Reference:
ERGO-SPB-D4000.4

Issue:
1.0

Date:
02/06/2021

ESE-ERGO

Executive Summary

Prepared by:

Y. Coene

Verified by:

Y. Coene

Approved by:

Y. Coene

Yves Coene
Digitally signed by Yves Coene
Date: 2021.06.03 09:02:26 +02'00'

Contract

Contractual	:	✓
Contract issuer	:	ESRIN
Contract n°	:	602 (Internal)

Classification

General Public	:	
Industry Programme	:	✓
Restricted Dispatching Programme	:	
Confidential Programme	:	

Configuration

Configured document	:	✓
Non-configured document	:	

References

Reference	:	ERGO-SPB-D4000.4
Issue	:	1 – 05/05/2021
Revision	:	0 – 05/05/2021

Internal Distribution

Name	*Dept.*	*Copies*	*For Information*	*For Action*	*Comments*
C. Bawin	QA	1		✓	
Y. Coene	Space	1		✓	
M. Gilles	Space	1		✓	
T. Tran Minh	Space	1	✓		
M. Nguyen Quang	Space	1	✓		
V. Demeuse	CM	1		✓	

External Distribution

Name	*Company*	*Copies*	*For Information*	*For Action*	*Comments*
A. Della Vecchia	Randstad	1		✓	
J. van Bemmelen	ESRIN	1		✓	

Document Change Log

Issue	Date	§ or Page	Relevant Information
1.0	02/06/2021	All	Original.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Purpose

This document is the Executive Summary of the ESE-ERGO (ESE Ergonomic User Interfaces) project. This project is carried out for ESRIN by SPACEBEL (Belgium).

The document is a formal deliverable D4000.4 of the ESE-ERGO project Work Package 4000.

1.2 Scope of the Document

This document covers all the activities of the ESE-ERGO project.

1.3 Readership of the Document

This document is meant for the ESA Technical Officer and other ESA staff interested in the outcome of the ESE-ERGO Project.

1.4 Organisation of the Document

This document contains the following chapters:

- Chapter 1 is the introduction to this document.
- Chapter 2 lists the applicable and reference documents.
- Chapter 3 lists terms, definitions and abbreviations used in this document.
- Chapter 4 lists the project objectives.
- Chapter 5 describes the project organisation in terms of Consortium and project logic.
- Chapter 6 shows the project schedule.
- Chapter 7 lists the results produced by the project.
- Chapter 8 provides feedback in terms of lessons learned and future work.

2 APPLICABLE AND REFERENCE DOCUMENTS

The following tables identify applicable and reference documents for the project. In the body of the text these documents are referenced as listed here below. Without any indication of issue, the latest issue is applicable. If one issue is indicated, only that issue of the document is applicable.

2.1 Applicable Documents

Ref.	*Code*	*Title*	*Issue*	*Date*
AD01	ESA-EOPG-EOPGM-3	Statement of Work for ESA Ergonomic User Interfaces (ESE-ERGO).	1.0	12/03/2019
AD02	ERGO-SPB-O-0602/V2	ESE-ERGO Detailed Proposal	1.0	11/05/2019
AD03	ECSS-E-ST-40C	ECSS – Space Engineering Standards.		06/03/2009
AD04	ERGO-SPB-D4000.1	ESE-ERGO Software Project Management Plan.	1.0	02/12/2019
AD05	ECSS-M-ST-40C	ECSS – Configuration and Information Management.		06/03/2009
AD06	4000129090/19/I-DT	ESA/Contract No. 4000129090/19/I-DT, GSTP – ESE-ERGO ESE Ergonomic User Interfaces.		

Table 1: Applicable Documents

2.2 Reference Documents

Ref.	*Code*	*Title*	*Issue*	*Date*
RD01	ERGO-SPB-CIDL	ESE-ERGO Configuration Item Data List	2.0	30/04/2021
RD02	OGC 17-069r3	OGC API – Features – Part 1: Core		
RD03	OGC 17-003r2	OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard, https://docs.opengeospatial.org/is/17-003r2/17-003r2.html [CIDL-948].	1.0	14/02/2020
RD04		OpenAPI Specification, https://github.com/OAI/OpenAPI-Specification/blob/master/versions/3.0.2.md, [CIDL-711].	3.0.2	
RD05	RFC 7159	The JavaScript Object Notation (JSON) Data Interchange Format, http://www.ietf.org/rfc/rfc7159.txt,.		
RD06	OGC 17-084r1	OGC EO Collection GeoJSON(-LD) Encoding, https://docs.ogc.org/bp/17-084r1/17-084r1.html	1.0	21/04/2021
RD07		Data on the Web Best Practices: Data Quality Vocabulary, W3C Working Group Note 15 December 2016, https://www.w3.org/TR/vocab-dqv/, [CIDL-715].		15/12/2016
RD08	OGC 17-047r1	OGC OpenSearch-EO GeoJSON(-LD) Response Encoding Standard, https://docs.opengeospatial.org/is/17-047r1/17-047r1.html	1.0	27/04/2020

Ref.	Code	Title	Issue	Date
RD09	OGC 10-032r8	OGC OpenSearch Geo and Time Extensions.		10/04/2014
RD10	OGC 13-026r9	OGC OpenSearch Earth Observation Extension, http://docs.opengeospatial.org/is/13-026r9/13-026r9.html	1.1	25/11/2019
RD11	ISO 19168-1:2020	ISO 19168-1:2020(en), Geographic information — Geospatial API for features — Part 1: Core, https://www.iso.org/obp/ui/#iso:std:iso:19168:-1:ed-1:v1:en		2020
RD12		https://github.com/dewitt/opensearch/blob/master/mediawiki/Community/Proposal/Specifications/OpenSearch/Extensions/SRU/1.0/Draft%201.wiki	1.0	

Table 2: Reference Documents

3 TERMS, DEFINITIONS AND ABBREVIATED TERMS

3.1 Definitions

Broker	A delivery broker can assist in making decisions both as to where users are directed upon requesting resources. A broker does not imply a translation from one protocol to another.
Cascading Map Server	A "Cascading Map Server" is a WMS that behaves like a client of other WMSes and behaves like a WMS to other clients. For example, a Cascading Map Server can aggregate the contents of several distinct map servers into one service. Furthermore, a Cascading Map Server can perform additional functions such as output format conversion or coordinate transformation on behalf of other servers. Source: WMS Implementation Specification OGC 01-068r3.
Collection [OGC]	A set of features from a dataset. Source: OGC 17-069r3 [CIDL-954]. Used as synonym for Feature Collection.
Collection [EO]	A collection5 is an aggregation of granules sharing the same product specification. A collection typically corresponds to the series of products derived from data acquired by a sensor on board a satellite and having the same mode of operation. Different agencies use different terms for "collection": collection (CNES, NASA), dataset (JAXA), dataset series (ESA), product (JAXA).
Container(s) as a Service (CaaS)	Containers-as-a-Service (CaaS)[1] is a cloud service model that allows users to manage and deploy containers, Linux applications and clusters through container-based virtualization. Container engines, orchestration and the underlying compute resources are delivered to users as a service from a cloud provider.
Dataset [OGC]	collection of data, published or curated by a single agent, and available for access or download in one or more formats. Source: OGC [RD56].
Dataset [EO]	Alternative term for (EO) collection or (EO) granule. Source: CEOS.
Docker container	A Docker container is a standardized, encapsulated environment that runs applications. A container is managed using the Docker API or CLI. A Docker image is a read-only template used to build containers
Docker image	A Docker image is an inert, immutable, file that's essentially a snapshot of a container. Images are created with the build command, and they'll produce a container when started with run. Images are stored in a Docker registry such as registry.hub.docker.com.
Dockerfile	A Dockerfile is a text document that contains all the commands a user could call on the command line to assemble an image[2].
Feature [OGC]	Aabstraction of real world phenomena [ISO 19101-1:2014]. Source: OGC OGC 17-069r3 [CIDL-954].
Gateway	Gateways (in data centers) are a transition point from one protocol to another, such as from Ipv6 to Ipv4[3]. A Gateway implies a protocol translation[4].
Granule	A granule[5] is the finest granularity of data that can be independently managed. A granule usually matches the individual file of EO satellite data. Different agencies use different terms for "collection": dataset (ESA), granule (NASA), product (ESA, CNES), scene (JAXA).
Pods	Pods are the smallest deployable units of computing that you can create and manage in Kubernetes. A Pod is a group of one or more containers, with shared storage/network resources, and a specification for how to run the containers.
Product	Same meaning as "granule".

[1] https://susedefines.suse.com/definition/containers-service-caas/
[2] https://docs.docker.com/engine/reference/builder/
[3] http://www.f5.com/pdf/white-papers/f5-vmware-integrating-cloud-white-paper.pdf
[4] http://en.wikipedia.org/wiki/Gateway_%28computer_program%29
[5] http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/OpenSearch/CEOS-OPENSEARCH-BP-V1.2.pdf

3.2 Abbreviated Terms

API	Application Programming Interface
CaaS	Container(s) as a Service
CEOS	Committee on Earth Observing Satellites
CMR	Common Metadata Repository
DCAT	Data Catalog Vocabulary
ECSS	European Collaboration on Space Standardisation
EO	Earth Observation
ESA	European Space Agency
ESRIN	European Space Research Institute
FedEO	Federated Earth Observation
GCMD	Global Change Master Directory
HATEOAS	Hypermedia as the Engine of Application State
IETF	Internet Engineering Task Force
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organisation for Standardisation
OASIS	Organization for the Advancement of Structured Information Standards
OGC	Open Geospatial Consortium
O&M	Observations and Measurements
OSDD	OpenSearch Description Document
RDF	Resource Description Framework
REST	Representational State Transfer
SOA	Service Oriented Architecture
SOW	Statement of Work
SRU	Search/Retrieval via URL
SPB	Spacebel
W3C	World Wide Web Consortium
UML	Unified Modelling Language
WCS	Web Coverage Service
WMC	Web Map Context
WMS	Web Map Service
WP	Work Package
WPS	Web Processing Service
XML	eXtensible Markup Language

4 OBJECTIVES

4.1 Project Overview

The overarching objective of the project is the evolution of the current FedEO components towards the next generation of ESA scalable Catalogue and FedEO Gateway components exploiting the latest JSON-based technology for metadata, internal and external interfaces and adoption of the latest evolution OGC OpenSearch interfaces aiming to optimise performance.

In addition, the Catalogue shall provide support for Collection and Product metadata, but also Service and application metadata as being defined in OGC Testbed-15. The faceted search and sorting capabilities have to be extended and the external interfaces of the Catalogue may exploit the emerging OGC API Common specification.

The FedEO Metadata Editor purpose is to allow management of the metadata in the catalogue through the Catalogue management interfaces and will be completed as part of this project to become usable by the ESA Catalogue operators. It may be extended to interface with ESA Web thesauri and NASA GCMD.

The main technical objectives are described in section 1.4.2 of the SOW [AD01] and are summarised in the next section.

The results of the project consist of three major functional components:

- **Cascading Catalogue**, combining the functionality of the existing FedEO catalogue and Gateway and aligned to the latest available standard interfaces.
- **Metadata Editor**, Web-based metadata management tool for creating, updating and validating collection and/or metadata records and insertion in connected catalogues.
- **Metrics Client Application**, Web-based user interface for presenting catalogue access metrics as agreed in the context of the CEOS WGISS System Level Team meetings.

