

# EARTH OBSERVATION PRODUCT METADATA MAPPING TO STAC

*C. Reck, J. Eberle, T. Heinen, M. Kunze, F. Feckler, J. Meyer-Arnek*

German Aerospace Center (DLR)

German Remote Sensing Data Center (DFD), Oberpfaffenhofen, Germany

*E-Mails: {Christoph.Reck, Jonas.Eberle, Torsten.Heinen, Markus.Kunze, Felix.Feckler, Julian.Meyer-Arnek}@dlr.de*

Keywords: Earth Observation, Catalogue, Metadata, EOP, STAC

## ABSTRACT

The standard for Spatio-Temporal Asset Catalogs (STAC <sup>[15]</sup>) has matured and is widely being adopted. Capabilities, extensibility and tool support go beyond the existing OGC EOP <sup>[11]</sup>, OpenSearch <sup>[10]</sup> and OData <sup>[26]</sup> standards. Implementing a catalogue with interfaces according to a standard requires good specifications and examples. This always has been an area for improvements. The OGC EOP standards have undergone a series of evolutions to become adult <sup>[9]</sup>, onto which young STAC standard is still on the way. With a little help, it will soon overtake its master. In this paper we present three areas of improvements and suitable solutions are presented:

- a) overview of metadata specifications
- b) STAC metadata mappings for Copernicus datasets
- c) generic tool for metadata extraction

## 1. INTRODUCTION

OGC provides a good path to setup, present and maintain the specifications for standards. On the other hand, good specifications on complex subjects become difficult to read, especially if the standard bases on other standards. Quickly the content gets obscured by a complexity of inheritance and distributed information. We extract the structure of the metadata specified over several documents into one table to serve as an overview. Such a table then also serves as a checklist and can be extended with mapping columns to other related standards. For example, in this case depicting the complete EOP metadata structure and providing a mapping to STAC covering the Copernicus satellite series.

## 2. METADATA STANDARDS

DLR has accompanied the evolution and implementation of many Earth Observation metadata standards for many different satellite missions.

The first mentionable standard was the Catalogue Interoperability Protocol (CIP <sup>[27]</sup>) offering search and order functionality. It was based on the z39.50 binary protocol, described in ASN1. It was used in the joint EU, JRC and ESA INFEO project, but suffered from the resource hunger within the implementation and poor performance of the servers at the time.

With success of the OGC WMS, WCS and WFS as browser and service interfaces, ESA incited the use of the Heterogeneous Mission Accessibility interface (HMA <sup>[28]</sup>), based on the OGC CSW interface. HMA offered discoverable collections and ordering, allowed mission specific metadata extensions, and based on the harmonized EOP metadata standard. The adoption suffered from using a flat and wordy eBRIM XML structure.

First widespread success was achieved with OpenSearch, offering a self-describing interface, collection discovery, and product searches. It used the EOP metadata standard, yet induced some evolutions like moving from XML to GeoJSON. Some good interfaces like FEDEO <sup>[24]</sup> and clients like the EOxClient <sup>[25]</sup> natively access OpenSearch catalogues.

STAC <sup>[15]</sup> followed to overcome the not so flexible EOP data model, to profit from linking collections, catalogues and datasets, and to improve search capabilities, with direct access to the data itself. The concept and model for STAC is well described [16]. As STAC acceptance increases, it would profit from an alignment with the harmonized metadata specification of the mature OGC EO products metadata standards (EOP).

## 2.1. EOP

The most comprehensive Earth Observation products metadata model has been designed by OGC contributors and evolved over several years:

- OGC 06-131 EO Products Extension Package for ebRIM <sup>[12]</sup>
- OGC 10-157 Earth Observation Metadata profile of Observations & Measurements <sup>[11]</sup>
- OGC 13-026 OpenSearch Extension for Earth Observation <sup>[10]</sup>
- OGC 17-003 EO Dataset Metadata GeoJSON(-LD) Encoding Standard <sup>[9]</sup>

The latter one provides a cross-reference of the EOP model to other standards it bases on. Note that the STAC standard was developed nearly at the same time and is based on similar concepts like GeoJSON-LD and OGC Features.

