

# GEOSPATIAL COVERAGES DATA CUBE COMMUNITY PRACTICE

---

COMMUNITY PRACTICE

PUBLISHED

**Version:** 1.0

**Submission Date:** 2020-01-31

**Approval Date:** 2020-08-28

**Publication Date:** 2020-10-14

**Editor:** George Percivall

Notice: This document is not an OGC Standard. This document is an OGC Community Practice and is therefore not an official position of the OGC membership. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an OGC Standard.

Further, an OGC Community Practice should not be referenced as required or mandatory technology in procurements.

## License Agreement

Permission is hereby granted by the Open Geospatial Consortium, ("Licensor"), free of charge and subject to the terms set forth below, to any person obtaining a copy of this Intellectual Property and any associated documentation, to deal in the Intellectual Property without restriction (except as set forth below), including without limitation the rights to implement, use, copy, modify, merge, publish, distribute, and/or sublicense copies of the Intellectual Property, and to permit persons to whom the Intellectual Property is furnished to do so, provided that all copyright notices on the intellectual property are retained intact and that each person to whom the Intellectual Property is furnished agrees to the terms of this Agreement.

If you modify the Intellectual Property, all copies of the modified Intellectual Property must include, in addition to the above copyright notice, a notice that the Intellectual Property includes modifications that have not been approved or adopted by LICENSOR.

THIS LICENSE IS A COPYRIGHT LICENSE ONLY, AND DOES NOT CONVEY ANY RIGHTS UNDER ANY PATENTS THAT MAY BE IN FORCE ANYWHERE IN THE WORLD. THE INTELLECTUAL PROPERTY IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF THIRD PARTY RIGHTS. THE COPYRIGHT HOLDER OR HOLDERS INCLUDED IN THIS NOTICE DO NOT WARRANT THAT THE FUNCTIONS CONTAINED IN THE INTELLECTUAL PROPERTY WILL MEET YOUR REQUIREMENTS OR THAT THE OPERATION OF THE INTELLECTUAL PROPERTY WILL BE UNINTERRUPTED OR ERROR FREE. ANY USE OF THE INTELLECTUAL PROPERTY SHALL BE MADE ENTIRELY AT THE USER'S OWN RISK. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR ANY CONTRIBUTOR OF INTELLECTUAL PROPERTY RIGHTS TO THE INTELLECTUAL PROPERTY BE LIABLE FOR ANY CLAIM, OR ANY DIRECT, SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES, OR ANY DAMAGES WHATSOEVER RESULTING FROM ANY ALLEGED INFRINGEMENT OR ANY LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR UNDER ANY OTHER LEGAL THEORY, ARISING OUT OF OR IN CONNECTION WITH THE IMPLEMENTATION, USE, COMMERCIALIZATION OR PERFORMANCE OF THIS INTELLECTUAL PROPERTY.

This license is effective until terminated. You may terminate it at any time by destroying the Intellectual Property together with all copies in any form. The license will also terminate if you fail to comply with any term or condition of this Agreement. Except as provided in the following sentence, no such termination of this license shall require the termination of any third party end-user sublicense to the Intellectual Property which is in force as of the date of notice of such termination. In addition, should the Intellectual Property, or the operation of the Intellectual Property, infringe, or in LICENSOR's sole opinion be likely to infringe, any patent, copyright, trademark or other right of a third party, you agree that LICENSOR, in its sole discretion, may terminate this license without any compensation or liability to you, your licensees or any other party. You agree upon termination of any kind to destroy or cause to be destroyed the Intellectual Property together with all copies in any form, whether held by you or by any third party.

Except as contained in this notice, the name of LICENSOR or of any other holder of a copyright in all or part of the Intellectual Property shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Intellectual Property without prior written authorization of LICENSOR or such copyright holder. LICENSOR is and shall at all times be the sole entity that may authorize you or any third party to use certification marks, trademarks or other special designations to indicate compliance with any LICENSOR standards or specifications. This Agreement is governed by the laws of the Commonwealth of Massachusetts. The application to this Agreement of the United Nations Convention on Contracts for the International Sale of Goods is hereby expressly excluded. In the event any provision of this Agreement shall be deemed unenforceable, void or invalid, such provision shall be modified so as to make it valid and enforceable, and as so modified the entire Agreement shall remain in full force and effect. No decision, action or inaction by LICENSOR shall be construed to be a waiver of any rights or remedies available to it.

None of the Intellectual Property or underlying information or technology may be downloaded or otherwise exported or reexported in violation of U.S. export laws and regulations. In addition, you are responsible for complying with any local laws in your jurisdiction which may impact your right to import, export or use the Intellectual Property, and you represent that you have complied with any regulations or registration procedures required by applicable law to make this license enforceable.

## Copyright notice

Copyright © 2022 Open Geospatial Consortium  
To obtain additional rights of use, visit <http://www.ogc.org/legal/>

## Note

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

# CONTENTS

---

I.	ABSTRACT .....	vi
II.	KEYWORDS .....	vi
III.	PREFACE .....	vii
IV.	SECURITY CONSIDERATIONS .....	viii
V.	SUBMITTERS .....	viii
1.	SCOPE .....	2
2.	CONFORMANCE .....	4
3.	NORMATIVE REFERENCES .....	6
4.	TERMS AND DEFINITIONS .....	8
5.	CONVENTIONS .....	11
5.1.	Identifiers .....	11
5.2.	Acronyms .....	11
6.	OBJECTIVES FOR GEOSPATIAL DATA CUBES .....	14
7.	INFORMATION MODEL: GEOSPATIAL COVERAGES .....	16
8.	DATA PREPARATION: TUNING AND METADATA .....	21
9.	ANALYTICS: QUERIES AND ACCESS .....	24
10.	INTEROPERABILITY: APIs AND ENCODINGS .....	28
10.1.	Core Access Interoperability .....	28
10.2.	Interoperability for Analysis .....	29
10.3.	Interoperability for Visualization .....	30
11.	FURTHER DEVELOPMENTS .....	33
11.1.	Federation of Data Cubes .....	33
11.2.	DGGS for Data Cubes .....	34
11.3.	Numerical Weather Prediction (NWP) coverages .....	35
11.4.	OGC APIs for geospatial resources .....	36
11.5.	Array Analytics .....	37

ANNEX A (NORMATIVE) CONFORMANCE CLASS ABSTRACT TEST SUITE .....	39
A.1. Conformance class: Core Geospatial Data Cube .....	39
A.2. Conformance class: Analytics Extension .....	42
A.3. Conformance class: Visualization Extension .....	43
ANNEX B (INFORMATIVE) DATA CUBE IMPLEMENTATION EXAMPLES .....	46
B.1. Introduction .....	46
B.2. OpenDataCube (Previously the Australian Geoscience Data Cube) .....	46
B.3. EarthServer-2 project .....	47
B.4. rasdaman .....	48
B.5. EOX Sentinel-2 cloudless .....	49
B.6. Advanced geospatial Data Management platform (ADAM) .....	50
B.7. Euro Data Cube .....	51
B.8. PCI Geomatics Tools for OpenDataCube implementation and data preparation .....	52
ANNEX C (INFORMATIVE) REVISION HISTORY .....	55
BIBLIOGRAPHY .....	57

## LIST OF TABLES

---

Table .....	
-------------	--

## LIST OF FIGURES

---

Figure 1 .....	11
Figure 2 – Example Grids consistent with CIS .....	18
Figure 3 – Data cube trimming (left) and slicing (right) [Baumann 2018] .....	25
Figure 4 – Data Cube Federation envisioned by CEOS [CEOS 2016] .....	33
Figure 5 – Data Cube Federation implemented with rasdaman [EarthServer 2017]: query path from EGU Vienna to ECMWF, subquery forked to NCI (center); query path from EGU Vienna to NCI, subquery forked to ECMWF (right); fusion query result (identical for both queries), displayed in NASA WebWorldWind (left) .....	34
Figure 6 – Visualization of a stack of 2D coverages as a 4D coverage data cube. ....	36
Figure B.1 – Left: Rasdaman architecture [Misev et al 2018]; right: Kaleidoscope of rasdaman-based datacube portals .....	49
Figure B.2 – Euro Data Cube .....	51

