
CEOS Working Group
on Information Systems and Services

**WGISS Connected Data Assets
FedEO Data Partner Guide (OpenSearch)**

Publication Date: 2021-09-14

Document version: V1.1

Category: WGISS Technical Document

Editors: Y. Coene, A. Della Vecchia

Document Status Sheet

Issue	Date	Comments	Editor
1.0	2019-04-18	Original	Y. Coene, A. Della Vecchia
1.1	2021-09-14	<p>Updated location of server endpoint to https://fedeo.ceos.org/.</p> <p>Location of ESA thesauri in validation environment changed from fedeo.spacebel.be to thesauri.spacebel.be and references updated to refer to SKOS scheme instead of concept.</p> <p>URL of CEOS Web pages changed to https://.</p> <p>GCMD URL updated to use https://gcmd.earthdata.nasa.gov. Location of GCMD .CSV files updated.</p> <p>Contact information updated.</p> <p>[AD4] updated.</p>	Y. Coene

Executive Summary

Recommendations for FedEO Data Partners

While FedEO can serve as an OpenSearch proxy to search almost any Internet-accessible inventory system for Data Partners, there are a small number of recommendations for Data Partners that will make the job vastly easier.

1. Register each distinct, searchable collection in the IDN. FedEO can also do this on your behalf if you provide the collection metadata in a supported format.
2. Provide a search interface accessible via a simple URL (*i.e.* HTTP/HTTPS GET), ideally including parameters for starting record number and number of records desired in the response. OGC CSW or OpenSearch (CEOS compliant) are preferred but not mandatory.
3. Support searching on spatial bounding box
4. Support searching on temporal extent, at least observation start and end dates
5. Identify and, ideally, filter for limitations on search extent (spatial and temporal) to prevent search timeouts
6. Provide search responses in well-structured text (XML, JSON, *etc.*) returning matching data granules
7. Identify each returned data granule by an identifier that is unique within the inventory system
8. Provide a capability for using the granule identifier to retrieve metadata about the granule
9. Return URLs for browse data and direct access to granule-level data (or to a data ordering system) in the search response

In general, these are common and widely implemented capabilities in almost any granule search system and should not represent an impediment to joining FedEO as a Data Partner. If any of these capabilities are not implemented, it is still possible to become a FedEO Data Partner – contact any of the FedEO team for details.

Table of Contents

1 Introduction.....1

2 Scope.....1

3 References.....2

 3.1 Applicable Documents.....2

 3.2 Reference Documents.....3

4 Before You Begin.....4

 4.1 FedEO Connected Data Asset Background4

 4.2 Search Concept and Design.....5

 4.3 WGISS Architecture for Connected Data Assets.....6

 4.4 Terms and Definitions.....6

 4.5 IDN Systems.....8

 4.6 FedEO Systems.....8

 4.7 Contact Information.....9

5 FedEO Data Partner Use Cases.....10

 5.1 Granule Metadata Publication10

 5.1.1 UC 1.1 - Online Granule Metadata.....10

 5.1.2 UC 1.2 - Offline Granule Metadata12

 5.1.3 Supported Granule Metadata Formats13

 5.1.4 Supported Catalog Protocols15

 5.2 Collection Metadata Publication.....23

 5.2.1 UC 2.1 - Online Collection Metadata23

 5.2.2 UC 2.2 - Offline Collection Metadata.....28

 5.2.3 Supported Collection Metadata Formats29

6 Common Partner Guidelines.....31

 6.1 Metadata & Semantic Mapping.....31

 6.1.1 HTTP/HTTPS access.....31

 6.1.2 Spatial search.....31

 6.1.3 Temporal search.....31

 6.1.4 Request for collection by ID.....32

 6.1.5 Request for granule by ID.....32

 6.1.6 Unique granule IDs.....32

 6.1.7 Record counts32

 6.1.8 Pagination32

 6.1.9 Other search parameters.....33

 6.2 Interaction & Services Model.....33

 6.2.1 Unique granule ID33

 6.2.2 Contact information33

 6.2.3 Browse URL33

 6.2.4 Order URL.....33

 6.2.5 Pagination support34

 6.2.6 Spatial search.....34

 6.2.7 Temporal search.....34

 6.3 Error Handling.....34

7	Sample Metadata Ingestion Report.....	1
7.1	Metadata Ingestion Report.....	1
7.1.1	Record Summary	1
7.1.2	Organisation Summary	2
7.1.3	Topic Category Summary	2
7.1.4	Keyword Details	3
7.1.5	Record Details	3
7.2	DIF-10 Metadata Validation Report.....	5

1 Introduction

This document contains the comprehensive data partner's guide for OpenSearch with FedEO, as adopted in the CEOS WGISS Connected Data Assets project. The document provides technical information about the interfaces required by the FedEO Granule Gateway to connect to external data provider systems and is intended to help data providers to make available their metadata and data through the FedEO Gateway.

The document covers the following use cases involving granules and their metadata which require interaction with FedEO:

- **UC1.1:** Data Provider wishes to provide an *online* access to his granules and their metadata through the FedEO Granule Gateway.
- **UC1.2:** Data Provider wishes to provide his granule metadata *off-line* to the FedEO team and use the FedEO Granule Gateway to provide an online access to his granules and their metadata.

In addition, the document covers two additional use cases which allow data providers to simplify the registration of their collection metadata in IDN by using built-in FedEO functionality for DIF-10 [AD.2] collection metadata generation.

- **UC2.1:** Data Provider wishes to provide an *online* access to his collection metadata via FedEO and use the FedEO DIF-10 [AD.2] export capability to get his collections registered in IDN (International Directory Network) by the FedEO team.
- **UC2.2:** Data Provider wishes to provide his collection metadata *off-line* to the FedEO team and use the FedEO DIF-10 export capability to get his collections registered in IDN (International Directory Network) by the FedEO team.

2 Scope

The Federated Earth Observation Missions (FedEO) initiative was initially an OGC pilot activity aimed to refine OGC specifications relevant to Earth Observation.

The same OGC specifications have subsequently been implemented in support of the Copernicus programme, in particular to provide interoperable access to catalogues from European Earth Observation Missions and support the subsequent ordering of data in support of the European Commissions' Copernicus Service Projects.

The resulting interoperable catalogue is presented here as the FedEO Clearinghouse, providing discovery, ordering and on-line data access services for space-based Earth Observation missions from over 25 years, as a potential contribution to GEO GEOSS. The FedEO

Clearinghouse brings together the catalogues from ESA, EUMETSAT, DLR, JAXA, CNES, ROSCOSMOS and VITO.

This data partner guide applies to any data provider who plans to join FedEO as one of the CEOS WGISS Connected Data Assets.

3 References

3.1 Applicable Documents

The following documents provide more background and supportive information.

Ref.	Title	Version	Date
[AD.1]	CEOS OpenSearch Best Practices OpenSearch, http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/OpenSearch/CEOS-OPENSEARCH-BP-V1.2.pdf	1.2	
[AD.2]	Directory Interchange Format (DIF) Standard (DIF-10), http://gcmd.gsfc.nasa.gov/DocumentBuilder/defaultDif10/guide/index.html	10.2	
[AD.3]	FedEO Client Partner Guide, PDGS-FEDEO-SPB-ICD-15-1174	1.1	
[AD.4]	https://github.com/dewitt/opensearch/blob/master/opensearch-1-1-draft-6.md	1.1	
[AD.5]	OGC 10-032r8, OGC OpenSearch GeoSpatial and Temporal Extensions, https://portal.opengeospatial.org/files/?artifact_id=56866 .	r8	14/04/2014
[AD.6]	OGC 13-026r8, OGC OpenSearch Extension for Earth Observation Products,	r8	
[AD.7]	RFC 6573, The Item and Collection Link Relations, http://www.ietf.org/rfc/rfc6573.txt .		April 2012
[AD.8]	RFC 4287, The Atom Syndication Format, http://www.ietf.org/rfc/rfc4287.txt .		
[AD.9]	OGC 10-157r3, Earth Observation Metadata profile of Observations & Measurements, https://portal.opengeospatial.org/files/?artifact_id=47040 .	1.0.0	
[AD.10]	OGC 10-157r4, Earth Observation Metadata profile of Observations & Measurements, https://portal.opengeospatial.org/files/61098 .	1.1	10/06/2014
[AD.11]	ISO 19139, Geographic Information – Metadata XML (ISO 19139:2007), http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32557 .		2007
[AD.12]	ISO 19139-2:2012, Geographic information -- Metadata -- XML schema implementation -- Part 2: Extensions for imagery and gridded data, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=57104		2012
[AD.13]	http://www.opensearch.org/Community/Proposal/Specifications/OpenSearch/Extensions/SRU/1.0/Draft_1	1.0 DRAFT 1	
[AD.14]	http://www.opensearch.org/Community/Proposal/Specifications/OpenSearch/Extensions/Semantic/1.0/Draft_1	1.0 DRAFT 1	
[AD.15]	searchRetrieve: Part 7. Explain Version 1.0, OASIS Standard, http://docs.oasis-		30 January 2013

Ref.	Title	Version	Date
	open.org/search-ws/searchRetrieve/v1.0/searchRetrieve-v1.0-part7-explain.html		
[AD.16]	searchRetrieve: Part 4. APD Binding for OpenSearch Version 1.0, OASIS Standard, http://docs.oasis-open.org/search-ws/searchRetrieve/v1.0/os/part4-opensearch/searchRetrieve-v1.0-os-part4-opensearch.html		30 January 2013
[AD.17]	OGC 06-121r9, OGC Web Services Common Standard, http://portal.opengeospatial.org/files/?artifact_id=38867 .	2.0	
[AD.18]	http://www.opensearch.org/Specifications/OpenSearch/Extensions/Parameter/1.0/Draft_2	1.0 DRAFT 2	
[AD.19]	Media RSS Specification, http://www.rssboard.org/media-rss .		