4.2 Project Objectives

The high-level objectives of the project are two-fold:

- Reviewing emerging IT solutions, technology trends and adopting to bring benefits to the EO community.
- Evolving necessary software components to improve future data access and exploitation mechanisms in terms of:
  - Performance and scalability
  - Widening applicability of the solution

4.2.1 Objective 1 – Adoption of emerging IT solutions

The prime objective of the activity is to adopt emerging IT solutions to enhance the FedEO software components in terms of performance, scalability and applicability. This can be further detailed into the following objectives:

- Adoption of OpenAPI, also encouraged by OGC, as standard interface for its services, in combination with JSON and JSON-LD. OGC API Common shall be considered for advertising a service landing page, API definition and conformance statements.
- Analyse and adopt new international guidelines and standards to improve gateway/catalogue functionalities, performance and interoperability. This includes:
  - Emerging GeoDCAT-AP specification to publish EO metadata
  - OGC 13-026r9 to introduce enhanced functionalities (e.g., flexible string search, results ranking, links to EO external services)

  - OGC 17-047r1 to provide a JSON based encoding response from OpenSearch interface (e.g., quicker catalogue response, ease User interface development)

- Transition to an internal metadata model of which other metadata models are views, to allow more flexibility and to align to similar approaches applied by NASA (e.g., Unified Metadata Model – UMM/-C/-G/-S/-JSON). This includes:
  - Recent OGC guidelines, in cooperation with CEOS/NASA, about collection and product metadata model based on GEO-JSON(-LD) encoding (e.g., OGC 17-084, OGC 17-003)
  - Generalization of the concept of catalogue entry in order to consider granules (/products), collections (aggregation of granules/products) and services related to both of them (e.g., remote processing, datacube, etc…) seamlessly
- Web-based metadata management tool to allow authorised users to add/modify collection catalogue entries.
- Advanced metrics collection for effective governance and consistent infrastructure sizing.

4.2.2 Objective 2 – Gradually enhance EOP-G infrastructure

Provide enhancements to FedEO software components in an incremental way and accompanied with a Transfer to Operations Plan allowing the gradual enhancement of the EOP-G infrastructure which is currently based on these components.

5 PROJECT ORGANISATION

5.1 SOFTWARE DEVELOPMENT APPROACH

5.1.1 Strategy to the Software Development

The software has been developed incrementally and an iterative development process has been used.

5.1.2 Strategy for Software Reuse

Before any open-source components are reused, their licenses have been be checked by QA to avoid contamination by license of other software components.

It was agreed during contract negotiation (letter YCE/ave/19-042, 4/11/2019) to publish the following software components with a non-contaminating open-source license:

- Collection Metadata Editor Tool,
- Metrics Client Application.

5.2 Software Project Development Life-cycle

5.2.1 Software Development Life-cycle Identification

The figure below shows the flow of activities in terms of work packages and review meetings. The Project Management (WP4000) is not shown, as it is a continuous activity that spans the duration of the project.

The project started with the definition of the scenarios (use cases) and requirements for the first iteration (WP1100), which were reviewed at the Requirement Review (RR-1).

The agreed requirements were refined and integrated in the design, implemented and tested at component level in WP2100 while incrementally update the software and the corresponding deliverables. In the Validation, Evaluation and Demonstration work package (WP3100), the validation plan and reports were produced. The evaluation by stakeholders and issues identified were an important input for the second iteration which started at the Mid Term Review (MTR) and had a similar structure as the first iteration.

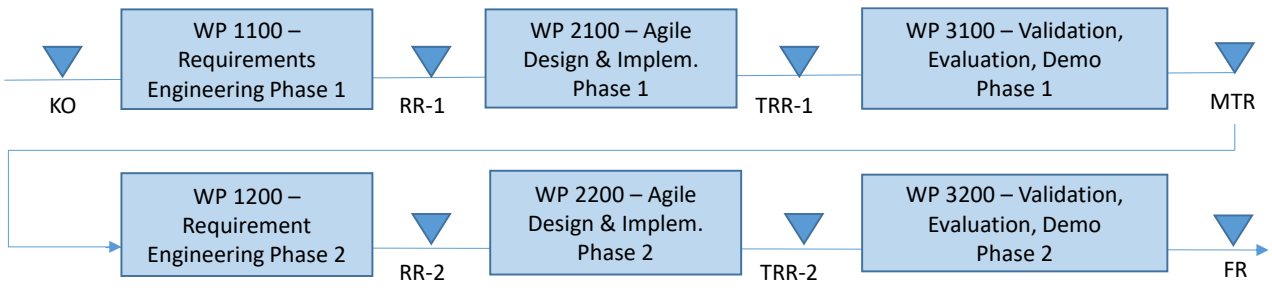


Figure 1: Project Work Logic

5.3 Deliverables

5.3.1 Minutes of Meeting

Date	Reference	Objective	Location
03/12/2019	ERGO-SPB-MOM-001	Kickoff	Teleconference
18/12/2019	ERGO-SPB-MOM-002	Requirements Technical Meeting	ESRIN
17/03/2020	ERGO-SPB-MOM-003	Requirements Review Meeting (RR-1)	Teleconference
11/08/2020	ERGO-SPB-MOM-004	Progress Meeting (PM-1)	Teleconference
16/09/2020	ERGO-SPB-MOM-005	Mid-term Review (MTR)	Teleconference
14/10/2020	ERGO-SPB-MOM-006	Requirements Review Meeting (RR-2)	Teleconference
16/02/2021	ERGO-SPB-MOM-007	Progress Meeting (PM-2)	Teleconference
28/05/2021	ERGO-SPB-MOM-008	Final Presentation (FP/FR)	Teleconference

Table 3: Minutes of Meeting

5.3.2 Project Deliverables

Reference	*Title*	*Issue*	*Date*
ERGO-SPB-D1000.3	ESE-ERGO Software Requirements Specification	1.4	29/01/2021
ERGO-SPB-D1000.3.1	Metadata Editor Software Requirements Specification	1.4	01/04/2021
ERGO-SPB-D1000.3.2	Metrics Client Application Software Requirements Specification	1.2	30/09/2020
ERGO-SPB-D2000.1	Interface Control Document	1.3	03/02/2021
ERGO-SPB-D2000.2	Software Design Document	1.2	01/04/2021
ERGO-SPB-D2000.2.1	Metadata Editor Software Design Document	2.0	01/04/2021
ERGO-SPB-D2000.2.2	Metrics Client Application Software Design Document	1.1	01/04/2021
ERGO-SPB-D2000.3.1	Metadata Editor Software Configuration File	1.1	01/04/2021
ERGO-SPB-D2000.3.2	Metrics Client Application Software Configuration File	1.2	01/04/2021
ERGO-SPB-D2000.3.3	Kubernetes Cluster Software Configuration File	1.0	29/01/2021
ERGO-SPB-D2000.5	Software Verification Plan	1.3	01/04/2021
ERGO-SPB-D2000.5.1	Metadata Editor Software Verification Plan	1.2	11/02/2021
ERGO-SPB-D2000.5.2	Metrics Client Application Software Verification Plan	1.2	29/01/2021
ERGO-SPB-D2000.7	Software Reuse File	1.0	17/07/2020
ERGO-SPB-D3000.2	Validation Test Report	1.1	01/04/2021
ERGO-SPB-D3000.2.1	Metadata Editor Validation Test Report	1.1	01/04/2021
ERGO-SPB-D3000.2.2	Metrics Client Application Validation Test Report	1.1	01/04/2021
ERGO-SPB-D3000.3	Transfer to Operations Recommendations	1.0	25/05/2021
ERGO-SPB-D4000.1	ESE ERGO Software Project Management Plan	1.0	02/12/2019
ERGO-SPB-D4000.2	Project Web Information Pages (PWIP) https://wiki.services.eoportal.org/tiki-index.php?page=ESE-ERGO.		14/04/2020
ERGO-SPB-D4000.4	Executive Summary	1.0	02/06/2021
ERGO-SPB-CIDL	Configuration Item Data List	2.0	30/04/2021

Table 4: Projects Deliverables

In addition to the above deliverables, two software packages are available as open-source on GitHub at the below locations. The software components are presented in more detail in the following subsections.

- https://github.com/spacebel/MDE (Metadata Editor)
- https://github.com/spacebel/MCA (Metrics Client Application)

6 SCHEDULE

The project started with a kick-off on 03/12/2019 and the Final Presentation was held on 28 May 2021.

6.1 Master Schedule

The bar chart below presents the scheduling of the activities.

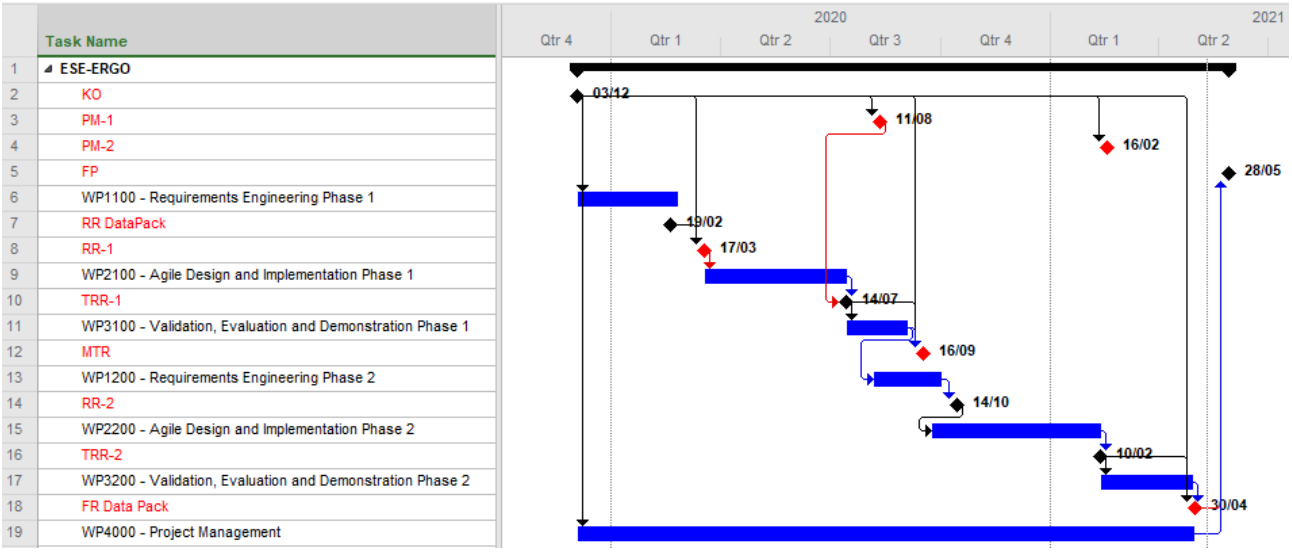


Figure 2: Final schedule

6.2 Meeting Schedule

Event	Location	Date	Original Date	Actual Date	Name
KO	Teleconf	T0		03/12/2019	Kick-off
	ESRIN			18/12/2019	Requirements Technical Meeting
RR-1	Teleconf	T0 + 2M	03/02/2020	17/03/2020	Requirements Review Phase 1
PM-1*	Teleconf	T0 + 6M	03/06/2020	11/08/2020	Progress Meeting 1
MTR-1	Teleconf	T0 + 9M	03/09/2020	16/09/2020	Mid-Term Review
(RR-2)	Teleconf	T0 + 10M	03/10/2020	14/10/2020	Requirements Review Phase 2
PM-2*	Teleconf	T0 + 13M	03/01/2021	16/02/2021	Progress Meeting 2
FR*	Teleconf	T0 + 18M	03/06/2021	28/05/2021	Final Review
PF	Teleconf	T0 + 18M	03/06/2021	28/05/2021	Final Presentation

Table 5: Meeting Schedule

7 TECHNICAL ACHIEVEMENTS

7.1 Overview

The project designed and implement enhancements to the FedEO software components in an incremental way and accompanied with a Transfer to Operations Plan (ERGO-SPB-D3000.3) allowing the gradual enhancement of the EOP-G infrastructure which is currently based on these components.

The following diagram depicts the configuration items (CI) part of the EOP-G (Common) Infrastructure affected by the ESE-ERGO project.

