costs are therefore computed in different ways to take into account these differences.

Funded by the EU
Copernicus esa European Space Agency

$$\mathbf{cost\_score} = \frac{\mathbf{max}(\mathbf{TC_{cost}}) - \mathbf{TC_{cost}}}{\mathbf{max}(\mathbf{TC_{cost}}) - \mathbf{min}(\mathbf{TC_{cost}})}$$

The cost_score value will be a number between 0 and 1, where 1 is for minimum processing cost and 0 for the maximum processing cost paid for TCs.

The sub-scores (**Q53** and **Q54**) used for the main Q5 Cloud Processing indicator are computed according to the following formulas:

$$\mathbf{Q53} = \mathbf{wd1} * (\mathbf{APDEXQ53M001}) + \mathbf{wd2} * (\mathbf{APDEXQ53M002}) + \mathbf{wd3} * (\mathbf{APDEXQ53M003}) + \mathbf{wd4} * (\mathbf{APDEXQ53M012})$$

Where the APDEXES are computed using the results obtained from Europe's Reference Test Sites during the reporting period and using the metrics, the tolerance zone parameters and the weights contained in the following table:

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weight
#M001 Average Response Time	Less than 2,5 s	More than 20 s	0.3
#M002 Peak Response Time	Less than 5 s	More than 20 s	0.1
#M003 Error Rate	Less than 1%	More than 5%	0.3
#M012 Results Error Rate	Less than 1%	More than 15%	0.3

Table 12: APDEX computation parameters used for the Q53 score. The elementary metrics are described in Annex E.

$$\mathbf{Q54} = \mathbf{wd1} * (\mathbf{APDEXQ54M001}) + \mathbf{wd2} * (\mathbf{APDEXQ54M002}) + \mathbf{wd3} * (\mathbf{APDEXQ54M003}) + \mathbf{wd4} * (\mathbf{APDEXQ54M005})$$

Where the APDEXES are computed using the results obtained from Europe's Reference Test Sites during the reporting period and using the metrics, the tolerance zone parameters and the weights contained in the following table:

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weight
#M001 Average Response Time	Less than 0.5 s	More than 2.0 s	0.2
#M002 Peak Response Time	Less than 1.0 s	More than 4.0 s	0.1
#M003 Error Rate	Less than 1%	More than 10%	0.2
#M005 Throughput	More than 16 MB/s	Less than 4 MB/s	0.5

Table 13: APDEX computation parameters used for the Q54 score. The elementary metrics are described in Annex E.

3.6 Q6: SITE STABILITY OVER TIME

Question: How stable are the above measured performances along time?

This aspect of target site quality concerns the variation of the obtained results over time. This concerns both the stability during the reporting period (i.e. hourly/daily/weekly variations within the 6 months under examination) and the variations with respect to previous reporting periods (i.e. long term trends). Although users might appreciate positive variations in the quality, the stability in the performances is a parameter to be taken into account in many operational applications.

Funded by the EU
opernicus esa European Space Agency

The time stability is evaluated by observing the variations of the indicators described in the previous sections along time. The QoE indicators from Q2 to Q5 used for this index are computed each week over a time range of last 7 days. The Q1 is computed each week over a time range of last 14 days. Since the QoE are values ranging from 0 to 100, the maximum expected standard deviation is 50. The formula for the standard deviation of QoE indicators is:

$$\sigma(Q[1..5]) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

Where N is the number of samples, x_i is the value of QoE at given sample time i, and $\bar{x}$ is the average value.

The above parameters are preliminarily considered as independent and the **Time Stability (Q6)** indicator is computed by combining the above metrics according to the following formula:

$$Q6 = 100 * \frac{\left(50 - \sum_i^M \frac{\sigma(Q_i)}{M}\right)}{\left(50 + \sum_i^M \frac{\sigma(Q_i)}{M}\right)}$$

Where M is the number of QoE indices that are taken into account for a given provider (i.e.: 4 for the ESA hubs, 5 for the DIASes).
The **Time Stability (Q6)** indicator assumes values between 0 (corresponding to an extremely variable profile) and 100 (corresponding to a flat profile).

The value of the QoE indicator drives the classification of the target site, which is established according to the following scheme:

Time Stability Classes				
A	**B**	**C**	**D**	**E**
90< *TimeStability* ≤100	60< *TimeStability* ≤90	40< *TimeStability* ≤60	10< *TimeStability* ≤40	0< *TimeStability* ≤10
Excellent stability. QoE indicators do not vary (or have very limited variations) over time.	Very good stability. QoE indicators have limited variations over time.	Fair stability. QoE indicators have variations over time that do not compromise the usability of the target site.	Limited stability. QoE indicators have strong variations over time that make the target site scarcely reliable.	Insufficient stability. QoE indicators have very strong variations over time that make the use of the target site not reliable.

3.7 Q7: SITE STABILITY WITH LOAD

Question: How stable are the above measured performances with respect to the user load?

This quality criterion concerns the stability of target site performances when accesses and requests increase in number. This criterion can be assessed only from the point of view of the marginal impact of

Funded by the EU
Copernicus esa European Space Agency

single (or a set of) test sites, through measuring the response to an increasing number of requests made from those sites.

This is evaluated by observing the target site reaction to an increase in load in the execution of different user operations. Concurrent requests are launched by a pre-defined increasing number of users (typically all test sites launching similar types of requests), according to a Load Factor which acts as a multiplier of the number of requests made to the target site. For instance, if load factor is 10, there will be 10 requests or 10 times a batch of concurrent requests (depending on the maximum parallelism set). This multiplies the stress imposed on each target site and the reaction of the target site is then measured. Details can be found in Annex C to this document.

Of course, there is considerable uncertainty in assigning the variations in the site response to the load variations: the overall number of users connected to a target site is not known a-priori and it might vary during the test execution, potentially impacting the measured performances. To minimise this effect, stress tests are launched at different times during the day.

Stress tests can put excessive load on the target sites and are then tested cautiously and not for all the target sites. Stress test campaigns are organised in collaboration with the target site operator through ESA.

The load stability is evaluated by observing the variations of the QoE indicators related to the reactivity, discoverability, download and cloud processing.

The detailed computation method of Load Stability is under development and will be reported in the next issue of this document. Diagrams representing Quality indicators versus load will be created to show how performances degrade with the number of accesses. From these diagrams some numerical indicators will be extracted, representing Load Stability Scores (SS):

- LSS(Reactivity)
- LSS(Discoverability)
- LSS(Download)
- LSS(CloudProcessing)

The above Load Stability Scores will be combined in a linear combination whose value will range between 0 and 100.

The value of the indicator drives the classification of the target site, which is established according to the following scheme:

Load Stability Classes				
A	**B**	**C**	**D**	**E**
90< *LoadStability* ≤100	60< *LoadStability* ≤90	40< *LoadStability* ≤60	10< *LoadStability* ≤40	0< *LoadStability* ≤10
Excellent stability. QoE indicators do not vary (or have very limited variations) with load.	Very good stability. QoE indicators have limited variations with load.	Fair stability. QoE indicators have variations with load. that do not compromise the usability of the target site.	Limited stability. QoE indicators have strong variations with load that make the target site scarcely reliable.	Insufficient stability. QoE indicators have very strong variations with load that make the target site not usable.

This indicator has not been tested so far and it is expected that significant adjustments may arise following our calibration process.

Funded by the EU
Copernicus esa European Space Agency

3.8 Q8: GEOGRAPHIC VARIABILITY

Question: How variable are the above measured performances with respect to the user location in Europe and around the world?

Users can experience, on average, very different performances when accessing the same target site, especially from different regions of the world. These imbalances are associated not only to the geography (i.e. or, to the physical distance from the target site) but more appropriately to the "network distance", which is impacted by additional parameters such as the local connectivity (which also depends on the network provider) and the characteristics of the HW infrastructure at the test site (i.e. the writing speed on the local disks).

This criterion is therefore assessed separately from the core benchmark parameters described in the previous Sections (i.e. Q1 to Q7) that are instead computed using a restricted set of reference test sites of known and guaranteed performances.

The geographic variability or, more appropriately, the "user infrastructure variability" is assessed via performing the user operations from the different test sites of the network and by computing the QoE indicators over each connection. A prospect of the results obtained for the different test sites is generated and the QoEs can then easily be compared.

It is worth noticing that, in addition to the QoE values, a score is also generated for each connection that represents the theoretical "network distance" between the test and the target site. This score can be used as an estimate of the specific user site performances with respect to the target site and it indicates, in other words, **its "intrinsic" potential capability to fully exploit the functionalities of the target site**. This score is regularly computed as part of the benchmarking calibration according to a procedure which is detailed in Annex B to this document.

3.9 Q9: COMPLEMENTARY OFFER

Question: How rich and attractive is the global offer from the provider in terms of e.g. additional data collections (i.e. non-Copernicus), SW tools and cloud service features?

In addition to the criteria described above, the availability of complementary features offered by a target site concur to the overall user perception and in some cases even drive the preferred choice (especially for newcomers). For example, the availability of additional data collections[8] can be of interest to many users although their relevance and significance largely depends on the actual users and on the actual collections at stake. Similarly, the free availability of additional processing and development tools and libraries[9] can be appealing and facilitate the exploitation of the Copernicus data products: when integrated within the cloud environment, these tools constitute a real asset and even a game-changer for the users, depending on the available features and the type of scenario of interest. For back-end services development in general, and for the benchmarking service in particular, the availability of powerful Application Programming Interfaces is fundamental. The work of the developer is made much easier by e.g. the compliance of the exposed interfaces with respect to widely used EO standards (e.g. OGC, OpenSearch, Odata) and the availability of EO-relevant filters therein. The availability of smart programming features (e.g. Kubernetes, JupyterHub, Jupyter Notebook, ontology-based vocabularies…) can also be of interest.

The evaluation of the overall offer relies on manual inspection, with a partial exception for the API completeness which is based on the experience gathered during automated tests for the Q1-Q5 indicators; the availability of pre-defined features of interest is systematically noted and rated according to a pre-

[8] In addition to the Q1 reference ones defined in Sub-Section 3.1 such as e.g. other free EO data collections such as Landsat, Envisat, Cryosat but also commercial Very High Resolution data and Digital Elevation Models.

[9] E.g. the SNAP toolbox, the Sen2Cor, the https://gdal.org , QGIS, Jupiter Hub, PostGIS…

Funded by the EU
Copernicus esa European Space Agency

defined scheme. This is essential to represent the offers on comparable footing, given the intrinsic subjectivity in the evaluation of heterogeneous features that widely vary with the provider and its main target audience. For this reason, no user satisfaction thresholds have been defined because the metrics are directly computed as a completion percentage over the maximum possible value.

This is the first release of the Q9 indicator, so the metrics and their schemes may be subject to review and improvements in the next benchmarking periods.

The specific metrics to be used in this quality indicator and the related APDEX computation parameters are listed in Table 14, and have been defined starting from the following prototypes:

- Availability of Copernicus Services Data (CAMS, CMEMS, etc.)
- Availability of additional non-Copernicus data collections (only EO data)
- Availability and completeness of APIs (including e.g. filtering capabilities)
- Availability of cloud services
- Availability of other services in the provider offering
- Availability of smart programming features
- Availability of data exploitation tools and SW integrated in the Cloud Services

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weights
#M041 copernicusCoreServices	Not Applicable	Not Applicable	0.10
#M042 otherData	Not Applicable	Not Applicable	0.10
#M043 apiCompleteness	Not Applicable	Not Applicable	0.15
#M044 cloudServices	Not Applicable	Not Applicable	0.20
#M045 otherServices	Not Applicable	Not Applicable	0.10
#M046 devTools	Not Applicable	Not Applicable	0.10
#M047 devServices	Not Applicable	Not Applicable	0.10
#M048 eoTools	Not Applicable	Not Applicable	0.15

Table 14: APDEX computation parameters used for the Complementary Offer QoE indicator. The elementary metrics are described in Annex E.

The QoE **Complementary Offer (Q9)** indicator is then computed by processing the above metrics according to the following formula:

$$\mathbf{Q9} = \mathbf{w1} * (\mathbf{Q9M041}) + \mathbf{w2} * (\mathbf{Q9M042}) + \mathbf{w3} * (\mathbf{Q9M043}) + \mathbf{w4} * (\mathbf{Q9M044}) + \mathbf{w5} * (\mathbf{Q9M045}) + \mathbf{w6} * (\mathbf{Q9M046}) + \mathbf{w7} * (\mathbf{Q9M047}) + \mathbf{w8} * (\mathbf{Q9M048})$$

The value of the QoE indicator drives the classification of the target site, which is established according to the following scheme:

Complementary Offer QoE Classes				
A	**B**	**C**	**D**	**E**
90< *QoE* ≤100	60< *QoE* ≤90	40< *QoE* ≤60	10< *D QoE* ≤40	0< *QoE* ≤10
Excellent complementary offer, full featured.	Very good complementary offer.	Fair complementary offer providing most common features.	Poor complementary offer providing just some features.	No complementary offer is present. Very few features are provided, or none at all.

3.10 Q10: SERVICE SUPPORT

Funded by the EU
Copernicus esa European Space Agency

Question: What is the level of support provided to allow an optimal user exploitation of the advertised services?

An appealing service needs to be exploitable: even potentially highly performing data distribution services can be frustrating for the users in case they are too cumbersome to be used and do not offer an adequate support. The user support should not be intended as limited to a responsive human help desk (e.g. replies to phone calls, emails or ticketing service) but may be including other concurring aspects such as e.g. a news service promptly reporting about service improvements and anomalies (e.g. interruption of service, change of format notifications…) and the availability of relevant Frequently Asked Questions databases. The availability of well-documented service specifications should be included as well as e.g. the ease of selection and provisioning of computing resources, together with the availability of user monitoring tools. Also of relevance may be the transparency and the availability of tools to monitor the user's financial information (e.g. activity reports, dashboards, billings).

The services support is normally leveraged when executing the benchmarking activities: the services support material is consulted and the help desk triggered whenever needed (under anonymity as much as possible). The experienced quality is noted and scored with respect to some pre-defined criteria that allow streamlined comparisons across different providers. This implies, however, that no extensive testing is performed (e.g. boots automatically and regularly polling opening tickets) and this constitutes an intrinsic limitation of the statistical significance of this indicator.

The metrics used in this quality indicator are generated through manual procedures (see section 8.4 procedures #TP01, #TP02, #TP03, #TP04 and #TP14).

Values taken during the first subscription phase of the project have been merged in the sub-indicator **Q101_serviceSubscription** and address the following metrics previously known as the whole Q10:

- User enrollment methods (e.g. effectiveness of registration procedure, effectiveness of payment schemes)
- Availability and Completeness of documentation for the registration
- Availability and Completeness of tutorials and training materials
- Registration/Subscription Efficiency

The sub-indicator **Q101_serviceSubscription** is computed by processing the above metrics according to the following formula:

$$\mathbf{Q101\_serviceSubscription} = \mathbf{wss1} * (\mathbf{APDEXQ10M201}) + \mathbf{wss2} * (\mathbf{Q10M202}) + \mathbf{wss3} * (\mathbf{APDEXQ10M203}) + \mathbf{wss4} * (\mathbf{APDEXQ10M204}) + \mathbf{wss5} * (\mathbf{APDEXQ10M205}) + \mathbf{wss6} * (\mathbf{APDEXQ10M206}) + \mathbf{wss7} * (\mathbf{APDEXQ10M207}) + \mathbf{wss8} * (\mathbf{APDEXQ10M208})$$

Where the APDEXES computation parameters are listed in Table 15: APDEX computation parameters used for the Q101_serviceSubscription indicator. The elementary metrics are described in Annex E.

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weights
#M201 registrationTimeSpan	Less than 60 minutes	More than 1440 minutes	0.10
#M202 subscriptionDocCompleteness	Not Applicable	Not Applicable	0.20
#M203 subscriptionSupportAnswerTime	Less than 60 minutes	More than 1440 minutes	0.10
#M204 subscriptionSupportThreadLength	Less than 3	More than 6	0.10
#M205 subscriptionSupportAppreciation	More than 4	Less than 2	0.10

Funded by the EU
opernicus esa European Space Agency

#M206 subscriptionTimeSpan	Less than 4320 minutes	More than 14400 minutes	0.10
#M207 subscriptionPaymentMethodPrioNumber	Less than 2	More than 4	0.20
#M208 subscriptionFailures	Less than 1	More than 3	0.10

Table 15: APDEX computation parameters used for the Q101_serviceSubscription indicator. The elementary metrics are described in Annex E.

The Q101_serviceSubscription score with the other following metrics creates the new revised Q10 indicator:

- Completeness of the service documentation
- Completeness of HelpDesk channels
- Service desk reactivity
- Number of replies in a thread
- Answer general satisfaction about proposed solution

The QoE **Service Support (Q10)** indicator is then computed by processing the above metrics according to the following formula:

$$\mathbf{Q10} = \mathbf{wss9} * (\mathbf{Q101\_serviceSubscription}) + \mathbf{wss10} * (\mathbf{Q10M230}) + \mathbf{wss11} * (\mathbf{Q10M231}) + \mathbf{wss12} * (\mathbf{APDEXQ10M232}) + \mathbf{wss13} * (\mathbf{APDEXQ10M233}) + \mathbf{wss14} * (\mathbf{APDEXQ10M234})$$

Where the APDEXES computation parameters are listed in Table 15: APDEX computation parameters used for the Q101_serviceSubscription indicator. The elementary metrics are described in Annex E.

Elementary Metrics of Relevance	User Satisfaction Zone	User Frustration Zone	Weights
Q101_serviceSubscription	See above	See above	0.20
#M230 serviceDocCompleteness	Not Applicable	Not Applicable	0.25
#M231 serviceSupportHelpDesk	Not Applicable	Not Applicable	0.25
#M232 serviceSupportAnswerTime	Less than 240 minutes	More than 960 minutes	0.10
#M233 serviceSupportThreadLength	Less than 3	More than 6	0.10
#M234 serviceSupportAnswerSatisfaction	More than 4	Less than 2	0.10

Table 16: APDEX computation parameters used for the Service Support QoE indicator. The elementary metrics are described in Annex E.

The value of the QoE indicator drives the classification of the target site, which is established according to the following scheme:

Service Support Classes				
A	**B**	**C**	**D**	**E**
90< *ServiceSupport* ≤100	60< *ServiceSupport* ≤90	40< *ServiceSupport* ≤60	10< *D ServiceSupport* ≤40	0< *ServiceSupport* ≤10
Excellent service support capabilities.	Very good service support capabilities.	Fair service support capabilities.	Poor service support capabilities.	Service support capabilities are definitely not acceptable.

Funded by the EU
Copernicus esa European Space Agency

4 THE END-TO-END SUITABILITY INDICATORS

In the following sub-sections, the End-to-End suitability indicators (shortened as E#) are described in detail for each scenario, together with the approach used to compute them.

For each scenario, the scene is set in terms of:

- the "typical user profile",
- the Sentinel data set of relevance and
- the value-adding processing to be performed on the data
- the scenario expected outcome.

These parameters will drive the definition of the criteria that will be used to compute a final end-to-end parameter (e.g. satisfaction and tolerance values, weights assigned to the different steps).

4.1 E1: NDVI MAPPING

Question: How suitable is the platform for a user wishing to generate NDVI maps from Sentinel-2 scenes?

This scenario concerns the evaluation of the suitability of the target site for a user wishing to compute value-added products relevant to monitor the status of vegetation e.g. agriculture and forestry. The scenario involves very simple processing or manipulation of same-product parts implying e.g. extraction of features from a given product and computation of a generic index (e.g. bands manipulation).

Emphasis is on the capability for an entry user to relocate a simple processing from the small program on the scientist laptop requiring long data downloads, to a cloud processing on demand. Data loads are not heavy sized (e.g. few tens of products) but would requires few hours for the user to get on a local computer. Time constraints are relaxed. The availability of pre-installed tools such as e.g. GDAL and numpy is considered an asset that greatly facilitate the work of the user

Setting the scene	
User Profile	The user is an environmental scientist with a basic computer science training. The User's programming language and Tools ability is limited computer but very sufficient for the use case. It can program in python using some Earth observation data manipulation libraries such as GDAL and numpy. It is not familiar with any other language or dedicated tool.
Value-adding processing	The user wishes to compute some normalized difference vegetation index (NDVI)[10] maps using a simple SW program in python.
Sentinel Data Set	The dataset of relevance for this scenario is limited to a few tiles from Sentinel-2 L1C data. The geographic location is variable over continental land (e.g. France). The time span covers from 1 day to 1 week and is limited to fresh data (i.e. no off-line data is involved). Filters to search for Sentinel2 Level 1 over a specific AOI for a given timespan over land area excluding too high cloud coverage
User's expected outputs	The user may wish to download the processed NDVI data but having the possibility to visualize and share the results on a map directly without downloading the product e.g. through results exploitation tools such as - Access to private storage area for saving (and sharing) the generated results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.)

[10] NDVI provides a measure of healthy vegetation and ranges in value from -1 to 1. Values closer to 1 represent healthy, green vegetation. NDVI can be calculated from Sentinel-2 data using band 4 (red) and band 8 (near-infrared). This is taken into account in Test scenario #15 of the cdab benchmark suite with the parameters relative to NDVI processing (Processing Scenario [ps] in cdab test suite)

Funded by the EU
Copernicus esa European Space Agency

	- Access (via API) to a catalogue index allowing publication of basic metadata about the generated results and allow later discovery and sharing (e.g. a OGC OpenSearch Catalogue) - Access to a OGC WMS (e.g. GeoServer) where to publish the generated results for visualization and sharing - Access to GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.) is considered an asset.

Table 17: E1 Scenario Customization

Variable	Value	Comment	Used in
Scenario Repository	https://github.com/esa-cdab/cdab-testsuite/tree/master/Use%20Cases/Scenario%201%20-%20NDVI%20Mapping	Github repo with all the resource for the scenario	All steps
Payment methods	#1 paypal, #2 credit card	user wants to use exclusively it's credit card to pay for it's account	Step #1
User's programming language and Tools ability	python (0.8), gdal (0.2)	Limited computer skills but very sufficient for the use case	
User's profile description	An environmental scientist that would like to integrate an NDVI program in python using some earth observation data, typically Sentinel-2. I need to use software like GDAL and python.	Description to be used in the exchange for describing the user	
Development Environment installation procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository)	Simple installation steps to have GDAL and python	Step #2
Integration procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository)	Some integration steps to reproduce	
Application build procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository)	recipe to build the user application	
Use Case Data Collection	Sentinel-2 L1C	Sentinel 2 Level 1 to process the NDVI	Step #1 Step #2 Step #3
Useful Data Access Filter	Mission, product type or level or collection, geographical AOI, sensing time span, Cloud Coverage, Land Coverage	Filters to search for Sentinel2 Level 1 over a specific AOI for a given timespan over land area excluding too high cloud coverage	Step #3
Processing Scenario	NDVI	Processing Scenario to execute	Step #4
Data visualization tools	WMS, geobrowser	Having the possibility to visualize the results on map directly without downloading the product is a bonus	Step #5

Table 18: E1 Scenario Procedures Variables

The E1 score is computed based on the weighted combination of scores, apdex or metrics listed in the following Table. Satisfaction and frustration zones are computed based on the scenario customization described above.

Funded by the EU
Copernicus esa European Space Agency

E1 term	Elementary Metrics of Relevance	Type	User Satisfaction Zone	User Frustration Zone	Weight
E11	Q1_score	V	See 3.1	See 3.1	1.00
E12	#M209 integrationToolsAdequacy	V			0.15
E12	#M210 integrationEnvSetupTimespan	A	60	1440	0.15
E12	#M211 integrationCompleteness	V			0.15
E12	#M213 integrationSupportAnswerTime	A	60	1440	0.05
E12	#M214 integrationSupportThreadLength	A	2	5	0.05
E12	#M215 integrationSupportAppreciation	A	4	2	0.05
E12	#M216 storageSetupTimeSpan	A	10	120	0.05
E12	#M217 storageCostPerGB	A	0.5	2	0.05
E12	#M219 appStorageAvailability	A	4	2	0.05
E12	#M220 integrationCost	A	5	20	0.05
E12	T701_score	V			0.10
T701	#M001 avgResponseTime	A	500	2000	0.30
T701	#M005 throughput	A	16000000	4000000	0.70
E12	T702_score	V			0.10
T702	#M001 avgResponseTime	A	500	2000	0.30
T702	#M005 throughput	A	16000000	4000000	0.70
E13	#M301 dataQueryFiltersAdequacy	V			0.40
E13	Q51_score	V			0.20
Q51	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q51	#M110 cost	A	0.5	2	0.50
E13	Q53_score	V	See 3.5	See 3.5	0.20
E13	Q54_score	V	See 3.5	See 3.5	0.20
E14	Q52_score	V			1.00
Q52	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q52	#M110 cost	A	0.5	2	0.50
E15	#M501 dataExploitationToolsAdequacy	V			0.60
E15	Q3_score	V	See 3.3	See 3.3	0.20
E15	Q4_score	V	See 3.4	See 3.4	0.20

Table 19: Computation parameters used for the E1 indicator. The Q scores and metrics are described in the Section 3. The Type column (V=value, A=apdex) define the applicability of the User Satisfaction/Frustration zone values, otherwise the value of the metric or score is used. To be noted that the sum of weights for a given step should be 1.

The value of the E2E indicator drives the classification of the target site, which is established according to the following scheme:

E1 Indicator Classes				
A	B	C	D	E
$90<E\leq100$	$60<E\leq90$	$40<E\leq60$	$10<E\leq40$	$0<E\leq10$

Funded by the EU
Copernicus esa European Space Agency

The platform offers overall excellent conditions to perform the NDVI mapping	Very good	Fair	Poor	The platform does not allow the NDVI scenario.

4.2 E2: RAPID MAPPING

Question: How suitable is the platform for a user wishing to generate burned area analysis maps from Sentinel-2 scenes?

This scenario concerns the evaluation of the suitability of the target site for a user wishing to compute value-added products relevant to create burned area analysis map and NIR/SVWI RGB composites after a wildfire event using Sentinel-2 MSI L2A products. The scenario involves the creation of a directed acyclic graph (DAG) for SNAP and to call the SNAP gpt utility to process it using a SW program in python from within a Jupyter Notebook.

Focus is on stringent timeliness and the capability of identification of best pre- and post- event images. Data loads are small size (e.g. a few products) and no heavy processing is involved. The availability of an interactive environment for fast step by step verification (such as e.g. the Jupyter Notebook) and quick parametrization is very important. Availability of NRT or STC datasets is obviously a requirement. The capability to access only parts of the data (e.g. single bands) is also important.

Setting the scene	
User Profile	The user is an environmental engineer with intermediate computer science training working for an organization involved in Disasters Risk Management. He can program in Python and has an intermediate knowledge of the SNAP toolbox. He is looking for an interactive development environment allowing to rapidly prototype the service with intermediate checkpoints.
Value-adding processing	The user wishes to calculate the Normalized Burn Ratio (dNBR) and the Relativized Burn Ratio (RBR) between pre-fire and post-fire acquisitions. The Burned Area map is obtained through the intersection between AOI and the pre-fire, post-fire Sentinel-2 chosen as input products. The formula used to estimate burn severity is the Normalized Band Ratio (NBR): $$\mathrm{NBR} = \frac{NIR - SWIR}{2NIT + SWIR}$$ A high NBR value indicates healthy vegetation while a low value indicates bare ground and recently burnt areas. Non-burnt areas are normally attributed to values close to zero. The Burned Area Analysis product output of the service provides the delta Normalized Burn Ratio (dNBR) and the Relativized Burn Ratio (RBR). The dNBR (the difference between pre-fire and post-fire) is useful to identify recently burned areas and differentiate them from bare soil and other non-vegetated areas. The RBR is advantageous when the absolute change between pre-fire and post-fire NBR is small.
Sentinel Data Set	The dataset of relevance for this scenario is limited to Sentinel-2 MSI L2A products over a continental land (e.g. Australia). The time span covers a few days and is limited to fresh data (i.e. no off-line data is involved). Timeliness of availability of latest Sentinel-2 acquisitions is very important. Filters to search for Sentinel-2 MSI

Funded by the EU

	L2A over a specific AOI/geoname for a given timespan over land area with the post-fire image very close in time to the fire date is important.
User's expected outputs	The user may wish to download the generated products but having the possibility to visualize them directly on a map without download (e.g. through WMS, geobrowser), to publish them in a catalogue and to share them in an online storage area to other users is considered an asset. Example of such results exploitation tools are: - Access to private storage area for saving (and sharing) the generated results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.) - Access (via API) to a catalogue index allowing publication of basic metadata about the generated results and allow later discovery and sharing (e.g. a OGC OpenSearch Catalogue) - Access to a OGC WMS (e.g. GeoServer) where to publish the generated results for visualization and sharing - Access to GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.).