# LIST OF RECOMMENDATIONS

---

REQUIREMENTS CLASS 1 .....	16
REQUIREMENTS CLASS 2 .....	29
REQUIREMENTS CLASS 3 .....	30
REQUIREMENT 1 .....	17
REQUIREMENT 2 .....	21
REQUIREMENT 3 .....	21
REQUIREMENT 4 .....	22
REQUIREMENT 5 .....	24
REQUIREMENT 6 .....	24
REQUIREMENT 7 .....	24
REQUIREMENT 8 .....	25
REQUIREMENT 9 .....	26
REQUIREMENT 10 .....	28
REQUIREMENT 11 .....	28
REQUIREMENT 12: DATA CUBE SUPPORT FOR OGC WCS-P OR OGC WPS .....	29
REQUIREMENT 13 .....	30
CONFORMANCE CLASS A.1: CORE .....	39
CONFORMANCE CLASS A.2: ANALYTICS EXTENSION .....	42
CONFORMANCE CLASS A.3: VISUALIZATION EXTENSION .....	43

## ABSTRACT

---

Data cubes for geospatial information provide the means to integrate observations and other types of geospatial data for use in multiple applications through simplified access and efficient analytics. Using the Geospatial Coverages data structure, this Community Practice defines requirements for a geospatial coverages data cube infrastructure and guidelines for enhancements and extensions to the basic core.

## KEYWORDS

---

The following are keywords to be used by search engines and document catalogues.

Data Cubes, Coverages, Spatial Analytics, Geospatial

This Community Practice documents aspects of operational systems that are agreed by the OGC members to be requirements for Geospatial Coverage Data Cubes. This document draws from the operational systems as listed in the references section including EarthServer, ESA, Open Data Cube, rasdaman, EOX, ADAM, and PCI Geomatics platforms.

With this paper and other activities, OGC is contributing to the advancement of Geospatial Data Cubes. A recent US National Geospatial Advisory Committee (NGAC) Task Team made this recommendation:

The NGAC Task Team recommends that OGC be encouraged to continue work on the standards that support the agile and reliable and consistent use of a data cube approach. This would help address this section's question about the efficient synergy between public and private sector use to meet customer/client requirements.

— NGAC 2018

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. The Open Geospatial Consortium shall not be held responsible for identifying any or all such patent rights.

Recipients of this document are requested to submit, with their comments, notification of any relevant patent claims or other intellectual property rights of which they may be aware that might be infringed by any implementation of the standard set forth in this document, and to provide supporting documentation.

## SECURITY CONSIDERATIONS

---

No security considerations have been made for this document.

## SUBMITTERS

---

All questions regarding this submission should be directed to the editor or the submitters:

NAME	AFFILIATION
George Percivall, editor	OGC
David Gavin	Digital Earth Australia
Peter Baumann	Jacobs University and rasdaman
Peter Strobl	EC-JRC
Syed R. Rizvi	ama-inc.com
Andrew Cherry	ama-inc.com
Simone Mantovani	MEEO
Grega Milcinski	Sinergise
Stephan Meissl	EOX
Arnold Hougham	PCI Geomatics

1

# SCOPE

---

## SCOPE

---

This document describes community practices for Geospatial Coverage Data Cubes as implemented by multiple communities and running as operational systems. An objective of this document is to promote coordination and common terminology leading to the possibility to federate data cube services.

This Community Practice represents a specific approach to making geospatial data available using open standards. Other approaches for geospatial data cubes, e.g., different data structures and access methods, will be investigated in OGC based on this Community Practice and other perspectives including promising research and innovative implementations.



2

# CONFORMANCE

---

# CONFORMANCE

---

This Community Practice defines requirements for a core Geospatial Coverages Data Cube and guidelines for enhancing the core. Requirements are placed in the following Conformance Classes:

- Core
- Analytics Extension
- Visualization Extension

Conformance with this Community Practice shall be checked using all the relevant tests specified in Annex A (normative) of this document. The framework, concepts, and methodology for testing, and the criteria to be achieved to claim conformance are specified in the OGC Compliance Testing Policies and Procedures and the OGC Compliance Testing web site<sup>1</sup>.

---

<sup>1</sup>[www.opengeospatial.org/cite](http://www.opengeospatial.org/cite)

3

# NORMATIVE REFERENCES

---

## NORMATIVE REFERENCES

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

OGC: OGC 07-011, *Topic 6 – Schema for coverage geometry and functions*. Open Geospatial Consortium (2007). [https://portal.ogc.org/files/?artifact\\_id=19820](https://portal.ogc.org/files/?artifact_id=19820).

Alexandre Robin: OGC 08-094r1, *OGC® SWE Common Data Model Encoding Standard*. Open Geospatial Consortium (2011). [https://portal.ogc.org/files/?artifact\\_id=41157](https://portal.ogc.org/files/?artifact_id=41157).

Peter Baumann, Eric Hirschorn, Joan Masó: OGC 09-146r6, *OGC Coverage Implementation Schema*. Open Geospatial Consortium (2017). <https://docs.ogc.org/is/09-146r6/09-146r6.html>.

Peter Baumann: OGC 17-089r1, *OGC Web Coverage Service (WCS) 2.1 Interface Standard – Core*. Open Geospatial Consortium (2018). <https://docs.ogc.org/is/17-089r1/17-089r1.html>.

Peter Baumann, Jinsongdi Yu: OGC 08-059r4, *OGC® Web Coverage Service WCS Interface Standard – Processing Extension*. Open Geospatial Consortium (2014). [https://portal.ogc.org/files/?artifact\\_id=54506](https://portal.ogc.org/files/?artifact_id=54506).

Peter Baumann: OGC 08-068r2, *OpenGIS Web Coverage Processing Service (WCPS) Language Interface Standard*. Open Geospatial Consortium (2009). [https://portal.ogc.org/files/?artifact\\_id=32319](https://portal.ogc.org/files/?artifact_id=32319).

Jeff de La Beaujardiere: OGC 06-042, *OpenGIS Web Map Service (WMS) Implementation Specification*. Open Geospatial Consortium (2006). [https://portal.ogc.org/files/?artifact\\_id=14416](https://portal.ogc.org/files/?artifact_id=14416).

Joan Masó, Keith Pomakis and Núria Julià: OGC 07-057r7, *OpenGIS Web Map Tile Service Implementation Standard*. Open Geospatial Consortium (2010). [https://portal.ogc.org/files/?artifact\\_id=35326](https://portal.ogc.org/files/?artifact_id=35326).

4

# TERMS AND DEFINITIONS

---

# TERMS AND DEFINITIONS

---

This document uses the terms defined in [OGC Policy Directive 49](#), which is based on the ISO/IEC Directives, Part 2, Rules for the structure and drafting of International Standards. In particular, the word “shall” (not “must”) is the verb form used to indicate a requirement to be strictly followed to conform to this document and OGC documents do not use the equivalent phrases in the ISO/IEC Directives, Part 2.

This document also uses terms defined in the OGC Standard for Modular specifications ([OGC 08-131r3](#)), also known as the ‘ModSpec’. The definitions of terms such as standard, specification, requirement, and conformance test are provided in the ModSpec.

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

## 4.1. Data Cube

---

Multi-dimensional (“n-D”) array of values. Even though it is called a ‘cube,’ it can be 1-dimensional, 2-dimensional, 3-dimensional, or higher-dimensional. The dimensions may be coordinates or enumerations, e.g., categories.

## 4.2. Geospatial Coverage Data Cube

---

A *Data Cube* (Clause 4.1) established on basis of a Coverage with a specific grid system for geospatial data with at least one dimension of spatial or temporal definition.