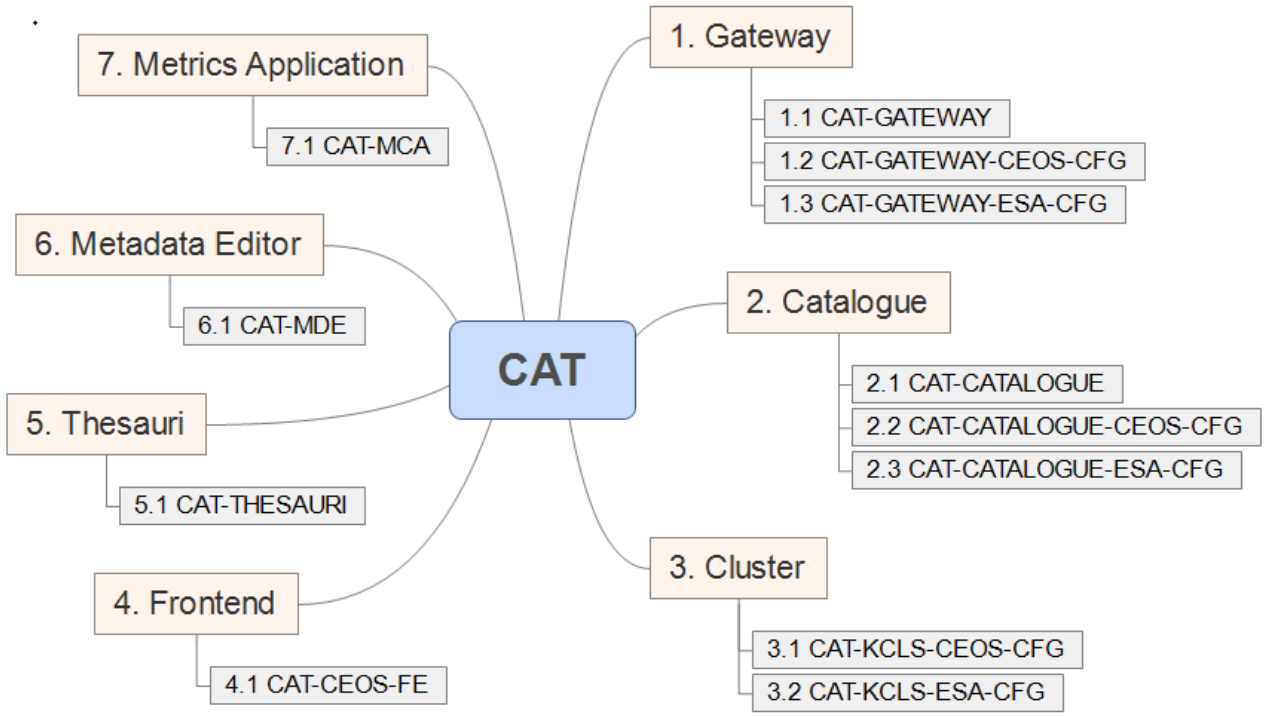


Figure 3: Configuration Items

The following new CI are an output of the project:

- 6. Metadata Editor
- 7. Metrics Client Application

In addition, the project delivered enhanced versions of:

- 1. Gateway
- 2. Catalogue

All components were hosted on a Kubernetes cluster (3) with an enhanced configuration to be able to capture access metrics in a non-intrusive way as documented in ERGO-SPB-D2000.3.3. The full set of components is depicted in *Figure 4*.

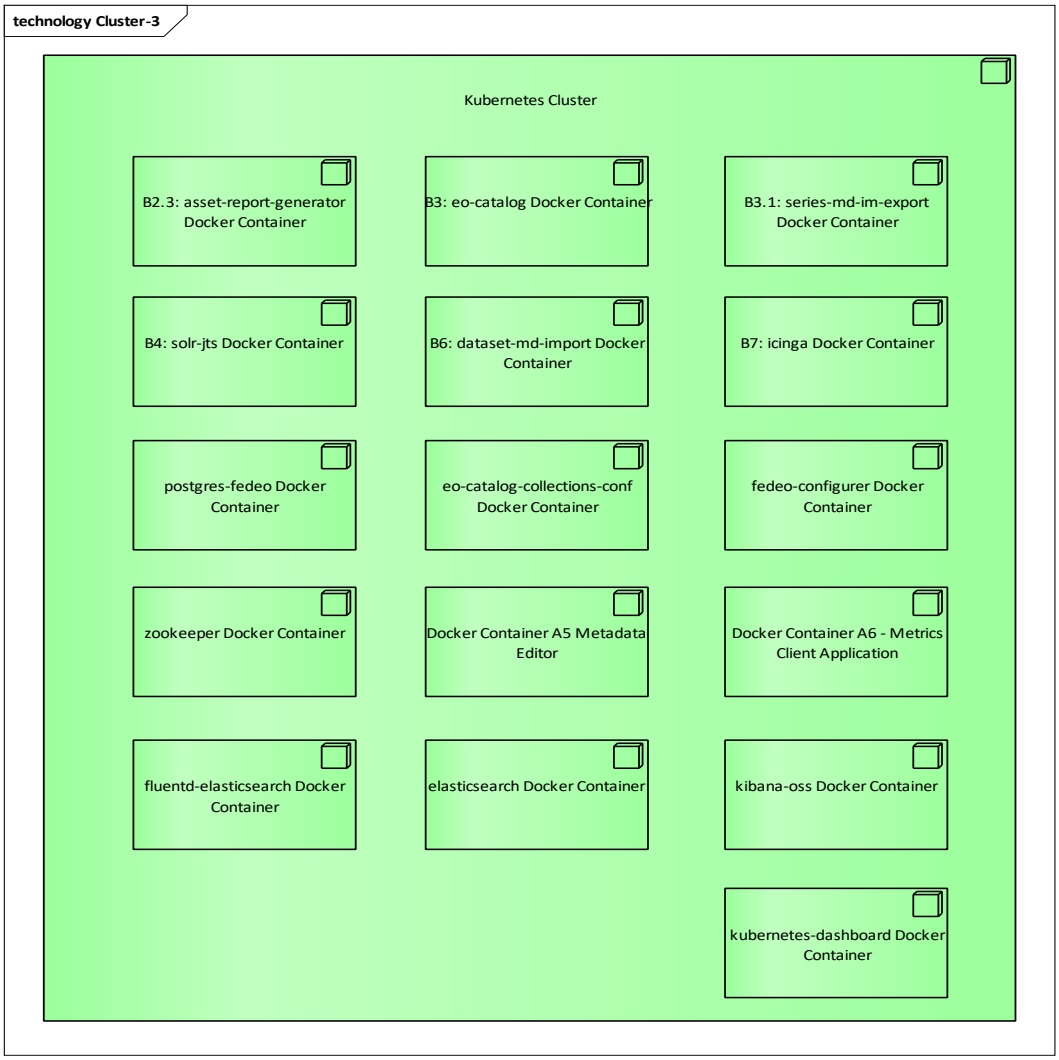


Figure 4: Kubernetes cluster hosting all ESE-ERGO components.

The diagram below lists the technical documents that were produced as part of the project for each of the above configuration items. The number between brackets refer to the corresponding CIDL [RD01] item number.

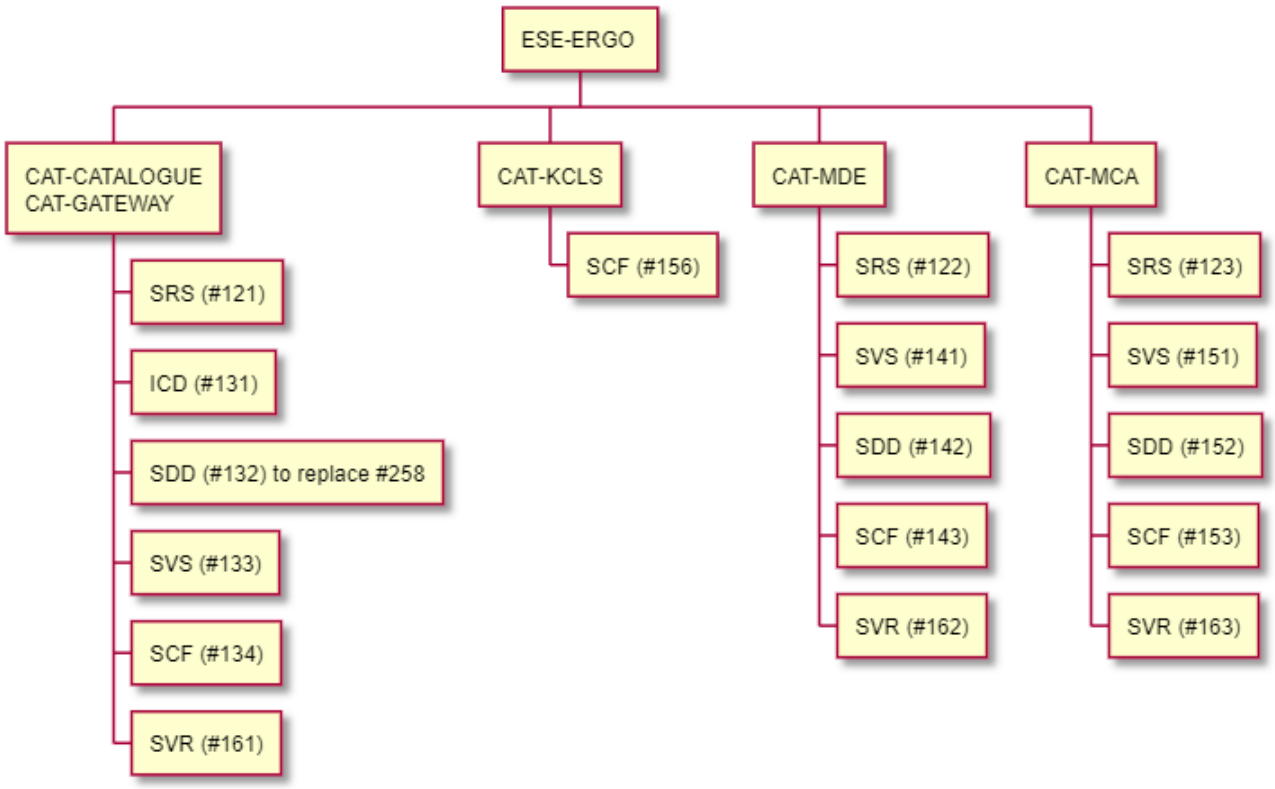


Figure 5: ECSS [AD03] deliverables per Configuration Item

7.2 Technology Assessment

As part of the project, a large number of technology candidates were assessed for adoption. The table below list the main technologies that were assessed and how they were applied in ESE-ERGO or proposed to be used in the future.