3.2 Reference Documents

Ref.	Title	Version	Date
[RD.1]	"Heterogeneous Missions Accessibility (HMA), Design methodology, Architecture and Use of geospatial Standards for the Ground Segment Support of Earth Observation Missions", European Space Agency, ESA TM-21, ISBN 978-92-9221-883-6, http://esamultimedia.esa.int/multimedia/publications/TM-21/TM-21.pdf		April 2012
[RD.2]	OGC 06-080r4, GML Application Schema for EO Products, Version 1.0.0, 25/02/2010, http://portal.opengeospatial.org/files/?artifact_id=31065		
[RD.3]	OGC 06-131r5, Catalogue Services Specification 2.0 Extension Package for ebRIM (ISO/TS 15000-3) Application Profile.	0.2.4	
[RD.4]	INSPIRE Technical Guidance for the Implementation of the INSPIRE Download Services, http://inspire.jrc.ec.europa.eu/documents/Network_Services/Technical_Guidance_Download_Services_v3.1.pdf .	3.1	09/08/2013
[RD.5]	OGC 06-141r2, Ordering Services for Earth Observation Products.	0.9.4	
[RD.6]	OGC 07-141r2, OGC KML		2008-04-14
[RD.7]	DCMI Metadata Terms, http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms		
[RD.8]	OGC 07-118r8, User Management for Earth Observation Services.	Version 1.0	08/09/2010
[RD.9]	OGC 07-038r3, OGC Cataloguing of ISO Metadata (CIM) – Using the ebRIM profile of CS-W.	Version 0.1.12	14/12/2009
[RD.10]	OGC 13-084r2, OGC I15 (ISO19115 Metadata) Extension Package of CS-W ebRIM Profile, https://portal.opengeospatial.org/files/?artifact_id=56905	Version 1.0 r2	28/04/2014
[RD.11]	OGC 11-035r1, EO Collection and Service Discovery using the ebRIM Application Profile of CSW 2.0, https://portal.opengeospatial.org/files/?artifact_id=53276 .	Version 1.0	26/03/2013

Ref.	Title	Version	Date
[RD.12]	http://www.loc.gov/standards/sru/recordSchemas/		
[RD.13]	OGC 06-131r5, OGC Catalogue Services Specification 2.0, Extension Package for ebRIM Application Profile: Earth Observation Products.	Version 0.2.4	
[RD.14]	CWIC Client Partner Guide, CWIC-DOC-12-006r1, CWIC Technical Document, http://www.ceos.org/images/WGISS/CWIC/CWIC-DOC-12-006r1_CWIC_Client_Partner_Guide.doc	V1.0	03/12/2012
[RD.15]	searchRetrieve: Part 0. Overview Version 1.0, OASIS Standard, http://docs.oasis-open.org/search-ws/searchRetrieve/v1.0/os/part0-overview/searchRetrieve-v1.0-os-part0-overview.html		30 January 2013
[RD.16]	searchRetrieve: Part 3. searchRetrieve Operation: APD Binding for SRU 2.0, Version 1.0, OASIS Standard, http://docs.oasis-open.org/search-ws/searchRetrieve/v1.0/cs01/part3-sru2.0/searchRetrieve-v1.0-cs01-part3-sru2.0.html		30 January 2013
[RD.17]	CWIC Client Partner Guide	V1.0	2012
[RD.18]	OGC 08-167r2, Semantic annotations in OGC standards, https://portal.opengeospatial.org/files/?artifact_id=47857 .	2.0	2012-10-10
[RD.19]	Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007	2.0.1	02/03/2017
[RD.20]	Google Data APIs, Protocol Reference, https://developers.google.com/gdata/docs/2.0/reference#query-requests		
[RD.21]	OGC 07-045, OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile, Version 1.0, http://portal.opengeospatial.org/files/?artifact_id=21460	Version 1.0	19/07/2007
[RD.22]	Metadata Vocabulary for Tabular Data, W3C Working Draft 08 January 2015, http://www.w3.org/TR/tabular-metadata/		
[RD.23]	Model for Tabular Data and Metadata on the Web, W3C Working Draft 08 January 2015, http://www.w3.org/TR/tabular-data-model/		

4 Before You Begin

This chapter introduces the background, concepts and architecture of IDN, CWIC, and FedEO, within the scope of the WGISS Connected Data Assets. The related skills you will need as a data partner are also discussed.

4.1 FedEO Connected Data Asset Background

For scientists who conduct multi-disciplinary research, there may be a need to search multiple catalogs in order to find the data they need. Such work can be very time-consuming and tedious, especially when different catalogs may use different metadata models and catalog interface protocols. It would be desirable, therefore, for those catalogs to be integrated into a catalog federation which will present a well-known and documented metadata model and interface protocol to users and hide the complexity and diversity of the affiliated catalogs behind the

interface. With such a federation, users only need to work with the federated catalog through the public interface or API to find the data they need instead of working with various catalogs individually.

The Committee on Earth Observation Satellite (CEOS) addresses coordination of the satellite Earth Observation (EO) programs of the world's government agencies, along with agencies that receive and process data acquired remotely from space. The Working Group on Information Systems and Services (WGISS) is a subgroup of CEOS, which aims to promote collaboration in the development of systems and services that manage and supply EO data to users worldwide.

NASA's contributions to the CEOS International Directory Network (IDN) provides access to more than 34,000 Earth science data set and service descriptions (stored in the NASA Common Metadata Repository [CMR]) which cover subject areas within Earth and environmental sciences. The IDN's mission is to assist researchers, policy makers, and the public in the discovery of and access to data and related services relevant to Earth science research.

To aid in the search and discovery effort, Global Change Master Directory (GCMD) controlled keywords have been developed and are regularly being refined and expanded. These keywords are also used in other applications within the broader scientific community. Users may perform searches through the IDN website and OpenSearch API using the controlled keywords, free-text searches, map/date searches, or any combination of the above; and may also search or refine a search by data center, instrument, platform, project, or temporal/spatial resolution.

To realize a federated catalogue for data discovery from multiple EO data centers, the CEOS WGISS Integrated Catalog (CWIC) system has been implemented. CWIC was initiated and supported by NASA, NOAA, and USGS as a contribution to CEOS. CWIC provides inventory search to WGISS agency catalog systems for EO data by distributing search requests to the appropriate server and sending search responses back to the requesting client. CWIC will provide translation from the CEOS OpenSearch search request to the native protocol used by the data partner server if the data partner system does not implement CEOS OpenSearch.

FedEO (Federated Earth Observation Gateway) 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.

WGISS is now coordinating efforts to connect CWIC and FedEO system with IDN through a common registration of metadata records to seamlessly provide search results for relevant data sets regardless of which system is used to access the granule level data.

4.2 Search Concept and Design

A two-step collection/granule search process, which separates discovery of data collections from searching within relevant data collections to retrieve specific data granules, has been adopted to realize the integrated access to heterogeneous, autonomous data sources.

The WGISS Connected Data Assets system is an implementation of this two-step process.

The IDN provides an OpenSearch front end to the collection search. The response from the collection search includes links to the OpenSearch Descriptor Documents (OSDD) at one of several Granule Gateways, providing search capability for granules at the relevant data providers. Current WGISS Connected Data Assets Granule Gateways include the CWIC and FedEO systems. Each of these systems provides access to different data archive systems using the same OpenSearch protocol. Spatial and temporal metadata are the only attributes guaranteed to be supported at all data providers. The CEOS OpenSearch Best Practice supports a lot of metadata attributes through the Parameter extensions, but they are not all universally supported across CWIC or FedEO.

4.3 WGISS Architecture for Connected Data Assets

At its core, the systems present to End Users and Clients an OpenSearch server compliant with the CEOS OpenSearch Best Practice. In this way, outside clients need to have no specific knowledge of the particular partner data systems and communicate only via OpenSearch.

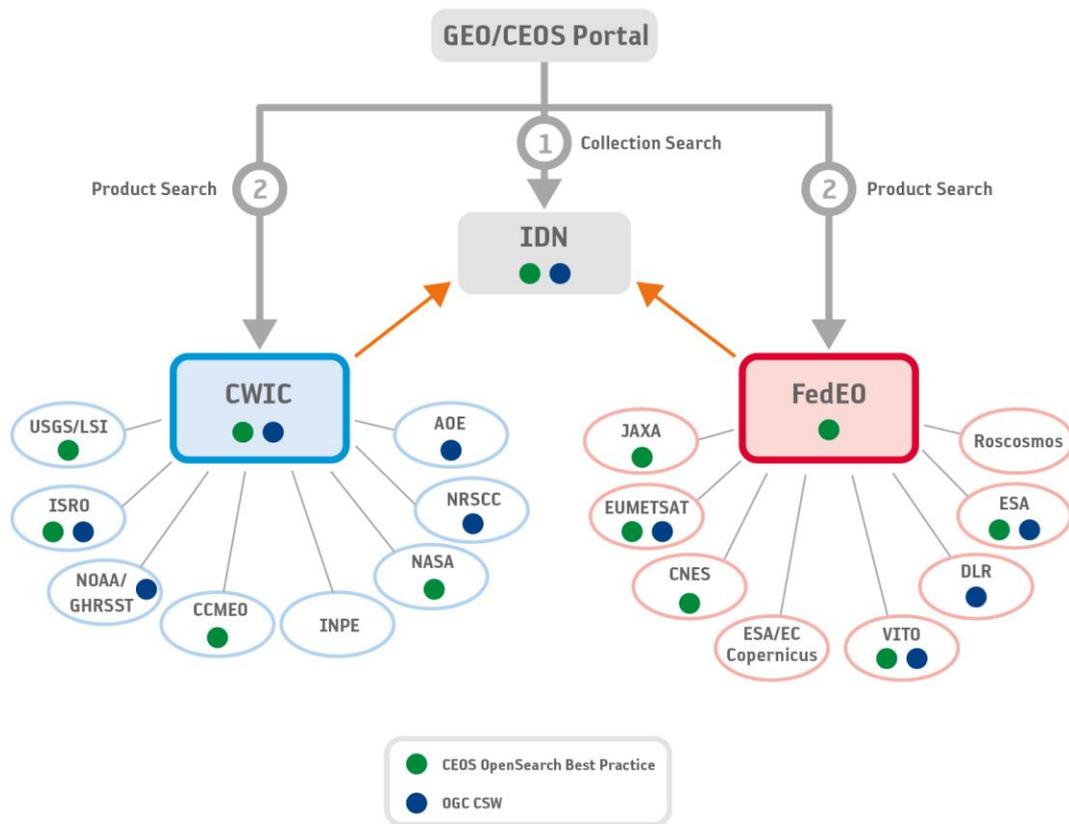


Figure 1: CEOS WGISS Connected Data Assets

4.4 Terms and Definitions

For the purposes of this document, the following terms and definitions apply:

Catalog ID	Identifier of data provider serving granule metadata
Client	A software component that can invoke an operation from a server.
CMR	The Common Metadata Repository (CMR) of National Aeronautics and Space Administration (NASA) is a high-performance, high-quality, continuously evolving metadata system that catalogs Earth Science data and associated service metadata records. See IDN definition for its relation to CWIC.
Collection	<p>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.</p> <p>A grouping of granules that all come from the same source, such as a modeling group or institution. Collections have information that is common across all the granules they “own” and a template for describing additional attributes not already part of the metadata model.</p>
Data Clearinghouse	The collection of institutions providing digital data, which can be searched through a single interface using a common metadata standard
GCMD	The Global Change Master Directory (GCMD) is a comprehensive directory of information about Earth science data.
Granule	The smallest aggregation of data that can be independently managed (described, inventoried, and retrieved). Granules have their own metadata model and support values associated with the additional attributes defined by the owning collection.
Granule Gateway	A CEOS OpenSearch compliant server providing access to remote data partner inventory systems.
Identifier	A character string that may be composed of numbers and characters that is exchanged between the client and the server with respect to a specific identity of a resource.
IDN	The CEOS International Directory Network is a gateway to the world of earth science data and services accessible at http://idn.ceos.org/portals/Home.do?Portal=idn_ceos
IDN dataset ID	Unique dataset identifier in IDN, returned from the IDN in response to the OSDD request.
native ID	Dataset identifier used by CWIC and FedEO to retrieve granule metadata through data provider API. This identifier is assigned by the data provider but may be the same as the IDN dataset ID.

operation	The specification of a transformation or query that an object may be called to execute
OSDD	OpenSearch Description Document
Product	Same meaning as granule.
profile	A set of one or more base standards and - where applicable - the identification of chosen clauses, classes, subsets, options and parameters of those base standards that are necessary for accomplishing a particular function.
request	The invocation of an operation by a client.
response	The result of an operation, returned from server to client.