**Note 1 to entry:** This definition is particular to the concept defined in this Community Practice.

## 4.3. Geospatial Coverage Data Cube Service

---

Online service that provides access and analytics on *Geospatial Coverage Data Cubes* (Clause 4.2).

## 4.4. Coverage

---

Feature that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain.

**Note 1 to entry:** Coverages include regular and irregular grids, point clouds, general meshes.

[SOURCE: OGC 07-011]

## 4.5. Grid

---

Network composed of two or more sets of curves in which the members of each set intersect the members of the other sets in an algorithmic way.

**Note 1 to entry:** This definition intends to include 1D.

[SOURCE: OGC 07-011]

## 4.6. Grid cell

---

Smallest polygonal area established by the intersecting curves of a grid.



5

# CONVENTIONS

---

# CONVENTIONS

This section provides details and examples for any conventions used in the document. Examples of conventions are symbols, abbreviations, use of XML schema, or special notes regarding how to read the document.

## 5.1. Identifiers

The normative provisions in this specification are denoted by the Uniform Resource Indicator (URI):

<http://www.opengis.net/bp/datacubes/ReqXX>

**Figure 1**

All requirements and conformance tests that appear in this document are denoted by partial URLs which are relative to this base.

## 5.2. Acronyms

API	Application Programming Interface
CEOS	Committee on Earth Observation Satellites
CIS	Coverage Implementation Schema
CRS	Coordinate Reference System
DGGS	Discrete Global Grid Systems
DIAS	Data and Information Access Services
ECMWF	European Centre for Medium-Range Weather Forecasts
ISO	International Organization for Standardization
NGAC	National Geospatial Advisory Committee
OGC	Open Geospatial Consortium,
OWS	OGC Web Service
SQL/MDA	Structured Query Language/Multi-Dimensional Arrays

SWE	Sensor Web Enablement
WCPS	Web Coverage Processing Service
WCS	Web Coverage Service
WMS	Web Map Service
WMTS	Web Map Tile Service
WPS	Web Processing Service

6

# OBJECTIVES FOR GEOSPATIAL DATA CUBES

---

# OBJECTIVES FOR GEOSPATIAL DATA CUBES

This document codifies a set of community practices regarding geospatial coverage data cubes in order to promote coordinated development, proliferation, and federation of data cubes for the purpose of rendering geospatial data more readily available for analysis and decision support.

Key themes in this document are:

- Multidimensional spatial-temporal data cube structure with interoperable services;
- Geospatial Analytics: A digital infrastructure to manage geospatial data and provide access and analytics services on the data; and
- Federation of Data Cubes: Coordination of accessibility and interoperability at multiple, distributed data cubes provided by different organizations using open consensus standards. Users should be able to access any data cube in the federation without knowing the location.

Data Cube related technologies are demonstrating to be an efficient technical mean to achieve the needed bridge from large amounts of observational (e.g., remote sensing) and other geospatial data to the wider geospatial domain. Data Cubes technology can support analytics through:

- Providing harmonized, aggregated access to large amounts of geospatial data by offering a logical view of single, multi-dimensional objects;
- Drilling through long time series of different satellite series more efficiently; and
- Efficiently merging different information sources (remote-sensing and in-situ, authoritative, and non-authoritative).

Historically, geospatial data cubes were introduced some time ago [Baumann 1992]. Recent activity by organizations listed in Annex C provide the support for this Geospatial Coverage Data Cube Community Practice.

7

# INFORMATION MODEL: GEOSPATIAL COVERAGES

---

# INFORMATION MODEL: GEOSPATIAL COVERAGES

To support analytics, a data cube shall use an information model that minimizes artifacts of how the data was created or acquired and support analytics on the information model that are of value to analysts. The Geographic Coverage concept provides such an information model. A Geographic Coverage associates positions within a bounded space (its domain) to attribute values (its range). OGC 07-011.

To be clear, this Community Practice defines three related concepts:

- Data Cube
  - Multi-dimensional (“n-D”) array of values. Even though it is called a ‘cube’, it can be 1-dimensional, 2-dimensional, 3-dimensional, or higher-dimensional.
- Geospatial Coverage Data Cube
  - A Data Cube established on basis of a Coverage with a specific grid system for geospatial data with at least one dimension of spatial or temporal definition.
- Geospatial Data Coverage Cube Service
  - Online service that provides access and analytics on Geospatial Data Cubes.

Although the name “data cube” suggests a 3-dimensional model, the geospatial data cube model is not restricted to that. The spectrum of dimensions includes 1-D sensor data, 2-D imagery, 3-D x/y/t image time-series and x/y/z subsurface voxel data, and 4-D x/y/z/t climate and ocean data cubes.

For the temporal dimension, a temporal coordinate system is required. A temporal coordinate system measures time system based on an interval scale on which distance is measured as a multiple of a single unit of time. Ordinal temporal reference systems and calendars need not be supported.

## REQUIREMENTS CLASS 1

IDENTIFIER	/req/core
TARGET TYPE	Implementation of Logical Model
CONFORMANCE CLASS	Conformance class A.1: /conf/core

## REQUIREMENTS CLASS 1

DESCRIPTION	Core requirements
NORMATIVE STATEMENTS	Requirement 1: /req/core/Req01 Requirement 2: /req/core/Req02 Requirement 3: /req/core/Req03 Requirement 4: /req/core/Req04 Requirement 5: /req/core/Req05 Requirement 6: /req/core/Req06 Requirement 7: /req/core/Req07 Requirement 8: /req/core/Req08 Requirement 9: /req/core/Req09 Requirement 10: /req/core/Req10 Requirement 11: /req/core/Req11

