Deliverable D1.2 – Report on CONCERTO Data Sharing Framework

Lead Beneficiary: MEEO

Author/s: Simone Mantovani, Damiano Barboni (MEEO)

30/06/2025

Funded by
the European Union

Prepared under contract from the European Commission

Grant agreement No. 101185000

Project acronym:	**CONCERTO**
Project full title:	**Improved CarbOn cycle represeNtation through multi-sCale models and Earth obseRvation for Terrestrial ecOsystems**
Start of the project:	January 2025
Duration:	48 months
Project coordinator:	Manuela Balzarolo
Deliverable title:	Report on CONCERTO Data Sharing Framework
Deliverable n°:	D1.2
Nature of the deliverable:	[Report]
Dissemination level:	[Public]
WP responsible:	WP1
Lead beneficiary:	[MEEO]
Author/reviewers	**Authors**: Simone Mantovani, Damiano Barb **Reviewers**: Filippo Malvolti (MEEO) Manuela Balzarolo (CMCC)
Citation:	Mantovani, S., Barboni, D. (2025) CONCERTO ***D1.2. Report on CONCERTO Data Sharing Framework***
Due date of deliverable:	Month n° 6
Actual submission date:	Month n° 6
Deliverable status:	Submitted

Version	Status	Date	Author(s)/Reviewer
0.1	Draft	15/05/2025	Simone Mantovani – MEEO – Initial draft and organisation of contents
0.2	Draft	21/05/2025	S. Mantovani, D. Barboni - MEEO
0.3	Draft	19/06/2025	S. Mantovani, D. Barboni - MEEO
1.0	Review	27/06/2025	Filippo Malvolti – MEEO
1.1	Review	28/06/2025	Manuela Balzarolo – CMCC
2.0	Final	29/06/2025	Sara Totaro – CMCC – proofreading, layout

Contents

List of Abbreviations 4
Executive summary 5
1. Introduction 6
2. Data Sharing Framework 8
2.1. Overall architecture 9
2.2. User roles 11
2.3. Infrastructure and deployment approach 12
3. User interfaces 13
3.1. Explorer 14
3.2. Dashboard 17
4. FAIR Services 18
4.1. S3 API 19
4.2. OGC Services API 19
4.3. STAC API 19
4.4. OpenSearch API 19
5. Data Access and Processing 20
6. Conclusions 21
7. References 22

List of Abbreviations

ARCO	Analysis-Ready Cloud-Optimised
API	Application Programming Interface
CLI	Command Line Interface
DAP	Data Access and Processing
DMP	Data Management Plan
DSF	Data Sharing Framework
FAIR	Findability, accessibility, interoperability, and reusability
FTP	File Transfer Protocol
FTPS	File Transfer Protocol Secure
HE	Horizon Europe
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
IAM	Identity Access Management
ISO	International Organization for Standardization
OGC	Open Geospatial Consortium
S3	Simple Storage Service
STAC	SpatioTemporal Asset Catalog
UI	User Interface
WCS	Web Coverage Service
WMS	Web Map Service
WP	Work package

Executive summary

This document reports on deliverable *1.2 – Report on CONCERTO Data Sharing Framework* to be submitted at month 6. It is a living document that evolves with technology and organisational needs, ensuring that CONCERTO data sharing remains aligned with the project strategic objectives and public trust.

> The **data sharing framework (DSF)** is a key element of the CONCERTO project. It establishes a **structured, secure, and scalable approach** for enabling efficient data exchange both among partners within the project Consortium and with external organisations, stakeholders, and the general public. It is designed to **promote interoperability, transparency, and responsible data use** while supporting organisational goals such as informed decision-making, innovation, and compliance with legal and ethical standards.

The DSF outlines key principles including data governance, access control, metadata standards, and data quality assurance. It encourages the use of open standards (such as Open Geospatial Consortium - OGC and International Organization for Standardization - ISO), ensures privacy and security by design, and promotes role-based access to safeguard sensitive information.

Furthermore, by fostering a culture of collaboration and data stewardship, this framework aims to reduce data silos, enhance cross-sector cooperation, and unlock the full value of data.

1. Introduction

The CONCERTO project - *Improved CarbOn cycle represeNtation through multi-sCale models and Earth obseRvation for Terrestrial ecOsystems* - was launched in January 2025 under the Horizon Europe (HE) Programme and brings together 13 partners spread across Europe and beyond[1].

Terrestrial ecosystems are essential to the functioning of the global carbon cycle and play a fundamental role in mitigating climate change. According to the Global Carbon Budget, approximately 30% of the atmospheric carbon dioxide (CO_2) emissions are removed through vegetation uptake via photosynthesis and biomass accumulation, forming terrestrial carbon sinks. However, this carbon sequestration potential can be constrained by land use change, nutrient limitations, and extreme events (e.g. droughts, fires and heat waves).

> CONCERTO aims to **improve the current understanding of the terrestrial carbon cycle** and accurately representing the underlying physical and ecological. Achieving this requires the **integration and exchange of a wide array of data**, including satellite-based products, model outputs and forecasts, and *in situ* measurements, across the project Consortium.