4.5 IDN Systems

The IDN, CMR OpenSearch (for IDN), and the GCMD's Keyword Management Service (KMS) only have operational systems which end-users can access.

- IDN site is available to all users.
Location: <https://idn.ceos.org/>
- OpenSearch API for IDN (via CMR) - production instance is available to all users.
Location: <https://cmr.earthdata.nasa.gov/opensearch/>
- KMS - production instance is available to all users.
Location:
<https://wiki.earthdata.nasa.gov/display/gcmdkey/Keyword+Management+Service+Application+Program+Interface#KeywordManagementServiceApplicationProgramInterface-1Introduction>

The IDN site search interface and the CMR OpenSearch production instances will provide access to all datasets which have been registered in the IDN. The KMS production instance will provide access to all approved GCMD keywords registered by IDN providers.

4.6 FedEO Systems

There are two operational FedEO systems to which end-users have access.

- FedEO PROD – this is FedEO production instance and is available to all users.
Location: <https://fedeo.ceos.org/opensearch/description.xml>
- FedEO TEST – this is FedEO testing instance used by data partners and external clients to perform testing before changes are made to the FedEO production instance.

Location: <https://geo.spacebel.be/opensearch/description.xml>

The production instance will provide access to collections registered in the IDN and verified to be at least minimally compliant. The TEST instance may provide access to additional collections (*e.g.*, new collections undergoing testing and not yet registered in the IDN), and capabilities which have not yet been tested sufficiently to move to the production system.

4.7 Contact Information

All the documents and information about FedEO are available at WGISS FedEO page at

<https://ceos.org/ourwork/workinggroups/wgiss/access/fedeo/>

Any questions regarding to FedEO, please send the email to

To: ceos@spacebel.be

Cc: joost.van.bemmelen@esa.int, giuseppe.troina@esa.int

5 FedEO Data Partner Use Cases

FedEO aims to harmonise the interface with the data partner systems by proposing a harmonised interaction model for collection and granule discovery. Data partners either support one of the already supported metadata models, or provide custom metadata in which case FedEO will allow its clients to either obtain this custom metadata or another metadata type FedEO was able to generate based on the custom metadata received from the data partner.

FedEO clients can use the `{sru:recordSchema}` parameter of its OpenSearch interface to obtain different supported metadata types embedded in the search response as explained in the FedEO Client Partner Guide section 3.3.3.

5.1 Granule Metadata Publication

5.1.1 UC 1.1 - Online Granule Metadata

5.1.1.1 Overview

Data Providers may provide an online access to their granule metadata. This allows the FedEO Gateway to access the Data Provider granule catalog for distributed searches or harvesting of the granule metadata.

The data partner is free to propose the set of search parameters supported by his granule catalog(s). This set is often known by FedEO if the data provider implements a standard catalog protocol. As a minimum, parameters equivalent to the following OpenSearch parameters are expected to exist:

- Spatial search: providing for instance an area of interest as a bounding box (`geo:box`).
- Temporal search: providing a time range which is to intersect with the start time / end time of individual granules (i.e. products).
- `eo:parentIdentifier`: identifier of the collections to be searched.
- `geo:uid`: a locally unique product identifier which can be used to retrieve all available metadata for a given granule (product), possibly in combination with the collections identifier.
- `startIndex`: start index of the first result to be returned.
- `count`: number of results to be retrieved.

The response format should at least contain:

- Total number of results matching the query

- Granule identifier
- Start time, End time
- Geographical extent (i.e. footprint information)
- Platform name (optional)
- Instrument name (optional)
- Orbit number (optional)
- URL to quicklook (optional)
- URL for granule download (optional)

It is highly recommended to include HTTP URL to (JPEG) thumbnail images, browse images (also called quicklooks) and possibly cloud masks for each granule (product) in the metadata. If the Data Partner provides these URL in the granule metadata then FedEO will include these URL in the Atom search responses. FedEO uses Atom link (rel="icon") and the "Media RSS Specification" [AD.19] to return this information inside Atom feeds.

If Data Partners can make available browse image URLs as WMS URLs, then FedEO will make these URL accessible in the Atom search responses and sophisticated catalog clients can exploit the WMS links to allow zooming in/out.

Data Partners typically make available their data products for download on an online server at their Product Facilities through HTTP, FTP, SFTP or other protocols and possibly apply local authentication before end-users can download their product.

We recommend partners in such case to include the download URL in the granule metadata record. FedEO will then extract this URL and include it in the corresponding Atom entry of the OpenSearch response as an atom:link "enclosure".

FedEO will return information to search clients according to CEOS OpenSearch Best Practices.

5.1.1.2 Summary

Use Case Overview	
Title	UC 1.1 – Online Granule Metadata
Description	This use case describes steps for connecting Data Partner assets to FedEO via an online Granule metadata access.
Actors	Data Provider FedEO Gateway
Initial Status and Preconditions	Data Provider has an accessible Granule Catalog with granule metadata in a structured format.
Postcondition	Data Provider granule metadata discoverable through FedEO.

Table 1 Use Case: Publish Granules provided as Online Granule Metadata

The following steps describe this use case.

Step 1: The Data Partner provides on-line access to his granule catalog via a standard protocol (See Section 5.1.4) or a custom HTTP-based interface.

Step 2: The Data Partner sends endpoint information to the ESA FedEO team for integration.

- Examples of successful search requests (including area of interest / time period).
- List of collection identifiers i.e. {eo:parentIdentifier}.
- List of supported search parameters if different from formal (OGC) specifications and actual values, including search parameter to be used to pass collections identifier (e.g. parentIdentifier) in request.
- Optional: mechanism to discover list of allowed collection identifiers automatically.

Step 3: FedEO team provides FedEO OpenSearch test endpoint to the Data Partner for verification and feedback.

After integrating the Data Partner catalog in the FedEO test environment, the FedEO support team will make available the OpenSearch endpoint to the Data Partner for verification and feedback. A list of findings, issues to be resolved or to improve integration will be provided as well.

Step 4: FedEO team publishes the validated Granule search interface for the Data Partner through the operational FedEO endpoint.

5.1.2 UC 1.2 - Offline Granule Metadata

5.1.2.1 Overview

Data Providers may provide granule metadata off-line to the FedEO team. FedEO can host the granule metadata on behalf of the Data Provider and allow using the FedEO Granule Gateway to provide an online access to the granules and their metadata.

5.1.2.2 Summary

Use Case Overview	
Title	UC1.2 - Offline Granule Metadata
Description	This use case describes steps for connecting Data Partner assets to FedEO via provision of offline Granule metadata.
Actors	Data Provider FedEO Gateway
Initial Status and Preconditions	Data Provider has granule metadata as a set of metadata files in a structured format.
Postcondition	Data Provider granule metadata discoverable through FedEO.

Table 2 Use Case: Publish Granules provided as Offline Granule Metadata

The following steps describe this use case.

Step 1: The Data Partner provides granule metadata to the FedEO team offline as a set of metadata files or index files.

Step 2: The ESA FedEO team will ingest the granule metadata in the local FedEO Granule catalog, possibly after conversion.

Step 3: The FedEO team provides the FedEO OpenSearch test endpoint to the Data Partner for verification and feedback.

After publishing the Data Partner metadata in the FedEO test environment, the FedEO support team will make available the OpenSearch endpoint to the Data Partner for verification and feedback. A list of findings, issues to be resolved or to improve integration will be provided as well.

Step 4: FedEO team publishes the Data Partner granule metadata through the operational FedEO endpoint.

5.1.3 Supported Granule Metadata Formats

The use of the following granule metadata formats by the data partners facilitates their integration in FedEO. In addition to these formats, data partners can also have their own custom format which can be considered for integration with some additional effort.

5.1.3.1 OGC 10-157r4 - EOP O&M Version 1.1

This format is defined in [RD.9] and is the recommended granule metadata format. It allows for specific attributes for optical, radar, atmospheric, altimetry, limb looking, synthesis and systematic products.

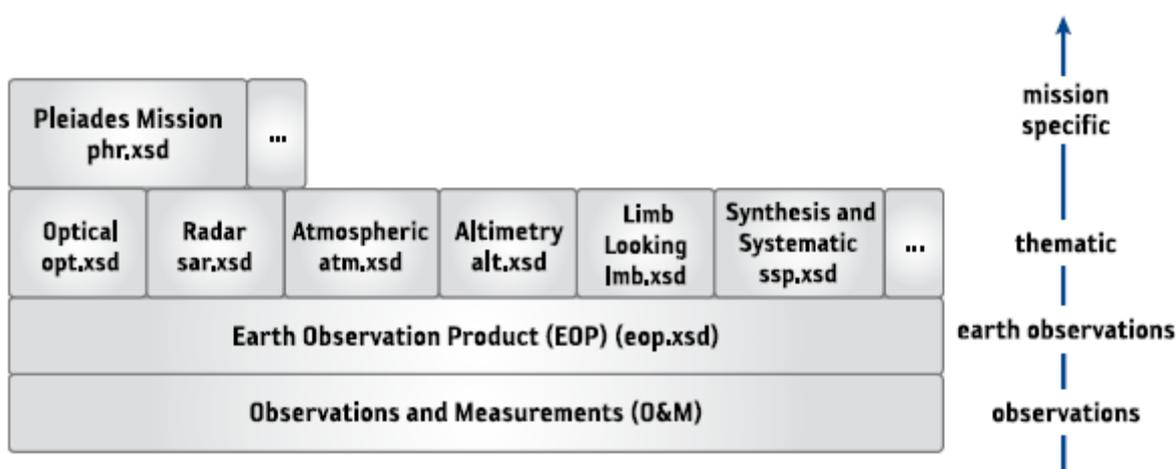


Figure 2: EOP O&M Metadata Model (source: [RD.1]).