Table 20: E2 Scenario Customization

Variable	Value	Comment	Used in
Scenario Repository	https://github.com/esa-cdab/cdab-testsuite/tree/master/Use%20Cases/Scenario%202%20-%20Rapid%20Mapping		All steps
Payment methods	#1 paypal, #2 credit card	User wants to use exclusively their credit card to pay for their account.	Step #1
User's programming language and Tools ability	Python (0.3), SNAP (0.3), OTB (0.1), Jupyter notebook (0.3)	Good computer skills and willing to use an interactive tool to preview integration results.	
User's profile description	The user is an environmental engineer with intermediate computer science training working for an organization involved in Disasters Risk Management. He can program in Python and has an intermediate knowledge of the SNAP toolbox. He is looking for an interactive development environment allowing to rapidly prototype the service with intermediate checkpoints	Description to be used in the exchange for describing the user.	
Development Environment installation procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository)	Installation steps to have a working Jupyter notebook in Python with the SNAP libraries.	Step #2
Integration procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository)	Some integration steps to reproduce.	
Application build procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository)	Recipe to build the user application.	
Use Case Data Collection	Sentinel-2 MSI L2A	Sentinel-2 MSI L2A pre- and post-event acquisitions to map the burnt area	Step #1 Step #2 Step #3
Useful Data Access Filter	Mission, product type or level or collection, geographical AOI, sensing time span, cloud coverage, land coverage, update datetime.	Filters to search for Sentinel-2MSI Level-2A acquisitions over a specific AOI in a recent timespan and as close as possible in time,	Step #3

Funded by the EU
Copernicus esa European Space Agency

		excluding too high cloud coverage.	
Processing Scenario	Rapid Mapping	Processing Scenario to execute.	Step #4
Data visualization tools	WMS, geobrowser, GIS	Having the possibility to visualize the results on map directly without downloading the product is a bonus.	Step #5

Table 21: E2 Scenario Procedures Variables

The E2 score is computed based on the weighted combination of scores, apdex or metrics listed in the following Table. Satisfaction and frustration zones are computed based on the scenario customization described above.

E2 term	Elementary Metrics of Relevance	Type	User Satisfaction Zone	User Frustration Zone	Weight
E21	Q1_score	V	See 3.1	See 3.1	1.00
E22	#M209 integrationToolsAdequacy	V			0.15
E22	#M210 integrationEnvSetupTimespan	A	60	1440	0.15
E22	#M211 integrationCompleteness	V			0.15
E22	#M213 integrationSupportAnswerTime	A	60	1440	0.05
E22	#M214 integrationSupportThreadLength	A	2	5	0.05
E22	#M215 integrationSupportAppreciation	A	4	2	0.05
E22	#M216 storageSetupTimeSpan	A	10	120	0.05
E22	#M217 storageCostPerGB	A	0.5	2	0.05
E22	#M219 appStorageAvailability	A	4	2	0.05
E22	#M220 integrationCost	A	5	20	0.05
E22	T701_score	V			0.10
T701	#M001 avgResponseTime	A	500	2000	0.30
T701	#M005 throughput	A	16000000	4000000	0.70
E22	T702_score	V			0.10
T702	#M001 avgResponseTime	A	500	2000	0.30
T702	#M005 throughput	A	16000000	4000000	0.70
E23	#M301 dataQueryFiltersAdequacy	V			0.40
E23	Q51_score	V			0.20
Q51	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q51	#M110 cost	A	0.5	2	0.50
E23	Q53_score	V	See 3.5	See 3.5	0.20
E23	Q54_score	V	See 3.5	See 3.5	0.20
E25	#M501 dataExploitationToolsAdequacy	V			0.60
E25	Q3_score	V	See 3.3	See 3.3	0.20
E25	Q4_score	V	See 3.4	See 3.4	0.20

Funded by the EU
opernicus esa European Space Agency

Table 22: Computation parameters used for the E2 indicator. The Q scores and metrics are described in the Section 3. The Type column (V=value, A=apdex) define the applicability of the User Satisfaction/Frustration zone values, otherwise the value of the metric or score is used. To be noted that the sum of weights for a given step should be 1.

The value of the E2E indicator drives the classification of the target site, which is established according to the following scheme:

E2 Indicator Classes				
A	**B**	**C**	**D**	**E**
$90<E\leq100$	$60<E\leq90$	$40<E\leq60$	$10<E\leq40$	$0<E\leq10$
The platform offers overall excellent conditions to perform rapid mapping	Very good	Fair	Poor	The platform does not allow the rapid mapping

4.3 E3: MOSAICKING / LAND MONITORING

Question: How suitable is the platform for a user wishing to mosaicking Sentinel-3 OLCI data for land monitoring?

This scenario concerns the computation of a generic raster derived from mosaicking Sentinel-3 L2 OLCI data of the same type over a regional size area (e.g. Western Africa) and within a given time window (e.g. 10 days composites). The purpose of this scenario is to evaluate the capability for the target site to handle a large number of datasets in a single processing to produce multi temporal products.

Focus is on timeliness and mosaicking of the recently acquired data but also on the processing performances when large quantities of data are involved. The other important aspect is the capacity for the target site to provide with ready to use time-driven processing scheduling/orchestration tools allowing to process systematically the mosaic every N days (where N is configurable).

Setting the scene	
User Profile	The user is a specialist in Earth system science, using satellite remote sensing techniques, responsible for scientific research projects in connection with satellite missions, with a scientific expertise in land, hydrology and forestry. The user and his team are often on the field for in-situ campaigns using satellite remote sensing data. They developed a quick and relatively simple program (black box) to mosaic OLCI data to produce products allowing a global monitoring of the land evolution (e.g. vegetation, desert extent, inland water bodies).
Value-adding processing	10 days mosaic Level 2 composite (NDVI) products
Sentinel Data Set	Sentinel-3 OLCI L2 data. Dataset exceeding a certain cloud coverage are discarded.
User's expected outputs	Systematic production of the 10-days (with configurable number of days) mosaic on a regular basis to monitor land evolution. Specific items intended to be benchmarked are: - Large AOI, sensor mode search filter - Thematic toolbox availability (e.g. SNAP, ORFEO)

Funded by the EU
Copernicus esa European Space Agency

	- Capability to process large amount of data at once
	- Management of heavy single process requiring lots of computer memory (scaling)
	- Availability at the provider of ready to use time-drive orchestration/scheduling of the processes, such as e.g. in reverse order of preference (leading to different scoring):
	  ○ crontab on a virtual machine
	  ○ dedicated scheduler triggering containers
	  ○ scheduler solution on infra (e.g. k8s cron jobs)
	  ○ fully integrated scheduler with dashboard
	  ○ DAG orchestrator with scheduler and dashboard
	- Availability of exploitation tools such as e.g.:
	  ○ Access to private storage area for saving (and sharing) the generated results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.)
	  ○ Access (via API) to a catalogue index allowing publication of basic metadata about the generated results and allow later discovery and sharing (e.g. a OGC OpenSearch Catalogue)
	  ○ Access to a OGC WMS (e.g. GeoServer, MapServer etc.) where to publish the generated results for visualization and sharing
	  ○ Access to GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.)

Table 23: E3 Scenario Customization

Variable	Value	Comment	Used in
Scenario Repository	https://github.com/esa-cdab/cdab-testsuite/tree/master/Use%20Cases/Scenario%202%20-%20Mosaicking		All steps
Payment methods	#1 Credit card, #1 Bank transfer	User wants to use exclusively her/his credit card to pay for the account. Bank transfer is acceptable but not preferred.	Step #1
User's programming language and Tools ability	Python (0.35), SNAP (0.35), GDAL (0.3)	Intermediate computer skills especially with SNAP toolbox.	
User's profile description	The user is a specialist in Earth system science, using satellite remote sensing techniques, responsible for scientific research projects in connection with satellite missions, with a scientific expertise in land, hydrology, and forestry. His team can program in python and have an intermediate knowledge of the SNAP toolbox and gdal. They want to easily upload and integrate a script developed at their premises.	User wants as much as possible to run a script as a black box in the new environment	
Development Environment installation procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Installation steps to have a working python env with the snap libraries.	Step #2
Integration procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Integration steps to prepare for the execution.	
Application build procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Recipe to build the user application.	
Use Case Data Collection	Sentinel-3 OLCI Level 2	Sentinel-3 OLCI L2 acquisitions over Large AOI. Temporal	Step #1 Step #2 Step #3

Funded by the EU
Copernicus esa European Space Agency

		range of N days (N =10, configurable).	
Useful Data Access Filter	Mission, product type and level, geographical AOI, sensing time span, track.	Filters to search for Sentinel-3 OLCI Level 2 intersecting the given AOI over last 10 days.	Step #3
Processing Scenario	10 days mosaic Level 2 composite (NDVI) products	Processing Scenario to execute.	Step #4
Exploitation tools	- Access to a OGC WMS (e.g. GeoServer, MapServer etc.) where to publish the generated results for visualization and sharing (0.3) - Access to GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.) (0.4) - Access to private storage area for saving (and sharing) the generated results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.) (0.2) - Access (via API) to a catalogue index allowing publication of basic metadata about the generated results and allow later discovery and sharing (e.g. a OGC OpenSearch Catalogue) (0.1)	Having the possibility to visualize the results on map directly without downloading the product is a bonus.	Step #5
Orchestration / scheduling tools	Time-drive orchestration/scheduling of the processes - no scheduler at all - crontab on a virtual machine - dedicated scheduler triggering containers - scheduler solution on infra (e.g. k8s cron jobs) - fully integrated scheduler with dashboar - DAG orchestrator with scheduler and dashboar	An important aspect is the capacity for the target site to provide with a time-driven processing scheduling mechanism allowing to re-process systematically the mosaic every 10 days	Step #5

Table 24: E3 Scenario Procedures Variables

The E3 score is computed based on the weighted combination of scores, apdex or metrics listed in the following Table. Satisfaction and frustration zones are computed based on the scenario customization described above.

E3 term	Elementary Metrics of Relevance	Type	User Satisfaction Zone	User Frustration Zone	Weight
E31	Q1_score	V	See 3.1	See 3.1	1.00
E32	#M209 integrationToolsAdequacy	V			0.15
E32	#M210 integrationEnvSetupTimespan	A	60	1440	0.15
E32	#M211 integrationCompleteness	V			0.15
E32	#M213 integrationSupportAnswerTime	A	60	1440	0.05
E32	#M214 integrationSupportThreadLength	A	2	5	0.05
E32	#M215 integrationSupportAppreciation	A	4	2	0.05
E32	#M216 storageSetupTimeSpan	A	10	120	0.05
E32	#M217 storageCostPerGB	A	0.5	2	0.05
E32	#M219 appStorageAvailability	A	4	2	0.05
E32	#M220 integrationCost	A	5	20	0.05
E32	T701_score	V			0.10
T701	#M001 avgResponseTime	A	500	2000	0.30

Funded by the EU
Copernicus esa European Space Agency

T701	#M005 throughput	A	16000000	4000000	0.70
E32	T702_score	V			0.10
T702	#M001 avgResponseTime	A	500	2000	0.30
T702	#M005 throughput	A	16000000	4000000	0.70
E33	#M301 dataQueryFiltersAdequacy	V			0.40
E33	Q51_score	V			0.20
Q51	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q51	#M110 cost	A	0.5	2	0.50
E33	Q53_score	V	See 3.5	See 3.5	0.20
E33	Q54_score	V	See 3.5	See 3.5	0.20
E35	#M501 dataExploitationToolsAdequacy	V			0.60
E35	#M049 wfTools	V			0.20
E35	Q3_score	V	See 3.3	See 3.3	0.10
E35	Q4_score	V	See 3.4	See 3.4	0.10

Table 25: Computation parameters used for the E3 indicator. The Q scores and metrics are described in the Section 3. The Type column (V=value, A=apdex) define the applicability of the User Satisfaction/Frustration zone values, otherwise the value of the metric or score is used. To be noted that the sum of weights for a given step should be 1.

The value of the E2E indicator drives the classification of the target site, which is established according to the following scheme:

E3 Indicator Classes				
A	**B**	**C**	**D**	**E**
$90<E\leq100$	$60<E\leq90$	$40<E\leq60$	$10<E\leq40$	$0<E\leq10$
The platform offers overall excellent conditions to perform mosaicking for land monitoring	Very good	Fair	Poor	The platform does not allow to perform mosaicking for land monitoring

4.4 E4: TRENDS ASSESSMENT

Question: How suitable is the platform for a user wishing to build long term land surface temperature (LST) time series from Sentinel-3 SLSTR data?

This scenario concerns processing of multiple data of the same type over long-time windows to build land surface temperature time series from Sentinel-3 SLTRS Level 2 data. Concerned time spans and area sizes can range from medium to very large ones.

Focus is on the capacity for the target site to provide with a data-driven processing scheduling mechanism allowing to process systematically new acquired data to produce quickly new complete time series.

Funded by the EU
opernicus esa European Space Agency

Options for storage and products exploitation/visualization (e.g. catalogue, WMS, Time Series viewer) may be at stake.

Setting the scene	
User Profile	The user is a scientist working in a Climate Monitoring Institute. He has strong thematic background and intermediate computer science training. He has developed in-house a LST processing service providing land surface temperature time series based on SLSTR instruments data that he wants to deploy on a Cloud environment.
Value-adding processing	Long time series of LST at regional scale (e.g. Iberian Peninsula, South Italy etc.).
Sentinel Data Set	Sentinel-3 SLSTR Level 2 products from the start of the mission. New acquisitions over the AOI every for each Satellite for a given track are considered for the processing.

User's expected outputs	Systematic production/update of the LST time series as soon as a new product over the AOI is available in the Service Provider catalogue.
	
	Specific items intended to be benchmarked:
	- Thematic toolbox availability (e.g. SNAP, ORFEO)
	- Management of long series of datasets (potentially from the start of the mission, i.e. including offline data)
	- Availability at the provider of ready to use data-driven Orchestration/Scheduling of the processes, such as e.g. in reverse order of preference (leading to different scoring):
	  ○ polling the catalogue from a VM
	  ○ dedicated scheduler triggering containers based on catalogue search
	  ○ as above but with management/monitoring Dashboard
	- Availability of exploitation tools such as e.g.:
	  ○ Access to private storage area for saving (and sharing) the generated results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.)
	  ○ Access (via API) to a catalogue index allowing publication of basic metadata about the generated results and allow later discovery and sharing (e.g. a OGC OpenSearch Catalogue)
	  ○ Access to a OGC WMS (e.g. GeoServer, MapServer etc.) where to publish the generated results for visualization and sharing
	  ○ Access to GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.)
	  ○ Visualization of time series (Web time series viewer)

Table 26: E4 Scenario Customization

Variable	Value	Comment	Used in
Scenario Repository	https://github.com/esa-cdab/cdab-testsuite/tree/master/Use%20Cases/Scenario%202%20-%20Trends		All steps
Payment methods	#1 Credit card, #1 Bank transfer	User wants to use exclusively her/his credit card to pay for the account. Bank transfer is acceptable but not preferred.	Step #1
User's programming language and Tools ability	Python (0.35), SNAP (0.35), GDAL (0.3)	Intermediate computer skills especially with SNAP toolbox.	

Funded by the EU
Copernicus esa European Space Agency

User's profile description	The user is a scientist working in a Climate Monitoring Institute. He has strong thematic background and intermediate computer science training. He has developed in-house a LST processing service providing land surface temperature time series based on SLSTR instruments data that he wants to deploy on a Cloud environment. He can program in python and has good knowledge of the SNAP toolbox and GDAL. He wants to reuse scripts already developed in-house and adapt them to run on the target site with minimal changes.	User wants as much as possible to reuse previously developed scripts in the target environment	
Development Environment installation procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Installation steps to have a working python env with the snap libraries.	Step #2
Integration procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Integration steps to prepare for the execution.	
Application build procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Recipe to build the user application.	
Use Case Data Collection	Sentinel-3 OLCI Level 2	Sentinel-3 SLSTR L2 acquisitions over AOI and from the start of the mission	Step #1 Step #2 Step #3
Useful Data Access Filter	Mission, product type and level, geographical AOI, sensing time span, track.	Filters to search for Sentinel-2 SLSTR Level 2 intersecting the given AOI and on a given track	Step #3
Processing Scenario	Time series of land surface temperature (LST) products	Processing Scenario to execute.	Step #4
Exploitation tools	- Access to a OGC WMS (e.g. GeoServer, MapServer etc.) where to publish the generated results for visualization and sharing (0.3) - Access to GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.) (0.4) - Access to private storage area for saving (and sharing) the generated results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.) (0.2) - Access (via API) to a catalogue index allowing publication of basic metadata about the generated results and allow later discovery and sharing (e.g. a OGC OpenSearch Catalogue) (0.1)	Having the possibility to visualize the results on map directly without downloading the product is a bonus.	Step #5
Orchestration / scheduling tools	Data-driven orchestration/scheduling of the processes - no solution - polling catalogue from a VM - dedicated scheduler triggering containers based on catalogue search - dedicated scheduler triggering containers based on catalogue search with dashboard	An important aspect is the capacity for the target site to provide with a data-driven processing scheduling mechanism allowing to process systematically the time series when a new acquisition is available	Step #5

Table 27: E4 Scenario Procedures Variables

The E4 score is computed based on the weighted combination of scores, apdex or metrics listed in the following Table. Satisfaction and frustration zones are computed based on the scenario customization described above.

Funded by the EU
Copernicus esa European Space Agency

E4 term	Elementary Metrics of Relevance	Type	User Satisfaction Zone	User Frustration Zone	Weight
E41	Q1_score	V	See 3.1	See 3.1	1.00
E42	#M209 integrationToolsAdequacy	V			0.15
E42	#M210 integrationEnvSetupTimespan	A	60	1440	0.15
E42	#M211 integrationCompleteness	V			0.15
E42	#M213 integrationSupportAnswerTime	A	60	1440	0.05
E42	#M214 integrationSupportThreadLength	A	2	5	0.05
E42	#M215 integrationSupportAppreciation	A	4	2	0.05
E42	#M216 storageSetupTimeSpan	A	10	120	0.05
E42	#M217 storageCostPerGB	A	0.5	2	0.05
E42	#M219 appStorageAvailability	A	4	2	0.05
E42	#M220 integrationCost	A	5	20	0.05
E42	T701_score	V			0.10
T701	#M001 avgResponseTime	A	500	2000	0.30
T701	#M005 throughput	A	16000000	4000000	0.70
E42	T702_score	V			0.10
T702	#M001 avgResponseTime	A	500	2000	0.30
T702	#M005 throughput	A	16000000	4000000	0.70
E43	#M301 dataQueryFiltersAdequacy	V			0.40
E43	Q51_score	V			0.20
Q51	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q51	#M110 cost	A	0.5	2	0.50
E43	Q53_score	V	See 3.5	See 3.5	0.20
E43	Q54_score	V	See 3.5	See 3.5	0.20
E45	#M501 dataExploitationToolsAdequacy	V			0.60
E45	#M049 wfTools	V			0.20
E45	Q3_score	V	See 3.3	See 3.3	0.10
E45	Q4_score	V	See 3.4	See 3.4	0.10

Table 28: Computation parameters used for the E4 indicator. The Q scores and metrics are described in the Section 3. The Type column (V=value, A=apdex) define the applicability of the User Satisfaction/Frustration zone values, otherwise the value of the metric or score is used. To be noted that the sum of weights for a given step should be 1.

The value of the E2E indicator drives the classification of the target site, which is established according to the following scheme:

E4 Indicator Classes				
A	B	C	D	E
$90<E\leq100$	$60<E\leq90$	$40<E\leq60$	$10<E\leq40$	$0<E\leq10$
The platform offers overall excellent conditions to	Very good	Fair	Poor	The platform does not allow to perform

Funded by the EU
Copernicus esa European Space Agency

perform trends assessment				trends assessment

4.5 E5: INTERFEROGRAMME COMPUTATION

Question: How suitable is the platform for a user wishing to create interferograms from Sentinel-1 scenes?

This scenario concerns the evaluation of the suitability of the target site for a user wishing to process Sentinel-1 SLC data[11] to generate an interferogram over an area of interest in a predetermined period corresponding to the occurrence of a well-known earthquake event. He uses the Sentinel Application Platform (SNAP) to generate the interferogram. He also wants to prepare a multi-temporal stacks of Sentinel-1 acquisitions for later generation of Earth deformation time series and mean velocity maps.

Focus is on tools facilitating the data search for the identification of interferometric pairs and for building and saving/sharing multi-temporal stacks of data. A powerful processing environment able to cope with Intensive I/O and high CPU load is needed to run the algorithm efficiently. Data loads are medium to large size (e.g. up to 150 Sentinel-1 scenes).

Setting the scene	
User Profile	The user is a remote sensing expert in SAR interferometry techniques with intermediate computer science training. He can program in bash and Python and has an advanced knowledge of the SNAP toolbox.
Value-adding processing	The user wishes to compute a directed acyclic graph (DAG) for SNAP and does a system call to the SNAP gpt utility to process it using a simple SW program in python. Interferometric pairs are selected as follows: the user selects Sentinel-1 the reference S1 SLC acquisition (referred to as reference), and then uses the catalog discovery mechanism to select the "paired" acquisition (referred to as secondary) setting the same orbit track and direction and an overlap of at least 80% between the two acquisitions. The DAG includes all the steps for performing phase unwrapping relying on a CLI utility called snaphu.
Sentinel Data Set	The dataset of relevance for this scenario is Sentinel-1 Level 1 SLC IW/EW mode data. The time span covers a few weeks for the interferogram but can also cover several years in the past for the multitemporal stack (i.e. off-line data is involved).
User's expected outputs	The user may wish to download the generated product but having the possibility to visualize them directly on a map without download (e.g. through WMS, geobrowser), is considered an asset. Having the possibility to save/share the built multitemporal stack (e.g. as a catalogue reference) is also important.

Table 29: E5 Scenario Customization

Variable	Value	Comment	Used in

[11] Single Look Complex data from Sentinel-1 correspond to Level-1 processing. For details, see Sentinel-1 user guide at https://sentinels.copernicus.eu

Funded by the EU
Copernicus esa European Space Agency

Scenario Repository	https://github.com/esa-cdab/cdab-testsuite/tree/master/tree/master/Use%20Cases/Scenario%205%20-%20Interferogram Computation		All steps
Payment methods	#1 Bank transfer, #2 Credit card	User wants to use exclusively their credit card to pay for their account.	Step #1
User's programming language and Tools ability	Java (0.3), SNAP (0.7)	Advanced computer skills especially with SNAP toolbox..	
User's profile description	The user is an environmental scientist with intermediate computer science training. He can program in bash and has an advanced knowledge of the SNAP toolbox.	Description to be used in the exchange for describing the user.	
Development Environment installation procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Installation steps to have a working jupyter notebook in python with the snap libraries.	Step #2
Integration procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Some integration steps to reproduce.	
Application build procedure	See integration.md in the test suite software package in the corresponding scenario folder (Scenario Repository).	Recipe to build the user application.	
Use Case Data Collection	Sentinel-1 SLC	Sentinel-1 SLC acquisitions pre and post event.	Step #1 Step #2 Step #3
Useful Data Access Filter	Mission, product type or level or collection, geographical AOI, sensing time span, track, land coverage.	Filters to search for Sentinel-1 SLC pairs on the same track for a given AOI.	Step #3
Processing Scenario	Interferogram Computation	Processing Scenario to execute.	Step #4
Data visualization tools	WMS, geobrowser	Having the possibility to visualize the results on map directly without downloading the product is a bonus.	Step #5

Table 30: E5 Scenario Procedures Variables

The E5 score is computed based on the weighted combination of scores, apdex or metrics listed in the following Table. Satisfaction and frustration zones are computed based on the scenario customization described above.

E5 term	Elementary Metrics of Relevance	Type	User Satisfaction Zone	User Frustration Zone	Weight
E51	Q1_score	V	See 3.1	See 3.1	1.00
E52	#M209 integrationToolsAdequacy	V			0.15
E52	#M210 integrationEnvSetupTimespan	A	60	1440	0.15
E52	#M211 integrationCompleteness	V			0.15
E52	#M213 integrationSupportAnswerTime	A	60	1440	0.05
E52	#M214 integrationSupportThreadLength	A	2	5	0.05
E52	#M215 integrationSupportAppreciation	A	4	2	0.05
E52	#M216 storageSetupTimeSpan	A	10	120	0.05
E52	#M217 storageCostPerGB	A	0.5	2	0.05

Funded by the EU
Copernicus esa European Space Agency

E52	#M219 appStorageAvailability	A	4	2	0.05
E52	#M220 integrationCost	A	5	20	0.05
E52	T701_score	V			0.10
T701	#M001 avgResponseTime	A	500	2000	0.30
T701	#M005 throughput	A	16000000	4000000	0.70
E52	T702_score	V			0.10
T702	#M001 avgResponseTime	A	500	2000	0.30
T702	#M005 throughput	A	16000000	4000000	0.70
E53	#M301 dataQueryFiltersAdequacy	V			0.40
E53	Q51_score	V			0.20
Q51	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q51	#M110 cost	A	0.5	2	0.50
E53	Q53_score	V	See 3.5	See 3.5	0.20
E53	Q54_score	V	See 3.5	See 3.5	0.20
E54	Q52_score	V			1.00
Q52	#M101 avgProvisioningLatency	A	60000	90000	0.50
Q52	#M110 cost	A	0.5	2	0.50
E55	#M501 dataExploitationToolsAdequacy	V			0.60
E55	Q3_score	V	See 3.3	See 3.3	0.20
E55	Q4_score	V	See 3.4	See 3.4	0.20

Table 31: Computation parameters used for the E5 indicator. The Q scores and metrics are described in the Section 3. The Type column (V=value, A=apdex) define the applicability of the User Satisfaction/Frustration zone values, otherwise the value of the metric or score is used. To be noted that the sum of weights for a given step should be 1.

The value of the E2E indicator drives the classification of the target site, which is established according to the following scheme:

E5 Indicator Classes				
A	**B**	**C**	**D**	**E**
$90<E\leq100$	$60<E\leq90$	$40<E\leq60$	$10<E\leq40$	$0<E\leq10$
The platform offers overall excellent conditions to perform interferogramme computation	Very good	Fair	Poor	The platform does not allow to perform interferogramme computation

4.6 E6: HIGH DEMANDING, MASSIVE PROCESSING WITH S-2 DATA

Funded by the EU
Copernicus esa European Space Agency

Question: How suitable is the platform for a user wishing to build long term trends of Sentinel-2 smoothed temporal and spatial gap filled NDVI over one or more geographic area(s) (e.g. drought affected areas)?

This scenario concerns processing of multiple data of the same type (Sentinel-2 MSI) over long time windows to build:

- NDVI time series with temporal and spatial gaps
- Smoothed temporal and spatial gap filled NDVI time series employing the Whittaker filetring method.

The following figure provides a sketch of the scenario:

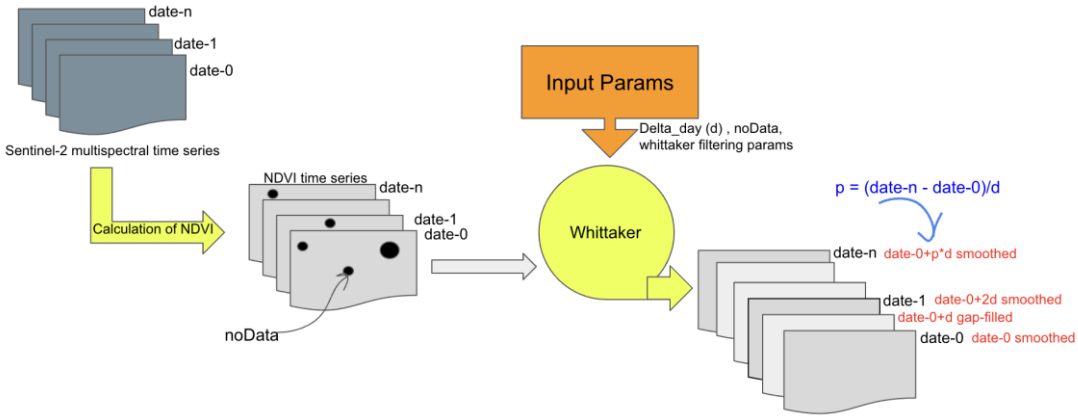


Figure 7 Smoothed and gap filled NDVI time series

- Given a time series of high spatial resolution data and input params, among which delta-day interval (d) and noData value, the processor generates a gap-filled and smoothed output time series, exploiting the Whittaker filter
- The output series contain smoothed data per each input which coincide with the delta-day interval dates and gap-filled (i.e. interpolated) data for missing days in our input time series;
- The interpolated value is calculated per pixel based on the vertical array of time series values after discarding the nodata pixels values.

Concerned time range can span from medium sizes to very large ones (but at least three years of data). A number of small AOIs (e.g. covered by a single Sentinel-2 tile) well spread over the world (e.g. one for each continent) will be considered. These could be e.g. selected drought affetcted areas.

Focus is on the capacity of the target site to provide efficient access to large stacks of Sentinel-2 data as well as Kubernetes computing clusters and S3 object storage for running resource savvy computational workflows to batch process new acquired data to produce quickly new, complete time series.
Options for storage and products exploitation/visualization (e.g. catalogue, WMTS) may be at stake.