Technology	*Ref*	*Description*	*Application in ESE-ERGO*
Kubernetes		Container orchestration technology.	Used in ESE-ERGO to deploy and host all applications including Catalogue, Metadata editor and Metrics client application. Deployment from EOCAT/FedEO was extended to host additional components and to have NGINX capture access metrics stored automatically in the local Elasticsearch database.
OpenAPI 3.0	[CIDL-1101]	Standard language-agnostic interface description for RESTful API. Can be used by documentation generation tools and code generation tools to generate server or client components in various languages. Is the interface description language underlying the new generation of OGC API including OGC API – Features.	Adopted in ESE-ERGO for describing the API of the catalogue allowing for CRUD operations (create, read, update, delete). Available via the /api path (e.g. https://ergo.spacebel.be/api). The "read" part of the API consists of an OGC API – Features interface.
OGC 17-069r3: OGC API – Features – Part 1: Core		OGC API – Features. Geospatial API for features (a.k.a. WFS 3.0) fully based on OpenAPI implementing new paradigm of OGC APIs.	Used in ESE-ERGO to implement the (read) API to the catalogue which is subsequently advertised in an OSDD document for access via OpenSearch.
ISO 19168-1:2020 [RD11]	[RD11]	Geospatial API for features.	Identical to OGC 17-069r3 (previous row).
OGC API - Records[6]		Emerging OGC specification for catalogue access defined as extension of OGC API-Features.	Not used in ESE-ERGO as not available in a stable version during the project. Is a candidate interface to be supported in a future version of ESE-ERGO. Is expected to have an OpenSearch conformance class.
OGC 13-026r9	[CIDL-953]	Latest version of OpenSearch EO extension published on 25 November 2019.	Implemented in ESE-ERGO as it is the most recent version of the interface available. In particular the sorting functionality is available in ESE-ERGO as proposed in the version approved by the OAB[7].
OGC 14-055r2	[CIDL-951]	OWS Context GeoJSON encoding.	Used for:

[6] https://ogcapi.ogc.org/records/
[7] https://portal.ogc.org/index.php?m=projects&a=view&project_id=442&tab=2&artifact_id=83291

			- Adding OGC service binding information in GeoJSON based metadata. - Supported in Metadata editor to attach service bindings to collection or service metadata records.
OGC 17-047r1	[CIDL-950]	GeoJSON(-LD) encoding of OpenSearch responses.	Used for GeoJSON encoding of OGC API-Features / OpenSearch responses.
STAC	[CIDL-1051]	- GeoJSON encoding of collection metadata - GeoJSON encoding of granule metadata - STAC API intended to support searches (competing with OpenSearch) - GeoJSON encoding of (granule) search results. Competes with OGC 17-047r1 responses.	Proposed for a future iteration as parallel interface as no final specification is available yet. API is an extension of the OGC API – Features API selected for ESE-ERGO. Deployment of the open-source STACBROWSER was performed to host a permanent GUI for accessing a prototype STAC interface to ESE-ERGO metadata records.
GeoJSON	[CIDL-1003]		Used as default encoding for search results (OpenSearch, OGC API – Features, STAC) and metadata formats. Lightweight encoding (compared with XML) with performance benefits (smaller payloads).
OGC 17-003r2	[CIDL-948]	- EO granule metadata encoding in GeoJSON(-LD).	Selected in ESE-ERGO for: - granule metadata format in GeoJSON returned by catalogue searches (OGC API – Features/OpenSearch).
OGC 17-084r1	[CIDL-952]	- EO collection metadata encoding in GeoJSON(-LD) published by OGC in April 2021 as Best Practice.	Selected in ESE-ERGO for: - Collection metadata format in GeoJSON returned by catalogue. - Internal exchange format by Metadata Editor
OGC 19-020r1	[CIDL-955]	- Service/application metadata encoding in GeoJSON(-LD)	Selected in ESE-ERGO for: - Service metadata format in GeoJSON returned by catalogue.

			- Encoding of faceted search results in GeoJSON - Internal exchange format by Metadata Editor
ISO 19115-3:2016	[CIDL-907]	XML schema implementation for ISO 19115-1 metadata (series, datasets, services). Replaces at ISO earlier metadata encodings ISO19139(-2). INSPIRE Technical guidance is still based on ISO19139(-2).	Not selected initially in ESE-ERGO to remain backward compatible with EOCAT and aligned with INSPIRE Technical Guidance [CIDL-961]. Proposed to be supported as an additional (richer) encoding for collection and service metadata in future version of the ESE-ERGO catalogue and Metadata Editor.
Markdown		Markup format (text/markdown) used in Jupyter notebooks, by GitHub etc. and more compact that HTML.	Selected in ESE-ERGO for insertion of rich text in collection metadata records as proposed in chapter 10 of OGC 17-084r1[8].
VoID	[CIDL-1000]	Is an RDF schema from W3C for expressing metadata about RDF datasets.	Is used by a configuration file of the Metadata Editor to describe the access methods to the ESA Thesauri and GCMD thesauri used and retrieved periodically by the Metadata editor.
JSON-LD 1.1	[CIDL-1004]	JSON encoding for Linked Data. Is one of the possible serializations for RDF and the encoding recommended for including schema.org annotations in HTML web pages. Is the encoding underlying the GeoJSON encodings in OGC 17-003r2, OGC 17-084r1, OGC 19-020r1 and OGC 17-047r1.	Proposed for Iteration I3 (EOVOC project) to fully support Linked Data.
GeoDCAT-AP	[CIDL-1100]	Collection metadata vocabularies for use with RDF encodings including JSON-LD.	Proposed for Iteration I3 (EOVOC project) to fully support Linked Data.
DCAT	[CIDL-999]	Metadata vocabulary for use with RDF encodings including JSON-LD.	Proposed for Iteration I3 (EOVOC project) to fully support Linked Data.
Schema.org		Vocabulary for use with RDF encodings including JSON-LD. When embedded as structured data inside HTML web pages, information is used by search engine crawlers to update their knowledge graph supporting searches and "dataset search" capabilities.	Proposed for Iteration I3 (EOVOC project).

[8] https://docs.ogc.org/bp/17-084r1/17-084r1.html#toc40

7.3 Cascading Catalogue

7.3.1 API

The Cascading Catalogue is the largest component resulting from the ESE-ERGO project. It persists metadata for collections, products and services. It provides a RESTful interface implementing all resources defined in the OGC API Features core specification. These resources are depicted below.

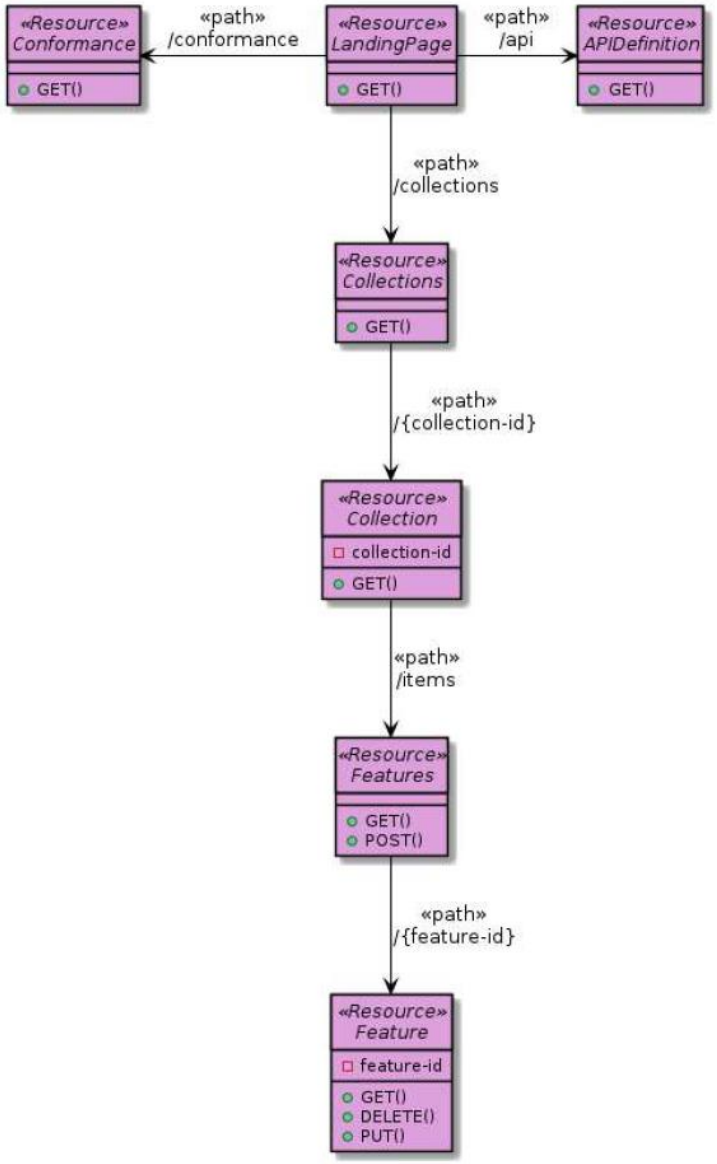


Figure 6: ESE-ERGO REST resources

The "Landing Page" provides access to the "readme" page below via the rel="service-doc" relation. The API supports three "feature collections" via the following paths:

- /collections/series/items (collections)
- /collections/datasets/items (products)
- /collections/services/items (services and applications)

For the complete overview of metadata encodings available for each of the feature collections, we refer to the OpenAPI definition or the ESE-ERGO ICD (ERGO-SPB-D2000.1).

The ICD describes the complete RESTful interface. An example of a resource description is given below. Within the project, we used a number of scripts that converted the OpenAPI definition of the interface or the OSDD document into UML models as depicted below.

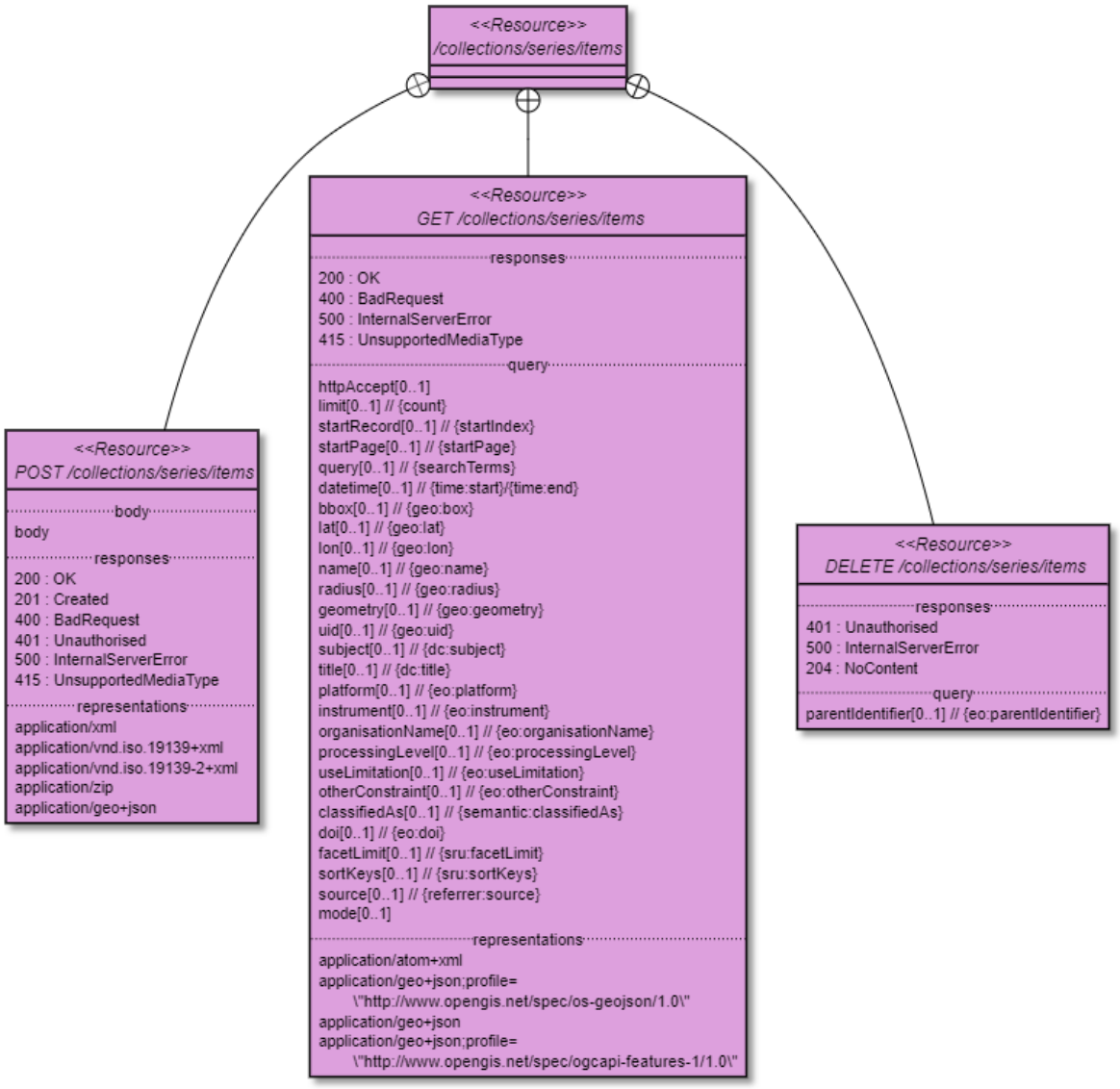


Figure 7: UML description "/collections/series/items"

An example of a generated UML object diagram representing the OSDD interface for collection search is included below.