5.1.3.2 OGC 10-157r3 - EOP O&M Version 1.0

This format is defined in [RD.8] and is slightly different from [RD.9] which is the recommended granule metadata format.

5.1.3.3 OGC 17-003 – GeoJSON

This format is based on GeoJSON and OGC OWS Context (OGC 14-055r2). The OGC 17-003 specification contains proposed mappings on UMM-G and OGC 10-157r4 granule metadata.

5.1.3.4 OGC 06-080 - EO GML

This format is defined in [RD.2] and contains similar attributes as [RD.9] which is the recommended granule metadata format.

5.1.3.5 ISO 19139

FedEO supports metadata as defined in [AD.11] [RD.11] and [RD.19]. OGC 11-035r1 [RD.11] is the recommended metadata model for collections metadata. It is a profile of ISO19139:2007 [AD.11] enriched with specific elements from the gmi namespace ISO19139-2:2012 [AD.12].

5.1.3.6 Dublin Core

A standard encoding schema for Dublin Core ¹ metadata is available at <https://www.loc.gov/standards/sru/recordSchemas/dc-schema.xsd>. It supports one or more of the following attributes:

- dc:title
- dc:creator
- dc:subject
- dc:description
- dc:publisher
- dc:contributor
- dc:date
- dc:type
- dc:format
- dc:identifier
- dc:source
- dc:language
- dc:relation
- dc:coverage
- dc:rights.

¹ <http://dublincore.org/documents/dces/>

5.1.3.7 Custom format

FedEO does also support custom metadata formats. Data partners can provide their metadata using their own XML, JSON, JSON-LD, RDF XML, Turtle or CSV (Comma-Separated-Values) formats. FedEO granule search responses will make available the original custom formats as atom:links.

5.1.4 Supported Catalog Protocols

FedEO provides a growing set of Connectors which aim to implement a common interaction model with the Data Provider backend systems. It is assumed that interactions between FedEO and the data partner infrastructure use the HTTP protocol.

The current section provides information about the protocols FedEO is able to use to communicate with online Data Partner catalogs. It allows Data Providers to understand which minimum information is required by FedEO and will be extracted from their OSDD or Capabilities documents.

5.1.4.1 OGC 06-131 – EOP EP of CSW

FedEO can connect directly to your OGC 06-131 catalog if a correct Capabilities document is available via HTTP GET. FedEO will extract the following information from this file:

- Location of the GetRecords and GetRecordById endpoints.
- Protocol binding for access: only <ows:HTTP><ows:Post> is allowed which is interpreted as a SOAP binding.
- List of supported collections from the <ows:OperationsMetadata> element in the Capabilities file as shown below.

Example 1: Extract of OGC 06-131 Capabilities Document

```
<?xml version="1.0" encoding="UTF-8"?>
<csw:Capabilities version="2.0.2"
  xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dct="http://purl.org/dc/terms/"
  xmlns:sch="http://www.ascc.net/xml/schematron"
  xmlns:gml="http://www.opengis.net/gml" xmlns:ogc="http://www.opengis.net/ogc"
  xmlns:ows="http://www.opengis.net/ows"
  xmlns:xlink="http://www.w3.org/1999/xlink">
  . . .
  <ows:OperationsMetadata>
    . . .
    <ows:Operation name="GetRecords">
      <ows:DCP>
        <ows:HTTP>
```

```

        <ows:Post
xlink:href="http://example1.com/gateway/services/Catalogue" xlink:type="simple"/>
        </ows:HTTP>
    </ows:DCP>

    . . .

</ows:Operation>
<ows:Operation name="GetRecordById">
    <ows:DCP>
        <ows:HTTP>
            <ows:Post
xlink:href="http://example1.com/gateway/services/Catalogue" xlink:type="simple"/>
            </ows:HTTP>
        </ows:DCP>
    . . .
</ows:Operation>
. . .
    <ows:ExtendedCapabilities xmlns:rim="urn:oasis:names:tc:ebxml-
regrep:xsd:rim:3.0">
        <rim:Slot name="urn:ogc:def:ebRIM-Slot:OGC-06-131:parentIdentifier"
slotType="urn:oasis:names:tc:ebxml-regrep:DataType:String">
            <rim:ValueList>
                <!-- VITO -->
                <rim:Value>urn:ogc:def:EOP:VITO:VGT_P</rim:Value>
                <rim:Value>urn:ogc:def:EOP:VITO:VGT_S10</rim:Value>
                <rim:Value>urn:ogc:def:EOP:VITO:VGT_D10</rim:Value>
            </rim:ValueList>
        </rim:Slot>
    </ows:ExtendedCapabilities>
</ows:OperationsMetadata>
. . .
</csw:Capabilities>

```

The above Capabilities file defines the collections EOP:VITO:VGT_P, EOP:VITO:VGT_S10, EOP:VITO:VGT_D10 as valid FedEO {eo:parentIdentifier} values. The prefix “urn:ogc:def” is ignored by FedEO.

The endpoint can be added to the FedEO Gateway by the FedEO team.

5.1.4.2 OGC 07-045 - ISO AP of CSW

FedEO can connect directly to your OGC 07-045 [RD.21] catalog if a correct Capabilities document is provided or accessible via HTTP GET. FedEO will extract the following information from this file:

- Location of the GetRecords and GetRecordById endpoints.
- Protocol binding: only <ows:HTTP><ows:Post> is allowed which is interpreted as a HTTP/Post binding

Example 2: Extract of OGC 07-045 Capabilities Document

```
<?xml version="1.0" encoding="UTF-8"?>
<csw:Capabilities version="2.0.2"
xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dct="http://purl.org/dc/terms/"
xmlns:sch="http://www.ascc.net/xml/schematron"
xmlns:gml="http://www.opengis.net/gml" xmlns:ogc="http://www.opengis.net/ogc"
xmlns:ows="http://www.opengis.net/ows"
xmlns:xlink="http://www.w3.org/1999/xlink">
  . . .
  . . .
  <ows:OperationsMetadata>
    . . .

    <ows:Operation name="GetRecords">
      <ows:DCP>
        <ows:HTTP>
          <ows:Post
xlink:href="http://example2.com/gateway/services/Catalogue" xlink:type="simple"/>
        </ows:HTTP>
      </ows:DCP>

      . . .
    </ows:Operation>
    <ows:Operation name="GetRecordById">
      <ows:DCP>
        <ows:HTTP>
          <ows:Post
xlink:href="http://example2.com/gateway/services/Catalogue" xlink:type="simple"/>
        </ows:HTTP>
      </ows:DCP>

      . . .
  </ows:OperationsMetadata>
</csw:Capabilities>
```

```
        </ows:Operation>
        . . .
    </ows:OperationsMetadata>
    . . .
</csw:Capabilities>
```

The endpoint can be added to the FedEO Gateway by the FedEO team.

Several open-source implementations of OGC 07-045 exist, for instance:

- GeoNetwork²

5.1.4.3 OpenSearch

Endpoints of an OpenSearch catalog can be added to the FedEO Gateway. However, OpenSearch catalogs exist in various flavours and all implement different portions of the related specifications OpenSearch 1.1 [AD.4], OGC 10-032r8 [AD.5] and OGC 13-026r8 [AD.6].

Several open-source implementations of OGC 13-026r8 exist which Data Partners can deploy, for instance:

- Earth Observation OpenSearch Catalog³.
- RESTo2⁴

Data Partners implementing their own OpenSearch interfaces may use the following open-source conformance tests

The endpoint of an OpenSearch catalog can be added to the FedEO Gateway by the FedEO team.

FedEO can connect to your OpenSearch catalog if a correct OSDD document is available via HTTP GET which complies with OGC 10-032r8 [AD.5] and OGC 13-026r8 [AD.6]. FedEO will extract the following information from this file:

- URL template of the OpenSearch endpoint (for collections) returning Atom responses: i.e. <Url rel="collections" template="https://...." type="application/atom+xml"/>. Is only required if the endpoint supports collections search.
- URL template of the OpenSearch endpoint (for granules) returning Atom responses: i.e. <Url rel="results" template="https://...." type="application/atom+xml"/>. Is only required if the endpoint supports granule search.

² <http://geonetwork-opensource.org/>

³ <https://github.com/IntecsSPA/eoos>

⁴ <https://github.com/jjrom/resto2/>

- Optionally: List of supported collections from the `<param:Parameter value="{eo:parentIdentifier}">` element in the OSDD file as shown below.

The collections which are searchable via this endpoint are extracted by FedEO from the OSDD description in which they can also be described using the Parameter extension [AD.18] as shown below.

Example 3: Extract of OSDD Document for OpenSearch Granule catalog

```
<opensearchdescription>

<Url rel="collection" template="https://...." type="application/atom+xml"/>

<Url rel="results" template="https://...." type="application/atom+xml">
  . . .
  <param:Parameter name="parentIdentifier" value="{eo:parentIdentifier}">
    <param:Option label="ASA_IM__0P" value="ASA_IM__0P"/>
    <param:Option label="TLM_MIRA0_" value="TLM_MIRA0_" />
  </param:Parameter>
</Url>
. . .

</opensearchdescription>
```

The above OSDD file defines the collections `ASA_IM__0P` and `TLM_MIRA0_` as valid FedEO `{eo:parentIdentifier}` values which can be used for granule search without a previous series search. If only a two-step search is required, then these identifiers do not need to be defined separately.

5.1.4.4 OGC 13-084r2 – I15 EP of CSW

FedEO can connect to your OGC 13-084r2 [RD.10] catalog if a correct Capabilities document is provided or made accessible via HTTP GET. FedEO will extract the following information from this file:

- Location and binding for the GetRecords endpoint. Only `<ows:HTTP><ows:Post>` is allowed which is interpreted as a SOAP binding

Example 4: Extract of OGC 13-084r2 Capabilities Document

```
<?xml version="1.0" encoding="UTF-8"?>
<ows:Capabilities version="2.0.2" xmlns:ows="http://www.opengis.net/ows" . . . >
  . . .
  . . .
  <ows:OperationsMetadata>
    . . .
```

```

        <ows:Operation name="GetRecords">
            <ows:DCP>
                <ows:HTTP>
                    <ows:Post
xlink:href="http://example2.com/gateway/services/Catalogue" xlink:type="simple"/>
                </ows:HTTP>
            </ows:DCP>
            . . .
        </ows:Operation>
        . . .
    </ows:OperationsMetadata>
    . . .
</csw:Capabilities>

```

The endpoint of an OGC 13-084r2 compliant (collections) catalog can be added to the FedEO Gateway by the Administrator.