To support this main objective, CONCERTO brings together a **multidisciplinary scientific community** comprising data modellers, machine learning researchers, and other domain experts. These groups will rely heavily on diverse datasets to develop and refine models, train machine learning algorithms, and validate simulation results.

Work Package 1 – Data sharing framework is the central component of CONCERTO responsible for data management. As illustrated in Figure 1, its primary objective is to **establish a comprehensive data sharing framework** that will host the datasets identified as relevant by CONCERTO researchers for carrying out project activities. WP1 also focuses on designing the framework architecture, ensuring the collection of appropriate datasets, and supporting data accessibility and interoperability across the project.

[1] 11 partners are part of EU, in addition 2 partners are from the UK.

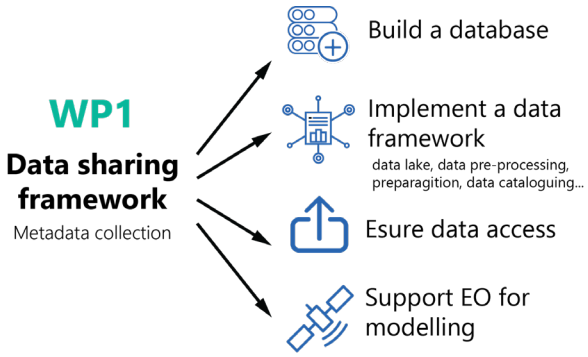

Figure 1. Main objectives of WP1.

D1.2 – Report on CONCERTO Data Sharing Framework has been developed alongside the *D8.6 – Data Management Plan* (DMP) and *D1.1 – Report on data requirements*, as shown in Figure 2.

While the DMP is intentionally structured to offer a clear and descriptive overview of the principles, approaches, and governance mechanisms guiding data use in CONCERTO, this report provides the strategic foundation and technical implementation details on the process of building the DSF. The deliverable links are illustrated in Figure 2. It specifically **supports both technical and policy objectives** by establishing data sharing agreements, usage licensing, and compliance monitoring mechanisms. It also provides guidance on implementing APIs, catalogue services, and cloud-based architectures for scalable and timely data delivery.

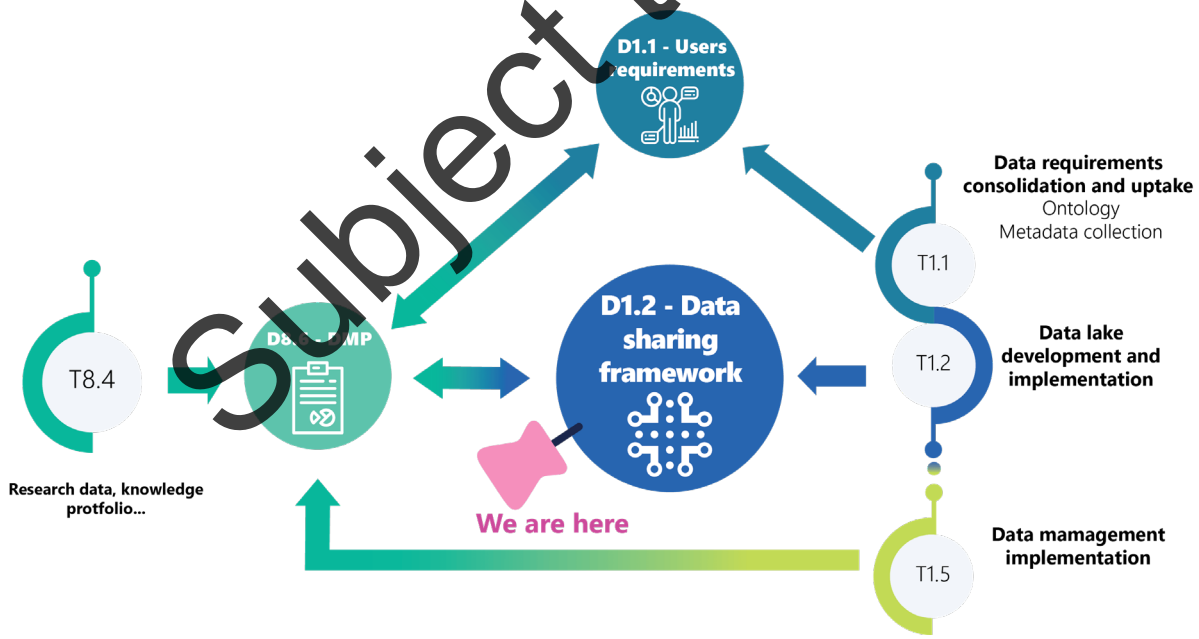

Figure 2. Visual representation of work packages, tasks and deliverables contributing to the Data Sharing Framework deliverable D1.2.

2. Data Sharing Framework

At the heart of CONCERTO's mission is the development and implementation of a robust data sharing platform that is strategically designed to integrate high-resolution Earth observation data, advanced land surface models, data assimilation, and machine learning techniques. This platform will improve data usability by removing access barriers to different data sources, maximise the exploitation of available data and extend the lifecycle of datasets beyond the project duration, ensuring continued value for the broader scientific and policy communities.