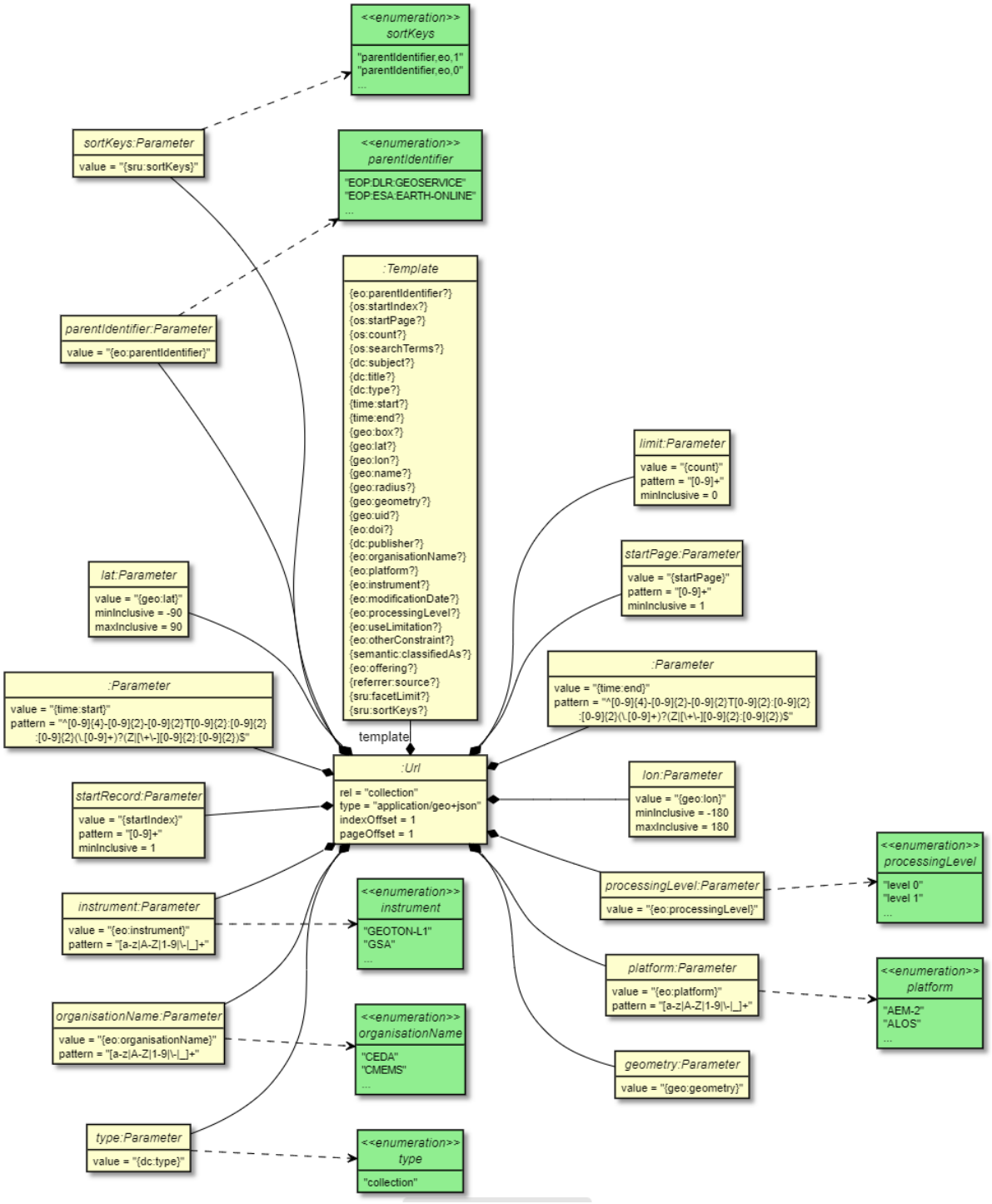
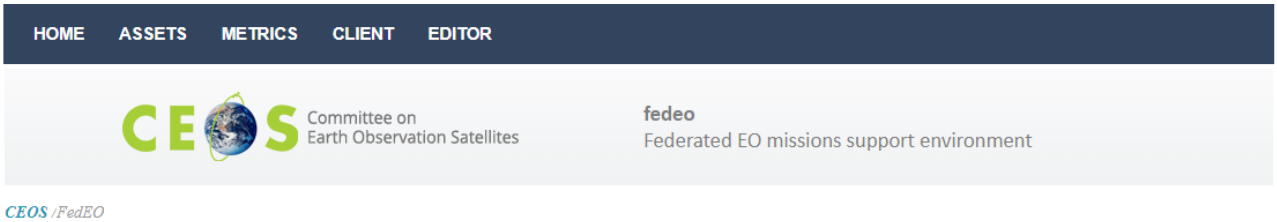


Figure 8: UML object diagram rel="collection" type="application/geo+json" URL template



CEOS /FedEO

FedEO

FedEO (Federated Earth Observation missions access) provides a unique entry point to a growing number of scientific catalogues and services for, but not limited to, EO European and Canadian missions. FedEO is deployed with ESA (European Space Agency) infrastructure as a gateway to:

- Provide brokered discovery, access and ordering capability to European/Canadian EO missions data based on HMA (Heterogeneous Missions Accessibility) interfaces;
- Implement the OpenSearch OGC (Open Geospatial Consortium) and other interfaces for an increased number of discoverable and accessible EO data collections;
- Implement the OpenSearch OGC interfaces for interfacing with CEOS Community Catalogues and Clients.

FedEO was initially developed as a prototype for GEO/GEOSS (Group on Earth Observations/Global Earth Observation System of Systems) and for EO-DAIL (Earth Observation Data Access & Integration Layer). It was consolidated after the CEOS Plenary in 2012 to facilitate access to European missions data in the international context and primarily in CEOS. See CEOS Documentation and CEOS Standardisation for more information.

The FedEO Clearinghouse provides access to the FedEO Clearinghouse Dataset and Dataset Series Catalogs via different interfaces:

- OpenSearch Description Document
- Explain Document
- OpenAPI Definition - Open with Swagger.io

The implementation of the above interfaces is based on:

- CEOS OpenSearch Best Practice v1.2
- WGISS CDA OpenSearch Client Guide v1.0
- WGISS FedEO Data Partner Guide-OpenSearch v1.0
- OGC 10-157r4, Earth Observation Metadata profile of Observations & Measurements, Version 1.1

7.3.2 OpenSearch

The above API is encapsulated in an OSDD document allowing OpenSearch client to consume the interface as well. A dedicated HTTP query parameter ("mode=owc") takes care of minor differences between the GeoJSON OpenSearch response encoding and the OGC API – Features response encoding.

The catalogue supported two-step search according to CEOS OpenSearch Best Practice version 1.2. The OSDD is considered a specific representation of the APIDefinition resource.

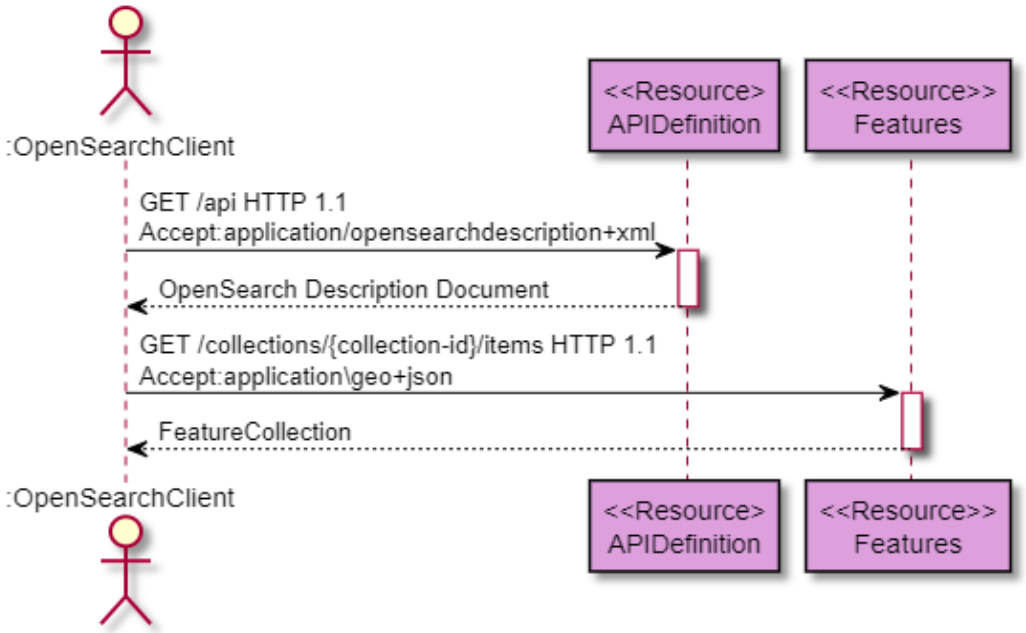


Figure 9: ESE-ERGO OpenSearch interaction

For backward compatibility, the existing Atom responses available in FedEO and EOCAT were preserved as far as possible. Specific differences are described in ERGO-SPB-D3000.3 and are mainly related to mandatory HTTP query parameter names imposed by OGC API – Features (e.g. limit and datetime).

Compared with the original OpenSearch catalogue, the new catalogue supports one additional feature type "service". OpenSearch clients use the URL template according to their preferred media type and the relation defining the feature type they expect to receive:

- rel="collection" (collections)
- rel="results" (granules)
- rel="service" (services and applications)

7.3.3 Performance improvements

The catalogue implementation combines the functionality of the original Catalogue and Gateway components from fedeo.esa.int and EOCAT. It avoids request and response conversion between these two components by providing a "cascading" interface, forwarding requests to external endpoints unless they can be processes locally.

To compare the ESE-ERGO implementation performance with the version currently in operations, we compared two installations at Spacebel hosted in an identical Kubernetes cluster with approximately the same number of metadata records (16.1 million). A stress test generated sequential random geographical searches on the largest collection for *n* threads in parallel.

Response time stress test	5 threads	10 threads	20 threads
ESE-ERGO Atom	425	875	1659
ESE-ERGO GeoJSON	400	833	1550
"old" catalogue Atom	521	1102	1986
"old" gateway Atom	650	1234	2312

Figure 10: Granule search response time comparison in milliseconds (stress test).

The response times measured for granule and collection search and how they compare with the same measurements on the software currently in operation are depicted on the three figures below. The first diagram indicated the performance gain for granule search, which is approximately 30 percent. The GeoJSON responses are slightly more efficient that the atom responses that are kept for backward compatibility.

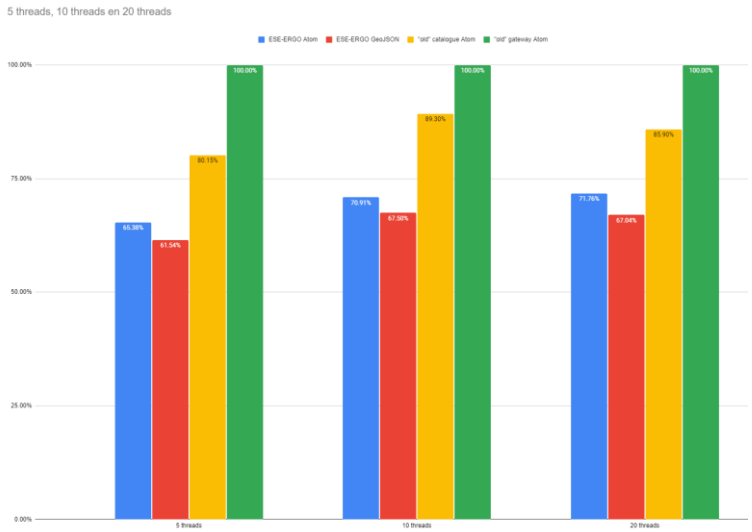


Figure 11: Granules search response time comparison in percentage (stress test).