Open-source implementations of OGC 13-084r2 exist, for instance:

- Buddata⁵.

5.1.4.5 OGC CSW 2.0.2

FedEO can connect to a CSW 2.0.2 catalog if a correct Capabilities document is available via HTTP GET. FedEO will extract the following information from this file:

- Location of the GetRecords and GetRecordById endpoints.
- Protocol binding: only <ows:HTTP><ows:Post> is allowed which is interpreted as a HTTP/Post binding
- List of supported collections from the <cwic:FederationMetadata> element in the Capabilities file as shown below.

Example 5: Extract of CWIC-Style CSW Capabilities Document

```

<?xml version="1.0" encoding="UTF-8"?>
<csw:Capabilities version="2.0.2"
xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dct="http://purl.org/dc/terms/"
xmlns:sch="http://www.ascc.net/xml/schematron"
xmlns:gml="http://www.opengis.net/gml" xmlns:ogc="http://www.opengis.net/ogc"

```

⁵ <https://github.com/IntecsSPA/buddata-ebxml-registry>

```

xmlns:ows="http://www.opengis.net/ows"
xmlns:xlink="http://www.w3.org/1999/xlink">
  . . .
  . . .
  <ows:OperationsMetadata>
    . . .

    <ows:Operation name="GetRecords">
      <ows:DCP>
        <ows:HTTP>
          <ows:Post
xlink:href="http://example4.com/gateway/services/Catalogue" xlink:type="simple"/>
          </ows:HTTP>
        </ows:DCP>

        . . .
      </ows:Operation>
      <ows:Operation name="GetRecordById">
        <ows:DCP>
          <ows:HTTP>
            <ows:Post
xlink:href="http://example4.com/gateway/services/Catalogue" xlink:type="simple"/>
            </ows:HTTP>
          </ows:DCP>

          . . .

        </ows:Operation>

    . . .

  <cwic:FederationMetadata>
    <cwic:catalog id="USGS">
      <cwic:granule entryId="LANDSAT_ETM_PLUS">Enhanced Thematic Mapper Plus
(Landsat 7)</cwic:granule>
      <cwic:granule entryId="LANDSAT_MSS">Landsat Multispectral Scanner (MSS)
Imagery </cwic:granule>
      <cwic:granule entryId="LANDSAT_TM">Landsat Thematic Mapper
Imagery</cwic:granule>
    </cwic:catalog>
    <cwic:catalog id="INPE">
      . . . . .
  </cwic:catalog>
  <cwic:catalog id="NOAA">
    . . . . .

```

```
</cwic:catalog>
<cwic:catalog id="NASA">
  . . .
</cwic:catalog>
</cwic:FederationMetadata>
. . .
</ows:OperationsMetadata>
. . .
</csw:Capabilities>
```

The above Capabilities file defines the collections LANDSAT_ETM_PLUS, LANDSAT_MSS and LANDSAT_TM as valid FedEO {eo:parentIdentifier} values.

The endpoint can be added to the FedEO Gateway by the FedEO team.

5.1.4.6 Custom HTTP-based API

Other API (over HTTP or HTTPS) can be considered on a case by case basis. The data partner is free to propose response formats (including Atom, RSS, RDF or custom XML) and can offer metadata in various formats (EOP O&M, EO GML, ISO19139, Dublin Core, .CSV, RDF or custom XML) as long as they can be mapped on the internal model.

5.2 Collection Metadata Publication

The publication of Data Partner collection metadata through FedEO can be used as an intermediate step to publish these metadata in the IDN.

5.2.1 UC 2.1 - Online Collection Metadata

5.2.1.1 Overview

Data Providers may provide an online access to their collection metadata. This allows the FedEO Gateway to access the Data Provider collection metadata and export the required DIF-10 encoding for publication of these metadata in IDN.

If the data partner hosts collection metadata online, then FedEO can host a mirror copy of these metadata in its FedEO Collection Catalog and synchronise the content regularly with the master copy residing at the Data Partner. The Data Partner can propose the protocol that FedEO should use to harvest the content of their Collection Catalog. FedEO supports among other the following interfaces for accessing backend Collection Catalogs:

- OpenSearch, e.g. OGC 10-032r8, OGC 13-026r8
- OGC 07-045 [RD.21].
- OGC 13-084r2 [RD.10].
- Custom API (over HTTP).

The original harvested collection metadata undergo two transformations when being ingested in FedEO:

- (1) If sufficient metadata elements are provided then FedEO will generate an ISO19139(-2) metadata file for the collections or perform a syntactical conversion to the internal ISO metadata format used by FedEO. This allows to offer a harmonised metadata model to external search clients.
- (2) Addition of information necessary to be able to republish the metadata as DIF-10, mainly insertion of controlled vocabulary elements from GCMD for science keywords, platforms, instruments, organisation names and topic categories.

If Start Date and End Date information for the collections is available, it will be used by FedEO to enrich the Granule search OSDD for this collections with constraints for the time:start and time:end queryables using the Parameter extension [AD.18] as shown below.

Example 6: Inclusion of time:start and time:end constraints in generated OSDD

```
<param:Parameter name="startDate" value="{time:start}"
```

```
        minimum="0"
        minInclusive="2009-01-01T00:00:00Z"
    />
    <param:Parameter name="endDate" value="{time:end}"
        minimum="0"
        maxInclusive="2011-01-01T00:00:00Z"
    />
```

During operations, FedEO will retrieve regularly the collection metadata from the Data Partner for FedEO to remain synchronised with the actual Data Partner Catalog content.

To allow for systematic synchronisation of collection metadata and export to IDN, the original collection metadata shall contain keywords that FedEO can extract and map to:

- GCMD platform names
- GCMD instrument names
- GCMD science keywords

For “platform” names, FedEO accepts the names used in the EO Web Thesauri (https://thesauri.spacebel.be/thesaurus/en/page/?uri=https://earth.esa.int/concepts/concept_scheme/platforms) or the names used in the GCMD platform thesaurus: https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/platforms/?format=csv

Not all platforms in this GCMD file are eligible. Only platforms in the “category” (column 1) called “Earth Observation Satellites” shall be considered. Column 3 “short_name” contains the labels to be used. Concepts with an empty third column, e.g. “METOP” or “FireBIRD” shall be ignored as they only contain a “Series_Entity” (column 2) and not a “Short_Name” (column 3). These correspond to SKOS concept which have “skos:narrower” links and are thus not “leaves” in the GCMD SKOS representation.

For “instrument” names, FedEO accepts the names used in the EO Web Thesauri (https://thesauri.spacebel.be/thesaurus/en/page/?uri=https://earth.esa.int/concepts/concept_scheme/instruments) or the names used in the GCMD instrument thesaurus https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/instruments/?format=csv

Not all instruments in this GCMD file are eligible.

Only instruments in the “category” (column 1) called “Earth Remote Sensing Instruments” shall be considered. Column 5 “short_name” contains the labels to be compared. Concepts with an empty column 5 shall be ignored as they correspond to SKOS concept which have “skos:narrower” links and are thus not “leaves” in the GCMD SKOS representation.

As an additional constraint, the metadata shall for each “platform keyword” contain at least one “instrument keyword” for an instrument which is hosted by the platform. The EO Web thesauri is used by the FedEO software to verify the platform/instrument “hosts” relation and will be extended by the FedEO team if additional platforms and/or instruments are to be included to cover your metadata content.

For “science keyword” names, FedEO accepts the names used in the GCMD science keyword thesaurus

https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/?format=csv.

Not all science keywords in this GCMD file are eligible: Only the following columns contain “valid” keywords which shall be used:

- Term,
- Variable_Level_1,
- Variable_Level_2,
- Variable_Level_3,
- Detailed_Variable.

If the original metadata format is ISO19139(-2), keywords appearing in two different ways will be extracted and analysed by FedEO. How to include keywords without anchors is explained in example 2.2 of the INSPIRE Technical Guidance [RD.19]. A similar example is included below.

Example 7: Alternative 1 - keyword without anchor

```
<gmd:keyword>
  <gco:CharacterString>TerraSAR-X</gco:CharacterString>
</gmd:keyword>
```

Keywords appearing without anchor will be processed as explained earlier in this section, i.e. compared with GCMD and ESA thesauri information, and is possible, automatically mapped on these.

How to include keywords with anchors is explained in example 2.3 of the INSPIRE Technical Guidance [RD.19] and corresponds to the Technical Guidance Recommendation C.8. A similar example is included below.

Example 8: Alternative 2 - keyword with anchor

```
<gmd:keyword>
  <gmx:Anchor xlink:href="http://vocab-
test.ceda.ac.uk/collection/cci/platformProg/plat_envisat_prog">Environmental
Satellite</gmx:Anchor>
</gmd:keyword>
```

If keywords are encoded with anchors, then the anchor will be used to map to a corresponding GCMD

platform/instrument or science keyword and when no match is found, the label will be compared in the same way as a keyword without anchor (i.e. alternative 1).

5.2.1.2 Summary

Use Case Overview	
Title	UC2.1 - Online Collection Metadata
Description	This use case describes steps for connecting Data Partner assets to FedEO via an online Collection metadata access.
Actors	Data Provider FedEO Gateway IDN
Initial Status and Preconditions	Data Provider has an accessible Collection Catalog with collection metadata in a structured format.
Postcondition	<ul style="list-style-type: none"> • Data Provider collection metadata discoverable through FedEO. • Data Provider collection metadata available in DIF-10 format for ingestion by IDN.

Table 3 Use Case: Publish Collection Metadata provided as Online Collection Metadata

The following steps describe this use case:

Step 1: The Data Partner provides on-line access to his collection catalog via a standard protocol (See Section 5.1.4) or a custom HTTP-based interface.

Step 2: The Data Partner sends endpoint information to the ESA FedEO team for integration.

Step 3: The FedEO team provides the FedEO OpenSearch test endpoint to the Data Partner for verification and feedback.

After integrating the Data Partner Collection catalog in the FedEO test environment for systematic, the FedEO support team will provide feedback to the Data Partner. A list of findings, issues to be resolved or to improve integration will be provided as well.

Step 4: FedEO team provides to the Data Partner access to the test data assets page accessible at <https://geo.spacebel.be/opensearch/asset/index.html> (See **Figure 3**). It provides the following reports:

A **metadata ingestion report** is provided to the Data Partner identifying which metadata records contain insufficient information to allow for DIF-10 export and which keywords appearing in the metadata could not be mapped to the reference thesauri. See also section 7.1.