As with most formally described standards, it is difficult to obtain a full overview of the definitions it encompasses and references. Distributed figures depict only parts of the whole, e.g. the figure on the right is an extract from OGC 17-003 <sup>[9]</sup>,

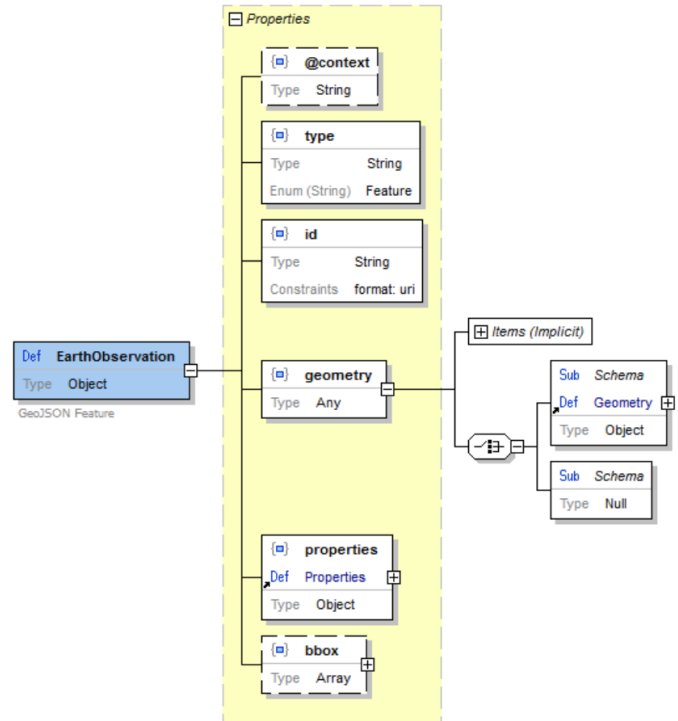


Figure 1: EarthObservation Schema (GeoJSON)

For our own developments at DLR, we have extracted the OGC EOP metadata model specification into one table see [2] (only 138 rows!), below we depict a part as an example:

OGC17-003 OSEO Attribute	Datatype	Cardinality	EO Vocabulary	Description	Valid
type	String	1		Type of the EarthObservation me (fixed value "Feature")	
id	String	1		Unique identifier for the EarthObservation element	
bbox	Bbox	0..1	gj:bbox	Information on the coordinate range of the geometry object representing t	
geometry	{}	1	gj:geometry	Contains the description of the target location observed during the EarthOb	
type	String	1	dct:title	Type of the geometry representi Point LineString Polygon MultiPoint Mu	
coordinates	[]	1	gj:coordinates	Array with structure determined by the geometry type to be represented (l	
properties	{}	1	gj:Properties (Class)	Groups all other properties of the EarthObservation not covered by the pro	
status	String	1	eop:status	Refers to product status	ARCHIVED ACQUIRED CANCELLED FAILE
doi	String	0..1	eop:doi	Digital Object Identifier identifying the product (see http://www.doi.org)	
title	String	1	dct:title	Human readable title given to th	productType + date
lang	String	0..1	dct:language	Metadata language, not empty with an RFC-3066 code as defined in [NRS]	
identifier	String	1	dct:identifier	Identifier given to the resource, typically the dataset (producer) identifier.	
parentIdentifier	String	0..1	eop:parentIdentifier	Collection identifier.	
date	RateRange	1	dct:date	Range of dates relevant for the resource (RFC-3339)	
created	DateTime	0..1	dct:created	Date of creation or last update of the EO Metadata Document	
published	DateTime	0..1	dct:issued	Date of first availability of the EO Metadata Document	
updated	DateTime	0..1	dct:modified	Date of creation or last update of the EO Metadata Document	
available	DateRange	0..1	eop:availabilityTime	Date range during which the resource will be available (RFC-3339)	
acquisitionInformation	[{}]	0..1	eop:AcquisitionInformation (Class)	Acquisition Information	
platform	{}	0..1	eop:platform	The Platform used for the acquisition	
id	String (URI)	0..1		URI of platform.	(e.g., http://gcmdservices.gsfc.nasa.gov/k
platformShortName	String	1	eop:platformShortName	Platform short name	(e.g., "Seasat" or "ENVISAT")
platformSerialIdentifier	String	0..1	eop:platformSerialIdentifier	Platform serial identifier	(e.g. 1: for Seasat)
orbitType	String	0..1	eop:orbitType	High level characterisation of ma	GEO LEO
instrument	{}	0..1	eop:instrument	The Instrument/Sensor used for the acquisition	
id	String (URI)	0..1		URI of instrument	(e.g., http://gcmdservices.gsfc.nasa.gov/k
sensorType	String	0..1	eop:sensorType	Sensor type based on codelist	OPTICAL RADAR ALTIMETRIC ATMOSPHE
instrumentShortName	String	1	eop:instrumentShortName	Instrument (Sensor) name	
description	String	0..1	dct:description	Instrument description	
acquisitionParameters	{}	0..1	eop:AcquisitionParameters (Class)	The acquisition parameters (i.e., pointing angles, etc.)	
beginningDateTime	Date	1	ical:dtstart	Acquisition start date time	