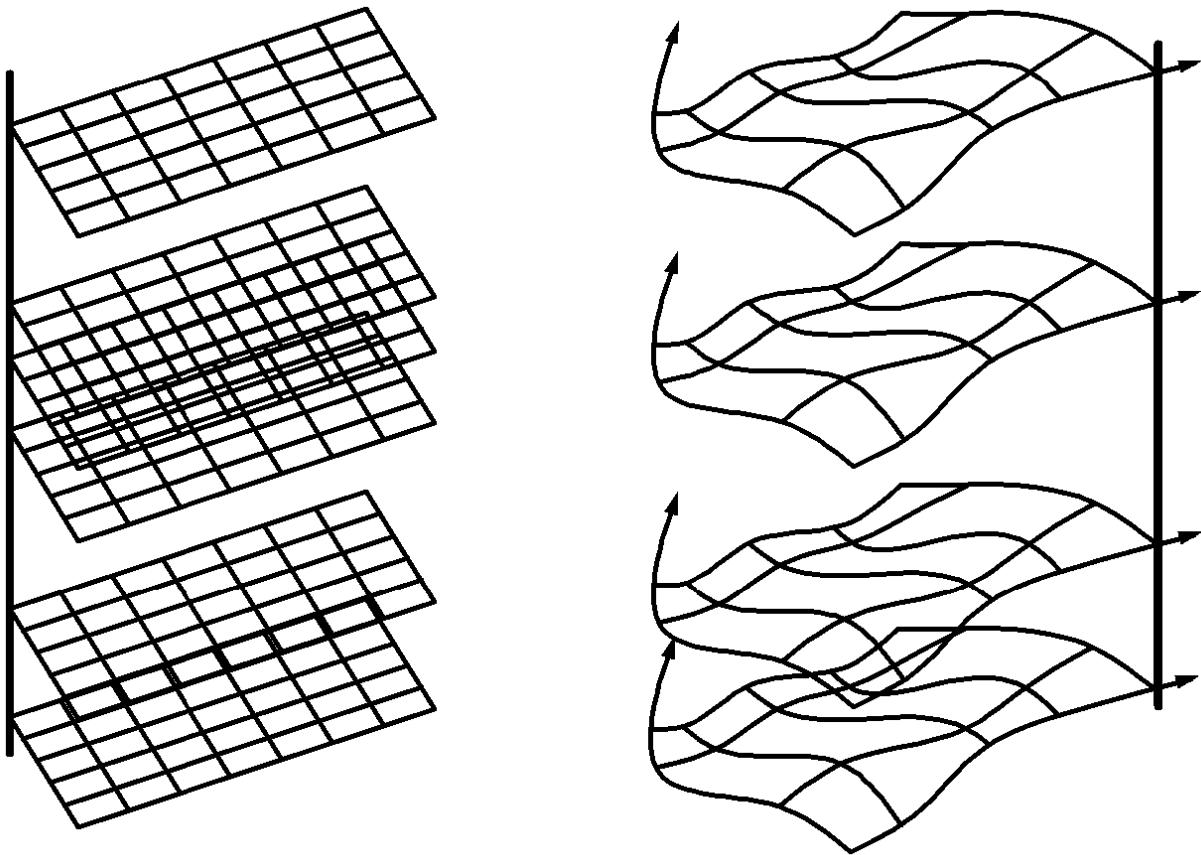
## REQUIREMENT 1

IDENTIFIED /req/core/Req01

INCLUDED Requirements class 1: /req/core  
IN

STATEMENT A Geospatial Data Cube shall use the Geographic Coverage model as defined in OGC Coverage Implementation Schema (CIS).

The Coverage Implementation Schema (CIS) specification [OGC 09-146r6] establishes a concrete, interoperable, conformance-testable Coverage. CIS is based on the abstract concepts of OGC Abstract Topic 6 [OGC 07-011]. CIS is interoperable in the sense that coverages can be conformance tested, regardless of their data format encoding, down to the level of single “pixels” or “voxels.” CIS provides for grids such as shown in Figure 2 including combinations of regular and irregular axis; and showing grids over time as a vertical axis.



**Figure 2 – Example Grids consistent with CIS**

As defined in CIS, the domain set of a Coverage determines the exact locations of a coverage overall and its set of direct positions. The domain set is defined through an ordered list of axes whose lower and upper bounds establish the extent along each axis. The axis sequence and their meaning is defined by the CRS. This domain set CRS is called the coverage's Native CRS. CIS supports all CRSs including Euclidean and spherical, and geographic. CRSs maybe defined by identifier and listing in a registry, e.g., EPSG, or by custom definition of the CRS.

Additionally, some redundant information is present in the domain set for efficiency reasons: the number of dimensions, axis labels, and UoM (Unit of Measure) labels simplify parsing the coverage as the parser does not have to retrieve the CRS definition, such as from the OGC CRS resolver at <http://www.opengis.net/def/crs> and <http://www.opengis.net/def/crs-compound>.

As defined in CIS, the RangeType component adds a structure description and technical metadata required for an appropriate (however, application independent) understanding of a coverage. For this structure description, the SWE Common DataRecord [OGC 08-094r1] is used. Optionally, interpolation directives can be added.

The CIS standard does not require uniform time. CIS Clause 8.2 on Irregular independent grid axes states that time axis may be irregular, e.g., supporting arbitrary image acquisition time.

Coverages can be encoded in any suitable format (such as GML, JSON, GeoTIFF, NetCDF, HDF) and can be partitioned, e.g., for a time-interleaved representation. Coverages are independent from service definitions and, therefore, can be accessed through a variety of OGC services

types, such as the Web Coverage Service (WCS) standards suite (see [Baumann 2018] for an introduction). The coverage structure can serve a wide range of coverage application domains, thereby contributing to harmonization and interoperability between and across these domains.

8

# DATA PREPARATION: TUNING AND METADATA

---

# DATA PREPARATION: TUNING AND METADATA

Data preparation and structuring involves processes that convert less structured data into organized information. These processes and resulting structured information enable more effective search, analytics, and visualization. Data structuring is performed mainly to support analytics (See Clause 9).

Objectives of Data Preparation are summarized in in “Frontiers in Massive Data Analysis” by the US National Academies as:

Although a picture may be worth a thousand words, a good representation of data is priceless: a single data representation, or sometimes multiple ones, allows one to carry out a large number of data processing and analysis tasks in a manner that is both algorithmically efficient and statistically meaningful.

## REQUIREMENT 2

**IDENTIFIED** /req/core/Req02

**INCLUDED IN** Requirements class 1: /req/core

**STATEMENT** A Geospatial Coverage Data Cube Service shall host geospatial data cubes and support services on the data when requested.

Data store optimization is performed when ingesting data into a Data Cube Service. This preparation may include gridding, tiling, compression, tuning, etc. These operations are done in advance of user query and access. This allows administrators to specify any (regular or irregular) spatio-temporal tiling structure, compression, and more settings as tuning parameters; on query level users are not bothered with such internal details, but it may have a substantial impact on performance.

## REQUIREMENT 3

**IDENTIFIED** /req/core/Req03

**INCLUDED IN** Requirements class 1: /req/core

**STATEMENT** When ingesting data, a Geospatial Coverage Data Cube shall provide methods for data storage and access optimizations consistent with the Geospatial Coverage Data Cube Information Model.

For an example of data store ingest optimization, see the rasdaman storage layout language Baumann 2010.

## REQUIREMENT 4

IDENTIFIED /req/core/Req04

INCLUDED Requirements class 1: /req/core  
IN

A Geospatial Coverage Data Cube shall provide metadata for the user that defines the provenance of all source data, data preparation methods applied and data structure of the cube. If applicable, the metadata shall be defined in accordance with ISO 19115.

Metadata are crucial to describe the data that are offered by a data cube. A user needs information to understand the data, for example, data provider, geospatial and temporal resolution, parameter and its unit. The same applies for data requests. Besides the actual data array, a user is also interested in the accompanied axes information for each data value, as this facilitates further data processing and visualization. [Wagemann 2017].

Metadata can be defined in two categories: General and Quality [CARD4L].

- **General Metadata** – Dataset and pixel descriptive information that establish the lineage of the product and provide confidence to the user that it is authoritative, by detailing the steps taken to produce the data in its current state. This includes satellite, instrument, acquisition date and time, spatial boundaries, pixel locations, mode, processing details, spectral or frequency response, and grid projection.
- **Quality Metadata** – Dataset and pixel descriptive information such as quality flags which allow users to make informed decisions about the suitability of the products for a particular use. For example, clouds, cloud shadows, missing data, saturation, and accuracy assessments.

9

# ANALYTICS: QUERIES AND ACCESS

---

# ANALYTICS: QUERIES AND ACCESS

The significant value of data cubes services is the ability to perform analytics that meet user needs. Key to the value is that the user need not have detailed knowledge of the source data structure and observation artifacts. This section defines Geospatial Coverage Data Cube analytics in terms of data access methods and queries.

## REQUIREMENT 5

**IDENTIFIED** /req/core/Req05

**INCLUDED IN** Requirements class 1: /req/core

**STATEMENT** A Geospatial Coverage Data Cube Service shall support analytics on all domain axes alike, irrespective of the nature of the dimension, e.g. spatial or temporal.

In the past, most standards as well as implementations had dedicated methods for spatial extraction and different ones for temporal extraction. Data cubes, on the contrary, offer the single concept of spatial-temporal axis described by ornamenting metadata. After such a unifying step, trimming and slicing (see Figure 3) can follow the same syntax along all axes. Note that this is notwithstanding different units of measure — Latitude might be measured in degrees like 42°30', time in days like 2017-06-06, height in flight levels like FL100.

## REQUIREMENT 6

**IDENTIFIED** /req/core/Req06

**INCLUDED IN** Requirements class 1: /req/core

**STATEMENT** A Geospatial Coverage Data Cube Service shall allow efficient trimming and slicing along any number of axes from a data cube in a single request by the user.