A **DIF-10 metadata validation report** is provided showing the result of the NASA DIF-10 validation⁶ of the auto-generated DIF-10 files. See also section 7.2.

⁶ <https://cmr.earthdata.nasa.gov/ingest/site/docs/ingest/api.html#validate-collection>

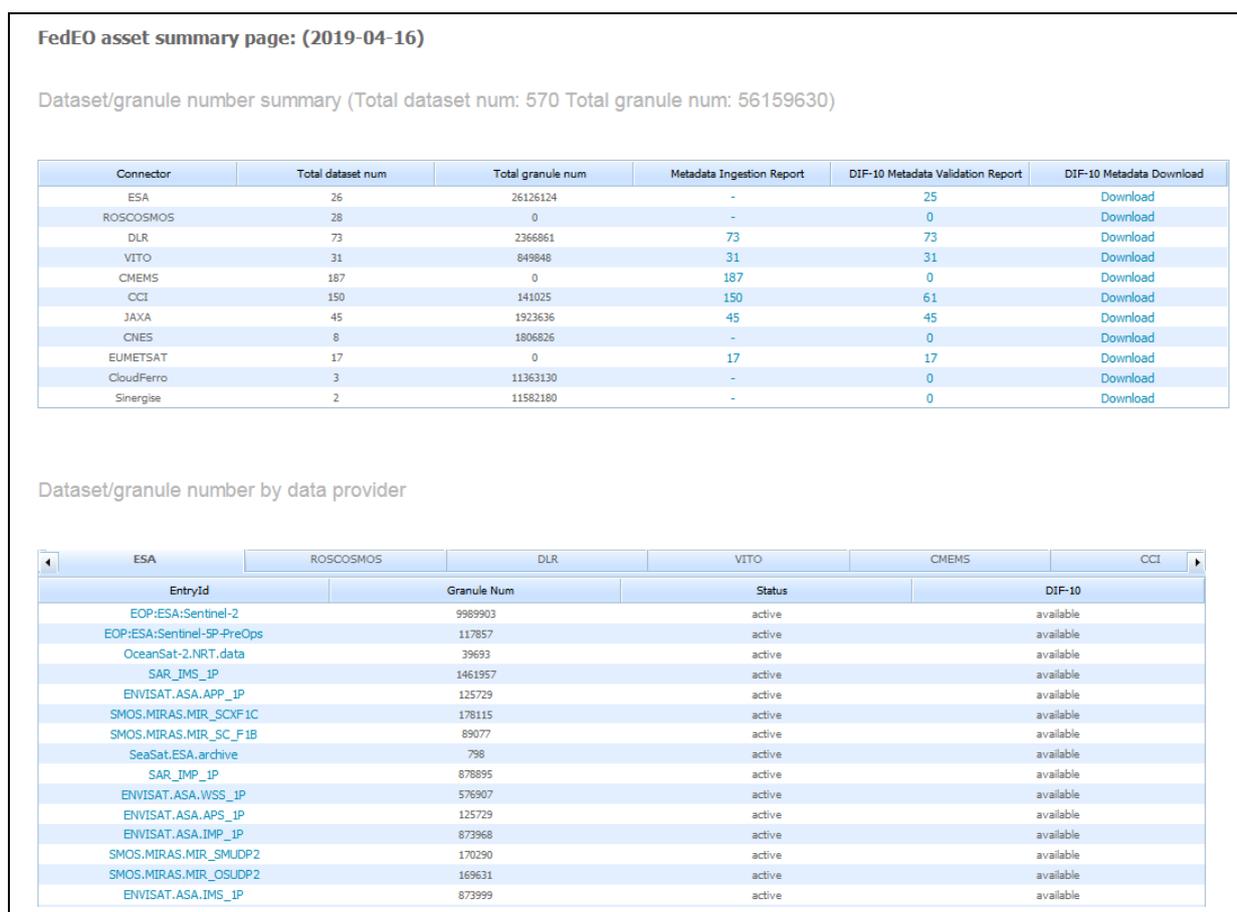


Figure 3: FedEO Assets Summary Page

Step 5: Data Partner makes adjustments to his collection metadata based on findings reported in the reports provided in STEP 4 and STEP4/STEP5 are repeated until the reports are satisfactory.

Step 6: FedEO team includes the Partner Collection Catalog endpoint in the operational FedEO configuration for systematic periodical harvesting.

Step 7: Data Partners can monitor the status of their collection metadata in FedEO via the data assets page accessible from <https://ceos.org/ourwork/workinggroups/wgiss/connected-data-assets/>.

5.2.2 UC 2.2 - Offline Collection Metadata

5.2.2.1 Overview

Data Providers may provide collection metadata off-line to the FedEO team. FedEO can host the collection metadata on behalf of the Data Provider and allow using the FedEO Gateway to provide an online access to the collection metadata. In addition, it allows exporting the required DIF-10 encoding for publication of the metadata in IDN.

If the data partner does not host his collection metadata online, then FedEO can host these metadata in its FedEO Collection Catalog on behalf of the Data Provider.

The set of metadata can be provided offline to the FedEO team (e.g. ISO metadata files in a .ZIP archive, by email). These metadata will then undergo similar enhancements as explained in UC 2.1 before they are published in the FedEO Collection Catalog and IDN.

Alternatively, the FedEO team can provide an Excel spreadsheet template that can be used by the Data Provider to capture the main characteristics of the collections. The filled spreadsheets can be sent to the FedEO team and will be converted into ISO19139-2 metadata for ingestion in the FedEO Collection Catalog.

Please contact the FedEO team for further information if you wish that FedEO hosts the collection metadata on your behalf and subsequently makes these available through IDN.

5.2.2.2 Summary

Use Case Overview	
Title	UC2.2 - Offline Collection Metadata
Description	This use case describes steps for connecting Data Partner assets to FedEO via provision of offline Collection metadata.
Actors	Data Provider FedEO Gateway IDN
Initial Status and Preconditions	Data Provider has collection metadata as a set of metadata files in a structured format.
Postcondition	<ul style="list-style-type: none"> Data Provider collection metadata discoverable through FedEO. Data Provider collection metadata available in DIF-10 format for ingestion by IDN.

Table 4 Use Case: Publish Collection Metadata provided as Offline Collection Metadata

The following steps describe this use case:

Step 1: The Data Partner provides collection metadata to the FedEO team offline as a set of metadata files or populated Excel collection templates.

Step 2: The ESA FedEO team will ingest the collection metadata in the local FedEO Collection catalog, possibly after conversion.

Step 3: The FedEO team provides the FedEO OpenSearch test endpoint to the Data Partner for verification and feedback.

After publishing the Data Partner metadata in the FedEO test environment, the FedEO support team will make available the OpenSearch endpoint to the Data Partner for verification and feedback. A list of findings, issues to be resolved or to improve integration will be provided as well.

Step 4: FedEO team provides to the Data Partner access to the test data assets page accessible at <https://geo.spacebel.be/opensearch/asset/index.html> (See **Figure 3**). It provides the following reports:

A **metadata ingestion report** is provided to the Data Partner identifying which metadata records contain insufficient information to allow for DIF-10 export and which keywords appearing in the metadata could not be mapped to the reference thesauri. See also section 7.1.

A **DIF-10 metadata validation report** is provided showing the result of the NASA DIF-10 validation⁷ of the auto-generated DIF-10 files. See also section 7.2.

Step 5: Data Partner makes adjustments to his collection metadata based on findings reported in the reports provided in STEP 4 and provides a set of updated metadata files STEP1 to STEP5 are repeated until the ingestion and validation reports are satisfactory.

Step 6: FedEO team publishes the Data Partner collection metadata through the operational FedEO endpoint.

5.2.3 Supported Collection Metadata Formats

FedEO can cope with various types of collection metadata made available by Data Providers either online or offline. The preferred format for collections metadata is OGC 11-035r1 [RD.11]. It is a profile of ISO19139:2007 [AD.11] enriched with specific elements from the gmi namespace ISO19139-2:2012 [AD.12].

Examples are available online: e.g. all collections returned by <https://fedeo.ceos.org/opensearch/request/?parentIdentifier=EOP:ESA:FEDEO:COLLECTI ONS> are described using OGC 10-035r1.

FedEO will convert the custom metadata to OGC 11-035r1 compliant metadata if sufficient metadata elements are available.

FedEO support staff can make available sample collections XML records for data partners to complete and return or prepare the metadata using information provided by Data Partners.

⁷ <https://cmr.earthdata.nasa.gov/ingest/site/docs/ingest/api.html#validate-collection>

The following options are available for data partners:

- Provide collections metadata online using one of the supported Protocol Connectors (See section 5.1.4).
- Provide collection metadata files (e.g. ISO19139) offline. FedEO will then host the collections metadata on behalf of the data partner.

The following metadata elements are of particular importance:

- The unique identifier for the collection. This identifier is returned by FedEO in collections OpenSearch responses as the <dc:identifier> element and is to be used in subsequent granule searches as {eo:parentIdentifier} search parameter.
- The temporal extent of the collection. This information is used by FedEO to return the <dc:date> element in the Atom search response. It is typically used by clients to constrain the search period for which granules may exist.
- The geographical extent of the collection.
- Abstract describing the collection.
- The satellite (platform) name. Ideally, provide the GCMD⁸ URI for the platform as explained in section 5.2.1. For example “Sentinel-1A” corresponds to <https://gcmd.earthdata.nasa.gov/kms/concept/c7279e54-f7c1-4ee7-a957-719d6021a3f6>.
- The instrument name. Ideally, provide the GCMD URI for the instrument as explained in section 5.2.1. For example “ASAR” corresponds to <https://gcmd.earthdata.nasa.gov/kms/concept/912c3308-23bc-4e12-b7fb-9d82e9fc5fe9>.
- Scientific keywords: Provide GCMD science keywords from https://gcmd.earthdata.nasa.gov/kms/concepts/concept_scheme/sciencekeywords/?format=csv as explained in section 5.2.1.

The “Metadata ingestion” report generated by the ingestion software will provide indications about missing metadata elements which are required to auto-generate DIF-10 metadata for publication in IDN.

⁸ <https://gcmd.earthdata.nasa.gov/KeywordViewer/> provides easy access to available GCMD concepts.

6 Common Partner Guidelines

These guidelines apply to both FedEO and CWIC Data Partner and are therefore “common” guidelines.

6.1 Metadata & Semantic Mapping

6.1.1 HTTP/HTTPS access

Although Granule Gateways such as FedEO and CWIC will attempt to use any mechanism available for connecting to Data Partners’ data management systems in order to access the available inventory search, there are a few specifics which make the process simpler and more robust.