Figure 2: Extract of the table depicting the OGC17-003 OSEO Attributes-Specification

This table provides an overview of commonly used properties (note the cardinalities show that most properties are optional). We use this for our own catalogue implementations, to ensure these are in-line with the harmonized model.

## 2.2. STAC

The SpatioTemporal Asset Catalog (STAC) specification is a “common language to describe geospatial information” [15]. The development started in 2017 and focuses on providing access not only to metadata but to the data itself. Thus, main objectives of STAC are crawlable indexes, standardized description of data based on feature objects, data discovery and a simplification of conducting data processing and analysis. The specification is divided into four elements: STAC Item representing one or more geospatial assets (files), STAC collection to bind STAC items together to a dataset series, STAC catalog to bind collections and items together, and STAC API to provide a RESTful endpoint for data discovery. Various software libraries and tools have already been developed around STAC allowing for standardized creation of STAC metadata (e.g., pystac, stactools, rio-stac) as well as tools for data discovery (e.g., stac-fastapi, stac-server, stac-browser), visualization (e.g., TiTiler), and analysis (e.g., StackSTAC, STAC front-end for Open Datacube, ... [20]).

Creating STAC metadata is simple, just run the specific stactool<sup>[18]</sup> on the target data product and voilà, it outputs the JSON file for it. This works for many earth observation product types from the Copernicus Sentinel

missions, over Planet datasets and many other datasets with a large community support [19]. But here it ends:

- the extracted STAC metadata has been implemented to suit specific needs, but covers only a fraction of metadata available for a product, e.g. absolute orbit number is not included in the produced sentinel2 metadata
- new products require coding and maintaining their own stac-tool, which can be a challenge if you have a variety of products to manage
- searching for exhaustive STAC examples is often fruitless, in many cases you only get excerpts of what is possible
- some of the examples stem from STACs early days
- the available stack utilities do not work with the original packed data, requiring extracted ZIP files in a directory

Having an overview of the full EOP or STAC metadata specification makes it easier to identify historically grown inconsistencies. The tabularized metadata specification allows cross-checking the available metadata for any product source and enables creating mapping tables to generate consistent and harmonized STAC metadata output.

## 3. METADATA MAPPING AND EXTRACTION TOOL

We completed the table in Figure 2 with the corresponding metadata specification taken from the multiple STAC specification sources. We use this table to enable metadata selection for specific EO datasets and to create a CSV file defining the mapping and extraction specification with XPath statements, e.g.:

metadata	file	mapping	datatype
<b>gml:beginPosition</b>	MTD_MSIL[12][ABC].xml	//PRODUCT_START_TIME/text()	DateTimeOffset
<b>gml:endPosition</b>	MTD_MSIL[12][ABC].xml	//PRODUCT_STOP_TIME/text()	DateTimeOffset
<b>eop:processingCenter</b>	MTD_MSIL[12][ABC].xml	//PROCESSING_CENTER/text()	String
<b>eop:platformShortName</b>	MTD_MSIL[12][ABC].xml	substring(//SPACECRAFT_NAME/text(),1,10)	String
<b>eop:orbitNumber</b>	manifest.safe	//safe:orbitNumber[@type="start"]/text()	Int64
<b>eop:sensorType</b>	static	OPTICAL	String
...	...	...	...

Figure 3: Extract of a metadata extraction table for Sentinel-2 products

Expressing the metadata extraction in a table, with rules in XPath syntax is very flexible, powerful, compact and a clean representation of the desired information. It can be easily placed side-by-side to the our OGC 17-003 table for completeness checks and allows comparison to other datasets.

DLR implemented a generic python tool and API [1] that:

- reads metadata from files in multiple formats like zip, gzip, tar, [SAFE](#), [NetCDF](#), or directory tree;
- extracts product metadata from one or more files;

- uses XPath expressions for the metadata extraction;
- adds XPath functions extensions for common text manipulations: join, regex, map, uppercase, date;
- creates any desired output format like STAC, EOP XML, JSON<sup>[23]</sup> using a Jinja<sup>[22]</sup> template;
- allows using the tool as a Python API to embed it in other applications.