> A DSF defines and implements a **structured set of guidelines that govern the secure and efficient exchange of data among partners**, including individuals, organisations, or services. It encompasses policies, procedures, standards, and technical protocols designed to ensure that data is shared in a consistent, compliant, and interoperable manner.

Establishing a DSF is particularly important when managing sensitive, personal, or proprietary data, as it establishes trust, transparency, and accountability.

Key components of a DSF are summarised in Figure 3.

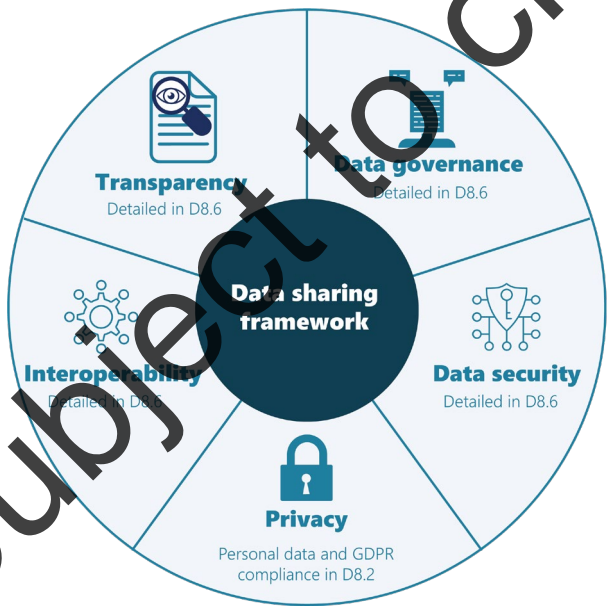

Figure 3. Components of the DSF. As explained before, D1.2 *is linked to other deliverables where the various aspects on data management have been explained. This figure points to the deliverables dealing with specific aspects of data management.*

1. Data governance – defines who owns the data, who can access them, and how they can be used – detailed explanation in Section 3.1 of *D8.6 – Data Management plan*.
2. Security – ensures that data are protected from unauthorised access, breaches, or misuse, through encryption, access controls, etc. – detailed explanation in Section 7 of *D8.6 – Data management plan*.
3. Privacy – addresses how personal or sensitive data is handled to comply with privacy regulations (e.g., GDPR) – detailed in *D8.2 – Ethics (gender, equality/persona data) requirements*
4. Interoperability – ensures that systems and platforms can communicate effectively when sharing data, often through standardized formats or APIs - – detailed explanation in Section 6.3 of *D8.6 – Data management plan*.
5. Transparency – providing clarity about what data is shared, who it is shared with, and for what purposes – detailed explanation in Section 5 of *D8.6 – Data Management plan*.

A well-designed data-sharing framework can be crucial for promoting collaboration, fostering innovation, and enabling data-driven decision-making while maintaining security and compliance standards.

In the following sections, the overall architecture, the different user roles of the DSF, the infrastructure and deployment approach are described.

2.1. Overall architecture

> CONCERTO's DSF aims to provide a unique endpoint for its community as well as for external users (Section 2.2), offering all the relevant actions for data handling, such as add, update and remove datasets and products, discover, access and explore the data collected and generated within the project.

The high-level architecture (Figure 4) shows the major components groups by functional capabilities, their relationships and interfaces, and how users with different roles can interact with the DSF.

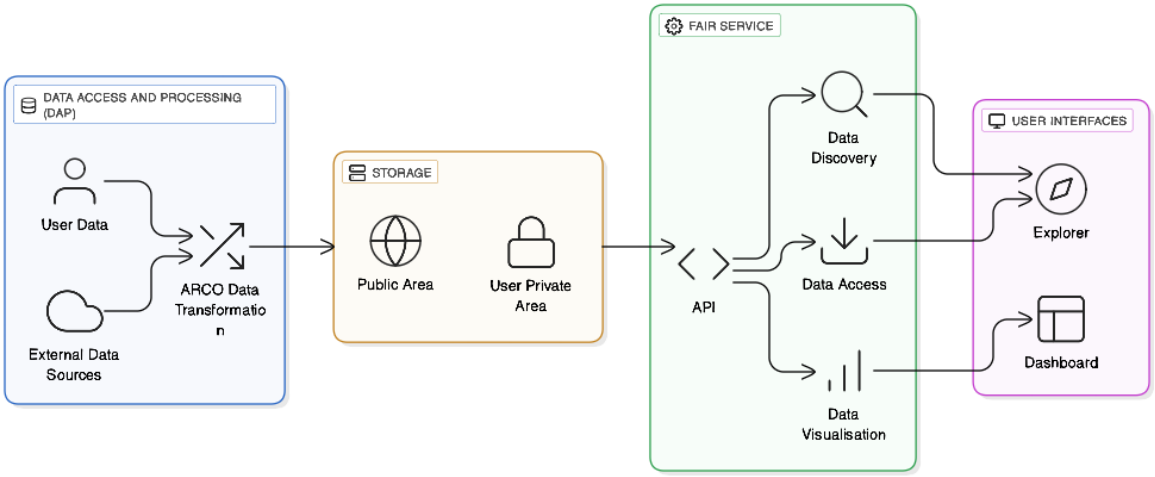

Figure 4. DSF: main components of its architecture.