## REQUIREMENT 7

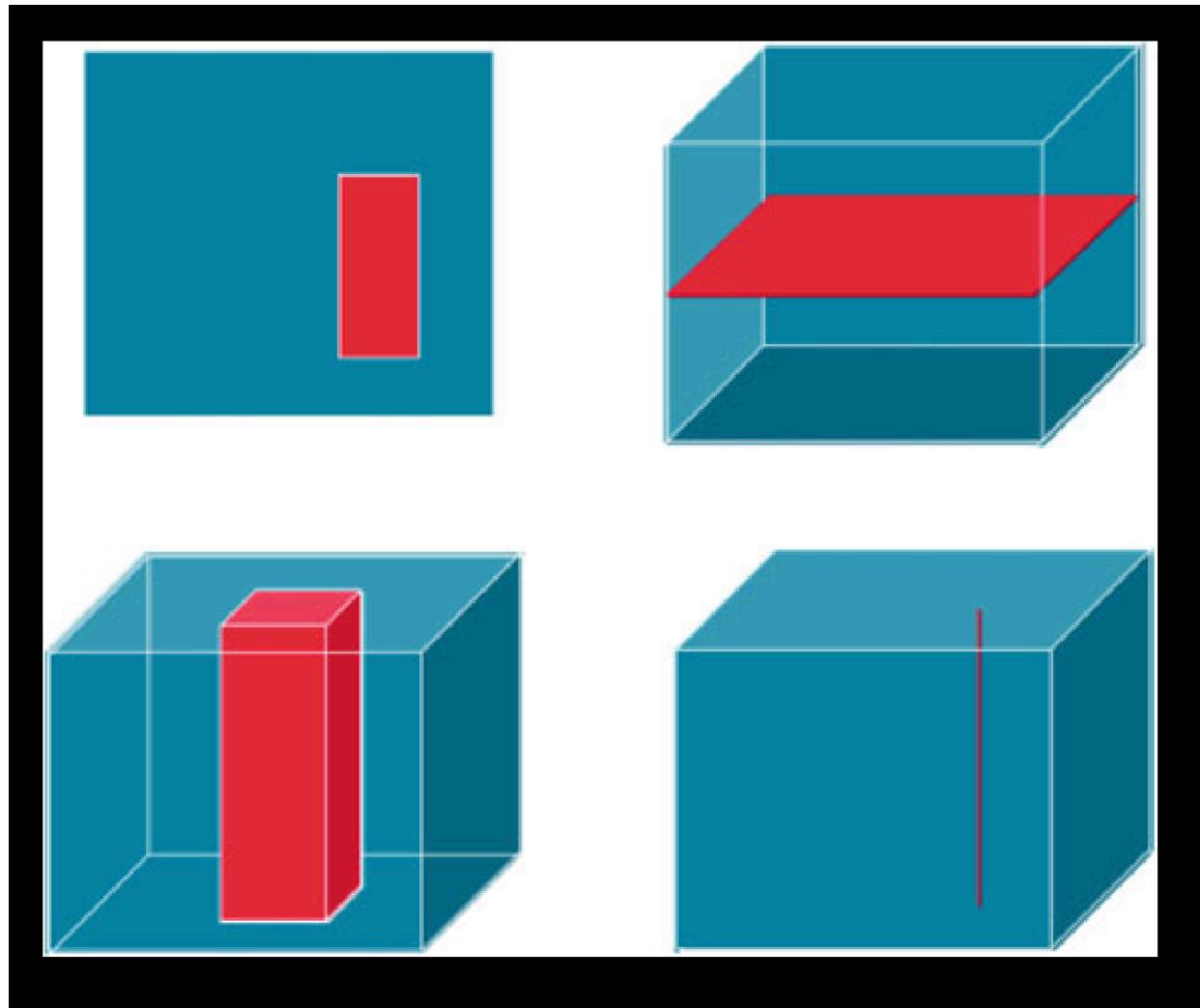
**IDENTIFIED** /req/core/Req07

**INCLUDED IN** Requirements class 1: /req/core

**STATEMENT** A Geospatial Coverage Data Cube Service shall not allow requests that require excessive resources, responding to the user with an estimate of the resource requirements and/or recommendations on how to reduce resources.

Operations on a data cube can be trim or slice operations or other analytics. Users from most Earth Science communities are interested in performing both request types efficiently and fast [Wagemann 2017].

- Trimming: a cut-out of the same dimensionality (e.g., a regional subset of a larger multiband image).
- Slice: a subset of reduced dimensionality (e.g., a single band from a multiband image, or a time series of feature values at a single location).
- Trimming and slicing can be combined in a single request.



**Figure 3 – Data cube trimming (left) and slicing (right) [Baumann 2018]**

## REQUIREMENT 8

IDENTIFIER /req/core/Req08

## REQUIREMENT 8

**INCLUDED  
IN**

Requirements class 1: /req/core

**STATEMENT**

A Geospatial Coverage Data Cube Service shall allow tuning of performance to anticipated user query patterns.

Traditional storage, for example, has been optimized for horizontal access, penalizing temporal and vertical extraction.

Data cubes should support dimension neutrality: meaning that the complexity of a specific processing request (e.g., a query) does not depend on the dimensions involved in the request. Data cubes should expose the dataset in a way that assure dimension neutrality [Stefano 2017].

This document makes no quantitative performance requirements. The emphasis is on functionality. Optimization as a capability shall be supported.

## REQUIREMENT 9

**IDENTIFIED** /req/core/Req09

**INCLUDED  
IN**

Requirements class 1: /req/core

**STATEMENT**

A Geospatial Coverage Data Cube Service shall provide server-side analytic processing on the data including composite extraction, processing, filtering, and fusion tasks in an ad-hoc fashion.

Different Earth Science disciplines require processing operations of different kinds. The climate science community, for example, is interested in generating anomalies and averages of data over a certain period of time. For the vegetation community, the query language defined in Web Coverage Processing Service (WCPS) is very useful to do band calculation, for example, to calculate on-demand the Normalized Difference Vegetation Index (NDVI) of a satellite image [Wagemann 2017]. WCPS is also useful for spatiotemporal statistics, etc.

Data cubes may need to do regridding and reprojection on the fly based on the datasets and the user requirements. Gridding to a common projection and tiling system comes with loss of original product information. As a single grid may not be suitable for the community's purposes, a datacube service may need to perform a regridding and reprojection on the fly.

10

# INTEROPERABILITY: APIs AND ENCODINGS

---

Interoperability and scalable fusion of spatial information across different data cubes is crucial and highly dependent on the use of robust international standards governing the access and transfer protocols for communication between client and server as well as among different servers [Strobl 2017].

For interoperability and federation, data cubes need to provide access and analytics using well-known and open interfaces. To serve the largest communities possible, Application Programming Interfaces (APIs) defined by open consensus standards are needed. The OGC Web Service standards have been implemented by most of the examples listed in Annex B. A Geospatial Coverage Data Cube Service may provide more than one interface.

## 10.1. Core Access Interoperability

### REQUIREMENT 10

IDENTIFIED /req/core/Req10

INCLUDED Requirements class 1: /req/core  
IN

STATEMENT A Geospatial Coverage Data Cube service interface for data access shall be compliant with OGC Web Coverage Service (WCS).

The primary access method to a geospatial data cube is using the interface defined in the OGC Web Coverage Service, Version 2.1 [OGC 17-089r1].

Compliance of the data cube to the WCS standard shall be certified using the OGC Compliance Test suite. <http://www.opengeospatial.org/compliance/getCertified>

Data cubes may choose to define a WCS profile for their community, e.g., WCS EO Profile.

Data and other information returned from a data cube to a client may be encoded in a variety of forms. There are many formats suitable for use in Geospatial Data Cubes. Preferences for a specific format tend to be related to the community to be served. Analysis Ready Data (ARD) data choices are based on the analysis needs agreed by their targeted user community. The Data Cube will need to support formats typically used by the ARD community.