Setting the scene	
User Profile	The user is Scientist working in a Food Security Monitoring Institute. He has strong thematic background and intermediate computer science training. He has developed in-house EO application packages using the Common Workflow Language (CWL) and containers for defining resource savvy computational workflows for detecting NDVI anomalies that he wants to deploy on a Cloud environment.
Value-adding processing	Long time series of Sentinel-2 smoothed temporal and spatial gap filled NDVI.
Sentinel Data Set	Sentinel-2 MSI Level 1C (or 2A) products from the start of the mission (or at least three years) over 5 AOIs spread in different continents. Each AOI is covered by a single Sentinel-2 tile.

Funded by the EU
opernicus esa European Space Agency

	New acquisitions over the AOIs every for each Satellite are considered for the processing (i.e. One new product per AOI every 5/10 days).
User's expected outputs	- Systematic production of Sentinel-2 NDVI. - Update of the Sentinel-2 smoothed temporal and spatial gap filled NDVI as soon as set of new products become available in the service provider catalogue. Specific items intended to be primarily benchmarked: - Main goal of the scenario is to evaluate the providers against their capability to provide "cloud optimised services", i.e.: ○ Provide Kubernetes resources as part of their offering (Kubernetes-as-a-service) ○ provide virtual machines that can be procured and configured as a Kubernetes cluster in case of lack of Kubernetes-as-a-service ○ provide storage mechanisms to support the volume claims for the Kubernetes cluster ○ provide S3 based object storage for holding the produced results (NDVI derived from Sentinel-2, Whittaker synthetic and gap filled NDVI, Static STAC catalog of generated results) ○ Provide autoscaling capacity ○ Support running of Common Workflow Language (CWL) documents using a Kubernetes CWL runner (calrissian) ○ Provide access to long series of Sentinel-2 data (SAFE or STAC/COG formats), potentially from the start of the mission (i.e. including offline data)

Table 32: E6 Scenario Customization

Variable	Value	Comment	Used in
Scenario Repository	https://github.com/esa-cdab/cdab-testsuite/tree/master/Use%20Cases/<TBW>		All steps
User's programming language and Tools ability	Python (0.35), GDALL (0.3), CWL	Intermediate computer skills especially with Python and GDAL.	
User's profile description	The user is Scientist working in a Food Security Monitoring Institute. He has strong thematic background and intermediate computer science training. He has developed in-house EO application packages using the Common Workflow Language (CWL) and containers for defining resource savvy computational workflows for detecting NDVI anomalies that he wants to deploy on a Cloud environment. He can program in python and has good knowledge of the GDAL, CWL and the Whittaker package. He wants to reuse scripts already developed in-house and adapt them to run on the target site with minimal changes.	User wants as much as possible to reuse previously developed scripts in the target environment	
Use Case Data Collection	Sentinel-2 MSI Level 1C (2A could also be used)	Sentinel-2 MSI L1 acquisitions over AOIs and from the start of the mission	Step #1
Application Integration	Processing cost estimation and evaluation (linked to Step #4)		Step #2
Data Access	Data Discovery and Access evaluation of the provider		Step #3
Processing Scenario	Long time series of Sentinel-2 smoothed temporal and spatial gap filled NDVI	Processing Scenario to execute.	Step #4
Cloud optimized services	Capability to provide cloud optimised services:	Offering of kubernetes computing clusters and S3 object storage	Step #5

Funded by the EU
Copernicus esa European Space Agency

	• provide Kubernetes resources as part of their offering (Kubernetes-as-a-service) • provide virtual machines that can be procured and configured as a Kubernetes cluster in case of lack of Kubernetes-as-a-service • provide storage mechanisms to support the volume claims for the Kubernetes cluster • provide S3 based object storage for holding the produced results • support autoscaling capacity for processing and storage • provide upstream API compatible with Terraform • provide access to data in STAC/COG formats	for running the processing workflow and storing related results is fundamental for this scenario	

Table 33: E6 Scenario Procedures Variables

The E6 score is computed based on the weighted combination of scores, apdex or metrics listed in the following Table. Satisfaction and frustration zones are computed based on the scenario customization described above.

E6 term	Elementary Metrics of Relevance	Type	User Satisfaction Zone	User Frustration Zone	Weight
E61	#M023 dataCoverage	V			0.70
E61	#M015 catalogueCoverage	V			0.30
E62	#M243 costPerformance	V			0.30
E62	#M244 devCostPerformance	V			0.30
E62	costScore (see #TP19)	V			0.40
E63	#M240 dataDiscoveryAndAccess	V			1.00
E64	#M245 testEfficiency	V			0.40
E64	#M241 processingSuccessRate	V			0.30
E64	#M246 processingTimeRate	V			0.30
E65	#M242 cloudOptimizedToolsAdequacy	V			0.50
E65	#M247 developerRank	V			0.50

Table 34: Computation parameters used for the E6 indicator. The Type column (V=value, A=apdex) define the applicability of the User Satisfaction/Frustration zone values, otherwise the value of the metric or score is used. To be noted that the sum of weights for a given step should be 1.

The value of the E2E indicator drives the classification of the target site, which is established according to the following scheme:

E6 Indicator Classes				
A	B	C	D	E
$90<E\leq100$	$60<E\leq90$	$40<E\leq60$	$10<E\leq40$	$0<E\leq10$
The platform offers overall excellent conditions to perform high demanding, massive	Very good	Fair	Poor	The platform does not allow to perform high demanding, massive

Funded by the EU
Copernicus esa European Space Agency

processing with S-2 data				processing with S-2 data

Funded by the EU
Copernicus
esa European Space Agency

5 CONCLUSIONS AND PLANNED EVOLUTIONS

In the first phase of the project, the benchmarking service was set-up and exercised through some preliminary testing with core target sites, which were mainly intended to support the consolidation of the benchmarking parameters. The attention was focused on the definition of a rigorous strategy that allows a comprehensive and easy to understand evaluation of the overall benchmarking results.

The approach was inspired by good practices found in the literature. However, these were applied in other domains and none could be fully transposed to the area of Earth Observation data. Thus, tailoring to the sector was undertaken and a dedicated calibration activity was defined.

In particular, the proposed approach based on Quality of Experience (QoE) indicators critically depends on the reliability and representativeness of a number of key reference parameters and these need extensive validation, especially with regards to:

- The definition of the tolerance zones for each elementary metrics used in the computation of the application indexes, as well as the weights used for the computation of the overall indexes.
- The threshold values assigned to the definition of the classes related to each QoE indicator.

Lately, application-specific indicators were introduced, that are computed by executing application-specific scenarios from first steps (e.g. user registration and evaluation of the data set) to to very final steps (i.e. download or publication of the resulting products).

Another aspect concerns the network of the test sites, and in particular the ones that are used as a reference to compute the key parameters (from Q2 to Q7). The current set is limited to European test sites characterised by high performances and excellent stability over time. To select them, a continuous calibration process has been designed that aims at identifying the best performing test sites in order to limit any bias in the target sites' evaluation.

The service scope has been progressively increased.

Robustness of the overall benchmarking has been improved through the availability of more statistically significant samples - not only in terms of the number and diversity of the test sites but also of the target sites. This will be further consolidated through interactions with key stakeholders as well as through the extension to additional third party target sites.

Funded by the EU
Copernicus esa European Space Agency

6 ANNEX A - SERVICE ARCHITECTURE

The main pillars of the benchmarking service are:

- a network of Test Sites, that perform the benchmarking operations towards the pre-defined target sites;
- the Orchestration function, in charge of managing the benchmarking operations;
- the Analysis and Reporting function;
- the Calibration and Validation function, ensuring a continuous monitoring of the service reliability;
- a Public Software Repository service; and
- the Service Management, in charge of the overall management including possible evolutions.

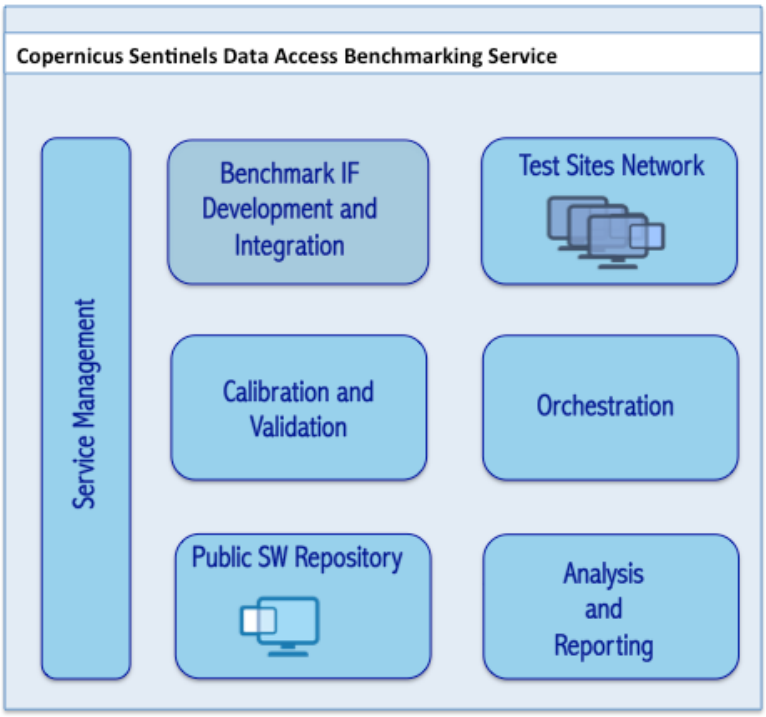


Figure 8: High-level Benchmarking Service Architecture

6.1 SERVICE COMPONENTS

6.1.1 The Test Sites Network

The Test Site network is a cloud of virtual nodes by different providers accurately chosen in order to be statistically and geographically representative of the Sentinels user segment.

On each node, through a centralized Test Sites management software tool, a software package (**Reference SW Suite**) is installed. The software package consists of a set of tools and scripts properly prepared for the execution of configurable user operations and the collection of the corresponding results into metrics that can be used to compute APDEXes and QoEs indicators for the benchmarks' comparison. Details on the Test Site Network can be found on ANNEX B - The Worldwide Test Sites Network.

Funded by the EU

6.1.2 The benchmark test tools development and integration function

This is the function in charge of development and maintenance of the core SW elements e.g. test tools to be deployed on the test sites network. When a new target site is to be added, the SW routines allowing this interface are developed, unit-tested and integrated as part of this function.

6.1.3 The orchestration function

This is the component that manages the cloud of Test Sites where the reference test suite will run, collecting the results from benchmarking operations towards a pre-defined set of Target Sites. Each node in the Test Sites network is configured in order to execute specific operational scenarios against specific Target Sites according to a pre-defined Connection Matrix. This configuration consists of the following steps:

- Definition of a Test Site-to-Test Scenario matrix
- Definition of a Test Scenario schedule configuration
- Definition of a Test Site-to-Target Site matrix (Connection matrix)
- Creation of scheduled jobs for Test Scenarios execution

During normal operations, the following steps are executed:

- Test Case deployment/run
  - At scheduled time the job identified by Test Scenario, Test Site and Target Site run
  - The job deploy, if updated, the new Docker image of the Reference SW Suite
  - The job run the Test Tools on the Test Site
  - The test results are archived
- Test results sent to the Dashboard, where they are stored, processed and visualized.

For the continuous integration/continuous delivery server, we make use of Jenkins[12]. This is a self-contained, open source automation server which can be used to automate all sorts of tasks related to building, testing, and delivering or deploying software. Jenkins can be installed through native system packages, Docker, or even run standalone by any machine with a Java Runtime Environment (JRE) installed. Jenkins can be easily set up and configured via its web interface, which includes on-the-fly error checks and built-in help. With hundreds of plugins in the Update Center, Jenkins integrates with practically every tool in the continuous integration and continuous delivery toolchain. Jenkins can be extended via its plugin architecture, providing nearly infinite possibilities for what Jenkins can do. Jenkins can easily distribute work across multiple machines, helping drive builds, tests and deployments across multiple platforms faster.

The Test Site Tools, aimed at gathering the appropriate metrics based on the applicable interface type (API or GUI) and/or the test functionality (reachability, download, process, etc.), are of different nature. To start, Selenium[13] (for GUIs) and the SW Suite package (for REST APIs) were used. Selenium is a suite of tools to automate web browsers across many platform. Primarily, it is for automating web applications for testing purposes, but is certainly not limited to just that. Boring web-based administration tasks can (and should!) be automated as well. Selenium has the support of some of the largest browser vendors who have taken (or are taking) steps to make Selenium a native part of their browser. It is also the core technology in countless other browser automation tools, APIs and frameworks.
The test case executed by the tools can sometimes overlap the functionality of other test tool, but what is important is the full coverage of all the metrics defined as base descriptors of the Service benchmarks. In addition, a set of Cloud Platform tools were developed by Terradue that allow to access the different endpoints to be tested[14]. These test tools are made publicly available as a Reference SW Suite to support self-benchmarking. Details are provided in Sect. 5.1.6.

[12] https://jenkins.io/

[13] https://www.seleniumhq.org/

[14] https://github.com/Terradue

Funded by the EU
opernicus esa European Space Agency

6.1.4 Analysis and Reporting

This is the function in charge of monitoring the low-level test result metrics (core benchmark) representative of the performance of the target sites for the functionalities under test over time through their full statistics, and of translating them into the high-level "Quality Of Experience" indicators.

This is implemented by using the Splunk platform which provides an easy, fast and secure way to search, analyze and visualize the massive streams of machine data generated by IT systems and technology infrastructure. The full stack of tools include:

- **Splunk Enterprise**[15], which monitors and analyzes machine data from any source to deliver Operational Intelligence. It is widely used in the IT market. The Splunk Enterprise COTS is a flexible platform that scales from focused use cases to an enterprise-wide analytics backbone. It speeds tactical troubleshooting by gathering real-time log data from distributed applications and infrastructure in one place to enable powerful searches, dynamic dashboards and alerts, and reporting for real-time/historical analysis; and
- **Splunk App for Jenkins** [16] that provides deep insights into Jenkins master and slave infrastructure, job and build details such as console logs, status, artifacts, and an incredibly efficient way to analyze test results. The app provides out-of-the-box dashboards and search capabilities to enable organizations to run a high performing Jenkins cluster and bring operational intelligence into the software development life cycle.

6.1.5 Calibration and Validation

This functionality is in charge of calibrating and validating the benchmark. This is achieved thorugh a set of semi-automated functions (e.g. for the calibration of the Test Sites Network through speedtest) and periodic sensitivity analyses aimed at adjusting the core parameter values used for the computation of the benchmark QoE indicators and set an appropriate scale of reference.

6.1.6 Reference SW Suite Public Repository

The Reference SW Suite is made of a set of routines that can be used by interested stakeholders to perform benchmarking according to the current framework, in view of comparing the achieved results. The Suite is derived from the operational Test Site tools that are deployed on the Test Sites. The Reference SW Suite is made available on a Public Software Repository (i.e.: GitHub). The Reference SW Suite will be packaged in a **Docker**[17] **container**. A container is a standard unit of software that packages up code and all its dependencies so the application runs quickly and reliably from one computing environment to another. A Docker container image is a lightweight, standalone, executable package of software that includes everything needed to run an application: code, runtime, system tools, system libraries and settings. Container images become containers at runtime and in the case of Docker containers - images become containers when they run on Docker Engine. Available for both Linux and Windows-based applications, containerized software have the advantage that they will always run the same, regardless of the infrastructure. Containers isolate software from its environment and ensure that it works uniformly despite differences for instance between development and staging. Docker containers that run on Docker Engine are:

- Standard: Docker created the industry standard for containers, so they could be portable anywhere
- Lightweight: Containers share the machine's OS system kernel and therefore do not require an OS per application, driving higher server efficiencies and reducing server and licensing costs
- Secure: Applications are safer in containers and Docker provides the strongest default isolation capabilities in the industry.

Details about the Reference SW Suite will be provided in Annex F to this document.

[15] https://www.splunk.com/

[16] https://splunkbase.splunk.com/app/3332/

[17] https://www.docker.com/

Funded by the EU
opernicus esa European Space Agency

6.1.7 Service Management

The Service Management includes activities related to the performance monitoring of the service. These are reported to ESA through a dedicated section of the dashboard, in charge of monitoring and reporting the service performance over a specific reference period (by default 1 month) and according to a set of Key Performance Indicators (KPIs) that are required for the service management.

Service Management functions include the planning and design of service evolutions as well as the interaction with the core stakeholders.

6.2 SERVICE WORKFLOW AND INTERFACES

The pre-requisite to start the service is the set-up of the Test Sites Network, through the procurement of the Virtual Machines from public commercial providers (i.e.: Amazon Web Service, Microsoft Azure, Google Cloud Platform, etc.) and/or at the public institutions (EGI). Then, after configuration of the orchestration function, benchmarking operations are performed against the configured Target Sites according to a pre-defined operations plan. The corresponding test results are collected, stored and analysed and reports are generated. The calibration and Validation activities are periodically executed.

The full integration of the previously described service components is shown in the following diagram, where the core interfaces are shown.

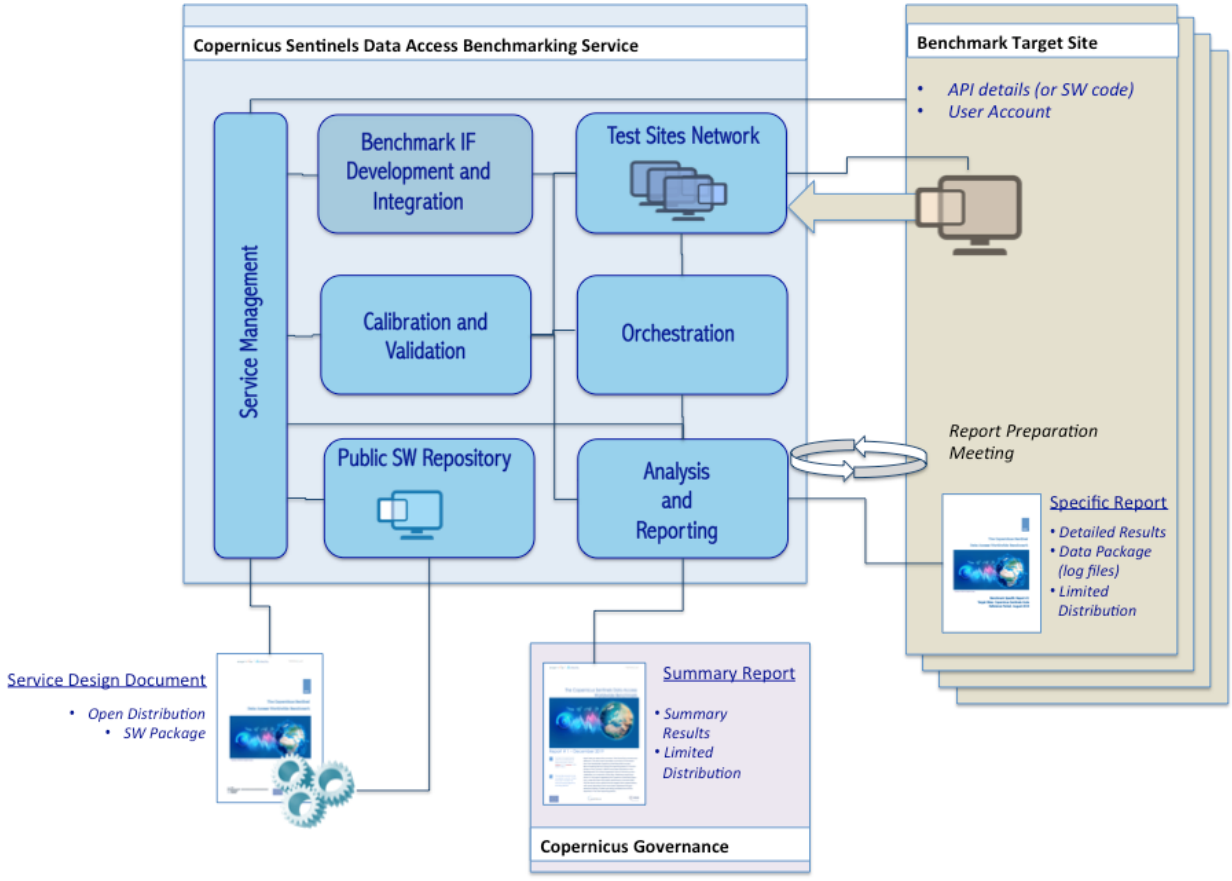


Figure 9 Benchmark functional elements and core interfaces.

Funded by the EU
Copernicus
esa European Space Agency

In particular, the interface with the Benchmark Target Site Operator is highlighted. Interactions include the creation of a user account, the benchmark operations (e.g. catalogue requests and Sentinels data downloads) as well as possible interactions to prepare the Specific Report that details the specific results.

Funded by the EU
Copernicus
esa European Space Agency

7 ANNEX B - THE WORLDWIDE TEST SITES NETWORK

In order to simulate different users for the QoE indicators computation, a set of virtual and hardware platforms has been procured in different areas of the world from various (commercial and academic) cloud providers. These constitute a network of User Test Sites that are independent from the operations of the Target Sites.

The test sites network is currently composed of **30 nodes**, that are:

Test Site ID	Location (country iso3)
ITA-1	ITA
LTU-1	LTU
ESP-2	ESP
ITA-2	ITA
ROU-2	ROU
FRA-1	FRA
IRL-1	IRL
NOR-1	NOR
FRA-3	FRA
DEU-1	DEU
SWE-1	SWE
POL-2	POL
GBR-1	GBR
BEL-1	BEL
BRG-1	BRG
GRC-1	GRC
W-UAE-1	UAE
W-COL-1	COL
W-CHL-1	CHL
W-MEX-1	MEX
W-ISR-1	ISR
W-AUS-1	AUS
W-NGA-1	NGA
W-ZAF-1	ZAF
W-IND-1	IND
W-USA-2	USA
W-SGP-1	SGP
W-JPN-1	JPN
W-CAN-1	CAN
W-BRA-1	BRA

This list must be considered a living document that evolves during the project lifetime according to the obtained benchmarking results, to cost/benefits considerations and to their performances.

The test sites performances are regularly monitored through a calibration process. A limited set of high-performing test sites is taken as the reference **benchmark trusted goldmarks** that are used

Funded by the EU
opernicus esa European Space Agency

to compute the scores for all the target sites. The test sites located in Europe obviously provide the best performance with respect to the core target sites when compared to test sites distributed worldwide. They constitute the trusted goldmarks that can be considered as a good representation of performance benchmarking of European hubs.

7.1 CLOUD PROVIDERS AND HARDWARE

In order to simulate different conditions for the user scenarios, a diversification was introduced in the procured hardware and network conditions.

Virtual machines (VMs) were procured from public commercial cloud providers to mimic as much as possible the resources available to the general public, while also using different providers to avoid biases due to the systematic use of a single provider and to avoid vendor lock-in with a single cloud service provider.

At the same time, test sites had to be spread across areas of Europe and the rest of the world, so even when using a single provider, there are differences in the available hardware and network performances.

With respect to the hardware performances, the nodes have been subdivided into the following "classes" according to their base specifications:

HW Class	Specifications
Low	2 CPU cores, 8 GB RAM
Medium	4 CPU cores, 8/16 GB RAM
High	8+ CPU cores, 32 GB RAM

With respect to the cloud and network service providers, the following were used:

Cloud Providers
EGI Federated Cloud
Amazon AWS
Google
Microsoft Azure
OneProvider
Orange
OVH
T-Systems
CloudFerro
GRNET
1&1 IONOS

Funded by the EU
Copernicus esa European Space Agency

The EGI Federated Cloud[18] services were selected to represent data access usage from research institutes, universities and so on. Commercial cloud providers like Amazon, Google and Microsoft were chosen because they offer the possibility to choose the hardware configuration of the virtualized platform and also due to their wide spread geographic distribution that allows a good variety of choices. Other commercial providers were added to the pool due to limited availability of commercial offers in some specific geographic areas.

As additional sites are procured, the spectrum of providers will be enlarged to allow more diversity.

7.2 NETWORK CALIBRATION

As presented above, the interpretation of the results obtained from the various test sites around the world needs to take into account both geographic considerations (i.e. the physical distance between the test site and the target site) and the connectivity which might also be related to the chosen provider and the HW performances. Oscillations over time might be due not only to variations on the target sites and of the internet but also reflect variations of the test sites. In order to disentangle all these components, and to guarantee an independent and absolute (as much as feasible) reference, a test sites calibration is periodically performed across the network.

The purpose of the test sites calibration is to evaluate the test sites on the basis of their intrinsic connectivity, independently from the Target sites that are being benchmarked. This is not trivial because the estimation of the network distance between two hosts is not straightforward given that it depends on several factors such as the chosen protocol, the router decision for path...

This kind of evaluations are generally accomplished by measuring the speed of data sent to/from given computers to a network of target servers of known and high performances located at different places around the world. Different services exist that provide these services. For the benchmarking project, the SpeedTest network has been chosen that offers accurate and high-quality performances thanks to a network of more than 8000 servers around the world having high upstream and downstream capacity (see https://www.speedtest.net/). This ensures that quasi-local readings can be obtained for almost all the test sites[19].

Speedtest retrieves the following metrics:

- The *Speed score*, which is a measure of the download and upload speeds ranked against the overall speed performance of the servers network.
- The *Ping (or Latency or RTT) score*, which is the amount of time it takes for a packet to go from the sender to the receiver and back (a "round trip"). This is the minimum amount of time for one host to get information back from the other host about data that was sent. Here as well, statistically significant values are computed.

In order to have a single indicator to compare the different test sites, the above metrics are combined to form a specific **Calibration Quality Indicator (CQI)** as follows:

[18] EGI is a federated e-Infrastructure set up to provide advanced computing services for research and innovation. The EGI e-infrastructure is publicly-funded and comprises hundreds of data centres and cloud providers spread across Europe and worldwide. The EGI Federated Cloud is a IaaS-type cloud, made of academic private clouds and virtualised resources and built around open standards. See https://www.egi.eu/services/

[19] Other network tools that have been evaluated are iperf3 and Traceroute. These were discarded because the first one provides info on speed download, but for upload input port on test site needs to be opened, while for Traceroute the RTT from test site to target site provides info on network connection which is not dependent on the test site but from routers on internet. Maybe only info on local network (first few hops) can be meaningful.

Funded by the EU
Copernicus esa European Space Agency

CQI = 80 * *TrimeanBalanced(Speed)* + 20 * *Trimean(RTT)*

Where:

- *TrimeanBalanced(Speed)* is composed by using the trimean[20] value of the speed score in which the download speed is weighted by 90% and the upload speed is weighted at 10%.
- *Trimean(RTT)* corresponds to a modified Trimean of the RTT, computed with a formula that takes values from the 10th percentile, 50th percentile (also known as the median), and 90th percentile, and combine them in a weighted average using a 1:2:1 ratio, respectively

C1 provides an estimate of the overall quality of a given test site, ranging between 0 (poorest) to 100 (highest).

It must be noted that the choice of the target server to be used among the ones provided by SpeedTest is not trivial. Generally, the nearest target server is chosen in order to minimize the degradation over distance and have the most reliable estimation of the "local connectivity" conditions. However, in addition to the nearest one, also the servers closest to the target sites are used, in order to evaluate the "connectivity distance" for each test-to-target connection. The corresponding Speedtest servers (i.e. closest to the test site and closest to the target site) are identified by minimizing the network distance. This is calculated as a function of the RTT as follows:

$$\text{d\_net (km)} = 0.7 \cdot 3 \cdot 10^5 \frac{km}{s} \cdot \frac{RTT}{2} \cdot 10^{-3} s = 105 \cdot RTT \text{ (km)}$$

Under the assumptions that:

- the propagation delay predominates,
- the bulk of cable connection is via optical fibre,
- the propagation within the fiber is 70% the speed of light.

This implies that for each test and each target site, two scores are obtained:

- One which is computed with respect to the closest speedtest server (CQI_Local) which provides an indication of the "local connectivity" of the test site and
- One which is computed with respect to the speedtest server which is closest to the target site (CQI_Target), which gives an indication of the "network distance" between the test site and the target site.

The two are then combined (50% + 50%) to provide a generic score CQI_Total that gives an estimation of the overall connectivity along the connection.

Calibration tests are regularly executed during a reporting period for periods sufficiently long to gather reliable statistics (e.g. at least one full week), and at different times of the day in order to avoid e.g. diurnal cycles due to network congestion, etc.

[20] https://en.wikipedia.org/wiki/Trimean

Funded by the EU
opernicus esa European Space Agency

8 ANNEX C – BENCHMARK USER SCENARII

This Annex describes:

- all the benchmark complex scenarii (or test scenario) Subsection C.1
- all the benchmark basic user scenario (or test cases) Subsection C.2
- all the benchmarking manual procedures (or test procedures) in Subsection C.3

Each scenario/procedure is linked to:

- a specific configuration, related to the input conditions of the scenario; and
- a set of metrics describing the outputs that are measured.

Configuration parameters are detailed in Annex D to this document while metrics are described in Annex E.