Leveraging on state-of-the-art cloud technologies and services to properly manage heterogeneous datasets distributed in and/or coming from different data sources and providers, the following modules are described below.

- **Identity access management** (IAM) serves to manage, control and monitor the accesses to the underlying platform, software stack layers and submodules.
- **User interfaces** (UI), including human-friendly (Explorer GUI, Dashboard, …) and machine-to-machine (API, S3, …) are devoted to the provision of an integrated environment to allow users developing their workflows along with a web portal for accessing the produced indicators.
- **Data access and processing** (DAP) for the end-to-end data management workflow, including:
    - user data provisioning – to provide users with functionality to upload and manage their own data in the data storage layer;
    - programmatic data retrieval – to provide the system operators with fully automated functionality to access data from remote sources, eventually apply data transformation from native to cloud-native data format, store in the data storage layer, index in the FAIR backend services in order to enable discovery, access and visualisation;
    - analysis-ready cloud-optimised (ARCO) data transformation – to prepare and convert native data into a format that is optimised for analysis in cloud environments. This transformation is crucial for enhancing the

accessibility, efficiency, and usability of data for scientific research and analysis.

- **Data storage** – to provide real-time, secure, and scalable access to data across various applications, users, and analytics tools. The data storage is based on object storage as the state-of-the-art technology adopted for storing large volumes of geospatial data across multiple locations for redundancy and availability.

- Interoperability and adherence to FAIR principles for discovery, access and visualisation (opensearch, WCS, WMS, etc.).

2.2. User roles

The DSF supports a range of user roles, each with distinct permissions and responsibilities. These roles ensure appropriate access control, system governance, and user experience tailored to different levels of engagement with the platform.

In Table 1 the four main roles identified for the CONCERTO project are listed, including information on the access level and description of the main functionalities.

Table 1. DSF: user roles.

User roles	Access level	Functionalities
Guest (not-registered)	Limited	• Browse publicly available datasets or metadata • View documentation and general platform information
Registered user	Standard	• Access additional datasets (subject to data access policies) • Download datasets or use APIs where permitted • Participate in data discussions or provide feedback • Manage a personal profile and data access requests
System operator	Elevated	• Maintain system functionality and uptime • Monitor system performance and troubleshoot issues • Enforce data governance policies at a technical level • Implement updates and manage integrations with other systems

Administrators	Full	• Manage user roles, permissions, and account statuses • Approve or reject dataset submissions and access requests • Oversee data licensing, compliance, and policy enforcement • Configure system-wide settings and access controls • Audit system activity and generate usage reports

2.3. Infrastructure and deployment approach

To ensure scalable, resilient, and automated deployment, the DSF is designed with high granularity and micro-services for deployment in a Kubernetes-based architecture[2] integrated with CI/CD pipelines[3].

Table 2 shows the list of CONCERTO components identified, including the sizing of resources broken down per component.

Table 2. CONCERTO components.

component	namespace	amount	core [max value]	RAM [max value]	Volume	Autoscaling
OpenSearch / STAC	concerto-catalog	1	1	1		FALSE
WMS / WCS	concerto-wxs	4	1	2		TRUE
mongo	concerto-mongo	2	4	8		FALSE
dap	concerto-dap	1	4	8		FALSE
Explorer	concerto-explorer	1	1	1		FALSE
Explorer backend	concerto-explorer	1	1	1		FALSE
Dashboard	concerto-dashboard	1	1	1		FALSE
Public data storage					TBD	
User private storage					TDB	

[2] https://kubernetes.io/docs/concepts/architecture/
[3] https://docs.gitlab.com/ci/pipelines/

This setup enables continuous integration, testing, and delivery of application updates, while providing robust cloud-native support for managing workloads and services. The continuous integration and continuous delivery CI/CD method is a key activity and with advent of the cloud services it has become central to define the portability strategy to easily deploy across different cloud infrastructures.

A series of tools and technologies will support the collaborative development of DSF components as well as the service deployment., including:

- GitLab[4], as code repository and collaborative software development platform,
- GitLab CI[3], as tool to support image building and test automation,
- Docker Repo[5], as repository to store and share docker-images with the project team,
- Helm[6], as tool to automate the creation, packaging, configuration, and deployment of Kubernetes applications into a reusable package,
- Kubernetes[7], as tool to manage containerized workloads and services.

The pairing of Kubernetes and Helm guarantees the portability between different cloud providers and a standardised integration and deployment procedure.

3. User interfaces

The DSF is the primary gateway for storing and exploring the data retrieved and generated by users.

Human-friendly and API / machine-to-machine interfaces are planned as two distinct access points:

- Explorer and Dashboard, the web-based interfaces granting a user-friendly access to users, operators and administrators
- FAIR services through APIs, for users seeking a more programmatic approach (see section 4).

[4] https://about.gitlab.com
[5] https://hub.docker.com
[6] https://helm.sh
[7] https://kubernetes.io

3.1. Explorer

The Explorer graphic user interface is a key element for users seeking an interactive and user-friendly approach to discover, explore, and access data belonging to the CONCERTO data portfolio.