### REQUIREMENT 11

IDENTIFIED /req/core/Req11

## REQUIREMENT 11

INCLUDED IN	Requirements class 1: /req/core
STATEMENT	A Geospatial Coverage Data Cube shall support encodings that are defined in open consensus standards and that are used by the community targeted.

To support the variety of data formats, the WCS standard includes extensions mapping the information model to the format, e.g., GeoTIFF, netCDF, JPEG2000, GMLJP2, JPIP, and GRIB2.

## 10.2. Interoperability for Analysis

### REQUIREMENTS CLASS 2

IDENTIFIER	/req/analytics
TARGET TYPE	Implementation of Logical Model
CONFORMANCE CLASS	Conformance class A.2: /conf/analytics
PREREQUISITE	Requirements class 1: /req/core
DESCRIPTION	Interoperability for analytics
NORMATIVE STATEMENT	Requirement 12: /req/analytics/Req12

### REQUIREMENT 12: DATA CUBE SUPPORT FOR OGC WCS-P OR OGC WPS

IDENTIFIER	/req/analytics/Req12
INCLUDED IN	Requirements class 2: /req/analytics
STATEMENT	A Geospatial Coverage Data Cube service interface for data analytics shall support OGC WCS Processing Extension (WCS-P) and/or OGC Web Processing Service (WPS).

When a geospatial data cube provides user-specified analytics on the data it holds, the data cube may use the WCS-P extension [WCS-P 2014] which extends the WCS using on the WCPS Language Interface Standard [OGC 08-068r2].

To support analytics, a data cube may accept requests for analysis via WCPS as it that expresses the analytics or by sending code to the server. Shipping procedural code to a data cube may be

undesirable when considering user friendliness, parallelizability, and security. Therefore, some high-level language must be supported where users describe what they want to get, not the detailed algorithm [Baumann 2017].

ISO IS 9075-15:2018 SQL/MDA (Multi-Dimensional Arrays) adds datacube modeling and querying to the SQL language in a domain-independent manner.

A Web Processing Service (WPS) may be used to run an analysis tool in the Geospatial Coverage Data Cube Service, and return the results to the user. The WPS processing on the data cube can bring the required information to a user, without having to download entire datasets. For example, WPS can be used to enable on-the-fly time series analysis and derived products. Exposing this functionality through intuitive workflows and reusable, standard interfaces will mean that these services can be consumed and enjoyed by a greater audience.

Analytics using server-side processing supported by Jupyter notebooks may also be used to perform data cube analytics. Further developments are described in Clause 11.5.

## 10.3. Interoperability for Visualization

---

### REQUIREMENTS CLASS 3

IDENTIFIER	/req/visualization
TARGET TYPE	Implementation of Logical Model
CONFORMANCE CLASS	Conformance class A.3: /conf/visualization
PREREQUISITE	Requirements class 1: /req/core
DESCRIPTION	Interoperability for Visualization
NORMATIVE STATEMENT	Requirement 13: /req/visualization/Req13

### REQUIREMENT 13

IDENTIFIER /req/visualization/Req13

INCLUDED IN Requirements class 3: /req/visualization

STATEMENT A Geospatial Coverage Data Cube service interface for maps shall be compliant with OGC Map Service (WMS) or OGC Web Map Tile Service (WMTS)

When a geospatial coverage data cube provides access to maps of the data it holds, the data cube shall use the using the interfaces defined in the OGC Web Map Service [OGC 06-042] or the OGC Web Map Tile Service [OGC 07-057r7].

Data cubes may choose to define a WMS profile for their community, e.g., WMS EO Profile, WMS profile for Time-Dependent or Elevation-Dependent Data.

11

# FURTHER DEVELOPMENTS

---

This section identifies topics that may affect the future development of Geospatial Coverage Data Cube services.

## 11.1. Federation of Data Cubes

Geospatial Coverage Data Cubes can be an important step toward establishing a common analytical framework to link very large, multi-resolution and multi-domain datasets together for analysis. The Geospatial Coverage Data Cube defined in this paper provides the basis for a network of data cubes operating on large geoscientific and geospatial datasets to address global as well as national challenges [Lewis 2017].

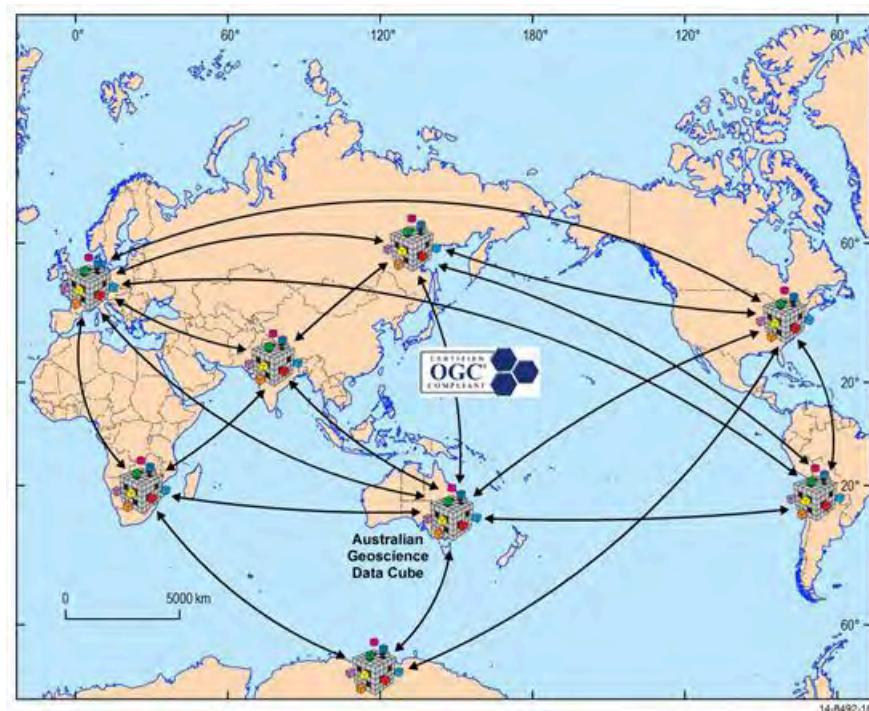


Figure 4 – Data Cube Federation envisioned by CEOS [CEOS 2016]

Data cubes are being established at Global, Regional, and National level, as in particular Regional and National data owners are constrained to publish their data on infrastructures on their territory. At the thematic level, it can be expected that different operators will appear and will want to operate their own instance of a data cube. Also, in this distributed scenario, applications shall be able to combine data layers held in different data cube instances, even if based on different technologies.

Federation of datacubes introduces location transparency of data sets hosted by geographically disparate data providers, as well as their processing. This offers several distinct advantages for users.

- Enhanced offering: Users can access the sum of all individual offerings of the participating data providers.
- Location transparency: When accessing a dataset, users can send the request to any federation member – users do not need to know the location of a particular data set. In other words, the federation offers a common, flat information space.
- Remote fusion: When combining two or more data sets in a request (such as a WCPS query) the federation partners determine an efficient way of solving this request and execute the request accordingly, without any extra client intervention. In other words, users get returned only the result requested, without any need for own source data download, reformatting, and processing.

Such federations have been implemented on rasdaman datacubes, demonstrated in EarthServer-2 [EarthServer 2017] on Petascale through intercontinental transparent federation connecting ECMWF/UK and NCI/Australia. Further, the German Sentinel hub, CODE-DE, is federated with an ESA DIAS and continuously being extended with further federation partners, including commercial offerings.