8.1 COMPLEX USER SCENARII (TEST SCENARII)

Test Scenario Identifier	#TS 01
Test Scenario Name	Simple data search and single download
Test Objective and Functionality	Test the target site general availability for a simple download
Description	This scenario running at high frequency aims at testing all basic data services: • Website availability • Search Service • Download capabilities
Applicable Interface Type	Application programming interface (API)
Indicative Test Frequency	Hourly
Indicative Configuration	Mission Configuration Dictionary Load factor = 3 max parallelism = [1-10] (depends on target site limitation) More info at ANNEX D – Operational Configuration
Test Cases orchestration	#TC101→ #TC201 → #TC301
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M026 totalSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	N/A
SW code references	cdab-tools / DotNetOpenSearchScihub

Funded by the EU
Copernicus esa European Space Agency

Test Scenario Identifier	#TS 02
Test Scenario Name	Complex data search and bulk download
Test Objective and Functionality	Test the target site general availability for multiple complex downloads
Description	This scenario running at medium frequency aims at testing all data services with a minimum of complexity: • Website availability • Search Service (complex) • Download capabilities (concurrent)
Applicable Interface Type	Application programming interface (API)
Indicative Test Frequency	Daily
Indicative Configuration	Mission Configuration Dictionary Load factor = 3 max parallelism = [1-10] (depends on target site limitation) More info at ANNEX D – Operational Configuration
Test Cases orchestration	#TC101→ #TC202 → #TC302
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M026 totalSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	N/A
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 03
Test Scenario Name	Systematic periodic data search and related remote data download
Test Objective and Functionality	Test the target site general availability for large systematic multiple complex download fixed in time
Description	This scenario running at low frequency aims at testing all data services with complexity and systematicity: • Website availability • Search Service (complex, large & fixed) • Download capabilities (concurrent & large)
Applicable Interface Type	Application programming interface (API)

Funded by the EU
opernicus
esa
European Space Agency

Indicative Test Frequency	Weekly
Indicative Configuration	Bulk Systematic Search Dictionary Load factor = 3 max parallelism = [1-10] (depends on target site limitation) More info at ANNEX D – Operational Configuration
Test Cases orchestration	#TC203 → #TC303
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M026 totalSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	N/A
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 04
Test Scenario Name	Offline data download
Test Objective and Functionality	Test the target site for downloading offline data
Description	This scenario running at high frequency aims at testing all data services focused on offline data • Search Service (fixed targeting offline) • Download capabilities (offline)
Applicable Interface Type	Application programming interface (API)
Indicative Test Frequency	Hourly
Indicative Configuration	Offline Data URLs states Offline Search Dictionary Load factor = 3 (needed to ensure finding offline data) max parallelism = 1 (to request only 1 offline data at a time) More info at ANNEX D – Operational Configuration
Test Cases orchestration	#TC204 → #TC304
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize

Funded by the EU
opernicus esa European Space Agency

	Metric #M009 maxSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M017 offlineDataAvailabilityLatency
Dependencies	N/A
Other comments	This test scenario keeps a state with the URLs of offline data requested but not yet downloaded in order to poll data availability at successive runs.
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 05
Test Scenario Name	Data Coverage Analysis
Test Objective and Functionality	Test and evaluate the catalogue coverage of a target site by collection and other filters (geographic, time…). It also measures the online data ratio between the specific hub being tested and the Open Hub
Description	This scenario running at low frequency aims at analysing the data coverage of the target site wrt to the reference site (API HUB)
Applicable Interface Type	Application programming interface (API)
Indicative Test Frequency	Weekly
Indicative Configuration	Catalogue sets (Global Baselines) Data provider Local Coverage Time Division Online/Offline Data (When Possible) Load factor = 20 (needed to ensure finding offline data) max parallelism = 1 (to request only 1 offline data at a time) More info at ANNEX D – Operational Configuration
Test Cases orchestration	#TC501 → #TC502
Relevant Metrics	Metric #M015 catalogueCoverage Metric #M016 dataCollectionDivision Metric #M022 totalReferenceResults Metric #M023 dataCoverage Metric #M024 dataOfferConsistency Metric #M025 totalResults
Dependencies	N/A
Other comments	This test scenario may run for several hours according to the size of the data coverage division (product types, geographical division, time division)
SW code references	cdab-tools / DotNetOpenSearchScihub

Funded by the EU
opernicus esa European Space Agency

Test Scenario Identifier	#TS 06
Test Scenario Name	Data Latency Analysis
Test Objective and Functionality	Test and evaluate the data latency (operational/availability) of a target site by collection
Description	This scenario running at high frequency aims at assessing the data availability latency
Applicable Interface Type	Application programming interface (API)
Indicative Test Frequency	Weekly
Indicative Configuration	Catalogue sets (Global Baselines) max parallelism = 2 More info at ANNEX D – Test Cases Configuration
Test Cases orchestration	#TC601 → #TC602
Relevant Metrics	Metric #M003 errorRate Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M013 avgDataAvailabilityLatency Metric #M014 avgDataOperationalLatency Metric #M016 dataCollectionDivision Metric #M018 maxDataOperationalLatency Metric #M019 maxDataAvailabilityLatency Metric #M020 totalValidatedResults Metric #M021 totalWrongResults Metric #M027 exception
Dependencies	N/A
Other comments	This test scenario can have different durations according to the size of the data coverage division (product types, geographical division, time division)
SW code references	cdab-tools / DotNetOpenSearchScihub

Funded by the EU
opernicus
esa
European Space Agency

Test Scenario Identifier	#TS 07
Test Scenario Name	Storage Upload and Download Performance
Test Objective and Functionality	Test and evaluate the performance of the target site's storage in terms of creation of storage locations and the upload and download to and from them
Description	This scenario running at medium frequency aims at assessing the storage performance
Applicable Interface Type	Application programming interface (API)
Indicative Test Frequency	Daily/Weekly
Indicative Configuration	More info at ANNEX D – Test Cases Configuration
Test Cases orchestration	#TC701 → #TC702
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize
Dependencies	N/A
Other comments	
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 09
Test Scenario Name	Users performing simple search and download of data by GUI
Test Objective and Functionality	Test the target site general availability for a simple download by Web site interface
Description	This scenario running at medium frequency aims at testing all basic data services: • Website availability • Search Service • Download capabilities
Applicable Interface Type	Graphical user interface (GUI)
Indicative Test Frequency	Daily
Indicative Configuration	Mission Configuration Dictionary Load factor = 3 max parallelism = [1] More info at ANNEX D – Operational Configuration
Test Cases orchestration	#TC209 → #TC309
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput

Funded by the EU
opernicus esa European Space Agency

	Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	N/A
SW code references	Cdab tools / Selenium

Test Scenario Identifier	#TS 11
Test Scenario Name	Cloud Services Simple local data search and single local download on single virtual machines
Test Objective and Functionality	Test DIAS target site cloud services availability by provisioning a virtual machine and running a test scenario similar to #TS01 from it.
Description	This scenario running at high frequency aims at testing all basic data services: • Cloud Service Availability • Search Service (Local) • Download capabilities (Local)
Applicable Interface Type	Application Programming Interface (API)
Indicative Test Frequency	Hourly
Indicative Configuration	Mission Configuration Dictionary Load factor = 10 max parallelism = [2-10] (depends on target site limitation) More info at ANNEX D – operational Configuration
Test Cases orchestration	#TC411 → (#TC211 → #TC311)
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M026 totalSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration

Funded by the EU
opernicus esa European Space Agency

Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote / cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 12
Test Scenario Name	Cloud Services Complex local data search and multiple local download on multiple virtual machines
Test Objective and Functionality	Test DIAS target site cloud services availability by provisioning several virtual machines and running a test scenario similar to #TS02 on each of them.
Description	This scenario running at medium frequency aims at testing all basic data services: • Cloud Service Availability • Search Service (Local) • Download capabilities (Local)
Applicable Interface Type	Application Programming Interface (API)
Indicative Test Frequency	Daily
Indicative Configuration	Mission Configuration Dictionary Load factor = 10 max parallelism = [2-10] (depends on target site limitation) virtual machines = [2-10] More info at ANNEX D – operational Configuration
Test Cases orchestration	#TC412 → (#TC212 → #TC312)
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M026 totalSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision Metric #M021 totalWrongResults Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration
Dependencies	N/A
Other comments	N/A

Funded by the EU
opernicus esa European Space Agency

SW code references	cdab-remote / cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 13
Test Scenario Name	Cloud Services Simple local data search, download and simple processing of downloaded data
Test Objective and Functionality	Test DIAS target site or third-party cloud provider (e.g. Amazon AWS or Google Cloud Platform) cloud services availability by provisioning several virtual machines and running a simple data transformation algorithm producing a new output, such as an image file.
Description	This scenario running at medium frequency aims at testing all basic data services: • Cloud Service Availability • Search Service (Local) • Download capabilities (Local) • Processing capabilities (Local)
Applicable Interface Type	Application Programming Interface (API)
Indicative Test Frequency	Daily
Indicative Configuration	Mission Configuration Dictionary Load factor = 10 max parallelism = [2-10] (depends on target site limitation) virtual machines = [2-10] More info at ANNEX D – operational Configuration
Test Cases orchestration	#TC413
Relevant Metrics	Metric #M003 errorRate Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration
Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote / cdab-tools / DotNetOpenSearchScihub

Test Scenario Identifier	#TS 15
Test Scenario Name	Cloud Services Processing of Specific Workflows
Test Objective and Functionality	Test DIAS target site or third-party cloud provider (e.g. Amazon AWS or Google Cloud Platform) cloud services availability by provisioning several virtual machines and running a specific transformation algorithm of typical EO applications. While the test involves also search and download, the focus is on the processing suitability for the selected areas of EO applications.

Funded by the EU
opernicus esa European Space Agency

Description	This scenario running at medium frequency aims at testing all basic data services: • Cloud Service Availability • Search Service (Local) • Download capabilities (Local) • Processing capabilities (Local)
Applicable Interface Type	Application Programming Interface (API)
Indicative Test Frequency	Daily
Indicative Configuration	Mission Configuration Dictionary Load factor = 10 max parallelism = [2-10] (depends on target site limitation) virtual machines = [2-10] More info at ANNEX D – operational Configuration
Test Cases orchestration	#TC415
Relevant Metrics	Metric #M003 errorRate Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration
Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote / cdab-tools / DotNetOpenSearchScihub

Funded by the EU
opernicus
esa
European Space Agency

8.2 BASIC USER SCENARII (TEST CASES)

Test Case Identifier	#TC 101
Test Case Name	Service Reachability
Test Objective and Functionality	Test the target site general availability
Description	Multiple concurrent remote HTTP web requests are performed to the front endpoint of the target site. Application metrics are reported.
Applicable Interface Type	Graphical user interface (GUI) and application programming interface (API)
Indicative Test Frequency	Hourly
Test Case configuration	No test data set Load factor and max parallelism parameter used for defining the total number of requests and the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M006 avgConcurrency Metric #M007 peakConcurrency
Dependencies	N/A
Other comments	This test case now includes formerly defined TC 103 (authentication) by sending authentication along with the request.
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 201
Test Case Name	Basic query
Test Objective and Functionality	Test a simple filtered search (e.g. mission, product type...) and validate the results
Description	Multiple concurrent remote HTTP web requests are sent to the front opensearch API of the target site using the opensearch mechanism to query and retrieve the search results. Searches are limited to simple filters (no spatial nor time filters) established randomly on the missions dictionary (see ANNEX D – Test Cases Configuration). Application metrics are reported.
Applicable Interface Type	Application programming interface (API) [OpenSearch]
Indicative Test Frequency	Hourly
Test Case configuration	Mission Configuration Dictionary Load factor defines the number of search queries to be performed Max parallelism parameter defines the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate

Funded by the EU
Copernicus esa European Space Agency

	Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	This test case may report systematic wrong result count (#M021 totalWrongResults) for a given target site if this target site does not implement all query filters
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 202
Test Case Name	Complex query (geo-time filter)
Test Objective and Functionality	Test a complex filtered search (e.g. geometry, acquisition period, ingestion date) and validate the results
Description	Multiple concurrent remote HTTP web requests are sent to the front catalogue search API (preferably opensearch API) of the target site using the search mechanism to query and retrieve the search results. N queries are prepared with all filters (spatial and time filters included) and composed with random filters from the missions dictionary (See ANNEX D – Operational Configuration). Application metrics are reported.
Applicable Interface Type	Application programming interface (API) [OpenSearch preferably]
Indicative Test Frequency	Daily
Test Case configuration	Mission Configuration Dictionary Load factor and max parallelism parameter used for defining the total number of requests and the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	This test case may report systematic wrong result count (#M021 totalWrongResults) for a given target site if this target site does not implement all query filters
SW code references	cdab-tools / DotNetOpenSearchScihub

Funded by the EU
Copernicus esa European Space Agency

Test Case Identifier	#TC 203
Test Case Name	Specific query (handle multiple results pages)
Test Objective and Functionality	Test a specific filtered search (e.g. geometry, acquisition period, ingestion date) with many results pages and validate the results
Description	Multiple concurrent remote HTTP web requests are sent to the front opensearch API of the target site using the opensearch mechanism to query and retrieve the search results over many results pages. Search filters are fixed (moving window in time). Application metrics are reported.
Applicable Interface Type	Application programming interface (API) [OpenSearch]
Indicative Test Frequency	Weekly
Test Case configuration	Specific Search Dictionary (bulk systematic/offline) Load factor and max parallelism parameter used for defining the total number of requests and the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	This test case may report systematic wrong result count (#M021 totalWrongResults) for a given target site if this target site does not implement all query filters
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 204
Test Case Name	Offline data query
Test Objective and Functionality	Test a simple filtered search (e.g. mission, product type...) for querying only offline data and validate the results
Description	Multiple concurrent remote HTTP web requests are sent to the front catalogue search API (preferably opensearch API) of the target site using the search mechanism to query and retrieve the search results. N queries are prepared with simple filters (no spatial nor time filters) + a specific filter to select offline data only and composed with random filters from the missions dictionary (see ANNEX D – Test Cases Configuration). Application metrics are reported.

Funded by the EU
opernicus
esa
European Space Agency

Applicable Interface Type	Application programming interface (API) [OpenSearch preferably]
Indicative Test Frequency	Hourly
Test Case configuration	Mission Configuration Dictionary Load factor defines the number of search queries to be performed Max parallelism parameter defines the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M008 avgSize Metric #M009 maxSize Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	This test case might be skipped if the target site does not offer a query filter to select only offline data.
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 211
Test Case Name	Basic query from cloud services
Other comments	All the same as #TC 201 except it is used in the cloud services test scenarios.

Test Case Identifier	#TC 212
Test Case Name	Complex query from cloud services
Other comments	All the same as #TC 202 except it is used in the cloud services test scenarios.

Test Case Identifier	#TC 209
Test Case Name	Basic query by GUI
Test Objective and Functionality	Test a simple filtered search (e.g. mission, product type...) and validate the results
Description	Single remote HTTP web requests are sent to the front web search page of the target site using the Selenium API to simulate a user's query and retrieve of the search results. Searches are limited to simple filters (time filters) established randomly on the missions dictionary (see ANNEX D – Test Cases Configuration). Application metrics are reported.
Applicable Interface Type	Graphical user interface (GUI) [Selenium API]

Funded by the EU
opernicus esa European Space Agency

Indicative Test Frequency	Daily
Test Case configuration	Mission Configuration Dictionary Load factor defines the number of search queries to be performed See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M016 dataCollectionDivision
Dependencies	N/A
Other comments	N/A
SW code references	cdab-tools / Selenium

Test Case Identifier	#TC 301
Test Case Name	Single remote online download
Test Objective and Functionality	Test the download service of the target site for online data
Description	Single remote download request is sent to retrieve a product file from a product URL. Application metrics are reported.
Applicable Interface Type	Application programming interface (API) [download]
Indicative Test Frequency	Hourly
Test Case configuration	Maximum Download Size
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize Metric #M016 dataCollectionDivision
Dependencies	If results are found from a previous test case such as #TC201 the URL for download is picked randomly from results of the search. Otherwise a new simple search is performed to find product to download
Other comments	Only online data is chosen for download
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 302
Test Case Name	Multiple remote online download
Test Objective and Functionality	Test the download capacity of the target site for online data using it maximum concurrent download capacity

Funded by the EU
opernicus esa European Space Agency

Description	Same as #TC 301 with as many concurrent download as the maximum allows
Applicable Interface Type	See #TC 301
Indicative Test Frequency	Daily
Test Case configuration	max_download_thread used for defining the total number of requests and the level of concurrency. Maximum Download Size. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize Metric #M016 dataCollectionDivision
Dependencies	Same as #TC301 with many URL picked randomly from results of previous searches
Other comments	Only online data are chosen for download
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 303
Test Case Name	Remote Bulk download
Test Objective and Functionality	Test the download capacity of the target site for downloading data in bulk
Description	Same as #TC 301 with as many download as the systematic search (#TC2013) returned
Applicable Interface Type	See #TC 301
Indicative Test Frequency	Weekly
Test Case configuration	max_download_thread used for defining the level of concurrency. Maximum Download Size.
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize Metric #M016 dataCollectionDivision
Dependencies	#TC203 All the URLs of the search are downloaded
Other comments	Only online data is chosen for download
SW code references	cdab-tools / DotNetOpenSearchScihub

Funded by the EU
opernicus esa European Space Agency

Test Case Identifier	#TC 304
Test Case Name	Offline download
Test Objective and Functionality	Test the capacity of the target site for downloading offline data
Description	Multiple concurrent remote HTTP web requests are sent to retrieve one or several product files from a set of selected URLs that are pointing to offline data. Application metrics are reported.
Applicable Interface Type	Application programming interface (API) [download]
Indicative Test Frequency	Hourly
Test Case configuration	Load factor defines the maximum total number of pending offline data waiting for download. max_download_thread used for defining the level of concurrency. Maximum Download Size.
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M012 resultsErrorRate Metric #M016 dataCollectionDivision Metric #M017 offlineDataAvailabilityLatency
Dependencies	#TC204 URLs pointing to an offline for download are picked randomly from results of the search
Other comments	Only offline data is chosen for download. A state file is maintained Test Case is skipped if no offline download is found from #TC204
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 309
Test Case Name	Single remote online download by GUI
Test Objective and Functionality	Test the download service of the target site for online data by web interface.
Description	Single remote download request is sent to retrieve a product file from a product URL. Application metrics are reported.
Applicable Interface Type	Graphical user interface (GUI) [Selenium API]
Indicative Test Frequency	Daily
Test Case configuration	Maximum Download Size
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency

Funded by the EU
opernicus esa European Space Agency

	Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize Metric #M016 dataCollectionDivision
Dependencies	The URL for download is picked randomly from results of previous test case #TC209.
Other comments	Only online data is chosen for download
SW code references	cdab-tools / Selenium

Test Case Identifier	#TC 311
Test Case Name	Single remote online download from cloud services
Other comments	All the same as #TC 301 except it is used in the cloud services test scenarios.

Test Case Identifier	#TC 312
Test Case Name	Multiple remote online download from cloud services
Other comments	All the same as #TC 302 except it is used in the cloud services test scenarios.

Test Case Identifier	#TC 411
Test Case Name	Cloud Services Single Virtual Machine Provisioning
Test Objective and Functionality	Test the cloud services capacity of the target site for provisioning a single virtual machine
Description	A remote web requests is sent using the cloud services API of the target site to request a typical virtual machine. Once the machine is ready, the test case executes a command within a docker container to start #TC211 and #TC311
Applicable Interface Type	Application programming interface (API) [cloud services]
Indicative Test Frequency	Hourly
Test Case configuration	N/A
Relevant Metrics	Metric #M003 errorRate Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration
Dependencies	N/A
Other comments	N/A

Funded by the EU
opernicus esa European Space Agency

SW code references	cdab-remote

Test Case Identifier	#TC 412
Test Case Name	Cloud Services Multiple Virtual Machine Provisioning
Test Objective and Functionality	Test the cloud services capacity of the target site for provisioning multiple virtual machines
Description	A remote web requests is sent using the cloud services API of the target site to request *n* typical virtual machines. Once the machine is ready, the test case executes a command within a docker container to start #TC212 and #TC312
Applicable Interface Type	Application programming interface (API) [cloud services]
Indicative Test Frequency	Daily
Test Case configuration	N/A
Relevant Metrics	Metric #M003 errorRate Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration
Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote

Test Case Identifier	#TC 413
Test Case Name	Cloud Services Virtual Machine Provisioning for Processing
Test Objective and Functionality	Test the cloud services capacity of the target site for provisioning virtual machines with the capability of running data-transforming algorithms
Description	A remote web requests is sent using the cloud services API of the target site to request a typical virtual machine. Once the machine is ready, the test case executes a command within a docker container to download a test product and run an algorithm to produce one or more outputs from it.
Applicable Interface Type	Application programming interface (API) [cloud services]
Indicative Test Frequency	Daily
Test Case configuration	N/A
Relevant Metrics	Metric #M003 errorRate Metric #M101 avgProvisioningLatency

Funded by the EU
opernicus esa European Space Agency

	Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration
Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote

Test Case Identifier	#TC 415
Test Case Name	Automated Processing of End-to-End Scenario of Specific Applications
Test Objective and Functionality	Test the cloud services capacity of the target site for provisioning virtual machines with the capability of running data-transforming pre-defined data-transforming algorithms in typical EO applications.
Description	A remote web requests is sent using the cloud services API of the target site to request a typical virtual machine. Once the machine is ready, the test case stages in its input data, executes a CWL workflow and compiles information about the success of the execution and related metrics.
Applicable Interface Type	Application programming interface (API) [cloud services]
Indicative Test Frequency	Daily
Test Case configuration	N/A
Relevant Metrics	Metric #M003 errorRate Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration Metric #M118 avgProcessDuration Metric #M119 processCount
Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote

Test Case Identifier	#TC 416

Funded by the EU
Copernicus esa European Space Agency

Test Case Name	Automated Processing of End-to-End Scenario of Specific Applications, additional metrics
Test Objective and Functionality	This test case can occur alongside #TC 415 to report additional metrics related to the end-to-end scenario in question.
Description	The test case builds on #TC415 and obtains additional metrics.
Applicable Interface Type	Application programming interface (API) [cloud services]
Indicative Test Frequency	Daily
Test Case configuration	N/A
Relevant Metrics	Metric #M003 errorRate Metric #M101 avgProvisioningLatency Metric #M110 cost Metric #M111 dataSummaryRun Metric #M112 flavorName Metric #M113 costHour Metric #M114 costMonth Metric #M115 duration Metric #M116 processDuration Metric #M117 totalDuration Metric #M118 avgProcessDuration Metric #M119 processCount
Dependencies	N/A
Other comments	N/A
SW code references	cdab-remote

Test Case Identifier	#TC 501
Test Case Name	Catalogue Coverage
Test Objective and Functionality	Test and evaluate the catalogue coverage of a target site by collection
Description	Multiple concurrent catalogue requests are sent to retrieve the total number of products for all the possible combinations of filters in configuration input When a timeliness is applicable on a collection, search are excluding the time critical items (e.g. NRT, STC) More details in ANNEX D - Test Cases Configuration - Data Coverage Benchmark Strategy
Applicable Interface Type	Catalogue programming interface (API) [search]
Indicative Test Frequency	Daily
Test Case configuration	Catalogue sets (Global Baselines) Geographic Areas Coverage Time Division Reference Target Site See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M015 catalogueCoverage Metric #M016 dataCollectionDivision Metric #M022 totalReferenceResults

Funded by the EU
Copernicus esa European Space Agency

	Metric #M025 totalResults
Dependencies	
Other comments	
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 502
Test Case Name	Local Data Coverage
Test Objective and Functionality	Test and evaluate the local data coverage of a target site for all product types collection
Description	Multiple concurrent catalogue web requests are sent to retrieve the total number of online and offline products for all the possible product types. More details in ANNEX D - Test Cases Configuration - Data Coverage Benchmark Strategy
Applicable Interface Type	Catalogue programming interface (API) [search]
Indicative Test Frequency	Daily
Test Case configuration	Catalogue sets (Global Baselines) Reference Target Site See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M016 dataCollectionDivision Metric #M022 totalReferenceResults Metric #M023 dataCoverage Metric #M025 totalResults
Dependencies	
Other comments	
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 503
Test Case Name	Data Offer Consistency
Test Objective and Functionality	Test and evaluate the local data consistency of a target site by data offer collection
Description	Multiple concurrent catalogue web requests are sent to retrieve the total number of online and offline products for all the possible product types. More details in ANNEX D - Test Cases Configuration - Data Coverage Benchmark Strategy
Applicable Interface Type	Catalogue programming interface (API) [search]
Indicative Test Frequency	Weekly
Test Case configuration	Local Collections Dictionary Reference Target Site See ANNEX D – Operational Configuration

Funded by the EU
opernicus esa European Space Agency

Relevant Metrics	Metric #M016 dataCollectionDivision Metric #M022 totalReferenceResults Metric #M024 dataOfferConsistency Metric #M025 totalResults
Dependencies	
Other comments	
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 601
Test Case Name	Data Operational Latency Analysis [Time Critical]
Test Objective and Functionality	Test and evaluate the data latency of a target site by collection
Description	Multiple concurrent remote HTTP web requests are sent to retrieve the latest products per collection and compare their data publication time to the sensing time. A timeliness is applied on a collection when applicable to limit the search to the time critical items (e.g. NRT, STC) More details in ANNEX D - Test Cases Configuration - Data Latency Benchmark Strategy
Applicable Interface Type	Application programming interface (API) [search]
Indicative Test Frequency	Hourly
Test Case configuration	Mission Configuration Dictionary Catalogue sets (Global Baselines) Max parallelism parameter defines the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M014 avgDataOperationalLatency Metric #M016 dataCollectionDivision Metric #M018 maxDataOperationalLatency Metric #M020 totalValidatedResults Metric #M021 totalWrongResults Metric #M027 exception
Dependencies	reuse of #TC201 to search for data
Other comments	This test case may be skipped if the target site does not offer a query filter to select the timeliness
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 602
Test Case Name	Data Availability Latency Analysis
Test Objective and Functionality	Test and evaluate the availability data latency of a target site by collection with respect to a reference target site
Description	Multiple concurrent remote HTTP web requests are sent to retrieve the latest products per collection and compare their data publication time to the sensing time. When a timeliness is applicable on a

Funded by the EU
opernicus esa European Space Agency

	collection, searches are excluding the time critical items (e.g. NRT, STC). More details in ANNEX D - Test Cases Configuration - Data Latency Benchmark Strategy
Applicable Interface Type	Application programming interface (API) [search]
Indicative Test Frequency	Hourly
Test Case configuration	Mission Configuration Dictionary Catalogue sets (Global Baselines) Max parallelism parameter defines the level of concurrency Reference Target Site See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M003 errorRate Metric #M010 totalReadResults Metric #M011 maxTotalResults Metric #M013 avgDataAvailabilityLatency Metric #M016 dataCollectionDivision Metric #M019 maxDataAvailabilityLatency Metric #M020 totalValidatedResults Metric #M021 totalWrongResults Metric #M027 exception
Dependencies	reuse of #TC201 to search for data
Other comments	This test case may be skipped if the target site does not offer a query filter to select the timeliness
SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 701
Test Case Name	Data Storage Upload Analysis
Test Objective and Functionality	Test and evaluate the upload performance on a target site's storage
Description	A large file is randomly generated and uploaded to a newly created storage. Typical transmission performance metrics are recorded during the process.
Applicable Interface Type	Application programming interface (API) [search]
Indicative Test Frequency	Daily
Test Case configuration	Max parallelism parameter defines the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize
Dependencies	
Other comments	

Funded by the EU
Copernicus esa European Space Agency

SW code references	cdab-tools / DotNetOpenSearchScihub

Test Case Identifier	#TC 702
Test Case Name	Data Storage Download Analysis
Test Objective and Functionality	Test and evaluate the upload performance on a target site's storage
Description	The file uploaded in #TC701 is downloaded from the storage which is deleted upon completion. Typical transmission performance metrics are recorded during the process.
Applicable Interface Type	Application programming interface (API) [search]
Indicative Test Frequency	Daily
Test Case configuration	Max parallelism parameter defines the level of concurrency. See ANNEX D – Operational Configuration
Relevant Metrics	Metric #M001 avgResponseTime Metric #M002 peakResponseTime Metric #M003 errorRate Metric #M005 throughput Metric #M006 avgConcurrency Metric #M007 peakConcurrency Metric #M009 maxSize Metric #M026 totalSize
Dependencies	requires #TC701 to have run previously
Other comments	
SW code references	cdab-tools / DotNetOpenSearchScihub