In the current phase of the project, it is planned to develop a generic user interface which supports the two-step search:

- 1st step: discover the list of available datasets;
- 2nd step: discover data belonging to a specific dataset.

In particular, users can search for datasets by scrolling the list of available collections (see Figure 5) and typing the collection name in the 'Title' field (red box) or by specifying temporal filtering criteria (blue box). Additional filterinng criteria can be defined according to the CONCERTO metadata and ontology as defined in *D8.6 - Data Management Plan* [2] and based on the information collected through the Data Mapping Form (see *D1.1 – Report on data requirements* for more details).

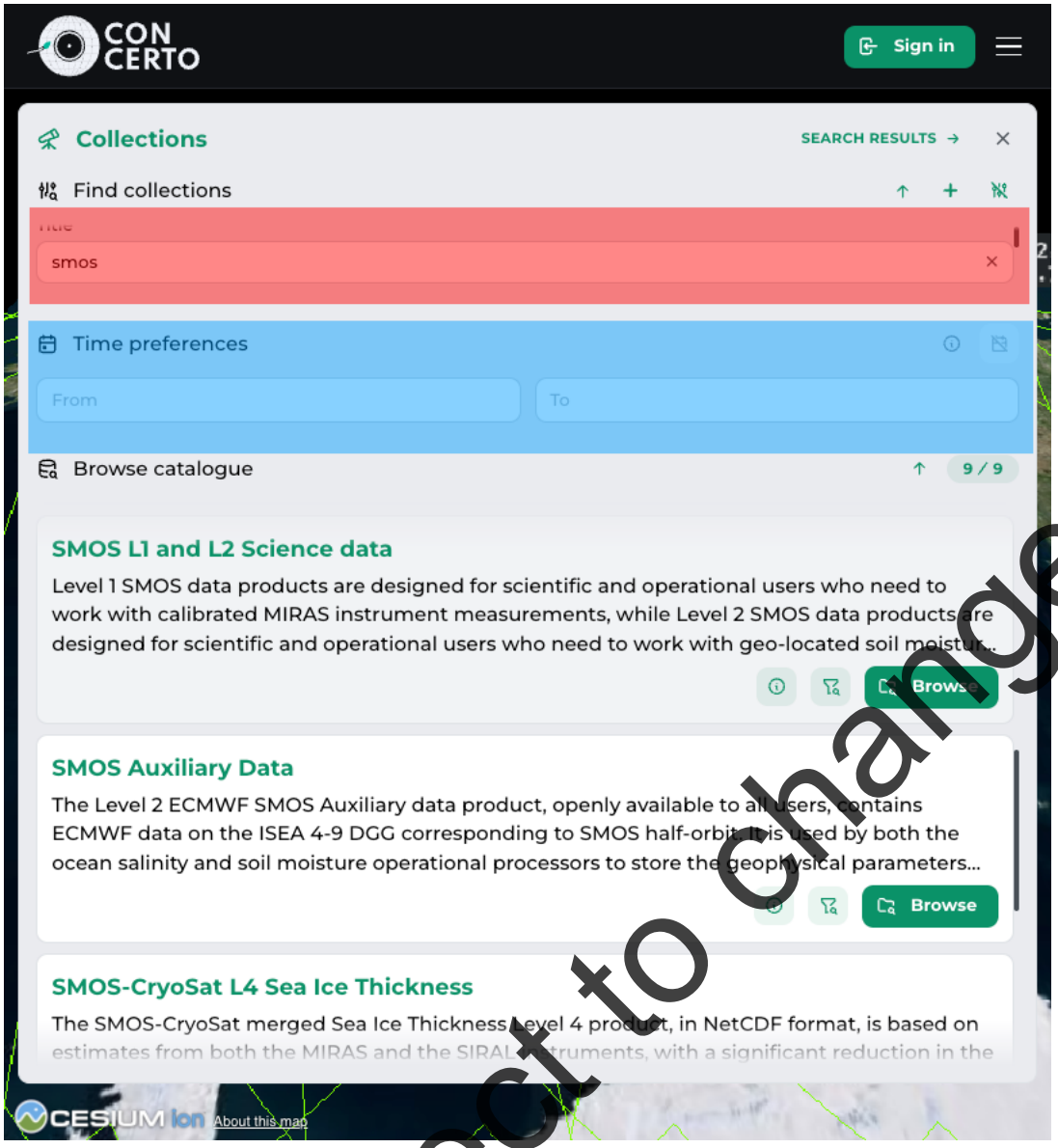

Figure 5. Explorer: dataset discovery (filtered search).

Once selected the dataset of interest, CONCERTO's users can search for data by applying any other '*queryable*' defined at data collection level such as product type, orbit direction, cloud cover, quality flag (see Figure 6).