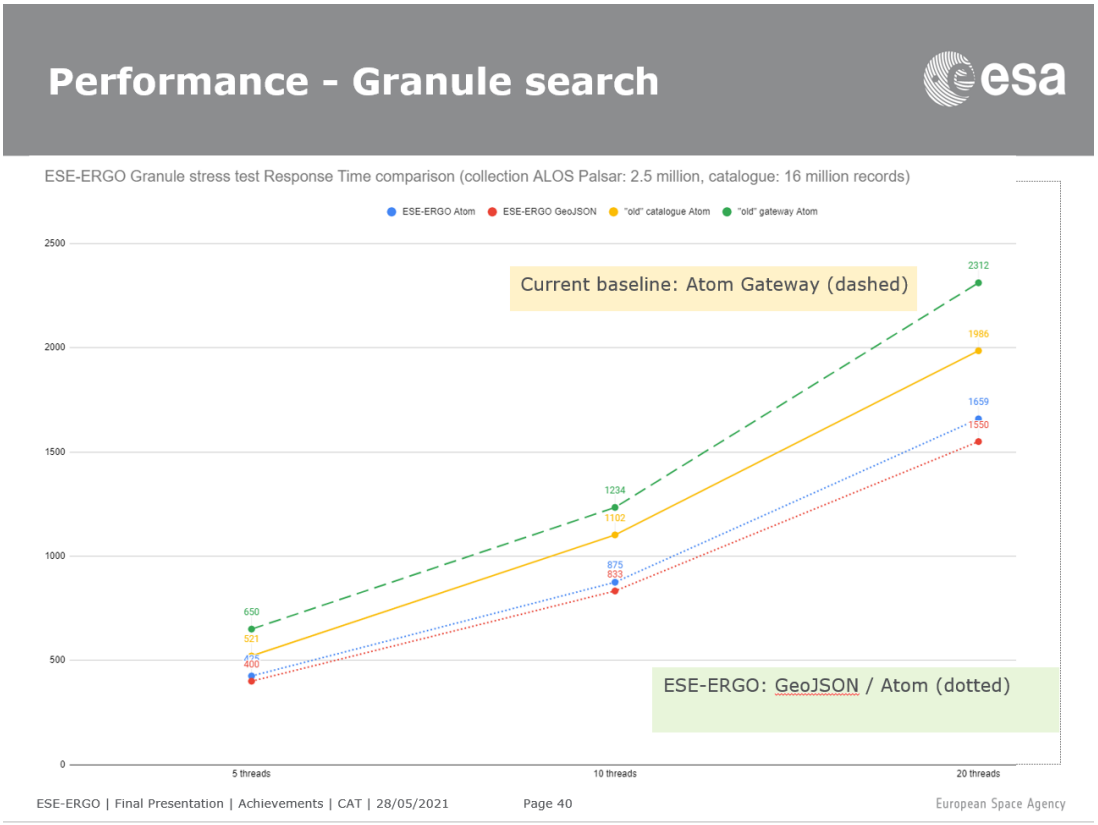


Figure 12: Granule search response time comparison in milliseconds (stress test).

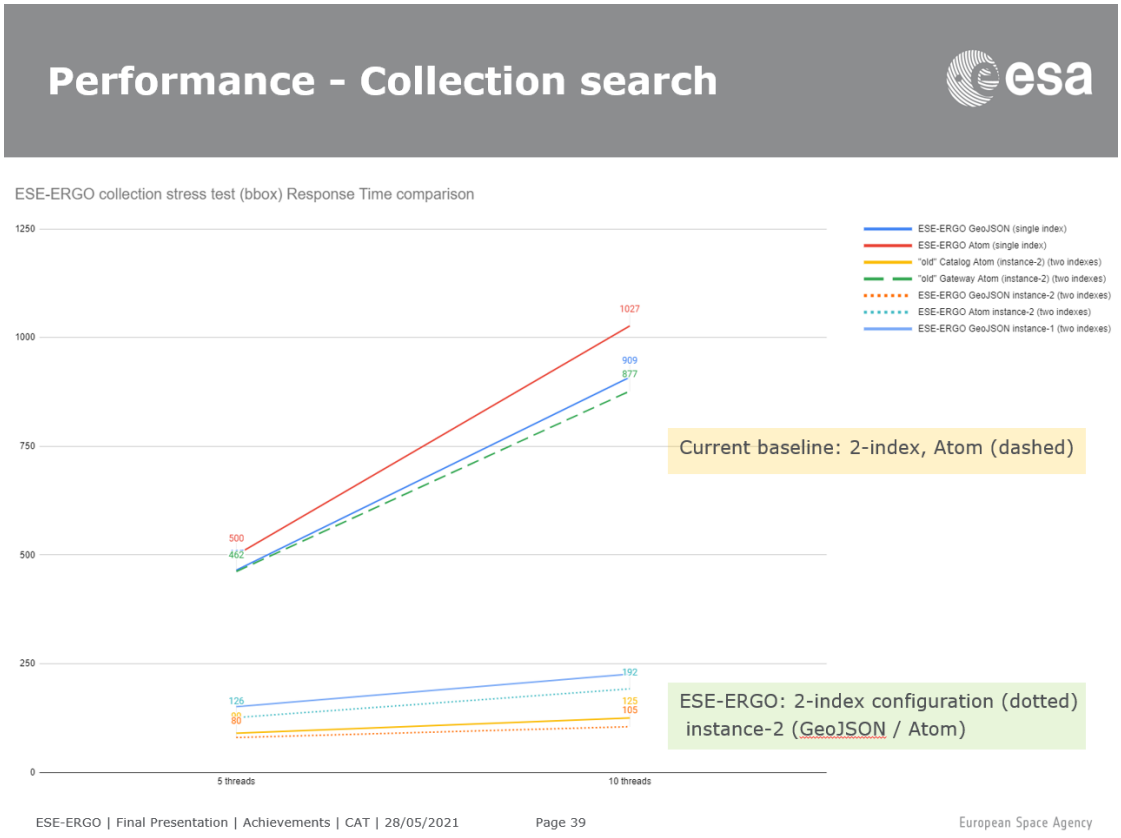


Figure 13: Collection search response time comparison in milliseconds (stress test).

7.4 Metadata Editor

7.4.1 Introduction

The Metadata Editor is published under open-source license. It provides a Web-based user interface allowing to create or update metadata records for collections and/or services/applications and interacting with one or more compatible catalogues (e.g. the ESE-ERGO Cascading Catalogue).

The Metadata Editor supports multiple metadata formats:

- Collection metadata: ISO19139-2 and OGC 17-084r1 (GeoJSON).
- Service metadata: ISO19139 and OGC 19-020r1 (GeoJSON).

It allows:

- Retrieving metadata from a catalogue supporting OGC 13-026r9 OpenSearch interfaces with Atom response.
- Storing metadata in a catalogue.
- Retrieve metadata from the local file system
- Save metadata to the local file system.
- Creating, updating, validating and visualising metadata in the user's private online workspace.

It supports multiple user roles:

- Anonymous (no persistent workspace, no catalogue write access)
- Registered user (persistent workspace, preferences, …)
- Administrator (add users, add catalogues, …)

7.4.2 External interfaces

The Metadata Editor is tightly integrated with the ESA SKOS Thesauri and the GCMD SKOS vocabularies. It supports users to find the appropriate platform, instrument and earth topics keywords from the ESA Thesauri and includes the corresponding GCMD concepts automatically. The tool automatically synchronises with the latest version of the thesauri and indicates discrepancies with the metadata records in the users' workspace.

The Metadata Editor interfaces with one or more catalogues to retrieve metadata records or store metadata records.

Finally, it interfaces with a NASA Web-service endpoint to validate the DIF-10 metadata encoding.

7.4.3 User Interface

The following screen capture give some examples of functionality available to the users through the Web-based user interface. For more details, we refer to the related ESE-ERGO deliverables.

The functionality of the tool allows the current manual capture of EO collection metadata through Excel files and subsequent conversion from Excel to ISO19139-2 metadata to be completely replaced by this interactive tool which in addition supports the update of the metadata in the catalogue and maintenance of the metadata in case of ESA thesauri and/or GCMD vocabulary changes. For example, the tool allows loading a .zip file with approximately 150 collection metadata records within 20 seconds in the user's workspace, including the time for validating against the applicable XML and JSON schemas.

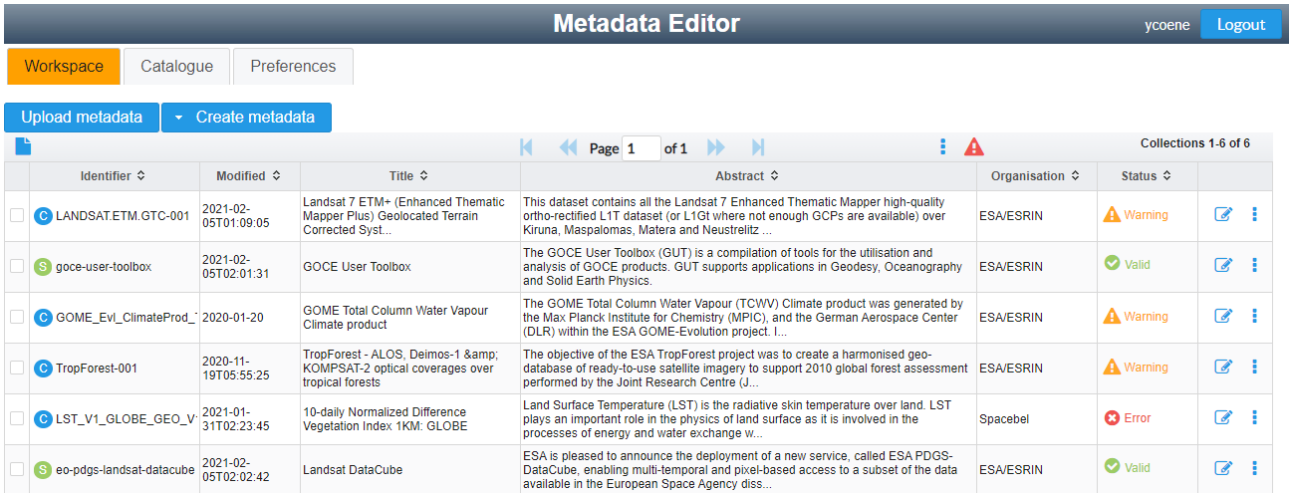


Figure 14: Metadata Editor user work space.

OGC service endpoints available for a particular collection or service/application can be associated to the collection or service metadata via offerings. They are subsequently encoded in the corresponding GeoJSON and ISO metadata formats which can be previewed within the tool's user interface.