Funded by the EU
Copernicus esa European Space Agency

8.3 END-TO-END SCENARII

User Scenario Identifier	#E1
User Scenario Name	NDVI Mapping
Test Objective and Functionality	This scenario concerns the evaluation of the suitability of the target site for a user wishing to compute value-added products relevant to monitor the status of vegetation e.g. agriculture and forestry. The scenario involves very simple processing or manipulation of same-product parts implying e.g. extraction of features from a given product and computation of a generic index (e.g. bands manipulation). Emphasis is on the capability for an entry user to relocate a simple processing from the small program on the scientist laptop requiring long data downloads, to a cloud processing on demand. Data loads are not heavy sized (e.g. few tens of products) but would require few hours for the user to get on a local computer. Time constraints are relaxed. The availability of pre-installed tools such as e.g. GDAL and numpy is considered an asset that greatly facilitate the work of the user
Description	The user wishes to compute some normalized difference vegetation index (NDVI) maps using a simple SW program in python. The dataset of relevance for this scenario is limited to a few tiles from Sentinel-2 L1C data. The geographic location is variable over continental land (e.g. France). The time span covers from 1 day to 1 week and is limited to fresh data (i.e. no off-line data is involved). Filters to search for Sentinel2 Level 1 over a specific AOI for a given timespan over land area excluding too high cloud coverage.
Applicable Interface Type	Application programming interface (API), Manual test procedures
Indicative Test Frequency	At every benchmarking period Monthly for the automated part
Indicative Configuration	More info at ANNEX D – Operational Configuration
Test Cases orchestration	See Activity diagram on 2.3 Automatic procedures: #TS05, #TS06, #TS07, #TS11, #TS12, #TS13, #TS15 Manual procedures: #TP01 .. #TP12
Relevant Metrics	#M209 integrationToolsAdequacy #M210 integrationEnvSetupTimespan #M211 integrationCompleteness #M213 integrationSupportAnswerTime #M214 integrationSupportThreadLength #M215 integrationSupportAppreciation #M216 storageSetupTimeSpan #M217 storageCostPerGB #M219 appStorageAvailability #M220 integrationCost #M001 avgResponseTime #M005 throughput #M001 avgResponseTime #M005 throughput #M301 dataQueryFiltersAdequacy #M101 avgProvisioningLatency

Funded by the EU
Copernicus esa European Space Agency

	#M110 cost #M501 dataExploitationToolsAdequacy Q1_score Q3_score Q4_score Q53_score Q54_score
Dependencies	N/A
Other comments	N/A
SW code references	https://github.com/esa-cdab/cdab-testsuite

User Scenario Identifier	#E2
User Scenario Name	Rapid Mapping
Test Objective and Functionality	This scenario concerns the evaluation of the suitability of the target site for a user wishing to compute value-added products relevant to create burned area analysis map and NIR/SVWI RGB composites after a wildfire event using Sentinel-2 MSI L2A products. The scenario involves the creation of a directed acyclic graph (DAG) for SNAP and to call the SNAP gpt utility to calculate the Normalized Burn Ratio (dNBR) and the Relativized Burn Ratio (RBR) between pre-fire and post-fire acquisitions from within a Jupyter Notebook. Focus is on stringent timeliness and the capability through catalogue search of identification of best pre- and post- event images.
Description	The user wishes to calculate the Normalized Burn Ratio (dNBR) and the Relativized Burn Ratio (RBR) between pre-fire and post-fire acquisitions. The Burned Area map is obtained through the intersection between AOI and the pre-fire, post-fire Sentinel-2 chosen as input products. The dataset of relevance for this scenario is limited to Sentinel-2 MSI L2A products over a continental land (e.g. Australia). The time span covers a few days and is limited to fresh data (i.e. no off-line data is involved). Timeliness of availability of latest Sentinel-2 acquisitions is very important. Filters to search for Sentinel-2 MSI L2A over a specific AOI/geoname for a given timespan over land area with the post-fire image very close in time to the fire date is important.
Applicable Interface Type	Application programming interface (API) Jupyter Notebook Manual test procedures
Indicative Test Frequency	At every benchmarking period Monthly for the automated part
Indicative Configuration	More info at ANNEX D – Operational Configuration
Test Cases orchestration	See Activity diagram on 2.3 Automatic procedures: #TS05, #TS06, #TS07, #TS11, #TS12, #TS13 Manual procedures: #TP01 .. #TP12

Funded by the EU
opernicus esa European Space Agency

Relevant Metrics	#M209 integrationToolsAdequacy #M210 integrationEnvSetupTimespan #M211 integrationCompleteness #M213 integrationSupportAnswerTime #M214 integrationSupportThreadLength #M215 integrationSupportAppreciation #M216 storageSetupTimeSpan #M217 storageCostPerGB #M219 appStorageAvailability #M220 integrationCost #M001 avgResponseTime #M005 throughput #M001 avgResponseTime #M005 throughput #M301 dataQueryFiltersAdequacy #M101 avgProvisioningLatency #M110 cost #M501 dataExploitationToolsAdequacy
Dependencies	N/A
Other comments	N/A
SW code references	https://github.com/esa-cdab/cdab-testsuite

User Scenario Identifier	#E3
User Scenario Name	Rapid mosaicking / Land monitoring
Test Objective and Functionality	This scenario concerns the computation of a generic raster derived from mosaicking Sentinel-3 L2 OLCI data of the same type over a regional size area (e.g. Western Africa) and within a given time window (e.g. 10 days composites). The purpose of this scenario is to evaluate the capability for the target site to handle a large number of datasets in a single processing to produce multi temporal products. Focus is on timeliness and mosaicking of the recently acquired data but also on the processing performances when large quantities of data are involved. The other important aspect is the capacity for the target site to provide with ready to use time-driven processing scheduling/orchestration tools allowing to process systematically the mosaic every N days.
Description	The user is a specialist in Earth system science, using satellite remote sensing techniques, responsible for scientific research projects in connection with satellite missions, with a scientific expertise in land, hydrology and forestry. The user and his team are often on the field for in-situ campaigns using satellite remote sensing data. They developed a quick and relatively simple program (black-box) to mosaic OLCI data to produce products allowing a global monitoring of the land evolution (e.g. vegetation, desert extent, inland water bodies).
Applicable Interface Type	Application programming interface (API), Manual test procedures
Indicative Test Frequency	At every benchmarking period Monthly for the automated part

Funded by the EU
opernicus esa European Space Agency

Indicative Configuration	TBD More info at ANNEX D – Operational Configuration
Test Cases orchestration	See Activity diagram on 2.3 Automatic procedures: #TS05, #TS06, #TS07, #TS11, #TS12, #TS13, #TS15 Manual procedures: #TP01 ... #TP12_10
Relevant Metrics	#M209 integrationToolsAdequacy #M210 integrationEnvSetupTimespan #M211 integrationCompleteness #M213 integrationSupportAnswerTime #M214 integrationSupportThreadLength #M215 integrationSupportAppreciation #M216 storageSetupTimeSpan #M217 storageCostPerGB #M219 appStorageAvailability #M220 integrationCost #M001 avgResponseTime #M005 throughput #M001 avgResponseTime #M005 throughput #M301 dataQueryFiltersAdequacy #M101 avgProvisioningLatency #M110 cost #M501 dataExploitationToolsAdequacy #M049 wfTools Q1_score Q3_score Q4_score Q53_score Q54_score
Dependencies	N/A
Other comments	N/A
SW code references	N/A

User Scenario Identifier	#E4
User Scenario Name	Trends Assessment
Test Objective and Functionality	This scenario concerns processing of multiple data of the same type over long-time windows to build land surface temperature time series from Sentinel-3 SLTRS Level 2 data. Concerned time spans and area sizes can range from medium to very large ones. Focus is on the capacity for the target site to provide with a data-driven processing scheduling mechanism allowing to process systematically new acquired data to produce quickly new complete time series. Options for storage and products exploitation/visualization (e.g. catalogue, WMS, Time Series viewer) may be at stake.

Funded by the EU
Copernicus esa European Space Agency

Description	The user is a scientist working in a Climate Monitoring Institute. He has strong thematic background and intermediate computer science training. He has developed in-house a LST processing service providing land surface temperature time series based on SLSTR instruments data that he wants to deploy in a Cloud environment.
Applicable Interface Type	Application programming interface (API) Manual test procedures
Indicative Test Frequency	At every benchmarking period Monthly for the automated part
Indicative Configuration	TBD More info at ANNEX D – Operational Configuration
Test Cases orchestration	See Activity diagram on 2.3 Automatic procedures: #TS05, #TS06, #TS07, #TS11, #TS12, #TS13, #TS15 Manual procedures: #TP01 .. #TP12_10
Relevant Metrics	#M209 integrationToolsAdequacy #M210 integrationEnvSetupTimespan #M211 integrationCompleteness #M213 integrationSupportAnswerTime #M214 integrationSupportThreadLength #M215 integrationSupportAppreciation #M216 storageSetupTimeSpan #M217 storageCostPerGB #M219 appStorageAvailability #M220 integrationCost #M001 avgResponseTime #M005 throughput #M001 avgResponseTime #M005 throughput #M301 dataQueryFiltersAdequacy #M101 avgProvisioningLatency #M110 cost #M501 dataExploitationToolsAdequacy #M049 wfTools Q1_score Q3_score Q4_score Q53_score Q54_score
Dependencies	N/A
Other comments	N/A
SW code references	N/A

User Scenario Identifier	#E5
User Scenario Name	Interferogramme computation
Test Objective and Functionality	This scenario concerns the evaluation of the suitability of the target site for a user wishing to process Sentinel-1 SLC data to generate

Funded by the EU
Copernicus esa European Space Agency

	an interferogram over an area of interest in a predetermined period corresponding to the occurrence of a well-known earthquake event. He uses the Sentinel Application Platform (SNAP) to generate the interferogram. He also wants to prepare a multi-temporal stacks of Sentinel-1 acquisitions for later generation of Earth deformation time series and mean velocity maps. Focus is on tools facilitating the data search for the identification of interferometric pairs and for building and saving/sharing multi-temporal stacks of data. A powerful processing environment able to cope with Intensive I/O and high CPU load is needed to run the algorithm efficiently. Data loads are medium to large size (e.g. up to 150 Sentinel-1 scenes).
Description	The user wishes to compute a directed acyclic graph (DAG) for SNAP and does a system call to the SNAP gpt utility to process it using a simple SW program in python. Interferometric pairs are selected as follows: the user selects Sentinel-1 the reference S1 SLC acquisition (referred to as reference), and then uses the catalog discovery mechanism to select the "paired" acquisition (referred to as secondary) setting the same orbit track and direction and an overlap of at least 80% between the two acquisitions. The DAG includes all the steps for performing phase unwrapping relying on a CLI utility called snaphu. The dataset of relevance for this scenario is Sentinel-1 Level 1 SLC IW/EW mode data. The time span covers a few weeks for the interferogram but can also cover several years in the past for the multitemporal stack (i.e. off-line data is involved).
Applicable Interface Type	Application programming interface (API) Manual test procedures
Indicative Test Frequency	At every benchmarking period Monthly for the automated part
Indicative Configuration	TBD More info at ANNEX D – Operational Configuration
Test Cases orchestration	See Activity diagram on 2.3 Automatic procedures: #TS05, #TS06, #TS07, #TS11, #TS12, #TS13, #TS15 Manual procedures: #TP01 .. #TP12
Relevant Metrics	#M209 integrationToolsAdequacy #M210 integrationEnvSetupTimespan #M211 integrationCompleteness #M213 integrationSupportAnswerTime #M214 integrationSupportThreadLength #M215 integrationSupportAppreciation #M216 storageSetupTimeSpan #M217 storageCostPerGB #M219 appStorageAvailability #M220 integrationCost #M001 avgResponseTime #M005 throughput #M001 avgResponseTime #M005 throughput #M301 dataQueryFiltersAdequacy

Funded by the EU
Copernicus
esa European Space Agency

	#M101 avgProvisioningLatency #M110 cost #M501 dataExploitationToolsAdequacy Q1_score Q3_score Q4_score Q53_score Q54_score
Dependencies	N/A
Other comments	N/A
SW code references	https://github.com/esa-cdab/cdab-testsuite

User Scenario Identifier	#E6
User Scenario Name	High-demanding, massive processing with S-2 data
Test Objective and Functionality	This scenario concerns processing of multiple data of the same type (Sentinel-2 MSI) over long time windows to build: • NDVI time series with temporal and spatial gaps • Smoothed temporal and spatial gap filled NDVI time series employing the Whittaker filetring method. Concerned time range can span from medium sizes to very large ones (but at least three years of data). A number of small AOIs (e.g. covered by a single Sentinel-2 tile) well spread over the world (e.g. one for each continent) will be considered. These could be e.g. selected drought affetcted areas. Focus is on the capacity of the target site to provide efficient access to large stacks of Sentinel-2 data as well as Kubernetes computing clusters and S3 object storage for running resource savvy computational workflows to batch process new acquired data to produce quickly new, complete time series. Options for storage and products exploitation/visualization (e.g. catalogue, WMTS) may be at stake.
Description	The user is Scientist working in a Food Security Monitoring Institute. He has strong thematic background and intermediate computer science training. He has developed in-house EO application packages using the Common Workflow Language (CWL) and containers for defining resource savvy computational workflows for detecting NDVI anomalies that he wants to deploy on a Cloud environment.
Applicable Interface Type	Application programming interface (API) Manual test procedures
Indicative Test Frequency	At a pre-defined benchmarking period Only once
Indicative Configuration	More info at ANNEX D – Operational Configuration
Test Cases orchestration	See Activity diagram on 2.3

Funded by the EU
opernicus esa European Space Agency

Relevant Metrics	#M023 dataCoverage #M240 dataDiscoveryAndAccess #M241 cloudProcessingScore #M242 cloudOptimizedToolsAdequacy
Dependencies	N/A
Other comments	N/A
SW code references	N/A

8.4 MANUAL PROCEDURES

Test Procedure Identifier	#TP01
Test Case Name	Register for a free account
Test Objective and Functionality	This procedure must be executed as a "normal user" that already visited the provider website. We assume the user can navigate through the provider portal and know the usual terminology necessary to sign up any usual services on the web.
Description	1. Go to the provider home page 2. Stop watch starts 3. Navigate to the registration page and sign up for a free account 4. Execute or wait all the required steps to complete the registration (e.g. email confirmation, admin approval...) 5. Once registration completed, login immediately to your personal space 6. Stop watch stops and record the **elapsed time [registrationTimeSpan]** (minutes)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	
Relevant Metrics	Metric #M201 registrationTimeSpan
Dependencies	
Other comments	

Test Procedure Identifier	#TP02
Test Case Name	Review Subscription documentation
Test Objective and Functionality	This procedure is a documentation review to assess the information provided to the user about the commercial subscription. This review consists of a series of questions that

Funded by the EU
opernicus esa European Space Agency

	can be answered by yes or no. All positive answers are summed to give the **Subscription Documentation Completeness** that must be reported by the operator.
Description	1. Login with the user account and go to the provider help and support 2. Use the different help pages, FAQs or support search to answer the following questions: a. Is there a page dedicated to subscription? b. Is there an explanation of the subscription cost model? (e.g. pay per use, monthly subscription, billing) c. Is there a list of the paying products with their price? d. Is there an how-to step-by-step to proceed with the subscription? e. Is there a list of the payment methods available? f. Is there one of the user's payment method (see variables) listed in the possible payment methods? g. Is there information about how how to check payment history, account balance (pay-per-use) or bills? h. Is there a contact form or a dedicated support category in the helpdesk for requesting support about the subscription or payments? 3. Sum up all the positive answers and split by the number of questions. Report the number in **Subscription Documentation Completeness [subscriptionDocCompleteness]** metric as percentage (%)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variables**: - An ordered set of payment's methods (see Annex E #M207)
Relevant Metrics	Metric #M202 subscriptionDocCompleteness
Dependencies	
Other comments	

Test Procedure Identifier	#TP03
Test Case Name	Request support for subscription
Test Objective and Functionality	This procedure must be executed as a "normal user" that is previously registered by the provider. We assume the user is familiar with the procedure to request for support to the provider's helpdesk.

Funded by the EU
Copernicus esa European Space Agency

	The aim is to report some metrics about the responsiveness on support request before subscription. It consists in contacting the helpdesk with a predefined message explaining the user needs in terms of paying services to integrate it's application with the payment methods at it's disposal
Description	Procedure: 1. Login with the user account and go to the provider helpdesk or support page 2. Create a request (ticket, form...) with the following message template replacing the placeholder with the variables Dear Support, I am *<User's profile description>*. What is the most adapted plan in your service offer that would meet my needs? If the proposed service requires a payment, what is the subscription procedure? Do you accept any of the following payment methods: *<Payment's methods>*. Thank you for your guidance. Best regards, <User name> 3. Submit the request and note the date and time 4. Wait for answers and reply until the answer is acceptable. If none of the proposed service is acceptable, close the request. 5. Report: – Time elapsed since step 3. Metric (minutes) **[subscriptionSupportAnswerTime]** – Number of iterations (number of reply) metric (#) **[subscriptionSupportThreadLength]** – Answers general appreciation metric (rank: 1 poor – 5 efficient) **[subscriptionSupportAppreciation]**
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variables**: – A description of the use case. – A description of the tools and services needed. – An ordered set of payment's methods (see Annex E #M207).
Relevant Metrics	Metric #M203 subscriptionSupportAnswerTime Metric #M204 subscriptionSupportThreadLength Metric #M205 subscriptionSupportAppreciation
Dependencies	
Other comments	

Funded by the EU
opernicus esa European Space Agency

Test Procedure Identifier	#TP04
Test Case Name	Subscribe for a commercial account
Test Objective and Functionality	This procedure measures the efficiency for a set of available payment methods proposed to fit user needs and priority. For that purpose, each use case defines an ordered set of payment methods. They should be as diverse as possible to cover as many as situation as possible and, for the scenario coherence, they should fit with the application purpose and scale. This scenario is intended to be executed at once to get the commercial account. The procedure must be adjusted in order to match the provider subscription procedure. If the procedure cannot be completed at once and requires to perform again some steps, the procedure is consider as 1 **failure** [**subscriptionFailures**]. If the provider's subscription procedure foresees a manual operation on it's side (e.g. approval, account upgrade...), this is not blocking and is not a failure.
Description	**Procedure**: 1. Login with the credentials used for the registration and go to your user space. 2. Stopwatch starts 3. Perform the subscription procedure as described by the provider either via it's documentation or it's answer to the question at the help-desk using the payment methods order priority. 4. The account demonstrates undoubtedly that a paying resources can be provisioned. For instance, account expose a credit or allows for starting a virtual machine. 5. Stop watch stops and record the **elasped time [registrationTimeSpan]** (minutes) 6. Record the **payment method priority order [subscriptionPaymentMethodPrioNumber]** (#) 7. Record the number of subscription **failures** [**subscriptionFailures**] (#)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variables**: - An ordered set of payment's methods (see Annex E #M207)
Relevant Metrics	Metric #M206 subscriptionTimeSpan Metric #M207 subscriptionPaymentMethodPrioNumber Metric #M208 subscriptionFailures
Dependencies	
Other comments	

Funded by the EU
opernicus esa European Space Agency

<table>
<tr><td>Test Procedure Identifier</td><td>#TP05</td></tr>
<tr><td>Test Case Name</td><td>Review application integration service offer</td></tr>
<tr><td>Test Objective and Functionality</td><td></td></tr>
<tr><td>Description</td><td>
Procedure:

Login with the credentials used for the registration and go to provider home page.
Find the section dedicated to the services and tools offer or the marketplace
Search one by one information about every tool or language defined in the set of tools and programming languages and fill a table similar to the sample below

<table>
<tr><td>Tool or language</td><td>Found in the doc [doc:0/0.3]</td><td>Related service [service:0/0.7]</td><td>Adequacy [adequacy]</td><td>Availability [# steps]</td></tr>
<tr><td>python</td><td>Yes by typing ‘python’ in the service search engine</td><td>Jupyter Notebook</td><td>1</td><td>Notebook immediately available on demand with login [1]</td></tr>
<tr><td>ESA SNAP toolbox</td><td>Yes by typing ‘esa snap’ in the service search engine</td><td>Docker container in a VM</td><td>0.7</td><td>As a precooked docker container that can be executed in a VM [3]</td></tr>
<tr><td>gdal</td><td>Yes in an application tutorial</td><td>GDAL package to be installed in VM</td><td>0.8</td><td>As an installable package in a VM [2]</td></tr>
<tr><td>java</td><td>Not found</td><td>Java openJdk package to be installed in VM</td><td>0.8</td><td>As an installable package in a VM [2]</td></tr>
</table>
The tool or languages can vary according to E2E scenario. For example NDVI Mapping only python and gdal are applicable (definitions in section 4).
The values in the table above are only as example.

Some notes for each column:

Found in the doc: 0.3 is granted if the doc related to the tool or language is findable quite easily (e.g. search engine, navigating the doc, listing the services...) < 10 minutes
Related service: 0.7 is granted if there is a related service that can enable the tool or the service. We consider there is a service even if the doc does not mention it but that it is implicitly available (e.g. installable via packager manager on the VM)
Adequacy: this is a assessment from 0 to 1 of the adequacy of the proposed services wrt to the need. For instance, if I need to work with ESA snap toolbox, if the related service if a docker container to be executed in a VM, this is not entirely adequate because this a GUI tool and proposed only as a command line. In addition the superposition of virtual layer (docker in VM) may affect performance.

</td></tr>
</table>

Funded by the EU
Copernicus esa European Space Agency

	– **Availability**: This is the number of macro-operations to be executed to have the services available. For instance, the jupyter notebook is available in 1 click. The ESA snap toolbox requires to start a VM, download the docker container and execute it to get to the tool. 4. Compute the following formula to obtain the adequacy of integration tools metric [**integrationToolsAdequacy**] $$\frac{\sum_i \frac{(doc + service) * adequacy}{\#steps}}{i}$$ Where i are the number of rows applicable for the specific scenario. Value as percentage (%)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	
Relevant Metrics	Metric #M209 integrationToolsAdequacy
Dependencies	
Other comments	

Test Procedure Identifier	#TP06
Test Case Name	Set-up Development Environment
Test Objective and Functionality	This procedure describes a manual provisioning for a development environment in the intent of integrating the user's application of the use case. We assume the use knows where to start with the procedure (e.g. starting a virtual machine) It includes starting the virtual machine or the notebook to work with but also the installation of all the tools necessary to perform the application integration procedure using the dependencies installation procedure. If at any moment, the operator cannot perform a step of the setup because – the operation has caused a bug or an unsolvable software error; – of a missing technical information of an insufficient documentation; – of a missing required software or toolbox; then the procedure is stopped with the and the procedure "Request support for application integration" is then started.
Description	1. Login with the credentials used for the registration and go to user dashboard

Funded by the EU
opernicus esa European Space Agency

	2. Stop watch starts 3. Proceed with the #TP06_10: *Development Environment installation procedure* 4. Stop watch stops and report the timeSpan as the metric [**integrationEnvSetupTimespan**] (minutes)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	**Variables**: Development Environment installation procedure (#TP06_10, #TP06_20, #TP06_30, #TP06_40, #TP06_50,)
Relevant Metrics	Metric #M210 integrationEnvSetupTimespan
Dependencies	
Other comments	

Test Procedure Identifier	#TP06_10
Test Case Name	Development Environment installation procedure
Test Objective and Functionality	This procedure details the steps needed to complete the development environment installation as described in procedure #TP06
Description	This end-to-end scenario requires the creation of a VM with at least 8 GB RAM and 100 GB free disk space. The installation procedure mainly involves the installation of the **conda** package manager and a number of packages, **snap** and otb among them. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP06_20
Test Case Name	Development Environment installation procedure
Test Objective and Functionality	This procedure details the steps needed to complete the developmente environment installation as described in procedure #TP06

Funded by the EU
opernicus esa European Space Agency

Description	This end-to-end scenario requires the creation of a VM with at least 8 GB RAM and 100 GB free disk space. The installation procedure mainly involves the installation of the **conda** package manager and a number of packages, **snap** and otb among them. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP06_30
Test Case Name	Development Environment installation procedure
Test Objective and Functionality	This procedure details the steps needed to complete the developmente environment installation as described in procedure #TP06
Description	This end-to-end scenario requires the creation of a VM with at least 8 GB RAM and 100 GB free disk space. The installation procedure mainly involves the installation of the **conda** package manager and a number of packages, **snap** and otb among them. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP06_40
Test Case Name	Development Environment installation procedure
Test Objective and Functionality	This procedure details the steps needed to complete the developmente environment installation as described in procedure #TP06

Funded by the EU
opernicus esa European Space Agency

Description	This end-to-end scenario requires the creation of a larger VM, preferably with 4 CPUs, 32 GB RAM and 100 GB free disk space. The installation procedure mainly involves the installation of the **conda** package manager and the **snap** package. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP06_50
Test Case Name	Development Environment installation procedure
Test Objective and Functionality	This procedure details the steps needed to complete the developmente environment installation as described in procedure #TP06
Description	This end-to-end scenario requires the creation of a larger VM, preferably with 4 CPUs, 32 GB RAM and 100 GB free disk space. The installation procedure mainly involves the installation of the **conda** package manager and the **snap** package. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Funded by the EU
Copernicus esa European Space Agency

Test Procedure Identifier	#TP07
Test Case Name	Perform Application Integration
Test Objective and Functionality	This procedure is completely defined by the use case and describe basic step to be reproduced on the development environment using the available tools. Each completed step of the procedure is point gained in the application integration score. When user is blocked, it should move to the request for support for application integration procedure. A blocking step is a step where the operator: – has caused a bug or an unsolvable software error – has not enough technical information in the documentation to perform the step does not have the adequate tool
Description	1. Login with the credentials used for the registration and go to user dashboard 2. Proceed with the #TP07_10: *Integration procedure* 3. Report metrics a. The % corresponding to the last completed step of the procedure [**integrationCompleteness**] b. The eventual blocking cause [**integrationBlockingCause**] (string)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variables**: Application Integration script
Relevant Metrics	Metric #M211 integrationCompleteness Metric #M212 integrationBlockingCause
Dependencies	
Other comments	

Test Procedure Identifier	#TP07_10
Test Case Name	Integration procedure
Test Objective and Functionality	This procedure details the steps needed to complete the integration procedure as described in procedure #TP07
Description	The integration is performed on a terminal on the virtual machine previously created. The procedure involves • The search and download of Sentinel-2 MSI L1C tiles over the AOI, • The execution of the NDVI script with the bands passed as inputs The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface

Funded by the EU
opernicus esa European Space Agency

Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP07_20
Test Case Name	Integration procedure
Test Objective and Functionality	This procedure details the steps needed to complete the integration procedure as described in procedure #TP07
Description	The integration is performed on a terminal on the virtual machine previously created. The procedure involves • The search and download of a pair of Sentinel-2 MSI L2A products (pre- and post-event), • The execution of a provided Jupyter notebook for the burnt area algorithm The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP07_30
Test Case Name	Integration procedure
Test Objective and Functionality	This procedure details the steps needed to complete the integration procedure as described in procedure #TP07
Description	The integration is performed on a terminal on the virtual machine previously created. The procedure involves The procedure involves • Searching on the target site's catalogue to select the Sentinel-3 OLCI Level 2 products for both Satellites, over a specific AOI from last 10 days, • The download of those products from the target site's data repository,

Funded by the EU
Copernicus esa European Space Agency

	• The execution of program generating the mosaic of the input scenes. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP07_40
Test Case Name	Integration procedure
Test Objective and Functionality	This procedure details the steps needed to complete the integration procedure as described in procedure #TP07
Description	The integration is performed on a terminal on the virtual machine previously created. The procedure involves • Searching on the target site's catalogue to select the stack of Sentinel-3 SLSTR Level 2 products for both Satellites, over a specific AOI and track from the beginning of the mission to-date, • The ordering of products (as the stack will include offline data) • The download of those products from the target site's data repository, • The execution of a python application generating the land surface temperature map for each date of the stack. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP07_50

Funded by the EU
Copernicus esa European Space Agency

Test Case Name	Integration procedure
Test Objective and Functionality	This procedure details the steps needed to complete the integration procedure as described in procedure #TP07
Description	The integration is performed on a terminal on the virtual machine previously created. The procedure involves • Searches on the target site's catalogue to select a pair of Sentinel-1 SLC products for the interferogram (pre- and post-event), • The download of those products from the target site's data repository, • The generation of the interferogram using the SNAP toolbox's graph processor **gpt**. • Catalogue searches for additional Sentinel-1 SLC products further in the past to build a stack of products, and the respective downloads. The steps are described in this section of the documentation.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP08
Test Case Name	Request support for application integration
Test Objective and Functionality	This procedure must be executed as a "normal user" that is previously registered by the provider. We assume the user is familiar with the procedure to request for support to the provider's helpdesk. The aim is to request support during an application integration. It should be executed when the procedure is blocked performing the application integration steps.
Description	1. Login with the user account and go to the provider helpdesk or support page 2. Create a request (ticket, form...) with the following message template replacing the placeholder with the variables

Funded by the EU
Copernicus esa European Space Agency

	Dear Support, I am *<User's profile description>*. I am currently integrating my application that <use case description> I encountered an issue during <setup or integration> I try to execute <step that is blocked> but <blocking cause> Could you please help me with this issue. Thank you in advance. Best regards, <User name> 3. Submit the request and note the date and time 4. Wait for answers and reply until the answer is acceptable. If none of the proposed service is acceptable, close the request. 5. Report: – Time elapsed since step 3. metric **[integrationSupportAnswerTime]** (minutes) – Number of iterations (number of reply) metric **[integrationSupportThreadLength]** – Answers general appreciation metric (rank: 1 poor – 5 efficient) **[integrationSupportAppreciation]**
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variables**: – A description of the use case. – The blocking step from procedure 4.2.6 or 4.2.7 – The blocking cause from procedure 4.2.6 or 4.2.7
Relevant Metrics	Metric #M213 integrationSupportAnswerTime Metric #M214 integrationSupportThreadLength Metric #M215 integrationSupportAppreciation
Dependencies	
Other comments	

Test Procedure Identifier	#TP09
Test Case Name	Setup User Storage
Test Objective and Functionality	This procedure describes a simple procedure to request or setup a user space storage dedicated for the use case. It can be a drive, a folder, a repository, any service proposed by the service provider to store persistently files on the platform. If possible this procedure should include the possibilty to store the integrated application of the user for a better usage in production (e.g. Docker hub registry)

Funded by the EU
Copernicus esa European Space Agency

Description	1. Go to the user dashboard where he can configure user storage 2. Stop watch starts 3. Start the storage setup following the documentation. For instance, create a drive or a repository. 4. Stop watch stops and report the timeSpan **[storageSetupTimeSpan]** (minutes) 5. Keep docker storage endpoint and credentials
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle (monthly/quarterly)
Test Case configuration	
Relevant Metrics	Metric #M216 storageSetupTimeSpan Metric #M217 storageCostPerGB
Dependencies	
Other comments	

Test Procedure Identifier	#TP10
Test Case Name	Store Application Container
Test Objective and Functionality	This procedure allows a user to store it's integrated application on the provider platform. Usually, this is a procedure to push a docker container in a docker registry. The procedure is necessary for the step "Data Processing". The procedure report the availability of this registry and the completeness of the container storage.
Description	1. Login with the credentials used for the registration and go to user dashboard *2.* Proceed with the #TP10_10: *Application build procedure* 3. Report the success and the docker id. **[appStorageAvailability]** (rank: 1 poor – 5 efficient)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variable**: – Application build procedure – Docker hub credentials (preferably from 4.2.9)
Relevant Metrics	Metric #218 **appStorageAvailability**
Dependencies	
Other comments	

Funded by the EU
opernicus esa European Space Agency

Test Procedure Identifier	#TP10_10
Test Case Name	Application build procedure
Test Objective and Functionality	This procedure details the steps needed to complete the application build procedure as described in procedure #TP10. Pre-requisite is to install the docker software on the virtual machine.