The DLR extraction tool has been tested for over 450 of the Copernicus Sentinel-1, -2, -3 and S5P mission products [14] during the qualification phase for the Long-term Archive tender. For this it generated JSON output with OData [26] structure and metadata names.

This DLR tool is capable of creating STAC output identical to the original *stactools* [18] or any other representation using own output templates.

Furthermore, the tool is being evaluated and extended in DLR's Earth Observation Center (EOC) Geoservice [3], [4] and [5] to facilitate the generation and registration of STAC compliant metadata in its EO product catalogue for a variety of data collections (ongoing as well as reprocessed missions). The catalogue is currently in incubation and is based on GeoServer's OpenSearch Plugin [8], providing both STAC as well as OpenSearch [7] endpoints to heterogeneous datasets of DLR's Earth Observation Center [6].

The EOP metadata overview and the generic metadata extract tool is available as open source at [1]. The approach harmonizes the procedure of identifying, specifying and generating valid STAC metadata for existing and future earth observation data products.

#### 4. FUTURE ACTIVITIES

The metadata extract tool has been released in an initial state for the public [1], it will need:

- mappings for further datasets
- code refactoring to be used as a *stac-util* replacement for existing a *stactool* sub-commands
- further elaboration of the documentation
- more test cases.

GitHub also hosts another project "*Mapping from OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard to STAC*" [30], that could be enhanced or aligned with the metadata table and tool described in this paper.

The success of STAC will massively profit from these related activities:

- Interface standardization  
<https://github.com/Open-EO/openeo-api>
- Ordering interface standardization  
<https://github.com/stac-extensions/order>
- Download Service standardization, here there are some existing and emerging standards, like: OGC 10-043<sup>[13]</sup>, [WCS](#), [S3](#), [GeoDataCube](#)
- Authentication and Authorization federations, which is part of the Open-EO API
- Sensor-planning, satellite tasking and future acquisition orders:  
<https://www.element84.com/blog/spatiotemporal-asset-tasking-sprint-2-recap>

#### 5. CONCLUSION

In this paper we revised the evolution of earth observation metadata standards and collected the latest OGC Earth Observation model into one table. This can be used in combination with the STAC mappings as a reference for the extraction metadata for new datasets. We published this EOP metadata reference table and our Copernicus dataset mappings in GitHub together with a Python metadata extraction tool that is configured with these mappings to extract and output STAC conform metadata.

The open source metadata extraction tool is based on simple configuration file utilizing the identified metadata mappings.

The unified EOP metadata reference table simplifies the implementation and adoption of STAC for existing and future EO datasets. It enables improvements of existing and harmonization of different EO metadata catalogues. Users will still need a good insight of the EO data products and the available metadata to design new mappings – having so many satellite missions, this is still work to do.

Future datasets will natively provide STAC conform metadata. With a harmonized metadata model like the OGC EOP standard and the STAC mappings available, will ease these implementations.

We expect that the STAC standard will be used for many years, harmonize EO dataset catalogue offerings, and provides the basis for further compatible extensions. With the widespread availability of STAC catalogues, we expect it being taken up in comfortable browser applications, as well as in powerful automated processing environments.