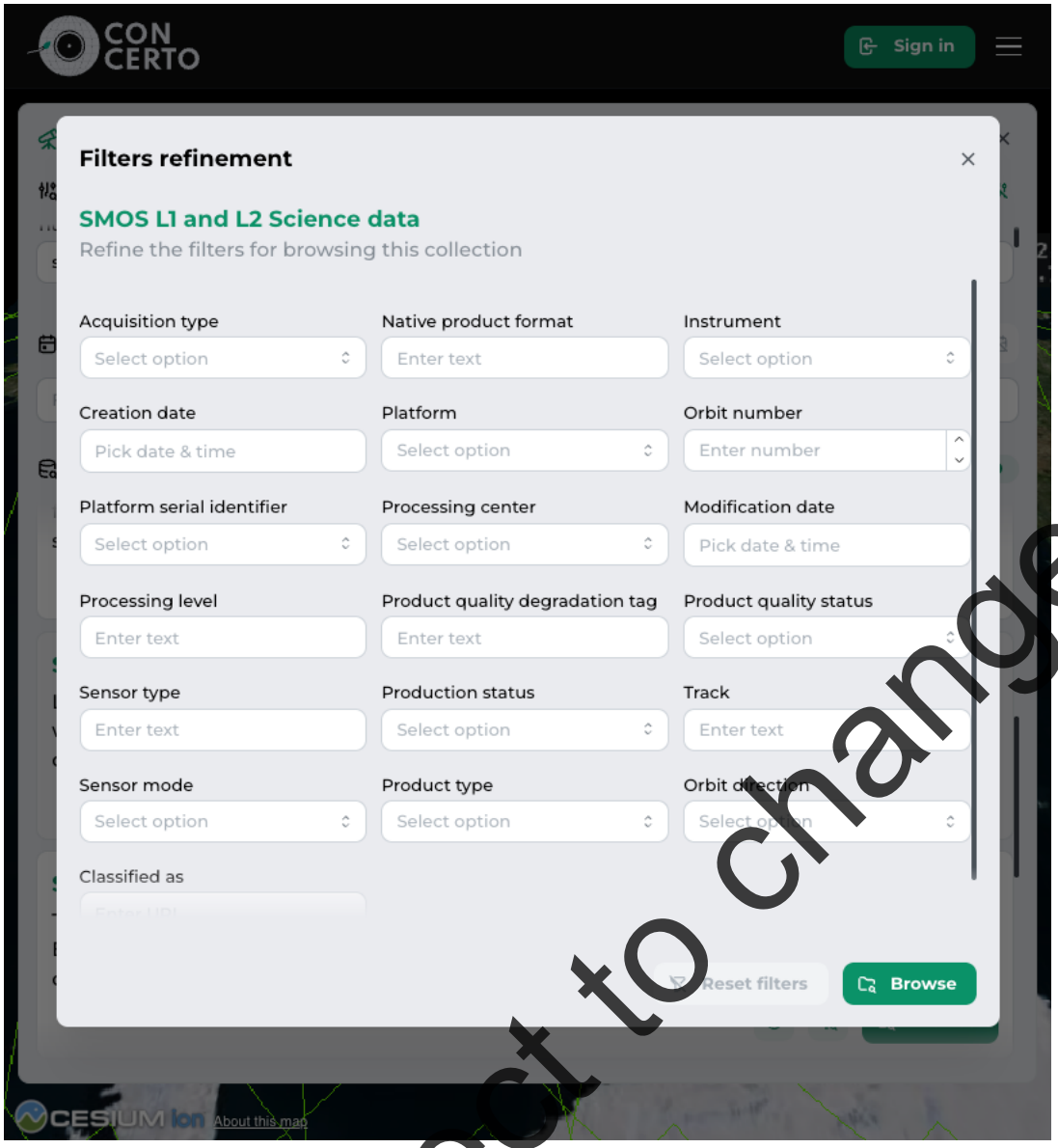

Figure 6. Explorer: data search (2nd step search).

As output, a paginated list of products and metadata corresponding to filtered criteria are returned to users (see Figure 7). Each item has a series of metadata and associated services.

The metadata, including product ID, start date, end date, quick look / thumbnail (when available), download URI, as well as all other metadata, attributes and services are directly retrieved from the data sources.