Description	1. Open a terminal on the previously set-up virtual machine
	2. Go to the Use case folder
	3. Launch docker build prefixing the name with the target site docker hub repository (here docker.terradue.com)
	`$ docker build -t docker.terradue.com/cdab-ndvi .`
	4. Push the docker to the hub
	`$ docker push docker.terradue.com/cdab-ndvi`

Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP10_20
Test Case Name	Application build procedure
Test Objective and Functionality	This procedure details the steps needed to complete the application build procedure as described in procedure #TP10. Pre-requisite is to install the docker software on the virtual machine.

Description	1. Open a terminal on the previously set-up virtual machine
	2. Go to the Use case folder
	3. Launch docker build prefixing the name with the target site docker hub repository (here docker.terradue.com)
	`$ docker build -t docker.terradue.com/cdab-ndvi .`
	4. Push the docker to the hub
	`$ docker push docker.terradue.com/cdab-ndvi`

Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	

Funded by the EU
Copernicus esa European Space Agency

Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP10_50
Test Case Name	Review application integration service offer
Test Objective and Functionality	This procedure details the steps needed to complete the application build procedure as described in procedure #TP10. Pre-requisite is to install the docker software on the virtual machine.

Description	1. Open a terminal on the previously set-up virtual machine
	2. Go to the Use case folder
	3. Launch docker build prefixing the name with the target site docker hub repository (here docker.terradue.com)
	`$ docker build -t docker.terradue.com/cdab-ndvi .`
	4. Push the docker to the hub
	`$ docker push docker.terradue.com/cdab-ndvi`

Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle
Test Case configuration	
Relevant Metrics	
Dependencies	
Other comments	

Test Procedure Identifier	#TP11
Test Case Name	Review Catalogue Data Access Adequacy
Test Objective and Functionality	This procedure allows a user to search for specific data using the data access service of the provider. More specifically, the procedure shall review the available filters that helps the user in its use cases. For instance, if the use case processes optical instrument dataset for processing reflectance over land, it is useful to have filters such as cloud coverage and land coverage.
Description	1. Login with the credentials used for the registration and go to the data access tool, catalogue GUI or API *2.* Make a data search in the use case data collection using all data filters

Funded by the EU
opernicus esa European Space Agency

	3. Report the % of available and effective filters as [**dataQueryFiltersAdequacy**] metric
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variable**: – Use Case Data Collection – Useful data filter list ○ AOI (point, bbox or wkt polygon) large and small ○ geoname ○ date / time ○ orbit direction / track ○ cloud coverage ○ land coverage ○ sensor mode, processing level
Relevant Metrics	Metric #M301 **dataQueryFiltersAdequacy**
Dependencies	
Other comments	

Test Procedure Identifier	#TP12
Test Case Name	Manual evaluation of the adequacy of the tool for data exploitation
Test Objective and Functionality	This procedure allows a user to use provided tools by the target site to visualize and exploit data. More specifically, the procedure shall review the available tools that helps the user in its results distribution. For instance, if the use case produces quicklooks, it could be useful to browse results on a map web app (e.g. geobrowser) with tiles (e.g. WMS).
Description	1. Login with the credentials used for the registration and go to the documentation 2. Make a search for every tool listed in the Set of Data visualization and distribution tools 3. Report the % of available and effective filters as [**dataExploitationToolsAdequacy**] metric
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	At least once per reporting cycle (monthly/quarterly)
Test Case configuration	**Variable**: – Set of Data visualization and distribution tools such as: ○ private storage area for sharing results (e.g. S3 bucket, HTTP file server, NextCloud, OwnCloud etc.)

Funded by the EU
opernicus esa European Space Agency

	○ access (via API) to a catalogue index allowing publication of basic metadata (e.g. a OGC OpenSearch Catalogue) ○ Access to a OGC WMS (e.g. GeoServer) where to publish the generated results for visualization and sharing ○ Access to interactive GIS tools to visualize and manipulate results (e.g. QGIS, ArcGIS etc.)
Relevant Metrics	Metric #M501 **dataExploitationToolsAdequacy**
Dependencies	
Other comments	

Test Procedure Identifier	#TP12_10
Test Case Name	Manual evaluation of the adequacy of the tool for process scheduling / orchestration
Test Objective and Functionality	The aim of this procedure is to inspect and search the provider environment to verify the availability of tools for orchestrating processes and applications (e.g. allowing the user to setup time driven or data driven workflows for events monitoring purposes)
Description	**wfTools** metric 1. Login with the credentials used for the registration and go to the documentation 2. Inspect the provider web site and offering, specifically regarding provided scheduling / workflow / orchestration tools, to verify if the following are available: ✓Crontab ✓Oozie ✓Airflow ✓Dagster ✓Apache NiFi ✓Prefect ✓NextFlow ✓Luigi 3. If the number of found tools is less than 2 assign a score of 10%. If it is 2 assign a score of 75%. If it is equal or greater than 3 assign a score of 100%. Report the assigned number in the metric.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	An active user account on the target provider might be needed to access certain parts of the web site.
Relevant Metrics	#M049 wfTools

Funded by the EU
opernicus esa European Space Agency

Dependencies	
Other comments	

Test Procedure Identifier	#TP13
Test Case Name	Complementary Offer inspection
Test Objective and Functionality	The aim of this procedure is to visually inspect and search the provider web site to verify the availability of the complementary offer elements defined in the elementary metrics which compose the Q9 indicator. For each metric, a series of questions will be presented that can be answered by yes or no, according to information gathered on the target web site. All positive answers are then summed to obtain the specific metric that must be reported by the operator.
Description	2. Open the provider's web site. 3. **copernicusCoreServices** metric: a. Go the pages describing the provider's data offer. b. Search which of the following Copernicus Core Services data are made available: ✓CAMS (Atmosphere) ✓CEMS (Emergency) ✓CLMS (Land) ✓CMEMS (Marine) ✓C3S (Climate Change) ✓Security c. Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%). 4. **otherData** metric: a. Go the pages describing the provider's data offer. b. Search which of the following non-Copernicus data (only EO data) are made available: ✓Landsat-5 ✓Landsat-7 ✓Landsat-8 ✓Envisat MERIS L1 ✓SMOS ✓S2GLC ✓Mapzen DEM ✓SRTM DEM ✓Jason-3 ✓Jilin-1 ✓KazEOSat ✓KOMPSAT ✓Sentinel-1 outside Open Hub scope ✓Sentinel-2 outside Open Hub scope ✓Sentinel-3 outside Open Hub scope ✓Sentinel-6A ✓Maxar GeoEye-1

Funded by the EU
Copernicus esa European Space Agency

	✓Maxar QuickBird ✓Maxar IKONOS ✓Maxar WorldView-1 ✓Maxar WorldView-2 ✓Maxar WorldView-3 ✓Maxar WorldView-4 ✓COSMO-SkyMed ✓Envisat ASAR L0 ✓Deimos-2 ✓SPOT 6 and 7 ✓Pleiades c. Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%). 5. **apiCompleteness** metric: a. Search which of the following functionalities are provided by the offered API, by searching the provider documentation or by relying on the results obtained in Q1-Q5 automated tests: ✓Full search filtering capabilities ✓Download online products ✓Request offline products for download ✓Products count b. Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%). 6. **cloudServices** metric: a. Inspect the provider web site and available documentation on service offering to find the answer to the following question: ✓Does the provider offer the possibility to access a cloud environment – for free or after payment – where users can create and manage virtual machines? ✓API for VM management ✓API for object storage management b. If the answer is "yes", the metric's value will be 100. If the answer is "no", the metric's value will be 0 (zero). 7. **otherServices** metric: a. Inspect the provider web site and available documentation on service offering to find the answers (yes/no) to the following questions: ✓Does the provider make available a marketplace where user-developed applications can be published? ✓Does the provider offer consulting services for special needs or more demanding projects? b. Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%). 8. **devTools** metric: a. Inspect the provider web site and offering, specifically regarding tools and aids for development (not specifically related to Earth

Funded by the EU
opernicus
esa
European Space Agency

	Observation), to verify if the following features are available: ✓ Jupyter Notebook ✓ Preconfigured development environments with preinstalled tools ✓ Big Data functions (e.g. Map Reduce Service) ✓ Time Series analysis b. Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%). 9. **devServices** metric: a. Inspect the provider web site and offering, specifically regarding advanced services for application development, to verify if the following features are available: ✓ Docker Hub ✓ Kubernetes ✓ Database as a service b. Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%). 10. **eoTools** metric: a. Inspect the provider web site and offering, specifically regarding provided Earth Observation tools, to verify if the following are available: ✓ Orfeo ToolBox ✓ SNAP ✓ Sen2Cor ✓ QGIS ✓ OGC Web Map Service (WMS) interface Sum up all the positive answers and divide by the number of questions. Report the number in the metric as percentage (%).
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per reporting cycle (monthly/quarterly)
Test Case configuration	An active user account on the target provider might be needed to access certain parts of the web site.
Relevant Metrics	#M041 copernicusCoreServices #M042 otherData #M043 apiCompleteness #M044 cloudServices #M045 otherServices #M046 devTools #M047 devServices #M048 eoTools
Dependencies	
Other comments	

Funded by the EU
copernicus esa European Space Agency

Test Procedure Identifier	#TP14
Test Case Name	Evaluate service support
Test Objective and Functionality	The aim of this procedure is to visually inspect and search the provider web site to evaluate the service support elements defined in the elementary metrics which compose the Q10 indicator. For each metric, a series of questions will be presented that can be answered by yes or no, according to information gathered on the target web site. All positive answers are then summed to obtain the specific metric that must be reported by the operator.
Description	1. Open the provider's web site. 2. **serviceDocCompleteness** metric: a. Login with the user account and answer the following questions: ✓Is there a link to documentation on the home page? ✓Is there a search engine for the documentation pages? ✓Is there a table of contents on the documentation page? ✓Is there a full API documentation? ✓Are tutorials available? b. Sum up all the positive answers and split by the number of questions. Report the number in **Service Documentation Completeness [serviceDocCompleteness]** metric as percentage (%) 3. **serviceSupportHelpDesk** metric: a. Check the Completeness of HelpDesk channels, verifying if the following communication channel exist: ✓contact form ✓phone support ✓email support ✓live chat ✓ticketing system b. Sum up all the positive answers and split by the number of questions. Report the number in Service Support Help Desk [serviceSupportHelpDesk] metric as percentage (%). 4. **serviceSupportAnswerTime** metric: a. For all the support requests, check the Time elapsed from first question and first human response on each support thread (considering only working hours 9-18), then compute the average and report in [**serviceSupportAnswerTime]** metric as minutes 5. **serviceSupportThreadLength** metric: a. For all the support requests, report the number of replies in a support thread), then compute the

Funded by the EU
opernicus esa European Space Agency

	average and report in [**serviceSupportThreadLength]** metric as number (#) 6. **serviceSupportAnswerSatisfaction** metric: a. For all the support requests, note the Answers general appreciation rank (1=poor – 5=efficient), then compute the average and report in [**serviceSupportAnswerSatisfaction]** metric as number (#)
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once per benchmarking campaign
Test Case configuration	An active user account on the target provider might be needed to access certain parts of the web site.
Relevant Metrics	#M230 serviceDocCompleteness #M231 serviceSupportHelpDesk #M232 serviceSupportAnswerTime #M233 serviceSupportThreadLength #M234 serviceSupportAnswerSatisfaction
Dependencies	
Other comments	

Test Procedure Identifier	#TP15
Test Case Name	Data Availability
Test Objective and Functionality	The aim of this procedure is to check the data coverage of the scenario. Data to be provided are Sentinel-2A/2B MSI L2A (and L1C) from start of the mission up to current date over 5 sites (one tile per site) that are subject to drought selected among the one listed below: - Lake Poyang - Donana National Park Spain - Horn of Africa - Parana' River basin Brazil - Gambela National Park Ethiopia
Description	For each of the previous site (Area Of Interest) retrieve the number of products to be processed from the target provider and from the ESA OpenHub. The following is just an example (L2A) on how to compute the **dataCoverage** metric: Area Of Interest | Tile ID | Relative orbit | Online products | API Hub query

Funded by the EU
opernicus esa European Space Agency

	<table><tr><td>Lake Poyang</td><td>50RMT</td><td>132</td><td>300</td><td>304</td></tr><tr><td>Doñana National Park</td><td>29SQB</td><td>137</td><td>354</td><td>357</td></tr><tr><td>Horn of Africa</td><td>39PVN</td><td>63</td><td>299</td><td>303</td></tr><tr><td>Paraná basin</td><td>21KZQ</td><td>124</td><td>302</td><td>305</td></tr><tr><td>Gambela National Park</td><td>36NXP</td><td>35</td><td>299</td><td>299</td></tr></table> This will lead to a metric of **dataCoverage** = 99,17 % respect to OpenHub. The same computation will be done for L1C products (leading for example to 92,05%) , then the max value between them will be taken as final metric. Further to the dataCoverage also the catalogueCoverage that counts also the Offline/orderable products will be measured. As before the max value between L2A and L1C product will be taken as final metric.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once
Test Case configuration	An active user account on the target provider is needed.
Relevant Metrics	#M023 dataCoverage #M015 catalogueCoverage
Dependencies	
Other comments	

Test Procedure Identifier	#TP16
Test Case Name	Data Discovery and Access
Test Objective and Functionality	The aim of this procedure is to check the data discovery and access for long series of Sentinel-2 data.
Description	The following is the procedure for the computation of the **dataDiscoveryAndAccess** metric. The answer to the question (1=true, 0=false) is multiplied by the weigth capped to 1 providing the partial term. The sum of partial terms divided by the number of questions and multiplied by 100, will provided the fine score as %. The answers in the below table is just an example of computation: <table><tr><td>1</td><td>Check data discovery and access for long series of Sentinel-2 data</td><td>(1=true, 0=false)</td><td>Weigth</td><td>Partial term</td></tr></table>

Funded by the EU
opernicus esa European Space Agency

	-	Provide access to historical Sentinel-2 data (SAFE or STAC/COG formats) [DATA-1] Does the provider expose a Sentinel-2 archive?	1	1	1
	-	Sentinel-2 discovery and access interfaces [DATA-2] Has the provider Sentinel-2 discovery and access interfaces?	1	1	1
	-	Sentinel-2 Collections [DATA-3]			
		Is Sentinel-2 L2A data discoverable and accessible?	1	1	
		No Sentinel-2 L2A, is Sentinel-2 L1C data discoverable and accessible?	0	0,5	1
	-	Data access and discovery tooling [DATA-4]			
		Can OpenSearch-client be used for discovery	1	0,5	
		Can STARS be used for data access	1	0,5	1
	-	Sentinel-2 format [DATA-5] - Which formats are available?			
		STAC/COG	0	1	
		Only SAFE	1	0,5	0,5
	2	Report the % of data discovery and access as [dataDiscoveryAndAccess] metric			90,0
Applicable Interface Type	Manual Human Machine Interface				
Indicative Test Frequency	Once				
Test Case configuration	An active user account on the target provider is needed.				
Relevant Metrics	#M240 dataDiscoveryAndAccess				
Dependencies					
Other comments					

Test Procedure Identifier	#TP17
Test Case Name	Processing Scenario
Test Objective and Functionality	The aim of this procedure is to compute a general score for the processing of multiple data of the same type (Sentinel-2 MSI) over long time windows.
Description	The procedure consists in: 1. Taking as input a Sentinel-2 acquisition it ingests and transforms it in a set of TOA/BOA reflectance single-band assets and a series of overview assets that provide three band combinations readily available in full

Funded by the EU
opernicus esa European Space Agency

	resolution. All assets are generated in COG format with associated STAC metadata 2. Generates seven radiometric indices (NDVI, NDMIR, NBR, NDWI, NDWI2, MNDWI, NDBI) for vegetation, water, fire and built-up classes derived from a combination of multiple optical bands of the previously pre-processed images. Results are stored in S3 3. Report the significative cluster processing metrics: - Report the % of test efficiency (total_job_runtime_hours/test_duration_hours) [**testEfficiency**] metric - Report the % of processing success rate (inputs_completed/inputs_processed) [**processing_success_rate**] metric - Report the % of processing time rate (total_processing_hours/total_pod_lifetime_hours) [**processing_time_rate**] metric The cluster processing metrics are extracted from the job logs and include the following eliglible for the score computation. A job is a sequential processing task for a given area of interest/time of interest (i.e.: target-lake-poyang-2018). A pod is a group of one or more containers, with shared storage and network resources, and a specification for how to run the containers (smallest deployable units of computing). **target:** target name **description:** description **data_catalog_provider**: name of catalog used for product search **input_data_provider**: name of provider for input data download (stage_in) **output_data_provider**: name of provider for output data upload (stage_out) **start_time**: start time of first job **finish_time**: end time of last job **test_duration_hours**: finish_time – start_time [hours] **quota**: info on provider's cluster quota (cpu cores, ram) **node**: info on VM flavour, #cpu cores and ram used **estimated_cost**: VM node resource unit costs **inputs_started**: # of images inputs **inputs_processed**: # of images processed **inputs_completed**: # of images completed **total_job_runtime_hours**: sum of elapsed time execution for all jobs [hours] **total_pod_lifetime_hours:** sum of lifetime of a single cluster pod (running in parallel) (hours) including the following phases: stage_in, calibration, spectral processing, stage out [hours] **total_processing_hours:** sum of only processing time for a single cluster pod (running in parallel): calibration and spectral processing [hours] **avg_processing_minutes:** total_processing_hours / inputs_processed [minutes]

Funded by the EU
opernicus
esa
European Space Agency

	avg_concurrency: average number of node used **avg_pod_concurrency:** average number of pods used **peak_concurrency:** max number of nodes used **total_s3_size_megabytes:** total size of output processed data [MB] **processing_success_rate:** inputs_completed/inputs_processed [%] **processing_time_rate:** total_processing_hours/total_pod_lifetime_hours [%] **test_efficiency_rate:** total_job_runtime_hours/test_duration_hours [%]
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once
Test Case configuration	An active user account on the target provider is needed.
Relevant Metrics	#M245 testEfficiency #M241 processingSuccessRate #M246 processingTimeRate
Dependencies	
Other comments	

Test Procedure Identifier	#TP18
Test Case Name	Cloud optimized services
Test Objective and Functionality	The aim of this procedure is to check the capability to provide cloud optimized services (such as Kubernates).
Description	The following is the procedure for the computation of the **cloudOptimizedToolsAdequacy** metric. The answer to the question (1=true, 0=false) is multiplied by the weigth capped to 1 providing the partial term. The sum of partial terms divided by the number of questions and multiplied by 100, will provided the fine score as %. The answers in the below table is just an example of computation:

1	Capability to provide cloud optimised services	(1=true, 0=false)	Weigth	Partial term
-	provide Kubernetes resources as part of their offering (Kubernetes-as-a-service) - [K8S-1]	0	1	
-	provide virtual machines that can be procured and configured as a Kubernetes cluster in case of lack of Kubernetes-as-a-service - [K8S-2]	1	0.5	0.5
-	provide upstream API compatible with Terraform	1	1	1
	provide storage mechanisms to			

Funded by the EU
Copernicus esa European Space Agency

		support the volume claims for the Kubernetes cluster (ReadWriteMany access mode) - [K8S-3]			
	-	ReadWriteMany access mode	1	1	
	-	ReadWriteOnce access mode	0	0.5	1
		support autoscaling capacity for processing and storage - [K8S-4]			
	-	autoscaling capacity	1	1	
	-	manual scaling capacity	0	0.5	1
	-	provide S3 based object storage for holding the produced results - [STO-1]	1	1	1
	2	Report the % of capability to provide cloud optimized services as [cloudOptimizedToolsAdequacy] metric			90.0

This will lead to a metric of **cloudOptimizedToolsAdequacy** = 90.0 % respect to OpenHub.

Moreover a **developerRank** measure is provided at the end of the scenario's execution based on the developer's experience in integrate and execute the scenario (0=poor – 5=efficient)

Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once
Test Case configuration	An active user account on the target provider is needed.
Relevant Metrics	#M242 cloudOptimizedToolsAdequacy #M247 developerRank
Dependencies	
Other comments	

Test Procedure Identifier	#TP19
Test Case Name	Processing Cost Estimation and Evaluation
Test Objective and Functionality	Report the Test scenario Costs
Description	The info on the VM resource unit cost (reported also into the estimated_cost in #TP17) extracted from the provider price list and/or web site are used to compute an estimation of expected cost before the run of E6 test scenario. For example:

flavour	n1-standard-8
cpu_hour	0,3962
disk_gb_month	0,0352
s3_storage_gb_month	0,0190
cluster_service	0,0948

Funded by the EU
Copernicus esa European Space Agency

	other: 0,0000 The foreseen metric values are used for the final cost estimation (**estimated_cost)** are: cluster_nodes: 20 disk_gb: depends on flavour total_s3_size_gb: 10000 total_processing_hours: 100 The actual test cost (**actual_cost**) is extracted from the provider's dashboard and/or invoice (including the start and end day of test, since most of the target providers do not show such detailed info) The Actual cost for development/dry-run (**dev_cost**) before test run is extracted from the provider's dashboard and/or invoice starting from the start of dry-runs to the day before the official E6 test run. The estimated_cost, actual_cost and dev_cost are used to report the metrics **costPerformance (#M243)** and **devCostPerformance (#M244)**. A further cost metric is computed in this procedure: **costScore**. The rationale is the same as the cost_score metric computed for the Q5 indicator. The algorithm for the computation is the following: 1. For each target provider the actual cost during a full day of processing is taken (including all resource usage). This cost is split by hour (**VM Cost hourly**). 2. The **runtime_cost** is computed multiplying such value for total_pod_lifetime_hours (see #TP17), then the min and max value for all the providers is extracted. 3. The following formula for each target is applied: $$\mathbf{costScore} = \mathbf{100} * \frac{\mathbf{max}(\boldsymbol{runtime}_{\mathbf{cost}}) - \boldsymbol{runtime}_{\mathbf{cost}}}{\mathbf{max}(\boldsymbol{runtime}_{\mathbf{cost}}) - \mathbf{min}(\boldsymbol{runtime}_{\mathbf{cost}})}$$ A score of 100 is assigned to the most convenient target, while a score of 0 is assigned to the less convenient target.
Applicable Interface Type	Manual Human Machine Interface
Indicative Test Frequency	Once
Test Case configuration	An active user account on the target provider is needed.
Relevant Metrics	#M243 costPerformance #M244 devCostPerformance costScore
Dependencies	

Funded by the EU
Copernicus esa European Space Agency

Other comments	

Funded by the EU
Copernicus
esa European Space Agency

9 ANNEX D – OPERATIONAL CONFIGURATION

This Annex describes the input parameters and/or conditions that are used to configure the benchmark user scenarios described in Annex C to this document.

9.1 MAIN CONFIGURATION PARAMETERS

Extended name	Maximum Catalogue Thread
Description	This is an integer number configured for each target site and applied to the test cases when applicable. It sets the concurrency level for the catalogue e.g. 2 means that maximum 2 catalogue requests are made contemporarily.
Typical values	Range values depend on the maximum concurrency level supported by the target site Many Sentinel Data Access Hub limits the number of concurrent requests to 2.
Test cases impact	*TC20X, TC50X:* Set the concurrency level for the catalogue requests. 2 means that 2 requests are made contemporarily.
Dependencies, variability and other comments	

Extended name	Maximum Download Thread
Description	This is an integer number configured for each target site and applied to the test cases when applicable. It sets the concurrency level for the download request e.g. 2 means that maximum 2 catalogue requests are made contemporarily.
Typical values	Range values depend on the maximum concurrency level supported by the target site Many Sentinel Data Access Hub limits the number of concurrent requests to 2.
Test cases impact	TC302, TC303: Set the concurrency level for the download requests. 2 means that 2 requests are made contemporarily. Note: Many Sentinel Data Access Hub limits the number of concurrent requests to 2.
Dependencies, variability and other comments	

Extended name	Maximum Download Size
Description	This is an integer number configured for each target site and applied to the test cases that perform downloads. It sets the maximum of the total download size in bytes for a dataset.
Typical values	2GBytes = 2147483648
Test cases impact	TC30X: Set the maximum download size when selecting downloads. Rejects downloads with a total size greater than the value

Funded by the EU
Copernicus esa European Space Agency

Dependencies, variability and other comments	

Extended name	**Load factor**
Description	This is an integer number configured for a test scenario and applied to the test cases. Act as a multiplier of the number of requests to be made to the target site.
Typical values	[1-10]
Test cases impact	*TC101, TC20x:* Act as a multiplier of the number of requests made to the target site. For instance, if load factor is 10, there will be 10 catalogue requests made *TC304:* Act as the number of pending offline data kept in state (recommended to 1)
Dependencies, variability and other comments	

Extended name	**Virtual Machines**
Description	This is an integer number configured for a test scenario and applied to the test cases. Act as a multiplier of the number of virtual machines requested and provisioned for the test scenario.
Typical values	[2-10]
Test cases impact	TC412
Dependencies, variability and other comments	

Extended name	**Mission Configuration Dictionary**
Description	This is a static configuration describing all the existing missions covered by the benchmarking tool and their valid filters combination. For each mission, it may define a set of filters with its namespace (see section 9.3), labels, values, options, ranges, steps... - platform (mission name) - lifetime : start, stop - platform identifiers (e.g 2014-016A, 2016-025A) - product types - polarisation - sensor mode - instruments - timeliness - product levels - relative orbit - cloud cover - platform name

Funded by the EU
Copernicus esa European Space Agency

	- online/offline - ingestion date - ...
Typical values	See section 9.2
Test CasesImpact	*TC2xx, TC5xx, TC6xx*
Dependencies, variability and other comments	More details on the usage of the MIssion Configuration Dictionary in the next sections

Extended name	**World Borders Dataset**
Description	This is a set of geometries representing the borders of world's countries
Typical values	Every world's country
Test CasesImpact	*TC202*
Dependencies, variability and other comments	A shapefile installed along with the CDAB tools called TN_WORLD_BORDERS-0.3

Extended name	**Geographic Areas Coverage**
Description	This is a list of geometries describing the geographic areas to be used to query the data coverage
Typical values	Europe, North America, Africa, Asia, Australia, Pacific Ocean, Atlantic Ocean, Indian Ocean, Artic Ocean, Antarctica
Relevant Tests	TC501
Dependencies, variability and other comments	The provided list is the initial one. Forthcoming refinements (e.g. by European countries) may be introduced as relevant. More details on the usage of the MIssion Configuration Dictionary in the next sections

Extended name	**Bulk Systematic Data Filters Definition**
Description	This is a static configuration file defining the filters to use for searching the systematic bulk data. It is a set of combination of filters from the Mission Configuration Dictionary
Typical values	Sentinel-1 GRD over Mopti floodable area in Mali ingested during last week Sentinel-2 MSI Level-1C over Etna ingested during last month
Relevant Tests	TC203
Dependencies, variability and other comments	

Extended name	**Catalogue sets (Global Baselines)**
Description	This is a dynamic configuration section in the configuration file defining all sets of catalogue necessary for the benchmark. It defines collections with product types to be used to query the data coverage

Funded by the EU
Copernicus esa European Space Agency

	and latency. Each product type is defined by specifying filters harmonized among all providers referencing the baseline.
Typical values	The Data hubs and the DIAS reference the Copernicus product types baselines:

Copernicus Baseline (without S5p)

Products	Sensor	Mission	Proc. Level
L1 GRDM	C-SAR	S-1	1
L0 RAW	C-SAR	S-1	0
L1 SLC	C-SAR	S-1	1
L2 OCN	C-SAR	S-1	2
L1C	MSI	S-2	1
L2A	MSI	S-2	2
OLCI L1 FR	OLCI	S-3	1
OLCI L1 RR	OLCI	S-3	1
OLCI L2 Land FR	OLCI	S-3	2
OLCI L2 Land RR	OLCI	S-3	2
SLSTR L1 RBT	SLSTR	S-3	1
SLSTR L2 Land	SLSTR	S-3	2
SRAL L1	SRAL	S-3	1
SRAL L1 A	SRAL	S-3	1
SRAL L1 BS	SRAL	S-3	1
SRAL L2 Land	SRAL	S-3	2

Copernicus S5p Baseline

Products	Sensor	Mission	Proc. Level
L2 AER AI	TROPOMI	S-5P	2
L2 CH4	TROPOMI	S-5P	2
L2 CLOUD	TROPOMI	S-5P	2
L2 CO	TROPOMI	S-5P	2
L2 HCHO	TROPOMI	S-5P	2
L2 NO2	TROPOMI	S-5P	2
L2 O3	TROPOMI	S-5P	2
L2 SO2	TROPOMI	S-5P	2
L2 NP BANDS	TROPOMI	S-5P	2
L1B RA BANDS	TROPOMI	S-5P	1
L2 AER AI	TROPOMI	S-5P	2
L2 CH4	TROPOMI	S-5P	2
L2 CLOUD	TROPOMI	S-5P	2
L2 CO	TROPOMI	S-5P	2
L2 HCHO	TROPOMI	S-5P	2
L2 NO2	TROPOMI	S-5P	2

Other baselines can be defined when the target site referencing them have a specific catalogue baseline. For instance Amazon, Has only Sentinel-1 GRD, Sentinel-2 and Sentinel-5p.