The use of HTTP/HTTPS for accessing the inventory search engine is strongly preferred. This is widely used already, as web browsers are nearly universal and provide an effective user interface for both human and automated access.

Similarly for results, FedEO will attempt to extract the relevant results from any responses the partner data system returns. However, structured text of some sort – XML, for example – is strongly preferred. The ability to easily and definitively parse results makes the process of mapping the metadata returned in the search response simpler and less error-prone. Other structured formats like comma-delimited tables or JSON are acceptable.

6.1.2 Spatial search

All data partners are expected to support some level of spatial search since all of the inventory data are anticipated to have a spatial component. For each granule, a simple bounding box, with the bounding coordinates individually identified, is the minimum required, although more complex spatial footprint geometries are possible in the future. The number of spatial footprints in data partner’s response is not limited to one.

It is desirable to have the API also support a dynamic call to return the limits of the spatial search, although not necessary. The presence of such a service can help the Granule Gateway avoid invalid or inappropriate search requests, such as those outside the spatial boundaries for specific data collections.

6.1.3 Temporal search

Similar to spatial search described in the previous section, all data partners are expected to support some level of temporal search since all of the inventory data is anticipated to have a temporal component. Simple temporal extent, with the start and end times individually identified is the minimum required, although more complex temporal relations are anticipated in the future. It is best to support some minimal subset of the ISO 8601 time specification for syntax – YYYY-MM-DD, at least.

It is desirable to have the API also support a dynamic call to return the limits of the temporal

extent search, although not necessary. The presence of such a service can help to avoid invalid or inappropriate search requests, such as those outside the existing temporal extent for specific data collections. Alternatively, it is expected that the corresponding collection metadata (in IDN) defines temporal constraints that match the limits supported by the granule search API.

6.1.4 Request for collection by ID

Collection ID (`eo:parentIdentifier`) serves as a mandatory parameter in the FedEO OpenSearch query interface. FedEO will query against data partner's search system, using the Collection ID, to retrieve the matching granule level metadata. From this point, all data partners are expected to support the query based on collection identifier. In particular, a globally unique dataset identifier (i.e. DIF Entry ID or IDN `datasetID`) is recommended to be implemented as query parameter on data partner side. However, other unique dataset identifiers within data partner system is also acceptable.

6.1.5 Request for granule by ID

FedEO OpenSearch supports searching for a single granule by granule ID. FedEO will query against data partner's search system with the granule ID (`geo:uid`) in order to retrieve the matching granule level metadata. From this point, all data partners are recommended to support the query based on granule identifier. FedEO will use granule IDs as they are advertised by the data partner in search responses which typically contain multiple matching granules, each identified by their granule ID.

6.1.6 Unique granule IDs

Each data granule returned in a search response should have an identifier associated with it which is unique within the collection. It is important that the search response include the unique identifier for each granule so that the full data on individual granules may be retrieved without re-executing a (potentially time-consuming) search.

6.1.7 Record counts

As part of the search response from the inventory system, it is highly desirable to have the total count of matching granules returned, even if the metadata for the granules is not contained in the search response. This parameter, coupled with the ability to specify the starting record number and number of desired records from the inventory system, will allow clients to implement results paging and reducing the load on both the Granule Gateway system and on the data partners.

6.1.8 Pagination

Pagination is supported (*i.e.* search by specifying start page and counts per page) in the FedEO

OpenSearch query interface. The search based on cursor parameter (e.g. start index and item per page) is expected to be implemented by the Data Partner.

6.1.9 Other search parameters

The Data Partner inventory systems may support search on additional parameters. FedEO can support any additional search parameter the Data Partner is ready to advertise. Preference is given to parameters which can be mapped on standard search parameters defined in OGC 13-026r8.

FedEO is also considering support for results sorting – by start time, by duration, by granule ID, etc. – and it would be useful, although not mandatory, to have a sort option available from the Partner inventory system.

6.2 Interaction & Services Model

6.2.1 Unique granule ID

As described above, each data granule should have a unique identifier which a) is passed back to the client as part of the search response and b) can be used as a key with which to retrieve that specific granule. The Granule Gateway components will manage the task of associating the identifier with the correct collection.

6.2.2 Contact information

Not applicable.

6.2.3 Browse URL

If browse images of the data granule are available, a valid URL to display the browse image should be included in the search response for each granule so that the client can display it as a link. While it is possible for the FedEO Gateway to build the URL based on some pre-defined, fixed pattern, this mechanism is not recommended because it removes control over the form of the URL from the Data Partner and changes may require modifications to the Gateway source code. This can lead to delays in the deployment of the correct URL when changes are implemented by the Data Partner.

6.2.4 Order URL

We recommend that, when the granule data can be downloaded from the data provider directly, a valid URL to retrieve the data be included in the search response for the granule.

Alternatively, the search response may contain a URL directing the user to a web site for ordering the data if this is the only option permitted by the data center. This is often necessary

even for freely available data if, for example, data center policies require user registration before downloads are made available. In such cases, we strongly recommends that the granule ID requested be cached at the ordering system so that whenever the data center requirements for downloading data are met, the user will be able to retrieve the data without re-entering the granule ID.

6.2.5 Pagination support

As stated above, there is no strict requirement in terms of pagination mechanism or pagination parameters imposed by FedEO. Various pagination mechanism and heterogeneous parameters are expected across data partners. To provide universal pagination, the FedEO Gateway translates `startIndex` and `count` (stipulated by CEOS OpenSearch Best Practices) to native pagination parameters for each Data Partner.

6.2.6 Spatial search

As stated above, a Data Partner is expected to support spatial search and include proper spatial component in each granule response. Native spatial components (one or many) from data partner are converted syntactically and semantically to proper spatial components (e.g. `georss:box` and `georss:polygon`) which are stipulated by CEOS OpenSearch Best Practices.

6.2.7 Temporal search

As stated above, a Data Partner is expected to support temporal search and include standard (ISO 8601 compliant) temporal elements in each granule response. Temporal elements with native format from data partner are converted syntactically and semantically to single temporal element (i.e. `dc:date`) which are stipulated by CEOS OpenSearch Best Practices.

With the knowledge of Data Partner specific limitations on temporal search expressed in the Collection metadata, FedEO will impose limitations on temporal request parameter by including minimum and maximum values for the time parameters in the Collection-specific OSDD it has registered at IDN.

6.3 Error Handling

Error handling in CEOS OpenSearch is based on standard HTTP/HTTPS status codes.

Useful status and error messages help the Gateway manage client sessions effectively. Any limitations on submitted search requests to the inventory systems should be noted in the response (e.g., “too many records requested”, “search timed out”) so that predictable error-handling can be managed by the Gateway. It is especially important to have a clear error response if there is any limitation on spatial and temporal values, which are two of the most common search specifications used against individual data collections.

The CEOS OpenSearch Best Practices (CEOS-OPENSEARCH-BP-V1.2) recommends the use

of specific HTTP/HTTPS status codes as following, 4xx for client errors, and 5xx for server errors.

CEOS OpenSearch implementations are recommended to support these codes. Please refer to [CEOS-OPENSEARCH-BP-V1.2] for more details.

7 Sample Metadata Ingestion Report

The “metadata ingestion” and “DIF-10 Metadata Validation” reports are made available online at the FedEO assets page (e.g. <https://geo.spacebel.be/opensearch/asset/index.html> or <https://fedeo.ceos.org/asset/index.html>). The current chapter contains an example of each.

7.1 Metadata Ingestion Report

This section contains a sample of a typical collection metadata ingestion report allowing Data Partners to improve the keyword information present in their collection metadata.

7.1.1 Record Summary

The “Record Summary” table provides the overview of all metadata records and keywords and their mapping status.

In the example below, the table shows that:

- 20 out of 31 metadata files are “complete”, i.e. they fulfil the minimal conditions to be exported as DIF-10 metadata to be published in IDN.
- 11 out of 31 metadata files are incomplete and require further work to enable their export as DIF-10 metadata.
- 20 out of 31 metadata records contain “platform” information that can be mapped to GCMD platform information.
- 20 out of 31 metadata records contain “instrument” information that can be mapped to GCMD instrument information.
- 24 out of 31 metadata records (or 77.4%) contain science keyword information.
- 29 out of 40 keywords appearing in the metadata records was not recognized as platform, instrument or science keyword.

Endpoint	EOP:VITO:PDF							
Ingestion Date	2018/10/26 11:30:16							
	Records		Keywords		Organisations		Topic Categories	
Total	31	-	40	-	03	-	02	-
Complete	20	64.5%	-	-	03	100.0%	02	100.0%
Incomplete	11	35.5%	-	-	00	0.0%	00	0.0%
No mapping	-	-	29	72.5%	-	-	-	-
Platform	20	64.5%	04	10.0%	-	-	-	-
Instrument	20	64.5%	03	7.5%	-	-	-	-
Science Keyword	24	77.4%	06	15.0%	-	-	-	-

7.1.2 Organisation Summary

Count	Organisation	GCMD UUID	GCMD Label
32	European Space Agency, Copernicus Operations	b60cff41-3717-4919-8efd-33ee6c173d19	VITO
32	Flemish Institute for Technological Research (VITO)	b60cff41-3717-4919-8efd-33ee6c173d19	VITO
32	VITO (Flemish Institute for Technological Research)	b60cff41-3717-4919-8efd-33ee6c173d19	VITO

7.1.3 Topic Category Summary

Count	Topic Category	DIF-10 Topic Category
18	environment	ENVIRONMENT
13	imageryBaseMapsEarthCover	IMAGERY/BASE MAPS/EARTH COVER

7.1.4 Keyword Details

Count	Keyword	Type	GCMD UUID	GCMD Label
01	METOP	platform	8c192c86-d07c-4e7b-af8f-92aa4b40fca7	
18	proba	platform	6f507389-2c7c-41b4-a638-95bdc73b63a3	Proba-V
01	MERIS	platform	a1498dff-002d-4d67-9091-16822c608221	ENVISAT
01	ENVISAT	platform	a1498dff-002d-4d67-9091-16822c608221	ENVISAT
18	proba	instrument	88094b8c-c68c-4e4f-a4d9-1d5b6e15ba5b	VGT-P
01	MERIS	instrument	fc3f3ca6-1578-47b2-9d99-bc35ca85e0ba	MERIS
01	AVHRR	instrument	e64e83bd-02b3-4a47-830d-00e1aa4b04d3	AVHRR
14	Snow	science	8ef6560e-c699-49b4-bcb3-6db68506ca22	SNOW COVER
03	Optical	science	c5ff6f39-0c35-488a-96f2-f3498c678e45	VISIBLE WAVELENGTHS
01	ndvi	science	e63844c1-015c-4776-b01c-e3e7d5dd3d0c	VEGETATION INDEX
15	Land cover	science	e5815f58-8232-4c7f-b50d-ea71d73891a9	LAND USE/LAND COVER
06	NDVI	science	e63844c1-015c-4776-b01c-e3e7d5dd3d0c	VEGETATION INDEX
18	land cover	science	e5815f58-8232-4c7f-b50d-ea71d73891a9	LAND USE/LAND COVER
05	biogeophysical			
14	Radiometric quality			

7.1.5 Record Details

The “Record Details” table provides the overview of all metadata records and their completeness status.