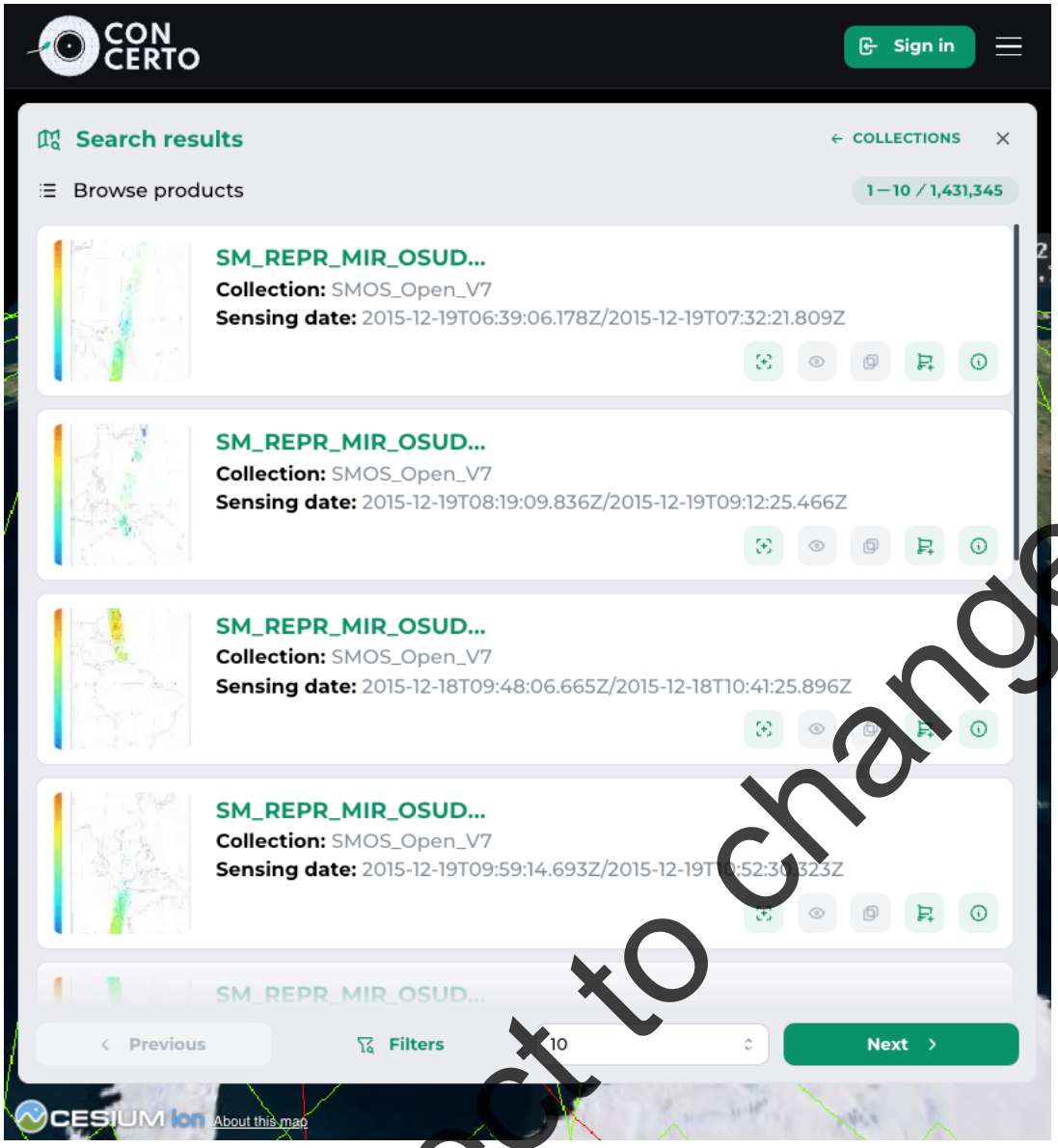

Figure 7. Explorer: data search (2nd step search) result page.

3.2. Dashboard

The Dashboard serves as the central hub for users to handle his/her own resources as well as for operators and administrators to oversee various aspects of the DSF.

The Dashboard user interface (see Figure 8) is a clean, modern, and minimalistic dashboard designed to help users interact with their object storage visually. It complements the CLI (mc) and API by offering an intuitive way to manage buckets, objects, and users.

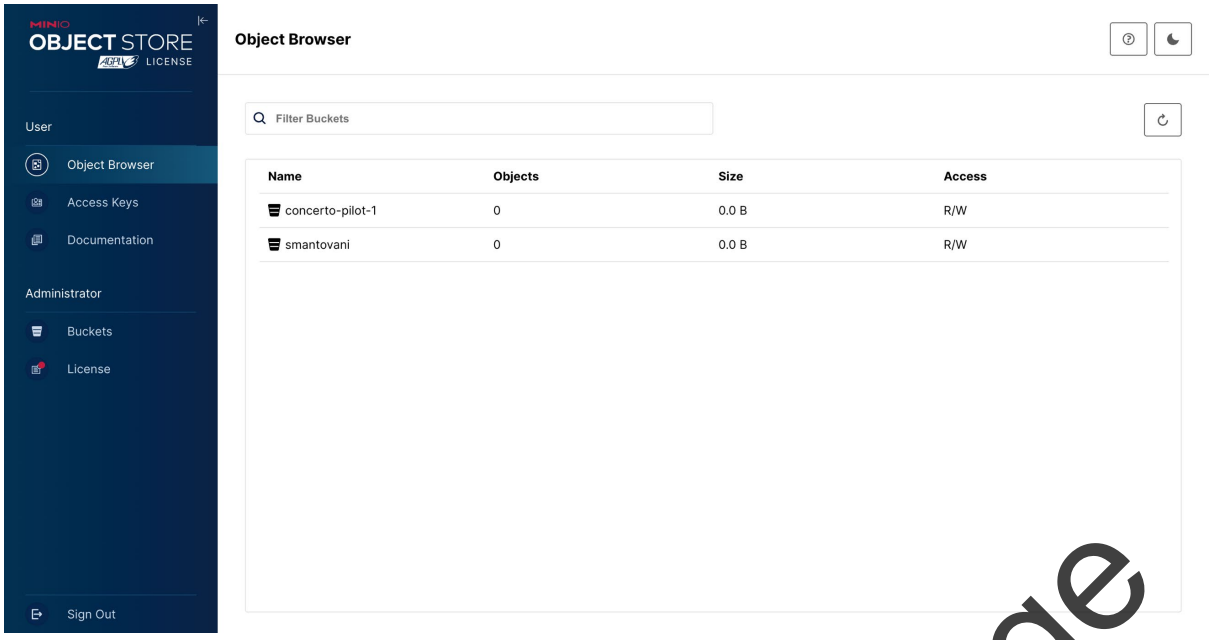

Figure 8. CONCERTO: Dashboard user interface.