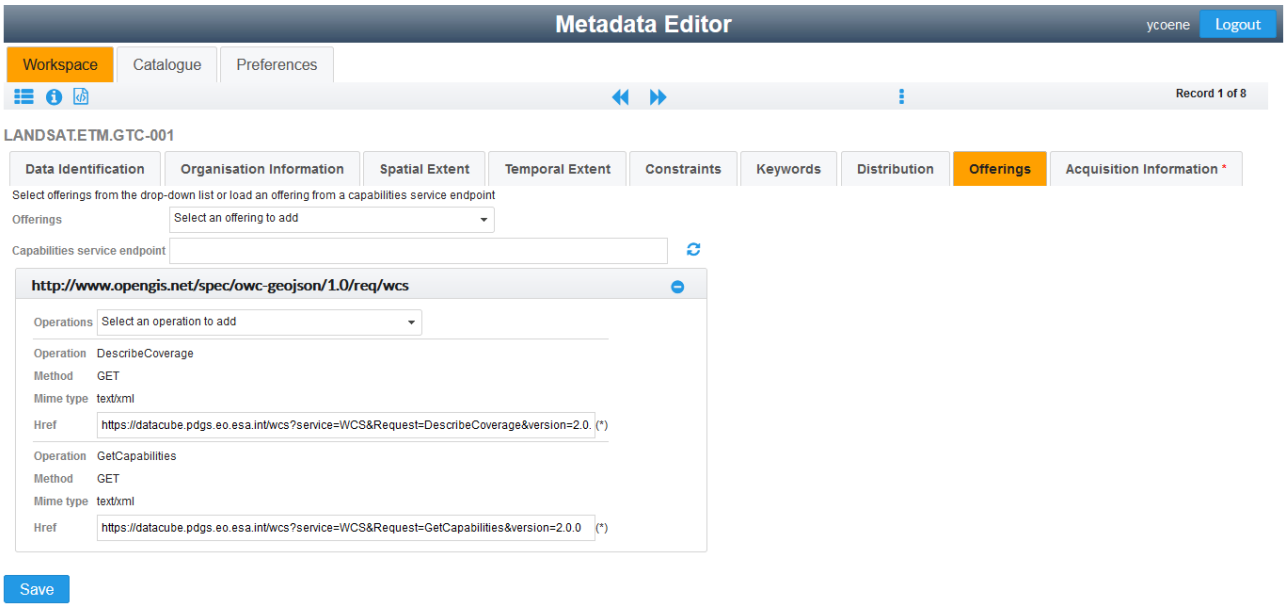


Figure 15: Metadata Editor support for offerings.

Users with administrator roles can enable/disable support for metadata formats, register multiple catalogue endpoints for read and/or write access and enable/disable automatic updates of ESA and GCMD thesauri. They also create user accounts and manage visibility of catalogue endpoints with registered users.

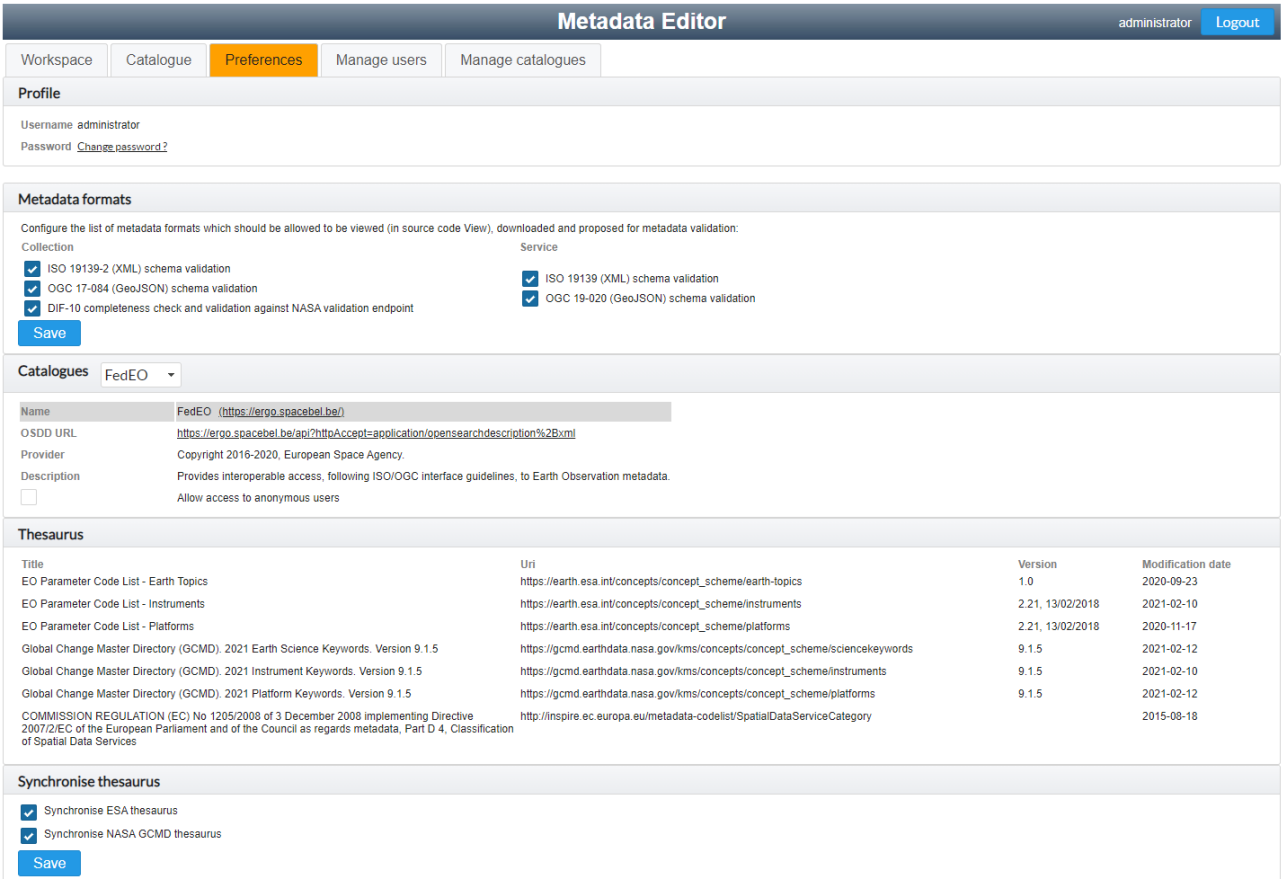


Figure 16: Metadata Editor | Administrator preferences.

7.5 Metrics Client Application

7.5.1 Introduction

The Metrics Client application is published under open-source license. It provides a Web-based user interface allowing to browse catalogue access metrics. The look-and-feel of the application was made consistent with the metrics Web interface available for CWIC. The metrics visualised comprise the following groups defined by the CEOS WGISS System Level Team:

- Global Core Metrics
- Data Provider Core Metrics
- Collection Core Metrics

7.5.2 External interfaces

The Metrics Client Application retrieves raw catalogue requests from the Elasticsearch engine. In the ESE-ERGO deployment, this search engine is part of the Kubernetes Cluster installation. The Elasticsearch database is populated with raw access information extracted from the NGINX load-balancer access.log file by a data collector (Fluentd). Information is captured about incoming catalogue requests (either requests for an OSDD or actual granule search requests).

In addition, it interfaces with an OpenSearch catalogue to obtain the list of collections available in the system and their mapping on "organisations".

Finally, it published daily summaries of metrics as CSV files on an FTP server for future consumption by a corporate dashboard (e.g. ESA TellUs).

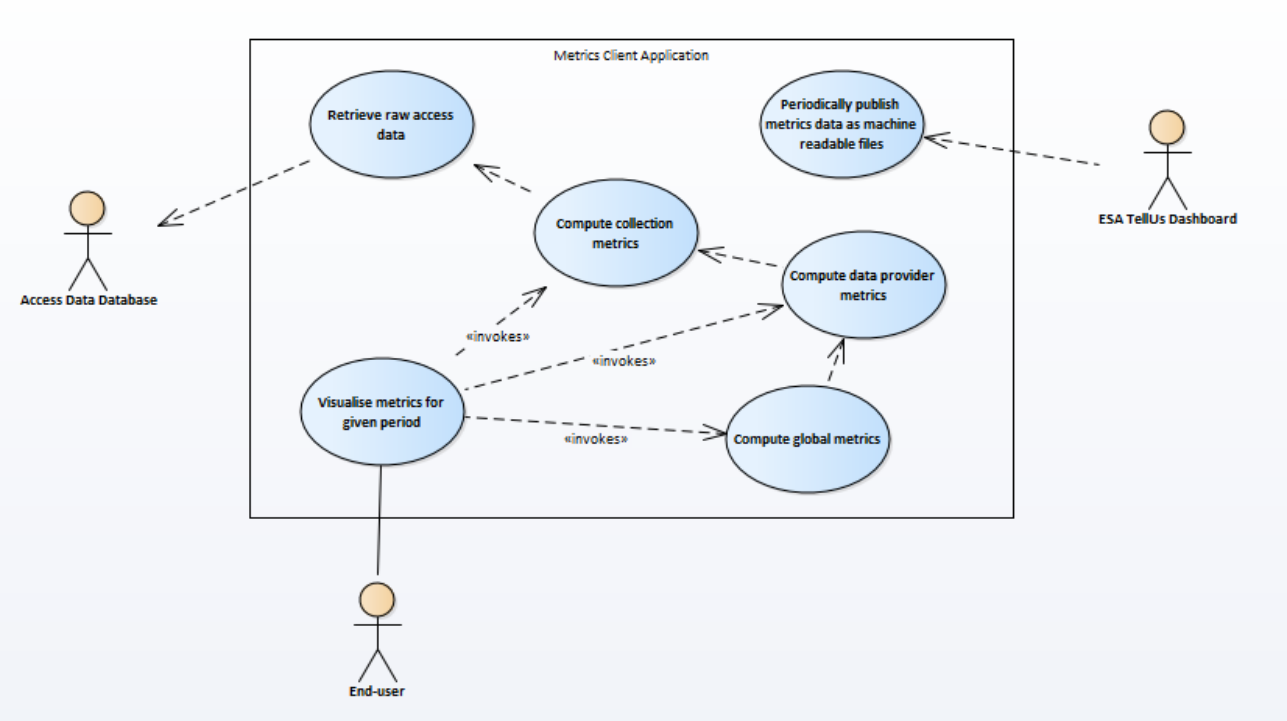


Figure 17: Metrics Client Application interfaces

7.5.3 User Interface

The following screen captures give some examples of functionality available to the users through the Web-based user interface. For more details, we refer to the related ESE-ERGO deliverables.