**Figure 5 – Data Cube Federation implemented with rasdaman [EarthServer 2017]: query path from EGU Vienna to ECMWF, subquery forked to NCI (center); query path from EGU Vienna to NCI, subquery forked to ECMWF (right); fusion query result (identical for both queries), displayed in NASA WebWorldWind (left)**

## 11.2. DGGS for Data Cubes

“Datacubes: A Discrete Global Grid Systems Perspective” [Purss 2019] discusses options and how it is possible to implement a data cube based on discrete global grid systems, while using the same topologies as conventional data cubes. These provide a flexible spatial data

infrastructure that leverages the same topological advantages as conventional geospatial data cubes, while reducing barriers to data interoperability of both raster and vector data and providing additional functionality. Also, they potentially provide a very efficient approach to connecting to big data sources in order to extract datasets on demand prior to proceeding to multi-level intelligent big data processing, mining, machine learning, and visualizations.

The OGC DGGS Abstract Specification Topic 21 [DGGS 2017] defines a DGGS as a spatial reference system that uses a hierarchical tessellation of cells to partition and address the globe. DGGS are characterized by the properties of their cell structure, geoencoding, quantization strategy, and associated mathematical functions. The OGC DGGS Abstract Specification supports the specification of standardized DGGS infrastructures that enable the integrated analysis of very large, multi-source, multiresolution, multi-dimensional, distributed geospatial data.

Data cubes can benefit from an underlying DGGS compliant data tiling and integration scheme as defined by the OGC DGGS specification. A data cube based on a DGGS enforces a more comprehensive set of spatial boundaries to the problem of global spatial data representation and integration. Further developments of DGGS aim to enable data cubes to be federated at a truly global scale, without the spatial limitations and constraints imposed by heterogenous data cubes.

### 11.3. Numerical Weather Prediction (NWP) coverages

---

Coverages as defined in CIS1.1 can describe the atmosphere (or ocean) as a multidimensional 4D cube that contains parameters, e.g., temperature, wind, humidity (salinity), etc. Coverages are defined by (axis) dimensions that can contain one or more of these meteorological/oceanographic parameters. A typical Numerical Weather Prediction (NWP) forecast simulation may be expressed as a set of 2D coverages, often, but not always based on rectified grids. A typical model run contains thousands of 2D files. This situation is managed by identifying, where possible, “4D Coverages” from many 2D files, as shown in Figure 6. [Met/Ocean 2018]

## Stack of 2D Coverages

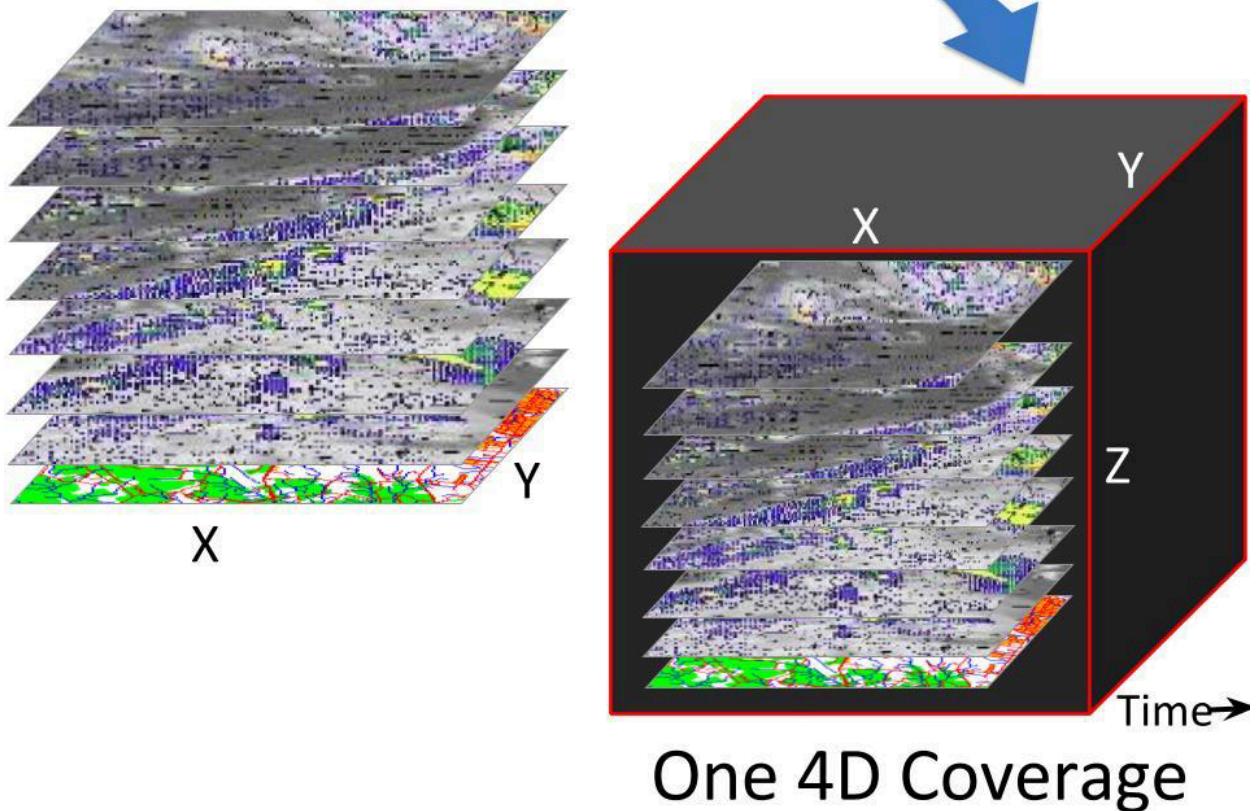


Figure 6 – Visualization of a stack of 2D coverages as a 4D coverage data cube.

### 11.4. OGC APIs for geospatial resources

OGC is defining web-friendly APIs that can be used to access geospatial and location information in any domain. This major evolution in the OGC Standards Baseline is underway with the development of several standards:

- OGC API-Features
- OGC API-Coverages
- OGC API-Maps
- OGC API-Tiles
- OGC API-Processes
- OGC API-Common

The OGC API standards define modular API building blocks to spatially enable Web APIs and be web-friendly for non-geospatial experts. The APIs are designed to allow developers to reuse and extend the APIs for any domain. The first consensus adoption of the OGC API-Features Core was published on 2019-10-14.

The OGC APIs are being developed using OpenAPI, GitHub, and Hackathons. OpenAPI is a standard, programming-language-agnostic interface description for Web APIs which allows both humans and computers to discover and understand the capabilities of a service. OGC uses GitHub as a collaborative platform for specification development and for OpenAPI artifacts. Hackathons are conducted to implement and test draft API specifications. Along with the Specifications, OGC members are developing Reference Implementations, Compliance Testing, e-Learning, and Outreach materials.

Once the OGC APIs specifications have been further developed, their relevance to the Geospatial Data Cubes can be better evaluated.

The SpatioTemporal Asset Catalog (STAC) specification is being developed in coordination with the OGC APIs. The STAC specification aims to standardize the way geospatial assets are exposed online and queried. A 'spatiotemporal asset' is any file that represents information about the earth captured in a certain space and time. The initial focus is primarily remotely-sensed optical imagery (from satellites, but also planes, drones, balloons, etc.), but the core is designed to be extensible to SAR, full motion video, point clouds, hyperspectral, LiDAR and derived data like NDVI, Digital Elevation Models, mosaics, etc.