Relevant Tests	TC501, T502, TC60X

Funded by the EU
Copernicus esa European Space Agency

Dependencies, variability and other comments	More details on the usage of the Mission Configuration Dictionary in the next sections

Extended name	**Time Division Coverage**
Description	This is a list of time range describing the time span of sensing time to be used to query the data coverage
Typical values	Last 24 months Last 12 months Last 3 months Last month
Relevant Tests	TC501
Dependencies, variability and other comments	More details on the usage of the Mission Configuration Dictionary in the next sections

Extended name	**Local collections dictionary**
Description	This is a set of catalogue collections defining the "data offer" by the target site
Typical values	Sentinel-1 RAW World last 6 months Sentinel-1 SLC Europe Sentinel-3 World Non Time Critical
Relevant Tests	TC503
Dependencies, variability and other comments	More details on the usage of the Local coverage collections in the next sections

9.2 MAIN TEST-SPECIFIC STRATEGIES

9.2.1 Catalogue Filters Randomization (#TS 01 and #TS 02)

In the Test Scenario #TS 01 and #TS 02, the test cases #TC 201 and #TC 202 performs a set of typical queries on the catalogue API of the target sites.
During the preparation phase of each test case, a function generates randomly those typical queries using the Mission Configuration Dictionary. It performs the following pseudo-algorithm:

1) picks up randomly a mission defined in the target site baseline
2) select randomly a collection in that baseline of the previously selected mission
3) For each selected filters, generate randomly a filter value according to the ranges or options defined in the dictionary (and in the World Borders Dataset for the complex)
4) combine all filters in a single query

All steps are performed N times for the N queries to be performed in the test cases according to the load factor.
Here are some examples of simple and complex randomly generated queries:

- "Mission Sentinel-1 Track [16 TO 169]",
- "Mission Sentinel-2 Count 86 Level-1C",
- "Mission Sentinel-1 Count 62 Stripmap",

Funded by the EU
Copernicus esa European Space Agency

- "Mission Sentinel-2 A Level-2A",
- "Mission Sentinel-3 B Near Real Time",
- "Mission Sentinel-2 A",
- "Mission Sentinel-3 Near Real Time Track [330 TO 355]",
- "Mission Sentinel-1 Count 47 Level-2 Ocean (OCN)",
- "Mission Sentinel-1 Count 73 B",
- "Mission Sentinel-3 Count 27 A"
- "Mission Sentinel-2 Level-2A From Sunday, October 22, 2017 To Thursday, June 13, 2019 intersecting United Republic of Tanzania",
- "Mission Sentinel-3 Count 18 From Sunday, October 22, 2017 intersecting Cape Verde",
- "Mission Sentinel-1 Count 13 From Sunday, January 22, 2017 To Tuesday, June 27, 2017 intersecting Cook Islands",
- "Mission Sentinel-3 A To Saturday, April 15, 2017 intersecting Jamaica"

For the spatial filtering, the benchmark client builds a WKT geometry filter from a world borders shapefile. Basically, the administrative boundary is first reduced to a precision model at 0.1 degree then simplified using Douglas-Peucker algorithm but down to ~10km accuracy preserving more initial topology. This method makes the geometry filter coarser then original but still representative of a user request.

"Count" means the number of items requested for a query. This is also randomized.

9.2.2 Validation Strategy

Each results returned during the main query phase of the test case is validated against every filter included in the search. For instance, each results of the search "Mission Sentinel-2 Level-2A To Thursday, June 13, 2019 intersecting United Republic of Tanzania" is verified that it is

- a Sentinel-2
- a Level-2A product type
- acquired between From Sunday, October 22, 2017 To Thursday June 13, 2019
- intersecting Tanzania border geometry

This randomization & validation process combined with a fairly significant load factor and the test scenarios frequency ensures

1. to query each time a new search to the target site catalogue and thus benchmark properly the search performance (no cache).
2. to benchmark properly the search functionality of each target site either on the filter capacity and on the results accuracy
3. to ensure a fair distribution of the search tests among the target sites

Concerning the validation of the accuracy of the results footprint against the geometry filter, the method also takes into account the data context of the target sites benchmarked. Indeed, the catalogues contain very large collections of low resolution product. This often leads the catalogue service to use a spatial index technology with a coarser accuracy to favor query speed. Thus, the validation method extends the validating geometry with a spatial buffer of 100km allowing close results to be validated.

Funded by the EU
opernicus esa European Space Agency

9.2.3 Offline download strategy (TS#04)

Offline download scenario aims at measuring the latency between the ordering of an offline data and its availability for direct download. To achieve that benchmark, a specific search for datasets marked as offline in the target site catalogue is first performed. Then a product order is placed using the API of the target site. The requested offline datasets are kept in a state file. Next runs of the scenario try to retrieve the ordered dataset when ready. The time span between the dataset order and the actual file download is the metric used to benchmark the offline availability latency.

9.2.4 Data Coverage Benchmark Strategy (#TS 05)

In test scenario #TS 05, test case #TC 501 aims at retrieving metrics about the catalogue coverage of a target site.
This test case uses the Catalogue sets (Global Baselines) referenced by the target site and are combined to form collections representing the local data collections.

Every combination is called a Data Collection Division. A search filter is generated accordingly. For instance, "Sentinel-1 Level-1 Single Look Complex (SLC) ".
At execution runtime, the test case 501 will be executed systematically on every Data Collection Division search requesting no result items but only the total number of results.

Test case 502 performs the same kind of request but requesting only the number of online data. Every Baseline Collection reports then the online/offline storage rate.

The test case results are reported in arrays of metrics sized to the number of data Collection Division. It is not possible to validate every item of the query thus the total results number returned by the search is assumed validated.

Test case 503 performs a similar test as per TC502 on collections defined by the target sites as the "data offer". This test case intent to measure how consistent is the online data locally managed by the service provider wrt to the "declared" data offer.

All TC50X test cases results are reported in arrays of metrics sized to the number of data Collection Division.
It is not possible to validate every item of the query thus the total results number returned by the search is assumed validated.

9.2.5 Data Latency Benchmark Strategy (#TS 06)

In test scenario #TS 06, test cases #TC 601 and #TC 602 aims at retrieving metrics about the data publication latency in the target sites catalogue.

Test case #TC 601 uses the Copernicus Product Types Coverage to generate a set of queries in order to retrieve the latest 100 results of the catalogue for every product type.
This test case aims at representing the latency of availability of the most recent and critical products with regards to the sensing time. It is therefore necessary to either select the time critical product or to discard the reprocessed data. So to ensure this time criticality filtering, the following strategy is performed:

1) When timeliness filter is available at target site and allows to select only time critical product, the filter is activated. Setting the timeliness to time critical products enable a better benchmarking of the operational data latency.

Funded by the EU
Copernicus esa European Space Agency

2) If no timeliness filter is applicable, a sensing time filter is applied to a time span covering 4 * the maximum expected timeliness of the product type as described in COPE-GSEG-EOPG-PD-14-0017.
3) By default, a sensing time filter of 40 days is applied

As a result, the following queries are the typical search made at the APIHUB:

- "Sentinel-1 Level-0 SAR raw data (RAW) last 4 days",
- "Sentinel-1 Level-1 Ground Range Detected (GRD) last 4 days",
- "Sentinel-1 Level-2 Ocean (OCN) last 12 hours",
- "Sentinel-1 Level-1 Single Look Complex (SLC) last 4 days",
- "Sentinel-2 Level-1C last 12 hours",
- "Sentinel-2 Level-2A last 4 days",
- "Sentinel-3 OLCI Level-1B EO FR last 12 hours",
- "Sentinel-3 OLCI Level-1B EO RR last 12 hours",
- "Sentinel-3 OLCI Level-2 Land RR last 12 hours",
- "Sentinel-3 OLCI Level-2 Land FR last 12 hours",
- "Sentinel-3 SLSTR Level-1B RBT last 12 hours",
- "Sentinel-3 SLSTR Level-2 Land Surface Temp last 12 hours",
- "Sentinel-3 Altimetry Level-1 SRA Near Real Time",
- "Sentinel-3 Altimetry Level-1 SRA Short Time Critical",
- "Sentinel-3 Altimetry Level-1 SRA_A Short Time Critical",
- "Sentinel-3 Altimetry Level-1 SRA_BS Short Time Critical",
- "Sentinel-3 Altimetry Level-2 Land Near Real Time",
- "Sentinel-3 Altimetry Level-2 Land Short Time Critical",
- "Sentinel-3 Altimetry Level-2 Water Short Time Critical",
- "Sentinel-3 Synergy Level-2 Surface Reflectance Short Time Critical"

Test case #602 uses the Copernicus Product Types CoverageCatalogue sets (Global Baselines) referenced by the target site to generate a set of queries in order to retrieve the latest 20 results of the catalogue for every product type.
This test case aims at representing the latency of availability of the most recent and critical products with regards to the Reference Target Site (APIHUB). It is desired to discard the time critical product to eliminate the time critical results that would be kept by a target site and not by the reference site.

As a result, the following typical queries are generated:

- "Sentinel-1 Level-0 SAR raw data (RAW)",
- "Sentinel-1 Level-1 Ground Range Detected (GRD)",
- "Sentinel-1 Level-2 Ocean (OCN)",
- "Sentinel-1 Level-1 Single Look Complex (SLC)",
- "Sentinel-2 Level-1C",
- "Sentinel-2 Level-2A",
- "Sentinel-3 OLCI Level-1B EO FR",
- "Sentinel-3 OLCI Level-1B EO RR",
- "Sentinel-3 OLCI Level-2 Land RR",
- "Sentinel-3 OLCI Level-2 Land FR",
- "Sentinel-3 SLSTR Level-1B RBT",
- "Sentinel-3 SLSTR Level-2 Land Surface Temp",
- "Sentinel-3 Altimetry Level-1 SRA Non Time Critical",
- "Sentinel-3 Altimetry Level-1 SRA_A Non Time Critical",
- "Sentinel-3 Altimetry Level-1 SRA_BS Non Time Critical",
- "Sentinel-3 Altimetry Level-2 Land Non Time Critical",
- "Sentinel-3 Altimetry Level-2 Water Non Time Critical",
- "Sentinel-3 Synergy Level-2 Surface Reflectance Non Time Critical"

Every results item is then queried on the reference target site and the 2 items' publication time is compared to extract the availability latency.

Funded by the EU
Copernicus esa European Space Agency

If the target site does not provide a method to get the publication date of products, the date when the query is performed is used instead as an approximate replacement.

9.2.6 Virtual Machines Provisioning Benchmark Strategy (#TS 1X)

In test scenario #TS 1X, all test cases are executed on remote hosted virtual machines provisioned on the infrastructure of the target site (which can be a DIAS or a third-party provider, such as Amazon AWS or Google Cloud Platform). Cdab-remote software, using the Cloud Services API, requests one or more virtual machines. Test cases #TC 41X monitor the provisioning sequence by measuring the time spans for the requested virtual machines to be up and ready for access.
For each target site, a virtual machine image and flavour are chosen and the following execution steps are performed to prepare and execute the remote test. It is also possible to create several virtual machines based on different flavours at the same time and run tests in parallel.

1. Install docker and start it;
2. Install CDAB docker image or another docker image for processing-related tests;
3. Run underlying test via cdab-client or the specific data transforming algorithm;
4. Retrieve results from underlying tests

Once all steps are performed, the virtual machine is shut down. The results files are amended with the results of the test case #TC41X.

9.3 PRODUCTS COLLECTIONS DEFINITION

9.3.1 Search Filters

In order to configure datasets, the benchmarking tools use a harmonized way of defining the products selection filters. They are based on the Earth Observation Extension for OpenSearch [RD 6] and use namespaces to specify filters that can be combined together to create specific filters set.

The next table lists the namespaces used for the filters definition and their related prefix used in the rest of this section for describing filters. Each data provider may define their own domain to add filters specific to their platform.

Namespace	Domain Description	Prefix
http://a9.com/-/spec/opensearch/1.1/	General OpenSearch filters	os
http://a9.com/-/opensearch/extensions/time/1.0/	Time filters	time
http://a9.com/-/opensearch/extensions/geo/1.0/	Geographical filters	geo
http://a9.com/-/opensearch/extensions/eo/1.0/	Earth Observation filters	eo

The next table lists the most common filters supported and used by the benchmarking software and some of their possible values.

Filter FQDN	Description	Options

Funded by the EU
opernicus esa European Space Agency

os:count	Number of items returned by the search	Integer (upper limit depends on the data provider)
os:searchTerms	Free text search filter	Any words
time:start	Items covering time after or equal the value	Date
time:end	Items covering time before or equal the value	Date
geo:geometry	Items intersecting the specified geometry	WKT geometry
eo:platform	Product satelitte platform identified by the mission name	- Sentinel-1 - Sentinel-2 - Sentinel-3
eo:platformSerialIdentifier	Product satelitte platform serial number	- 2014-016A (Sentinel-1A) - 2016-025A (Sentinel-1B) - 2015-028A (Sentinel-2A) - 2017-013A (Sentinel-2B) - 2016-011A (Sentinel-3A) - 2018-039A (Sentinel-3B)
eo:productType	Product type identifier	See next table for all Copernicus product types
eo:processingLevel	Product processing level	L0,L1,L2,L3
eo:instrument	Product intrument	SAR,MSI,OLCI,SLSTR,SRAL,SYNERGY
eo:polarizationChannels	Product polarizationChannels	HH,HV,VV,VH,HH+HV,VV+VH
eo:sensorMode	Product sensor mode	SM,IW,EW,WV,EO,OBS
eo:track	Product relative orbit (track)	Integer or range
eo:orbit	Product absolute orbit	Integer or range
eo:timeliness	Product availability timeliness	NRT,STC,NTC
eo:statusSubType	Product archive availability	Online,Offline

The next table list the Copernicus Product Types described in the previous section and their related filters set. The filters values used are independent from the data provider and follow the nomenclature used by the ESA Data Hub.

Products	**Sensor**	**Mission**	**Proc. Level**	**Filters used**
L1 GRD	C-SAR	S-1	1	eo:platform=Sentinel-1 eo:productType=GRD eo:productLevel=L1
L0 RAW	C-SAR	S-1	0	eo:platform=Sentinel-1 eo:productType=RAW
L1 SLC	C-SAR	S-1	1	eo:platform=Sentinel-1 eo:productType=SLC eo:productLevel=L1
L2 OCN	C-SAR	S-1	2	eo:platform=Sentinel-1 eo:productType=OCN
L1C	MSI	S-2	1	eo:platform=Sentinel-2 eo:productType=S2MSI1C
L2A	MSI	S-2	2	eo:platform=Sentinel-2 eo:productType=S2MSI2A
OLCI L1 FR	OLCI	S-3	1	eo:platform=Sentinel-3 eo:productType=OL_1_EFR___
OLCI L1 RR	OLCI	S-3	1	eo:platform=Sentinel-3 eo:productType=OL_1_ERR___
OLCI L2 Land FR	OLCI	S-3	2	eo:platform=Sentinel-3 eo:productType=OL_1_LFR___
OLCI L2 Land RR	OLCI	S-3	2	eo:platform=Sentinel-3 eo:productType=OL_1_LRR___

Funded by the EU
Copernicus
esa European Space Agency

SLSTR L1 RBT	SLSTR	S-3	1	eo:platform=Sentinel-3 eo:productType=SL_1_RBT___
SLSTR L2 Land	SLSTR	S-3	2	eo:platform=Sentinel-3 eo:productType=SL_2_LST___
SRAL L1	SRAL	S-3	1	eo:platform=Sentinel-3 eo:productType=SR_1_SRA___
SRAL L1 A	SRAL	S-3	1	eo:platform=Sentinel-3 eo:productType=SR_1_SRA_A_
SRAL L1 BS	SRAL	S-3	1	eo:platform=Sentinel-3 eo:productType=SR_1_SRA_BS
SRAL L2 Land	SRAL	S-3	2	eo:platform=Sentinel-3 eo:productType=SR_2_LAN___
SYN L2	SYNERGY	S-3	2	eo:platform=Sentinel-3 eo:productType=SY_2_SYN___
SYN L2 VGP	SYNERGY	S-3	2	eo:platform=Sentinel-3 eo:productType=SY_2_VGP___
L2 AER AI	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__AER_AI
L2 CH4	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__CH4___
L2 CLOUD	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__CLOUD_
L2 CO	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__CO_____
L2 HCHO	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__HCHO___
L2 NO2	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__NO2___
L2 O3	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__O3____
L2 SO2	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__SO2___
L2 NP BANDS	TROPOMI	S-5P	2	eo:platform=Sentinel-5P eo:productType=L2__NP_BD3 / eo:productType=L2__NP_BD6 / eo:productType=L2__NP_BD7
L1B RA BANDS	TROPOMI	S-5P	1	eo:platform=Sentinel-5P eo:productType=L2__RA_BD[1-8]

9.3.2 Data Providers Individual Configurations

For some test cases (TC50X and TC60X), the benchmarking tool requires some filters definitions specific to the data provider. Basically, the benchmarking tool must know:

- What are the filters to query the "reference" catalogue items? This is a set of filters that searches only for products that are in the ESA baseline (OpenHub), filtering out any products not available in the ESA Open Hub, like e.g. provider-specific custom products (Note: NRT data is already filtered out internally). These filters are used whenever a comparison between target and reference collections needs to be performed, and are defined in the *"catalogue_coverage"* section of the data provider in the configuration file. Information on how to configure them should be provided by the data providers themselves.
- What is the data offering for a data provider? A set of filters is needed to search only for products that are managed locally by the provider (online, or archived but still inside its own infrastructure), plus several filter sets defining the data offering of each data provider. All those are defined in the *"local_coverage"* section of the data provider in the configuration file. TC502 will use only the *parameters* sub-section to compute the *dataCoverage* metric, based on the Open Hub mission dictionary, while TC503 will use a combination of both the *parameters* and the *collections* sub-sections to verify the declared data offer.

Funded by the EU
Copernicus esa European Space Agency

All the filters make use of the harmonized filters definition described in the previous section.

Funded by the EU
Copernicus
esa
European Space Agency

10ANNEX E - METRICS DESCRIPTION

A set of core metrics have been elaborated, to be used to describe the benchmark operations results. These are described in the following, in terms of their measurement strategies, applicable test cases and indicative range values.

Metric #M001	**avgResponseTime**	
Average Response Time	This metric is the average of all measured time spans from the last byte of the request sent to the time the first byte is received. This metric should reflect the average total time the server has processed your request and come with an answer. This measure should not be impacted by the network speed since it does not include transport time This measure may be impacted by network latency	
Measurement Unit	ms	
Range values	The values depend on the QoE Indicator under examination. Typical values range from < 100 msecond (very good) to > 5 seconds (very bad) for Reactivity and from < 1000 msecond (very good) to > 50 seconds (very bad) for Discoverability.	
Measurement Strategy	A low level HTTP request is performed with the following pseudo logic (all steps are synchronous): request <- HttpRequest.Send() timer <- StopWatch.Start() request.WaitForResponse() timer.Stop() **averageResponseTime** <- timer.TotalMilliseconds response()<-request.GetResponse() then an average of all the measurements is performed	
Dependencies and Variability	This metric is depending on the *load factor* and the *maximum parallelism* set for the test. Highest is the load factor more tests are performed and thus smoother the average and probably more representative.	
Limitations / Other comments	N/A	
Test Categories	Availability, Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC10X, TC20X, TC21X	Dataset search query speed performance. How long does the server take to perform a complex query (e.g. geo&time)
	TC30X, TC31X	Product file retrieval speed performance How long does the server take to serve the client (remote or local) with the requested file and start transferring it

Metric #M002	**peakResponseTime**
Peak Response Time	This metric is the maximum of all measured time spans from the last byte of the request sent to the time the first byte is received. This metric should reflect the maximum total time the server has processed your request and come with an answer. This measure should not be impacted by the network speed since it does not include transport time

Funded by the EU
Copernicus esa European Space Agency

	This measure may be impacted by network latency
Measurement Unit	ms
Range values	The values depend on the QoE Indicator under examination. Typical values range from < 200 msecond (very good) to > 10 seconds (very bad) for Reactivity and from < 2000 msecond (very good) to > 100 seconds (very bad) for Discoverability.
Measurement Strategy	A low level HTTP request is performed with the following pseudo logic (all steps are synchronous): request <- HttpRequest.Send() timer <- StopWatch.Start() request.WaitForResponse() timer.Stop() **averageResponseTime** <- timer.TotalMilliseconds response()<-request.GetResponse() then the maximum of all the measurements is used
Dependencies and Variability	This metric is depending on the *load factor* and the *maximum parallelism* set for the test. Highest is the load factor more tests are performed and thus more likely a high peak response time may occur, especially with high max parallelism value
Limitations / Other comments	N/A
Test Categories	Availability, Search & View, Download/Upload
Test Cases	See #M001

Metric #M003	**errorRate**	
Error Rate	This metric is the percentage of requests that ended with an error. This metric should reflect the proportion of request failure due to a server error	
Measurement Unit	%	
Range values	0% is perfect <10% is acceptable >10% is unacceptable	
Measurement Strategy	A 5XX HTTP Web response status code for a request is considered as an error.	
Dependencies and Variability	This metric is depending on the *load factor* and the *maximum parallelism* set for the test. Highest is the load factor more tests are performed and thus more likely a high peak error rate may occur, especially with high max parallelism value	
Limitations / Other comments	N/A	
Test Categories	Availability, Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC10X, TC20X, TC302, TC304, TC21X	Error rate for all requests made per test cases
	TC301, TC303, TC311, TC312	No concurrency nor bulk request so 0% or 100% are the possible measurement values
	TC41X	Error rate for all requests made per test cases
	TC60X	Error rate for all requests made per test cases

Metric #M005	**throughput**

Funded by the EU
Copernicus esa European Space Agency

Throughput	This metric is the average of the requests' ratio between a volume of data transported and the time for this volume of data to be transferred from a system to another (client/server) This metric should reflect the speed for data transmission in a given test case. This metric is likely impacted by test site and/or target site network performance and punctual infrastructure issues	
Measurement Unit	Bytes/second	
Range values	Value subject to interpretation regarding test site	
Measurement Strategy	The transfer time of the relevant volume of data of the test case is measured on the time from the first byte to the last byte successfully transferred.	
Dependencies and Variability	This metric is depending on the load factor and the maximum parallelism set for the test. Highest is the load factor more are tests performed and thus smoother is the mean and probably more representative.	
Limitations / Other comments	N/A	
Test Categories	Availability, Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC30X	Product file download or upload speed performance. How fast was it for the client to download/upload a complete file?

Metric #M006	**avgConcurrency**	
Average Concurrency	This metric reports the average parallelism capacity sustained by the server during a batch of requests	
Measurement Unit	Float number	
Range values	The range is from 1 to *maximum parallelism or virtual machines* value configured. The closest to this latter value is the best.	
Measurement Strategy	All requests to a given service of the target site start and end time are sampled (sampling rate wrt TC) in the total test case time span. This measures the number of ongoing concurrent requests. An average is performed out of this sampling.	
Dependencies and Variability	This metric is depending on the load factor, the maximum parallelism and the number of virtual machines set for the test. Highest is the load factor more are tests performed and thus smoother is the mean and probably more representative.	
Limitations / Other comments	N/A	
Test Categories	Availability, Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC10X	Capacity for the target site to answer concurrently to the test site
	TC20X, TC21X	Capacity for the target site to answer search query concurrently to the test site
	TC302, TC304, TC312	Capacity for the target site to product download concurrently to the test site

Funded by the EU
Copernicus esa European Space Agency

	TC301, TC311	No concurrency nor bulk request so always 1
	TC41X	Average capacity vs. time for the target site to make available a certain amount of VM at the same time.

Metric #M007	**peakConcurrency**	
Peak Concurrency	This metric reports the maximum parallelism capacity sustained by the server during a batch of requests	
Measurement Unit	Float number	
Range values	The range is from 1 to *maximum parallelism or virtual machines* value configured. A Value lower than this latter value is not good.	
Measurement Strategy	All requests to a given service of the target site start and end time are sampled (sampling rate wrt TC) in the total test case time span. This measures the number of ongoing concurrent requests. The maximum is retrieved out of this sampling.	
Dependencies and Variability	This metric is depending on the load factor and the maximum parallelism set for the test. Highest is the load factor more are tests performed and thus more likely is the maximum concurrency to be lower or closer than the max parallelism.	
Limitations / Other comments	N/A	
Test Categories	Availability, Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC10X	Max capacity for the target site to answer concurrently to the test site
	TC20X, TC21X	Max capacity for the target site to answer search query concurrently to the test site
	TC302, TC304, TC312	Max Capacity for the target site to product download concurrently to the test site
	TC301, TC311	No concurrency nor bulk request so always 1
	TC41X	Maximum capacity vs. time for the target site to make available a certain amount of VM at the same time.

Metric #M008	**avgSize**
Average Size	This metric reports the average size of data volume transferred during a test case for each test unit
Measurement Unit	Byte number (long)
Range values	N/A
Measurement Strategy	Average of the data volume of the data unit transferred. For search results, it is the size of the feed of results returned by the search API. For downloaded/uploaded products, it is the size of the product file.

Funded by the EU
Copernicus esa European Space Agency

Dependencies and Variability	The size depends on the time when the test case is executed and on the time set for search query. Indeed, search results and thus downloaded files may largely differ if only small files are available at the selected time than large file (see other comments)	
Limitations / Other comments	Most of the tests cases make a random selection in the data chosen to be searched and/or downloaded. Thus this metric may be important to weight other metrics (e.g. throughput may largely differ if only small data files are downloaded than larger files)	
Test Categories	Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC20X, TC21X	Search results size. How big are the results of the search?
	TC302, TC303, TC304, TC312	Average product file size downloaded How big is the batch of download/upload?
	TC301, TC311	No concurrency nor batch request thus average size is the size of the single file downloaded/uploaded

Metric #M009	**maxSize**	
Max Size	This metric reports the maximum size of units data volume transferred during a test case for each test unit	
Measurement Unit	Byte number (long)	
Range values	N/A	
Measurement Strategy	Maximum of the data volume of the data unit transferred. For search results, it is the size of the feed of results returned by the search API. For downloaded/uploaded products, it is the size of the product files.	
Dependencies and Variability	The size depends on the time when the test case is executed and on the time set for search query. Indeed, search results and thus downloaded files may largely differ if only small files are available at the selected time than large file (see other comments)	
Limitations / Other comments	Most of the tests cases make a random selection in the data chosen to be searched and/or downloaded. Thus this metric may be important to weight other metrics (e.g. throughput may largely differ if only small data files are downloaded than larger files)	
Test Categories	Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC20X, TC21X	Search results size. How big is the largest results feed of the search?
	TC302, TC303, TC304, TC312	Largest product file size downloaded
	TC301, TC311	No concurrency nor batch request thus max size is the size of the single file downloaded/uploaded

Funded by the EU
Copernicus esa European Space Agency

Metric #M010	**totalReadResults**	
Total Read Results Count	This metric reports the total number of results retrieved. Beware this is different from #M011 (maxTotalResults).	
Measurement Unit	Integer number	
Range values	N/A	
Measurement Strategy	When a search query is performed, the total number of results items retrieved are counted. The counts are summed up as the test case result metric.	
Dependencies and Variability	N/A	
Limitations / Other comments	Most of the relevant tests cases queries a random count of data results per query. This metric may be used to weight the average response time in the benchmarking.	
Test Categories	Search & View	
Test Cases	**Identifier**	**Interpretation**
	TC20X, TC21X	Search results items retrieved count. How Many results items retrieved in the search query?
	TC60X	Search results items retrieved count. How Many results items retrieved in the search query?