On the left sidebar, the Navigation panel offers the access to all the functionalities:

1. object browser:
    - browse buckets,
    - browse bucket contents (folders and objects) via a file explorer-like interface,
    - upload/download/delete files,
    - drag-and-drop support for uploads;
2. bucket management (administrators only):
    - create, delete, and manage buckets;
    - configure bucket policies (read/write access control).

4. FAIR Services

The FAIR Services offers the core APIs that allow access to datasets distributed across several data sources, regardless of the heterogeneous data types. The FAIR principles and how they will be implemented in CONCERTO have been largely discussed in *D8.6 – Data management plan*.

4.1. S3 API

The object storages of CONCERTO data storage service are accessible through the S3 standard API. No public and open access to this storage is foreseen. Data downloads are made through S3 pre-signed URL.

4.2. OGC Services API

The Web Map Service (WMS) is an OGC standard for serving georeferenced map images over the web. It delivers rendered maps (usually in PNG, JPEG, or SVG formats) generated from spatial data on the server.

The Web Coverage Service (WCS) is an OGC standard for delivering raster data (also called "coverages") such as satellite imagery, elevation data, or climate grids.

Table 3. OGC Services API.

API	Description
GET /wcs	Web Coverage Service data access
GET /wms	Web Map Service data access

4.3. STAC API

The SpatioTemporal Asset Catalog (STAC) is an open specification to organise and share geospatial assets (like satellite imagery, aerial photos, and other raster data) in a consistent, searchable way.

Table 4. STAC API.

API	Description
GET /stac	STAC catalog interface to the cached data

4.4. OpenSearch API

OpenSearch is a simple, extensible standard for querying and retrieving search results, originally developed for web search but widely adopted in geospatial applications.

Table 5. OpenSearch API.

API	Description
GET /opensearch/datasets	Get list of available datasets in cache

GET /opensearch/search/{dataset_id}	Filter granule/products available for a given dataset

5. Data Access and Processing

The Data Access and Processing (DAP) component offers the core functionalities that allow to handle the end-to-end data management workflow, including user data and external data sources, regardless of the heterogeneous data types.

This interface handles the retrieval of data from external sources into the system. It ensures secure data transfer to the DSF for subsequent ingestion and storage into the system. As part of the data retrieval process, the data transformation can also be applied to generate analysis-ready cloud-optimised (ARCO) data products, which consist in data that have been organised and processed to a minimum set of requirements to allow immediate analysis. The generic process consists of:

- re-projection, to apply the target coordinate reference system (e.g. to move from satellite coordinate system to EPSG 4326);
- re-gridding, to resample the data at the target spatial resolution;
- masking, to apply quality flags and / or cloud mask and / or land/sea mask.
- aggregating, to images at the target temporal resolution (e.g. monthly average, quarterly composites, …)
- encoding, to apply the target data format and data packaging (e.g. and cloud-optimized GeoTIFF, Zarr, …).

To effectively manage the multiple external data services offered by the remote data sources and provide flexible, secure and scalable solution, the data retrieval consists of a series *ad hoc* tools implemented as k8s cronjob[8].

Each DAP subcomponent exploits the data source services, such as HTTP/HTTPS and RESTful API in case of web access services, FTP/FTS, other protocols and interfaces as required.

[8] https://kubernetes.io/docs/concepts/workloads/controllers/cron-jobs/

Products are stored in the CONCERTO public or user private storage according to the following archiving structure:

<root directory>/, this is the object storage

[WP##/Task##/], this is optional to keep data generated within different work packages and tasks well organised in the cloud storage

<dataset ID>/, it refers to the data collection

<variable ID>/, it refers to the variable of the dataset

[<YYYY>/<MM>/<DD>/], this is optional and the nested levels depends on the data characteristics on a case by case

<product ID>, it refers to the physical file

6. Conclusions

This document reports on the CONCERTO's DSF which aims to reduce data silos, enhance cross-sector cooperation, and unlock the full value of data while adhering to the FAIR principles.

It is conceived as a living document that will evolve with technology and organisational needs, ensuring CONCERTO's data sharing remains aligned with strategic objectives and public trust, maximising data exploitation, and extending the lifecycle of collected and generated datasets beyond the project's duration.

7. References

REF NUMBER	DOCUMENT / LINK
1	L. Cupelli, (2025) CONCERTO ***D8.2. Ethics (gender, equality, personal data) requirements***.
2	S. Totaro, M.L. Quarta., F. Malvolti, N. (2025) CONCERTO ***D8.6. Data management plan***