## 6. REFERENCES

- [1] Earth Observation Metadata Extraction Tool, <https://github.com/dlr-eoc/EOMETADATATool>
- [2] Earth Observation Metadata mapping table, <https://github.com/dlr-eoc/EOMETADATATool/.../OGC17-003-OSEO-Attributes-Specification.md>
- [3] DLR Earth Observation Center (EOC) discovery, visualization, and direct download services, <https://geoservice.dlr.de>
- [4] Heinen, Torsten; Buckl, Bernhard; Giannecchini, Simone; Kiemle, Stephan; Stephan, Meißl (2016) *Evolution of Earth Observation Online Data Access* Proceedings of 2016 conference on Big Data from Space (BiDS'16), Pages 31-34. Publications Office of the European Union. Conference on Big Data from Space - BiDS'16, 2016-02-15 to -17, Santa Cruz de Tenerife, Spain. doi: [10.2788/854791](https://doi.org/10.2788/854791). ISBN 978-92-79-56980-7. ISSN 1831-9424
- [5] Heinen, Torsten; Kiemle, Stephan; Buckl, Bernhard; Mikusch, Eberhard; Loyola, Diego (2009) *The Geospatial Service Infrastructure for DLR's National Remote Sensing Data Library* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume (4), Pages 260-269. IEEE - Institute of Electrical and Electronics Engineers. doi: [10.1109/JSTARS.2010.2040505](https://doi.org/10.1109/JSTARS.2010.2040505). ISSN 1939-1404
- [6] Dengler, Klaus; Heinen, Torsten; Huber, Albert; Molch, Katrin; Mikusch, Eberhard (2013) *The EOC Geoservice: Standardized Access to Earth Observation Data Sets and Value Added Products*. PV2013, 2013-11-04 to -06, Frascati, Italy
- [7] DLR geoservice extensions <https://geoservice.dlr.de/eoc/ogc/stac/v1/> <https://geoservice.dlr.de/eoc/oseo/description>
- [8] GeoServer, open source server for sharing geospatial data, <https://docs.geoserver.org/main/en/user/community/opensearch-geo/>
- [9] OGC 17-003 OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard, OGC Standard, 2020-02-14, <https://docs.ogc.org/is/17-003r2/17-003r2.html>
- [10] OGC 13-026 OGC OpenSearch Extension for Earth Observation, OGC Standard, 2016-12-16, <https://docs.ogc.org/is/13-026r8/13-026r8.html>
- [11] OGC 10-157 OGC Earth Observation Metadata profile of Observations & Measurements, OGC Standard, 2016-06-09 <http://docs.opengeospatial.org/is/10-157r4/10-157r4.html>
- [12] OGC 06-131 OGC CSW EO Products Extension Package for ebRIM, Application Profile, Discussion Paper, revision 6, 2009-05-07 [https://portal.ogc.org/files/?artifact\\_id=17689](https://portal.ogc.org/files/?artifact_id=17689)
- [13] OGC 13-043 Download Service for Earth Observation Products, Best Practice, 2014-01-31 [https://portal.ogc.org/files/?artifact\\_id=55210](https://portal.ogc.org/files/?artifact_id=55210)
- [14] Products of the Copernicus Sentinel missions <https://sentinel.esa.int/web/sentinel/sentinel-data-access/sentinel-products>
- [15] SpatioTemporal Asset Catalogs, <https://stacs-spec.org/>
- [16] STAC specifications, <https://github.com/radiantearth/stac-spec>
- [17] STAC Browser, <https://github.com/radiantearth/stac-browser>
- [18] STAC basic command line interface and API <https://github.com/stac-utils/stactools>
- [19] STAC tools implementation packages, <https://github.com/stactools-packages/stactools-packages.github.io>
- [20] STAC Index of Catalogs, Collections, APIs, Software and Tools, <https://stacindex.org>
- [21] Open Data Cube, <https://www.opendatacube.org/>
- [22] Jinja templating engine <https://jinja.palletsprojects.com/>
- [23] JavaScript Object Notation (JSON) is a lightweight, text-based, language-independent data interchange format. <https://www.rfc-editor.org/rfc/rfc7159>
- [24] Federated EO Gateway (FedEO), <https://ceos.org/ourwork/workinggroups/wgiss/access/fedeo/>
- [25] EOxClient webclient for map applications <https://eox-a.github.io/EOxClient/>
- [26] OData Specification, <https://www.odata.org/>
- [27] Catalogue Interoperability Protocol (CIP) <https://loc.gov/z3950/agency/profiles/cip.html> <https://www.loc.gov/z3950/agency/asn1.html>
- [28] Heterogeneous Missions Accessibility (HMA) collaborative project <https://earth.esa.int/eogateway/activities/gscb-and-ldp/heterogeneous-missions-accessibility-hma> <https://esamultimedia.esa.int/multimedia/publications/TM-21/TM-21.pdf>
- [29] OGC 10-189r2 Cataloguing Earth Observation Products for ebXML Registry Information Model 3.0 based Catalogues, OGC Application Profile, Best Practice, 2012-06-12 [https://portal.ogc.org/files/?artifact\\_id=47409](https://portal.ogc.org/files/?artifact_id=47409)
- [30] Mapping from OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard to STAC, [https://github.com/stac-utils/stac-crosswalks/tree/master/OGC\\_17-003r2](https://github.com/stac-utils/stac-crosswalks/tree/master/OGC_17-003r2)