The status information is encoded as follows:

Row 1	<p>COMPLETE: the minimum conditions to allow DIF-10 export to IDN are fulfilled.</p> <p>INCOMPLETE: a necessary condition to allow DIF-10 export to IDN is not fulfilled. See Row 2 and Row 3 for additional details.</p>
Row 2	<p>Format: <n1> <n2> <n3> <n4> <n5></p> <ul style="list-style-type: none"> • <n1> : Number of keywords mapped to GCMD platforms. Must be at least 1. • <n2> : Number of keywords mapped to GCMD instruments. Must be at least 1. • <n3> : Number of keywords mapped to GCMD science keywords. Must be at least 1.

	<ul style="list-style-type: none">• <n4> : ISO Organisation mapped to GCMD organization name. Must be 1.• <n5> : ISO Topic category matching allowed topic categories. Must be 1 (TBC).
Row 3	<p>Format: <n1> <n2> <n3></p> <ul style="list-style-type: none">• <n1> : Number of keywords mapped to GCMD platforms, but not having associated instruments in this metadata record. These platforms are labelled with an "*" in the table. Must be 0.• <n2> : Number of keywords mapped to GCMD instruments but not associated with a corresponding platform in this metadata record. These instruments are labelled with an "*" in the table. Must be 0.• <n3> : Number of keywords not used in any mapping. This number has no impact on the validity of the generated DIF-10. It is however recommended to have as many keywords as possible which correspond to GCMD (science) keywords.

Record identifier	Status	Keyword	Type	GCMD UUID	GCMD Label
urn:ogc:def:EOP:VITO:PROBAV_S1-TOA_1KM_V001	COMPLETE 01 01 02 05 01 01	proba	platform	6f507389-2c7c-41b4-a638-95bdc73b63a3	Proba-V
		proba	instrument	88094b8c-c68c-4e4f-a4d9-1d5b6e15ba5b	VGT-P
		land cover	science	e5815f58-8232-4c7f-b50d-ea71d73891a9	LAND USE/LAND COVER
		Snow	science	8ef6560e-c699-49b4-bcb3-6db68506ca22	SNOW COVER
		Cloud			
		Radiometric quality			
		Shadow			
		land use			
		environmental monitoring facilities			
		European Space Agency, Copernicus Operations	organisation	b60cff41-3717-4919-8efd-33ee6c173d19	VITO
		environment	topic category		ENVIRONMENT
urn:ogc:def:EOP:VITO:PROBAV_S1-TOC_1KM_V001	COMPLETE 01 01 02 05 01 01	proba	platform	6f507389-2c7c-41b4-a638-95bdc73b63a3	Proba-V
		proba	instrument	88094b8c-c68c-4e4f-a4d9-1d5b6e15ba5b	VGT-P
		land cover	science	e5815f58-8232-4c7f-b50d-ea71d73891a9	LAND USE/LAND COVER
		Snow	science	8ef6560e-c699-49b4-bcb3-6db68506ca22	SNOW COVER
		Cloud			
		Radiometric quality			
		Shadow			
		land use			
		environmental monitoring facilities			
		European Space Agency, Copernicus Operations	organisation	b60cff41-3717-4919-8efd-33ee6c173d19	VITO
		environment	topic category		ENVIRONMENT

7.2 DIF-10 Metadata Validation Report

This section contains a sample of a DIF-10 metadata validation report allowing Data Partners to verify the validation result generated by the NASA Online DIF-10

validator for the DIF-10 metadata files generated automatically by FedEO. Only metadata records which are indicated as “complete” in the “Ingestion Report” are included in the DIF-10 metadata validation report.

The following two figures provide examples of Errors and Warnings reported by this validator. Warning messages not related to DIF-10 (e.g. UMM-C) can be safely ignored.

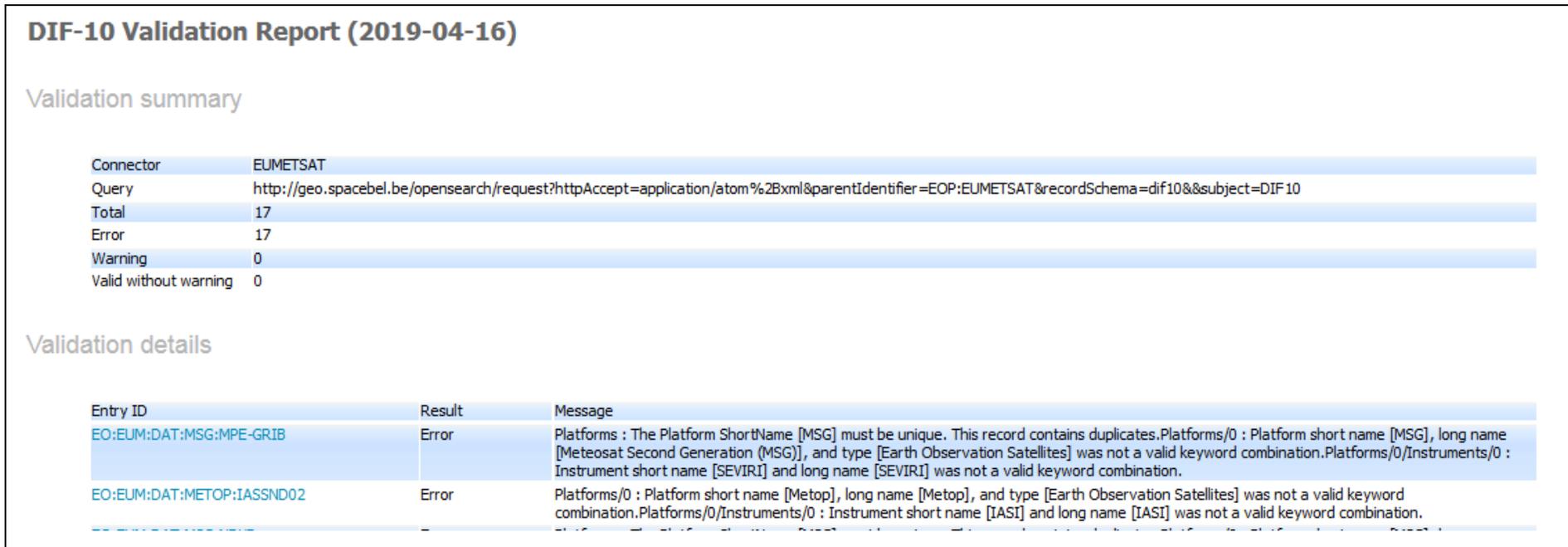


Figure 4: DIF-10 Metadata Validation Report with Errors

DIF-10 Validation Report (2019-04-16)

Validation summary

Connector	VITO
Query	http://geo.spacebel.be/opensearch/request?httpAccept=application/atom%2Bxml&parentIdentifier=EOP:VITO:PDF&recordSchema=dif10&&subject=DIF10
Total	31
Error	0
Warning	31
Valid without warning	0

Validation details

Entry ID	Result	Message
urn:ogc:def:EOP:VITO:PROBAV_S5-TOC-NDVI_100M_V001	Warning	After translating item to UMM-C the metadata had the following issue: #/DataCenters/0/ContactGroups/0: required key [GroupName] not found. #/RelatedUrls/0/GetData: required key [Format] not found. #/RelatedUrls/2/GetData: required key [Format] not found. #/RelatedUrls/1/GetService: required key [Protocol] not found. #/RelatedUrls/1/GetService: required key [FullName] not found. #/RelatedUrls/1/GetService: required key [DataID] not found. #/RelatedUrls/1/GetService: required key [DataType] not found. #/RelatedUrls/1/GetService/MimeType: application/opensearchdescription+xml is not a valid enum value. #/RelatedUrls/3/GetData: required key [Format] not found. #/RelatedUrls/3/GetData/MimeType: application/vnd.iso.19139-2%2Bxml is not a valid enum value. #/MetadataDates/0: required key [Date] not found. #/MetadataDates/1: required key [Date] not found. [:DataCenters 0 :ContactInformation :RelatedUrls 0] RelatedUrl does not have a description.
urn:ogc:def:EOP:VITO:PROBAV_S10-TOC_333M_V001	Warning	After translating item to UMM-C the metadata had the following issue: #/DataCenters/0/ContactGroups/0: required key [GroupName] not found. #/RelatedUrls/0/GetData: required key [Format] not found. #/RelatedUrls/2/GetData: required key [Format] not found. #/RelatedUrls/1/GetService: required key [Protocol] not found. #/RelatedUrls/1/GetService: required key [FullName] not found. #/RelatedUrls/1/GetService: required key [DataID] not found. #/RelatedUrls/1/GetService: required key [DataType] not found. #/RelatedUrls/1/GetService/MimeType: application/opensearchdescription+xml is not a valid enum value. #/RelatedUrls/3/GetData: required key [Format] not found. #/RelatedUrls/3/GetData/MimeType: application/vnd.iso.19139-2%2Bxml is not a valid enum value. #/MetadataDates/0: required key [Date] not found. #/MetadataDates/1: required key [Date] not found. [:DataCenters 0 :ContactInformation :RelatedUrls 0] RelatedUrl does not have a description.
95c6d610-ba36-11de-8a39-0800200c9a66	Warning	After translating item to UMM-C the metadata had the following issue: #/DataCenters/0/ContactGroups/0: required key [GroupName] not found. #/MetadataDates/0: required key [Date] not found. #/RelatedUrls/0/GetData: required key [Format] not found. #/RelatedUrls/2/GetData: required key [Format] not found. #/RelatedUrls/1/GetService: required key [Protocol] not found. #/RelatedUrls/1/GetService: required key [FullName] not found. #/RelatedUrls/1/GetService: required key [DataID] not found. #/RelatedUrls/1/GetService: required key [DataType] not found. #/RelatedUrls/1/GetService/MimeType: application/opensearchdescription+xml is not a valid enum value. #/RelatedUrls/3/GetData: required key [Format] not found.

Figure 5: DIF-10 Metadata Validation Report with Warnings