Metric #M011	**maxTotalResults**	
Max Total Results Count	This metric reports the total number of results available	
Measurement Unit	Integer number	
Range values	N/A	
Measurement Strategy	When a search query is performed, when available in the results feed, the total results items available in the target site is reported	
Dependencies and Variability	When available on the target site.	
Limitations / Other comments	This metric is available on the reference target site (Copernicus Scientific Data Hub). It is then used to compute the ratio of collection completeness. This metric may be used to weight the average response time in the benchmarking.	
Test Categories	Search & View, Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC10X	Not Reported
	TC20X, TC21X	Search results size. Among how many results items available were the search queries performed

Metric #M012	**resultsErrorRate**
Results Error Rate	This metric reports the ratio of results not corresponding to a search query or faulty downloaded products file.
Measurement Unit	%
Range values	N/A

Funded by the EU
Copernicus esa European Space Agency

Measurement Strategy	When a search query is performed, every item metadata is validated against the search query filters. This ratio is the sum number of not successfully validated data item on the total results count	
Dependencies and Variability	When metadata available in the results feed of the target site.	
Limitations / Other comments	N/A	
Test Categories	Search & View, Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC20X, TC21X	Search results consistency How good is the search query of the target site?
	TC302, TC303, TC304, TC312	Product file consistency How valid are the downloaded results?

Metric #M013	**avgDataAvailabilityLatency**	
Average Data Availability Latency	This metric reports average latency of the availability of the data with respect to the availability on the reference target site	
Measurement Unit	Seconds	
Range values	N/A	
Measurement Strategy	An average time is computed from the 20 latest products publication time on the target site then the average of the same 20 products publication time on the reference target site (scihub). The average time difference is reported.	
Dependencies and Variability	When publication metadata available in the results feed of the target site.	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC602	[Array] Data availability average latency per data collection division

Metric #M014	**avgDataOperationalLatency**	
Average Data Operational Latency	This metric reports average latency of the availability of the data with respect to the sensing time of the product.	
Measurement Unit	Seconds	
Range values	N/A	
Measurement Strategy	An average time is computed from the 20 latest products publication time on the target site then the average of the sensing end time is computed. The average time difference is reported.	
Dependencies and Variability	When metadata available in the results feed of the target site.	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC601	[Array] Data operational average latency per data collection division

Funded by the EU
opernicus esa European Space Agency

Metric #M015	catalogueCoverage	
Data Coverage	This metric reports the ratio between the number of relevant data products made available from the target site wrt to the reference target site (Open Hub).	
Measurement Unit	%	
Range values	0-100	
Measurement Strategy	This is a vector of the values covering all the available catalogue entries. It takes the form of a matrix having the following structure: • Rows correspond to a specific Product Type (Product, Sensor, Mission, Processing level… see Metrics Copernicus Product Types) or coherent collections of them. • Columns represent configurable pre-defined geographic areas or time spans. • Each cell of the matrix has a value that ranges from 0 (= no item available in this class) to 1 (same number of items as the Open Access Hub for this class).	
Dependencies and Variability	When metadata available in the results feed of the target site.	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC501	[Array] Data coverage per data collection division

	Europe	Africa	.....	Last month	Last 12 months
Product Type 1	*0 to 1*	*0 to 1*	*…*	*0 to 1*	*0 to 1*
…	*…*	*…*	*…*	*…*	*…*
Product Type N	*0 to 1*	*0 to 1*	*…*	*0 to 1*	*0 to 1*

Metric #M016	dataCollectionDivision	
Data Collection Division	This metric reports the data collection subset search during the test case.	
Measurement Unit	string	
Range values	N/A	
Measurement Strategy	This is a vector of string described all the searches performed during the test case	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Search, Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC20X, TC21X, TC30X, TC31X	[Array] Random Data search description
	TC50X, TC60X	[Array] Data coverage per data collection division

Funded by the EU
opernicus
esa
European Space Agency

Metric #M017	offlineDataAvailabilityLatency	
Offline Data Latency	This metric reports the time spent for an offline product to become available for download from the moment it is requested	
Measurement Unit	minutes	
Range values	10 to 10000	
Measurement Strategy	Time is measured from the first web request to download the data until it is downloading the first byte	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Download	
Test Cases	Identifier	Interpretation
	TC304	Time for download available

Metric #M018	maxDataOperationalLatency	
Max Data Operational Latency	This metric reports maximum latency of the availability of the data with respect to the sensing time of the product.	
Measurement Unit	minutes	
Range values	N/A	
Measurement Strategy	An average time is computed from the 20 latest products publication time on the target site then the average of the sensing end time is computed. The maximum time difference is reported.	
Dependencies and Variability	When metadata available in the results feed of the target site.	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC601	[Array] Data operational maximum latency per data collection division

Metric #M019	maxDataAvailabilityLatency	
Max Data Availability Latency	This metric reports maximum latency of the availability of the data with respect to the availability on the reference target site	
Measurement Unit	Minutes	
Range values	N/A	
Measurement Strategy	An average time is computed from the 20 latest products publication time on the target site then the average of the same 20 products publication time on the reference target site (scihub). The maximum time difference is reported.	
Dependencies and Variability	When publication metadata available in the results feed of the target site.	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC602	[Array] Data availability maximum latency per data collection division

Metric #M020	totalValidatedResults

Funded by the EU
Copernicus esa European Space Agency

Total Validated Results Count	This metric reports the total number of valid results retrieved. Beware this is different from #M010 (totalResults).	
Measurement Unit	Integer number	
Range values	0 to #M010 (totalResults).	
Measurement Strategy	When a search query is performed, retrieved results items are validated against the search filter. The valid items are counted and are summed up as the test case result metric.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Search & View	
Test Cases	**Identifier**	**Interpretation**
	TC60X	Valid results retrieved

Metric #M021	**totalWrongResults**	
Total Wrong Results Count	This metric reports the total number of wrong results retrieved.	
Measurement Unit	Integer number	
Range values	0 to #M010 (totalResults). More than 0 is not good.	
Measurement Strategy	When a search query is performed, retrieved results items are validated against the search filter. The wrong items are counted and are summed up as the test case result metric.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Search & View	
Test Cases	**Identifier**	**Interpretation**
	TC20X, TC21X	Wrong results retrieved

Metric #M022	**totalReferenceResults**	
Total Online Results	This metric reports the total number of products in the reference site catalogue for a given data collection.	
Measurement Unit	Integer number	
Range values	N/A	
Measurement Strategy	A search for all available products is performed on the reference site and the returned total is read.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC501, TC502, TC503	How many online results in the catalogue

Metric #M023	**dataCoverage**
Online Data Rate	This metric reports for a collection the ratio of products in the target site catalogue that are locally managed and thus immediately available for download wrt to the same collection in the reference site.
Measurement Unit	Integer number
Range values	0 to 100%. The higher the better.
Measurement Strategy	A ratio of Metric #M025 totalResults on #M022 totalReferenceResults is calculated

Funded by the EU
opernicus esa European Space Agency

Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC502	How large is the coverage of online results in the catalogue

Metric #M024	**dataOfferConsistency**	
Online Data Rate	This metric reports for a set of collection defined by the service provider as the "data offer", the ratio of products in the target site catalogue that are locally managed and thus immediately available for download with respect to to the same collection set and corresponding timespans and geographic coverages in the reference site.	
Measurement Unit	Integer number	
Range values	0 to 100%. The higher the better.	
Measurement Strategy	A ratio of Metric #M025 totalResults on #M022 totalReferenceResults is calculated	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC503	How consistent is the coverage of the data offer declared.

Metric #M025	**totalResults**	
Online Data Rate	This metric reports the total number of products available in the target site for a specific collection.	
Measurement Unit	Integer number	
Range values	N/A	
Measurement Strategy	A search query is performed and the total number of items available in the target site is reported	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Dataset Completeness	
Test Cases	**Identifier**	**Interpretation**
	TC501, TC502, TC503	Number of items available in the target site for the searched collections.

Metric #M026	**totalSize**
Max Size	This metric reports the total size of data volume transferred during a test case
Measurement Unit	Byte number (long)
Range values	N/A
Measurement Strategy	Total volume of transferred data for downloaded/uploaded products, it is the sum of the sizes of the product files.

Funded by the EU
Copernicus esa European Space Agency

Dependencies and Variability	Search results and thus downloaded files may largely differ according to the performed query.	
Limitations / Other comments	Most of the tests cases make a random selection in the data chosen to be searched and/or downloaded. Thus this metric may be important to weight other metrics (e.g. throughput may largely differ if only small data files are downloaded than larger files)	
Test Categories	Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC301, TC302, TC303, TC311, TC312	Total size of all downloaded products files.

Metric #M027	**exception**	
Max Size	This metric reports eventual exceptions that occurred during an action. It gives information on the nature of the caused error.	
Measurement Unit	Text	
Range values	N/A	
Measurement Strategy	error catching on service api call	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Search & View, Download/Upload	
Test Cases	**Identifier**	**Interpretation**
	TC602	used to report the type of issue on a target site service.

Metric #M041	**copernicusCoreServices**	
Description	This metric reports the completeness of the Copernicus Core Services data complementary offer on the provider, as ratio between positive answered questions over the total from the Test Procedure.	
Measurement Unit	%	
Range values	0-100	
Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M042	**otherData**
Description	This metric reports the completeness of the non-Copernicus data (only EO data) complementary offer on the provider, as ratio between positive answered questions over the total from the Test Procedure.
Measurement Unit	%
Range values	0-100
Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.

Funded by the EU
Copernicus esa European Space Agency

Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M043	**apiCompleteness**	
Description	This metric reports the completeness of the available APIs on the provider, as ratio between positive answered questions over the total from the Test Procedure.	
Measurement Unit	%	
Range values	0-100	
Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M044	**cloudServices**	
Description	This metric indicates the availability of a cloud environment on the provider, where users can create and manage virtual machines. The metric only evaluates the presence of basic cloud services so it can only have values YES (=100) or NO (=0).	
Measurement Unit	%	
Range values	0 or 100	
Measurement Strategy	Inspection of the provider offer.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M045	**otherServices**
Description	This metric reports the availability of other services not taken into account in other metrics, such as a Marketplace, or consulting, as ratio between positive answered questions over the total from the Test Procedure.
Measurement Unit	%
Range values	0-100

Funded by the EU
Copernicus esa European Space Agency

Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M046	**devTools**	
Description	This metric reports the completeness of the development tools complementary offer on the provider, as ratio between positive answered questions over the total from the Test Procedure.	
Measurement Unit	%	
Range values	0-100	
Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M047	**devServices**	
Description	This metric reports the availability of services oriented at supporting application development, e.g. Docker or Kubernetes, as ratio between positive answered questions over the total from the Test Procedure.	
Measurement Unit	%	
Range values	0-100	
Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M048	**eoTools**
Description	This metric reports the availability of EO-specific tools in the provider offering, e.g. SNAP or Sen2Cor, as ratio between positive answered questions over the total from the Test Procedure.
Measurement Unit	%

Funded by the EU
Copernicus esa European Space Agency

Range values	0-100	
Measurement Strategy	Add all positive answers to a set of predefined questions and then compute the ratio to the total number of questions.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP13	Review complementary offer on the provider website.

Metric #M101	**avgProvisioningLatency**	
Average Provisioning Latency	This metric reports the time for provisioning one or more virtual machine in from a cloud service provider	
Measurement Unit	milliseconds	
Range values	Typical values range from < 60.000 millisecond (very good) to > 3.600.000 seconds (very bad).	
Measurement Strategy	An average time is computed if several machines are requested simultaneously (#TS12) otherwise (#TS11), the unique provisioning time is taken.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Measure of the time for provisioning virtual machines

Metric #M110	**cost**	
Cost	This metric reports the cost of the services for executing the test on a target site.	
Measurement Unit	euros	
Range values	The values depend on the QoE Indicator under examination.	
Measurement Strategy	The cost is calculated based on the billing information provided by the target site. If a calculator is available through the Cloud Services API, it is used. Otherwise, a calculation is based on the customer information provided.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Cost of provisioning Virtual Machines and their uptime.

Metric #M111	**dataSummaryRun**
Summary for individual run	This value is an array containing identifiers for each individual execution of the test case sequence on a virtual machine. The size

Funded by the EU
Copernicus esa European Space Agency

	of the array corresponds to the number of virtual machines that are instantiated.	
Measurement Unit	string	
Range values	N/A	
Measurement Strategy	N/A	
Dependencies and Variability	N/A	
Limitations / Other comments	This is not a measurement, but reported along with the actual metrics.	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Identifiers of test runs (reference only).

Metric #M112	**flavorName**	
Flavour name	This value is an array containing the names of the virtual machine configurations (flavours) for each individual execution of the test case sequence on a virtual machine. The size of the array corresponds to the number of virtual machines that are instantiated.	
Measurement Unit	string	
Range values	N/A	
Measurement Strategy	N/A	
Dependencies and Variability	N/A	
Limitations / Other comments	This is not a measurement, but reported along with the actual metrics.	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Names of VM flavours used for test runs (reference only).

Metric #M113	**costHour**	
Hourly cost of VM	This value is an array containing the hourly cost of each virtual machine that is temporarily used for the execution of the test case sequence. The size of the array corresponds to the number of virtual machines that are instantiated. As different flavours may be used in the same test scenario execution, the values may vary.	
Measurement Unit	Euros	
Range values	N/A	
Measurement Strategy	N/A	
Dependencies and Variability	N/A	
Limitations / Other comments	This is not a measurement, but reported along with the actual metrics.	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Hourly cost of VMs used for test runs (reference only).

Metric #M114	**costMonth**

Funded by the EU
Copernicus esa European Space Agency

Hourly cost of VM	This value is an array containing the monthly cost of each virtual machine that is temporarily used for the execution of the test case sequence. The size of the array corresponds to the number of virtual machines that are instantiated. As different flavours may be used in the same test scenario execution, the values may vary.	
Measurement Unit	Euros	
Range values	N/A	
Measurement Strategy	N/A	
Dependencies and Variability	N/A	
Limitations / Other comments	This is not a measurement, but reported along with the actual metrics.	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Monthly cost of VMs used for test runs (reference only).

Metric #M115	**duration**	
Duration of VM	This metric is an array reporting the lifetimes of the virtual machines created temporarily for the purpose of individual test executions. The size of the array corresponds to the number of virtual machines that are instantiated.	
Measurement Unit	milliseconds	
Range values	The values typically range from 300,000 to over 600,000, but could also be significantly higher, depending on the nature of the test (e.g. download only or processing of downloaded data).	
Measurement Strategy	The time that passes between the request for creating the virtual machine and its confirmed deletion is measured for each individual execution on a virtual machine and placed in the value array.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Lifetime of VMs used for test runs.

Metric #M116	**processDuration**
Process duration on VM	This metric is an array reporting the effective duration of the test executions on the virtual machines created temporarily for the purpose of individual test executions. The size of the array corresponds to the number of virtual machines that are instantiated.
Measurement Unit	milliseconds
Range values	The values typically range from 60,000 to 300,000, but could also be significantly higher, depending on the nature of the test (e.g. download only or processing of downloaded data).
Measurement Strategy	The time that passes between the start of the processing command, such as the instantiation of the Docker container, and the completion of the command is measured for each individual execution on a virtual machine and placed in the value array.
Dependencies and Variability	N/A
Limitations / Other comments	N/A

Funded by the EU
Copernicus esa European Space Agency

Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Effective duration of test executions on VMs.

Metric #M117	**totalDuration**	
Total test duration	This metric is an the sum of total lifetimes of the virtual machines created temporarily for the purpose of individual test executions.	
Measurement Unit	milliseconds	
Range values	The values typically range from 300,000 to over 600,000 for each run, but could also be significantly higher, depending on the nature of the test (e.g. download only or processing of downloaded data). Given that it is the sum of all several VMs' durations, the value is a multiple of that and depends on the number of parallel tests.	
Measurement Strategy	The time that passes between the request for creating the virtual machine and its confirmed deletion is measured for each individual execution on a virtual machine and added to the total.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Total lifetime of VMs used for test runs. It corresponds to the hypothetical duration of the test if VMs were used sequentially rather than in parallel.

Metric #M118	**avgProcessDuration**	
Average process duration on VM	This metric is an array reporting the average effective duration of the test executions on the virtual machines created temporarily for the purpose of individual test executions. This is relevant for test cases where several serial processes of the same kind are executed on the same virtual machine. The size of the array corresponds to the number of virtual machines that are instantiated.	
Measurement Unit	milliseconds	
Range values	The values typically range from 60,000 to 300,000, but could also be significantly higher, depending on the nature of the test (e.g. download only or processing of downloaded data).	
Measurement Strategy	The time that passes between the start of the processing command, such as the instantiation of the Docker container, and the completion of the command is measured for each individual execution on a virtual machine, and the averages per virtual machine and placed in the value array.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Effective duration of test executions on VMs.

Funded by the EU
Copernicus esa European Space Agency

Metric #M119	**processCount**	
Average process duration on VM	This metric is an array reporting the number of serial processes of the same kind, e.g. data transformations of a series of products, that are executed on the same virtual machine. The size of the array corresponds to the number of virtual machines that are instantiated.	
Measurement Unit	no unit	
Range values	The values range from 1 to low two-figure values	
Measurement Strategy	The value is a simple count.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Cloud Services	
Test Cases	**Identifier**	**Interpretation**
	TC41X	Number of serial process within test executions on VMs.

Metric #M201	**registrationTimeSpan**	
Description	Time to get a free account	
Measurement Unit	minutes	
Range values	From a few minutes to hours/days	
Measurement Strategy	Register for a free account	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP01	Register for a free account

Metric #M202	**subscriptionDocCompleteness**	
Description	This metric provides an indication of the documentation completeness as ratio between positive answered questions over the total on the #TP02	
Measurement Unit	Counts	
Range values	From a few to several units	
Measurement Strategy	Add all positive answers to a set of predefined questions (see TP)	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP02	Review Subscription documentation

Metric #M203	**subscriptionSupportAnswerTime**

Funded by the EU
Copernicus
esa European Space Agency

Description	Time for getting an acceptable answer from the provider's help-desk for a question relevant to the subscription	
Measurement Unit	minutes	
Range values	From a few minutes to hours/days	
Measurement Strategy	Request support for subscription	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP03	Request support for subscription

Metric #M204	**subscriptionSupportThreadLength**	
Description	Number of messages exchanged between the user and the help-desk to get an acceptable answer	
Measurement Unit	Counts	
Range values	From a few units to several	
Measurement Strategy	Request support for subscription	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP03	Request support for subscription

Metric #M205	**subscriptionSupportAppreciation**	
Description	Answers general appreciation	
Measurement Unit	Counts	
Range values	From a few units to several	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP03	Request support for subscription

Metric #M206	**subscriptionTimeSpan**
Description	Time to get a commercial account
Measurement Unit	minutes
Range values	From a few minutes to hours/days

Funded by the EU
Copernicus esa European Space Agency

Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP04	Subscribe for a commercial account

Metric #M207	**subscriptionPaymentMethodPrioNumber**	
Description	Priority order of the payment method chosen	
Measurement Unit	Number	
Range values	#1 Paypal #2 Credit Card #3 Bank Transfer (predefined ceiling) #4 Bank Transfer (no roof)	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP04	Subscribe for a commercial account

Metric #M208	**subscriptionFailures**	
Description	Number of execution the subscription procedure that did not get a commercial account properly	
Measurement Unit	Counts	
Range values	A few units	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP04	Subscribe for a commercial account

Funded by the EU
Copernicus esa European Space Agency

Metric #M209	integrationToolsAdequacy	
Description	Sum of the scores corresponding to an available tool from a list of expected tools with regard to user's skills	
Measurement Unit	Counts	
Range values	From a few to several	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP05	Review application integration service offer

Metric #M210	integrationEnvSetupTimespan	
Description	Elapsed Time to setup the priviledged development environment	
Measurement Unit	minutes	
Range values		
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP06	Setup Development Environment

Metric #M211	integrationCompleteness	
Description	Application Integration Achievement	
Measurement Unit	%	
Range values	0 - 100	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP07	Perform Application Integration

Metric #M212	integrationBlockingCause
Description	Type of blocking cause of the application integration
Measurement Unit	string
Range values	Support ticket identifier
Measurement Strategy	

Funded by the EU
Copernicus esa European Space Agency

Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP07	– Setup Development Environment - Perform Application Integration

Metric #M213	**integrationSupportAnswerTime**	
Description	Time for getting an acceptable answer from the provider's help-desk for a question relevant to the application integration	
Measurement Unit	minutes	
Range values		
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP08	Request support for application integration

Metric #M214	**integrationSupportThreadLength**	
Description	Number of messages exchanged between the user and the help-desk to get an acceptable answer	
Measurement Unit	Count	
Range values		
Measurement Strategy	Number of iterations (number of reply) metric	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP08	Request support for application integration

Metric #M215	**integrationSupportAppreciation**
Description	Level of satisfaction with respect to the received answers
Measurement Unit	Rank
Range values	1 poor – 5 efficient
Measurement Strategy	Number of iterations (number of reply) metric
Dependencies and Variability	N/A
Limitations / Other comments	N/A
Test Categories	Manual Procedures

Funded by the EU
Copernicus esa European Space Agency

Test Procedures	**Identifier**	**Interpretation**
	#TP08	Request support for application integration

Metric #M216	**storageSetupTimeSpan**	
Description	Elapsed Time to setup the user storage	
Measurement Unit	minutes	
Range values		
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP09	Setup User Storage

Metric #M217	**storageCostPerGB**	
Description	Cost of a Gigabyte of persistent storage	
Measurement Unit	euro	
Range values		
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	Setup User Storage	

Metric #M220	**integrationCost**	
Description	cost of the application integration covering all the operations from the procedures	
Measurement Unit	euro	
Range values		
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	Setup User Storage	

Metric #M301	**dataQueryFiltersAdequacy**

Funded by the EU
Copernicus esa European Space Agency

Description	Adequacy of the data query filters when searching for data for processing	
Measurement Unit	%	
Range values		
Measurement Strategy	% of available and effective filters	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP11	Review Catalogue Data Access Adequacy

Metric #M219	**appStorageAvailability**	
Description	Availability of a storage for the integrated application	
Measurement Unit	Rank	
Range values	1 poor – 5 efficient	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP10	Store Application Container

Metric #M501	**dataExploitationToolsAdequacy**	
Description	Adequacy of the data query filters when searching for data for processing	
Measurement Unit	%	
Range values		
Measurement Strategy	% of available and effective filters	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	**Interpretation**
	#TP11	Review Catalogue Data Access Adequacy

Metric #M230	**serviceDocCompleteness**
Description	This metric provides an indication of the service documentation completeness as ratio between positive answered questions over the total on the #TP14
Measurement Unit	%
Range values	From a 0 to 100

Funded by the EU
Copernicus esa European Space Agency

Measurement Strategy	Add all positive answers to a set of predefined questions (see TP)	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP14	

Metric #M231	**serviceSupportHelpDesk**	
Description	This metric provides an indication of the service support channels completeness as ratio between positive answered questions over the total on the #TP14	
Measurement Unit	%	
Range values	From a 0 to 100	
Measurement Strategy	Add all positive answers to a set of predefined questions (see TP)	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP14	

Metric #M232	**serviceSupportAnswerTime**	
Description	Average Time elapsed from question and first human response on each support thread (considering only working hours 9-18)	
Measurement Unit	minutes	
Range values	From a few minutes to hours/days	
Measurement Strategy	Add all positive answers to a set of predefined questions (see TP)	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP14	

Metric #M233	**serviceSupportThreadLength**
Description	Average number of replies on each mail thread
Measurement Unit	Count
Range values	
Measurement Strategy	Number of iterations (number of reply) metric
Dependencies and Variability	N/A
Limitations / Other comments	N/A
Test Categories	Manual Procedures

Funded by the EU
opernicus esa European Space Agency

Test Procedures	**Identifier**	Test Procedures
	#TP14	

Metric #M234	**serviceSupportAnswerSatisfaction**	
Description	Average of level of satisfaction with respect to the received answers to a support thread	
Measurement Unit	Rank	
Range values	1 poor – 5 efficient	
Measurement Strategy		
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP14	

Metric #M240	**dataDiscoveryAndAccess**	
Description	This metric provides an indication of data discovery and access for long series of Sentinel-2 data as ratio between positive answered questions over the total on the #TP16	
Measurement Unit	%	
Range values	From a 0 to 100	
Measurement Strategy	Add all positive answers to a set of predefined questions (see TP)	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP16	

Metric #M241	**processingSuccessRate**	
Description	This metric provides an indication of processing of long series of Sentinel-2 data as described in #TP17	
Measurement Unit	%	
Range values	From a 0 to 100	
Measurement Strategy	Ratio of Sentinel-2 images completed on the processed ones: inputs_completed/inputs_processed	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP17	

Metric #M242	**cloudOptimizedToolsAdequacy**
Description	This metric provides an indication of the capability to provide cloud optimized service as described in #TP18
Measurement Unit	%

Funded by the EU
Copernicus esa European Space Agency

Range values	From a 0 to 100	
Measurement Strategy	Add all positive answers to a set of predefined questions (see TP)	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP18	

Metric #M243	**costPerformance**	
Description	This metric provides an indication of processing cost of long series of Sentinel-2 data as described in #TP19	
Measurement Unit	%	
Range values	From a 0 to 100	
Measurement Strategy	Test Cost Performance as: 1 - ((actual_cost - estimated_cost) / actual_cost) cap to 100%, where: estimated_cost is based on unit costs from the provider's website and the actual_cost is the test cost from the provider's dashboard and/or invoice.	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP19	

Metric #M244	**devCostPerformance**	
Description	This metric provides an indication of processing developing cost of long series of Sentinel-2 data as described in #TP19	
Measurement Unit	%	
Range values	From a 0 to 100	
Measurement Strategy	Developing Cost Performance as: 1 - dev_cost/actual_cost, where dev_cost is the actual cost for development/dry-run cost before test run	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP19	

Metric #M245	**testEfficiency**
Description	This metric provides an indication of processing of long series of Sentinel-2 data as described in #TP17
Measurement Unit	%
Range values	From a 0 to 100

Funded by the EU
Copernicus esa European Space Agency

Measurement Strategy	Ratio of the time elapsed from the jobs on the effective duration test: total_job_runtime_hours/test_duration_hours	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP17	

Metric #M246	**processingTimeRate**	
Description	This metric provides an indication of processing of long series of Sentinel-2 data as described in #TP17	
Measurement Unit	%	
Range values	From a 0 to 100	
Measurement Strategy	Ratio of the time elapsed for the processing on the time elapsed for the total parallel jobs: total_processing_hours/total_pod_lifetime_hours	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP17	

Metric #M247	**developerRank**	
Description	Level of developer satisfaction with respect to the development and integration of the scenario (see TP18)	
Measurement Unit	% - Rank (normalized to 100)	
Range values	0 poor – 5 efficient (From a 0 to 100)	
Measurement Strategy	Supported by the developer's comment during the setup and integration of scenario	
Dependencies and Variability	N/A	
Limitations / Other comments	N/A	
Test Categories	Manual Procedures	
Test Procedures	**Identifier**	Test Procedures
	#TP18	

Funded by the EU
Copernicus esa European Space Agency

11ANNEX F - THE BENCHMARKING SOFTWARE SUITE

The Copernicus Sentinels Data Access Worldwide Benchmark (CDAB) Test Suite is the software suite used to run Test Scenarios for benchmarking various Copernicus Data Provider targets. The suite is released and progressively enhanced after consolidation in operations.

The currently supported Target Sites are the Data Access Hubs:

- Copernicus Open Access Hub (aka SciHub)
- Copernicus Open Access Hub API (aka APIHub)
- Copernicus Collaborative Data Hub (aka ColHub)
- Copernicus Sentinels International Access Hub (aka IntHub)
- Any other Data Access Hubs using DHuS Data Hub software

and the DIASes:

- CREODIAS
- Mundi Web Services
- ONDA
- Sobloo
- WEkEO

The CDAB Test Suite is an ensemble of Software components to measure metrics in a set of Test Cases organised in Test Scenarios. It is provided in a public repository available at this URL:

- https://github.com/esa-cdab/cdab-testsuite

that contains in the **src** folder the source code of the two Command Line Tools

  - **cdab-client** for local Test Scenarios (#TS0X)
  - **cdab-remote-client** for remote Test Scenarios (#TS1X)

and in the **Use Cases** folder a set of end to end scenarios sub-folders including

  - stories, user profiles, scenario variables
  - procedures for application integration

The CDAB Test Suite is built automatically providing a docker image available publicly at esacdab/testsuite that can be used as Test Site.

The CDAB Test Suite including build, configuration and usage examples is provided at the Getting Started wiki at this URL:

- https://github.com/esa-cdab/cdab-testsuite/wiki

Funded by the EU
Copernicus esa European Space Agency