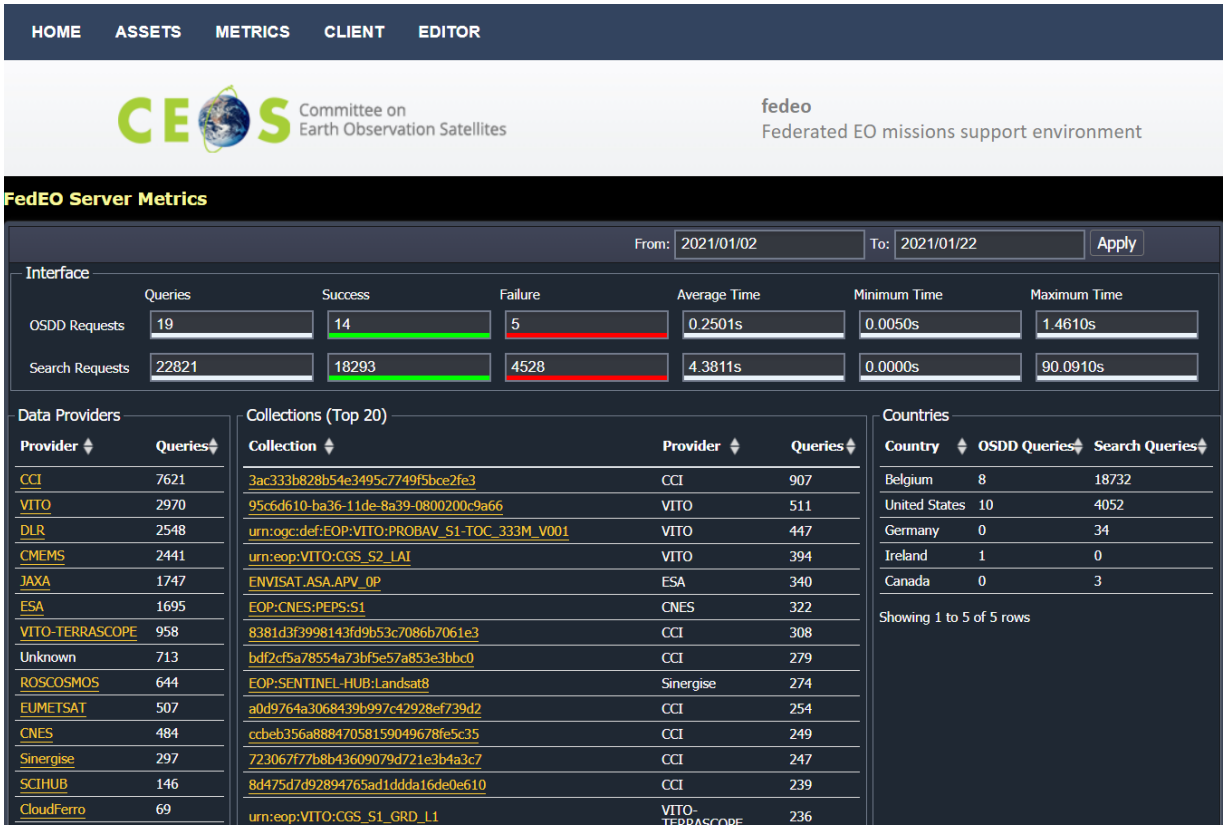


Figure 18: Metrics Client Application – Global metrics

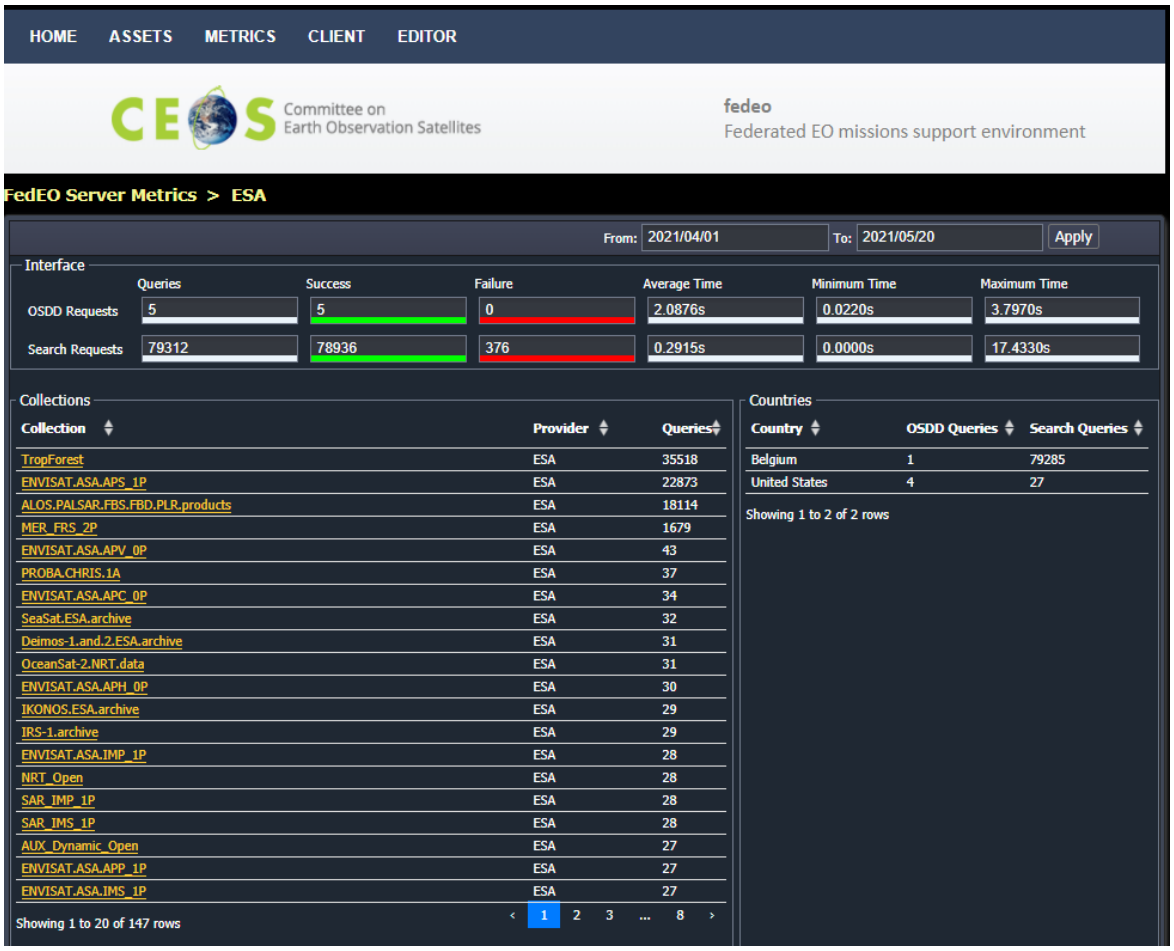


Figure 19: Metrics Client Application – Provider metrics

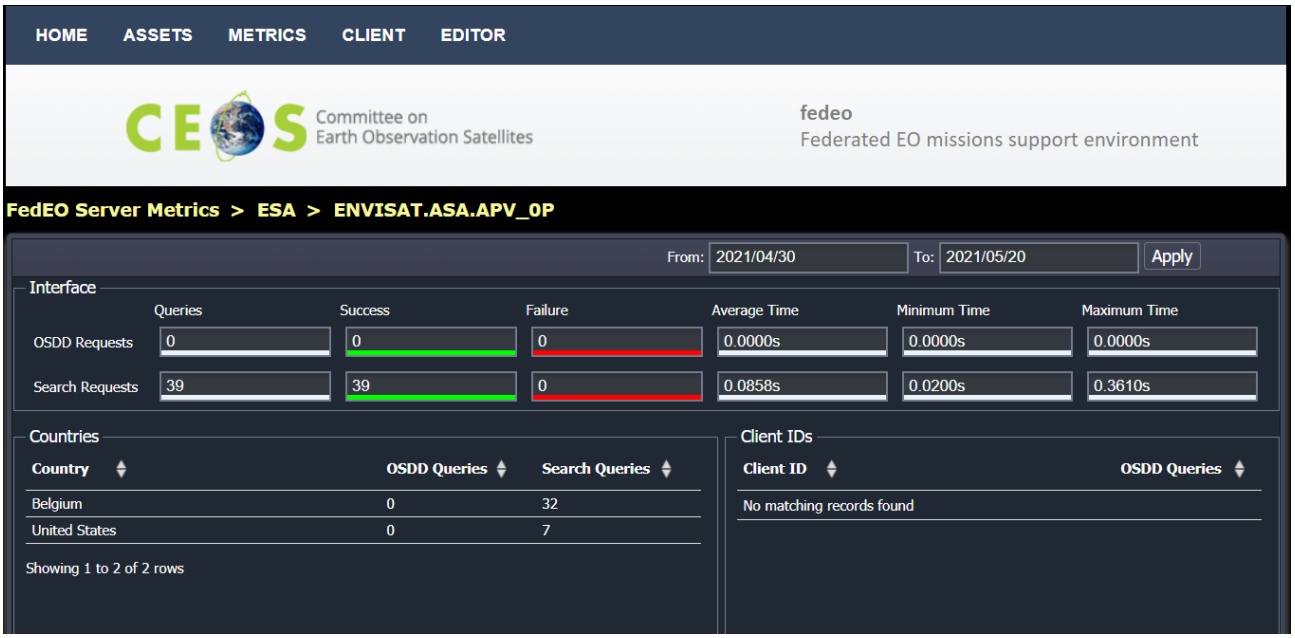


Figure 20: Metrics Client Application – Collection metrics

8 CONCLUSION

8.1 Lessons learned and Limitations

8.1.1 Cascading Catalogue

1. The GeoJSON response representations of OGC API – Features and OpenSearch (OGC 17-047r1) (FeatureCollection and Feature) differ in a small number of detailed points, including use of id/identifier and the link model which was already identified in OGC 19-020r1[9] but not yet resolved. Our implementation uses a query parameter (mode=owc) to select one or the other response format. Updating the OGC 17-047r1, OGC 17-003r2 and OGC 17-084r1 to adopt the same conventions as OGC API – Features (now also an ISO standard [RD11]) would remove this incompatibility.
2. Performance measurements at the end of the project demonstrated that using a single SolR index for all feature collections has a negative impact on performance. The current solution is to preserve two separate indexes. This observation showed that current performance for granule search could be further enhanced by using a separate index for each EO collection, as presented in the "Further work" section.
3. An OpenAPI description cannot accurately represent that a GET method on a resource returns a GeoJSON representation which differs depending on a query parameter in the request (e.g. mode=owc).
4. A number of OGC specifications on which the ESE-ERGO work is based were not yet published by OGC at the start of the project and we used the versions approved by the OAB until publication. We have observed recently that OGC has by mistake published an incomplete version of the OGC 13-026r9 specification, in which the definition of the sorting functionality implemented in ESE-ERGO is completely missing. OGC was first contacted in November 2020 about this issue (by Uwe Voges), but the mistake is not yet corrected.
5. The documentation set for the Operational Catalogue and Gateway and the newer ESE-ERGO catalogue documentation (in particular SRS, SDD, SVS) will need further work to have a unique set of documents replacing the current set.

8.1.2 Metadata Editor

1. The transactional interface used to insert metadata in the catalogue is an obvious use of the HTTP POST method using the OGC API – features specification. The behaviour of the POST method itself and/or the specification of a transactional interface is expected to be defined in a future part of the OGC API – Features set of documents and/or a STAC interface extension and may require a future software update.
2. Supporting an internal metadata model and allowing the Metadata Editor to "transport" it to the catalogue without affecting the normal interfaces of the Catalogue available to other clients was a challenge. It was finally decided to use GeoJSON with additional properties for populating the catalogue and keeping the read interface unchanged. A first implementation using gmd:supplementalInformation was not possible for service metadata and was abandoned.
3. The keyword insertion functionality for service metadata (in ISO) did not allow to use the same syntax as used for collection metadata (ISO19139-2) due to a problem in the official ISO19119 schemas also reported in the INSPIRE Technical Guidance [CIDL-961] not allowing the use of gmx:Anchor. The "patch" proposed by JRC/INSPIRE[10] to allow this nevertheless was identified

[9] https://docs.ogc.org/per/19-020r1.html#_alignment_with_the_ogc_api_features_standard
[10] https://inspire.ec.europa.eu/draft-schemas/inspire-md-schemas/srv/1.0/

too late in the project to be applied but would be an obvious enhancement to the current implementation.

4. Supporting metadata properties beyond what is currently available in the supported metadata representations requires adding support for additional representations in the metadata editor (internal model), for instance ISO19115-3 is an obvious candidate. The internal (export) model could allow passing the supported representations as a MIME bundle to the catalogue inside the GeoJSON representation currently used.

8.1.3 Metrics Client Application

1. The Metrics Client Application uses raw access information captured in an Elasticsearch database by other components. To allow preserving the raw access information when (Cluster) software is reinstalled and/or updated, a backup/restore functionality may be required to ensure that "old" access information is properly preserved. This was not yet implemented.
2. The Metrics client Application does currently not allow to easily drill down into search requests or OSDD requests while filtering to only visualise requests that returned errors to facilitate investigating which provider and/or collection is producing erroneous responses. This functionality would be helpful to track down issues with individual collections. Ranking providers and/or collections based on "number of errors" instead of total "number of requests" would be useful. Allowing to drill down (for errors) and see the original request generating the error, for instance for users with administration rights may be another interesting tool to monitor the behaviour of the catalogue.

8.2 Future Work

The Final Presentation (ERGO-SPB-MOM-008) contains as Annex A.5 a comprehensive PowerPoint presentation of future work including short term activities that would be required in case selected components would be transferred into operations.

In addition, the deliverable ERGO-SPB-D3000.3 "Transfer to Operations Recommendations" documents the same recommendations in more detail.

We refer to the deliverables ERGO-SPB-MOM-008 and ERGO-SPB-D3000.3 for more information about proposed Future Work.