## 11.5. Array Analytics

---

Several approaches to APIs for analytics are discussed in Clause 10.2. OGC held a Coverage and Processing Analytics Sprint in January 2020 to investigate the further development of technologies for analytic interoperability on data cubes, such as:

- Jupyter notebooks and Python scripting language
- Array Databases, e.g., ISO IS 9075-15 SQL/MDA (Multi-Dimensional Arrays)
- xarray, GeoXarray, ZARR
- application of R scripting language for statistics to data cubes, e.g. [Appel 2019]
- OGC Web Coverage Processing Service (WCPS) Language Interface Standard [OGC 08-068r2]
- OGC API – Coverages



A

# ANNEX A (NORMATIVE) CONFORMANCE CLASS ABSTRACT TEST SUITE

---

# ANNEX A (NORMATIVE)

## CONFORMANCE CLASS ABSTRACT TEST SUITE

Compliance of the data cube to OGC Web Service Standards shall be certified using the OGC Compliance Test suite. <http://www.opengeospatial.org/compliance/getCertified>

### A.1. Conformance class: Core Geospatial Data Cube

#### CONFORMANCE CLASS A.1: CORE

IDENTIFIER	/conf/core
REQUIREMENTS CLASS	Requirements class 1: /req/core
DESCRIPTION	Implementations claiming Core Conformance shall be conformant with the requirements listed in this clause.
CONFORMANCE TESTS	Conformance test A.1: /conf/core/Req01 Conformance test A.2: /conf/core/Req02 Conformance test A.3: /conf/core/Req03 Conformance test A.4: /conf/core/Req04 Conformance test A.5: /conf/core/Req05 Conformance test A.6: /conf/core/Req06 Conformance test A.7: /conf/core/Req07 Conformance test A.8: /conf/core/Req08 Conformance test A.9: /conf/core/Req09 Conformance test A.10: /conf/core/Req10 Conformance test A.11: /conf/core/Req11

#### CONFORMANCE TEST A.1

IDENTIFIER	/conf/core/Req01
------------	------------------

## CONFORMANCE TEST A.1

**REQUIREMENT** Requirement 1: /req/core/Req01

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.2

**IDENTIFIER** /conf/core/Req02

**REQUIREMENT** Requirement 2: /req/core/Req02

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.3

**IDENTIFIER** /conf/core/Req03

**REQUIREMENT** Requirement 3: /req/core/Req03

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.4

**IDENTIFIER** /conf/core/Req04

**REQUIREMENT** Requirement 4: /req/core/Req04

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.5

**IDENTIFIER** /conf/core/Req05

**REQUIREMENT** Requirement 5: /req/core/Req05

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.6

**IDENTIFIER** /conf/core/Req06

**REQUIREMENT** Requirement 6: /req/core/Req06

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.7

**IDENTIFIER** /conf/core/Req07

**REQUIREMENT** Requirement 7: /req/core/Req07

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.8

**IDENTIFIER** /conf/core/Req08

**REQUIREMENT** Requirement 8: /req/core/Req08

**INCLUDED IN** Conformance class A.1: /conf/core

**DESCRIPTION** TODO

## CONFORMANCE TEST A.9

IDENTIFIER /conf/core/Req09

REQUIREMENT Requirement 9: /req/core/Req09

INCLUDED IN Conformance class A.1: /conf/core

DESCRIPTION TODO

## CONFORMANCE TEST A.10

IDENTIFIER /conf/core/Req10

REQUIREMENT Requirement 10: /req/core/Req10

INCLUDED IN Conformance class A.1: /conf/core

DESCRIPTION TODO

## CONFORMANCE TEST A.11

IDENTIFIER /conf/core/Req11

REQUIREMENT Requirement 11: /req/core/Req11

INCLUDED IN Conformance class A.1: /conf/core

DESCRIPTION TODO

## A.2. Conformance class: Analytics Extension

### CONFORMANCE CLASS A.2: ANALYTICS EXTENSION

IDENTIFIER /conf/analytics

## CONFORMANCE CLASS A.2: ANALYTICS EXTENSION

REQUIREMENTS CLASS	Requirements class 2: /req/analytics
PREREQUISITE	Conformance class A.1: /conf/core
DESCRIPTION	Implementations claiming Analytics Extension Conformance shall be conformant with the requirements listed in the Core Conformance and in this clause.
CONFORMANCE TEST	Conformance test A.12: /conf/analytics/Req12

## CONFORMANCE TEST A.12

IDENTIFIER	/conf/analytics/Req12
REQUIREMENT	Requirement 12: /req/analytics/Req12
INCLUDED IN	Conformance class A.2: /conf/analytics
DESCRIPTION	TODO

## A.3. Conformance class: Visualization Extension

---

## CONFORMANCE CLASS A.3: VISUALIZATION EXTENSION

IDENTIFIER	/conf/visualization
REQUIREMENTS CLASS	Requirements class 3: /req/visualization
PREREQUISITE	Conformance class A.1: /conf/core
DESCRIPTION	Implementations claiming Visualization Extension Conformance shall be conformant with the requirements listed in the Core Conformance and in this clause.
CONFORMANCE TEST	Conformance test A.13: /conf/visualization/Req13

## CONFORMANCE TEST A.13

**IDENTIFIER** /conf/visualization/Req13

**REQUIREMENT** Requirement 13: /req/visualization/Req13

**INCLUDED IN** Conformance class A.3: /conf/visualization

**DESCRIPTION** TODO



B

# ANNEX B (INFORMATIVE) DATA CUBE IMPLEMENTATION EXAMPLES

# ANNEX B (INFORMATIVE) DATA CUBE IMPLEMENTATION EXAMPLES

---

## B.1. Introduction

---

This Annex includes non-normative examples of data cube implementations that adhere to most of the items in the body of this document. The implementations are listed in an order that balances historical development and listing data cubes that meet the body of the text higher in the order.

## B.2. OpenDataCube (Previously the Australian Geoscience Data Cube)

---

The Open Data Cube provides an integrated gridded data analysis environment for decades of analysis ready earth observation satellite and related data from multiple satellite and other acquisition systems [CEOS 2018].

The Open Data Cube is a system designed to:

- Catalogue large amounts of Earth Observation data;
- Provide a Python based API for high performance querying and data access;
- Give scientists and other users easy ability to perform Exploratory Data Analysis;
- Allow scalable continent scale processing of the stored data; and
- Track the provenance of all the contained data to allow for quality control and updates.

The Open Data Cube provides these APIs and community developed extensions:

- Loading Data from a Cube
- Searching and Indexing Datasets

- Model
- Bit Masking
- Analytics and Execution Engines
- Array IO on Amazon S3
- WMS and WCS
- WPS
- QGIS integration
- A UI for visualising its contents
- A UI for executing queries
- For **Exploratory Data Analysis** see Datacube Class for more details
- For **Writing Large Scale Workflows** see GridWorkflow Class for more details

## B.3. EarthServer-2 project

---

The EU FP7 EarthServer and EU H2020-funded EarthServer-2 [EarthServer 2017] projects have set up WCSs for different scientific domains: ocean science, earth observation, climate science and planetary science. The architectural setup follows the web services concept and bases on rasdaman (<http://rasdaman.org/>), an intelligent arraybased server technology, in combination with the OGC standard protocol WCS. Four WCSs are exploring the possibilities and challenges of providing access to data sizes beyond 1 PB of 3D to 4D Earth Science Data [Wagemann 2017] .

- Marine Science Data Service (<http://earthserver.pml.ac.uk/>) is developed by the Plymouth Marine Laboratory (PML) and makes available multiple hundred terabytes of ocean color data, produced by ESA's climate change initiative (CCI). The service will further offer Sentinel-3 data once the data products are disseminated.
- Climate Science Data Service was developed by ECMWF to demonstrate the ability to provide access to multiple petabytes of ERA-Interim reanalysis data (Dee et al. 2011) stored in ECMWF's archive. This was achieved by serving data from the Meteorological and Archival System (MARS archive) via a web service in a standardized way. The service was discontinued after the end of the project.
- Earth Observation (EO) Data Service (<http://eodataservice.org/>) is jointly developed by MEOO s.r.l. and the National Computing Infrastructure (NCI) Australia. The service will offer EO data from the Sentinel family, primarily data from Sentinel-2 satellites. NCI will eventually offer a WCS for multiple hundreds of terabytes of Landsat data.

- Planetary Science Data Service (<http://planetserver.eu/>) is developed by the Jacobs University Bremen and will provide access to tens of TB of data (topographic, multi- and hyperspectral) of the solar system bodies Mars, Mercury and Moon.

Since 2012, the intercontinental EarthServer initiative (<http://www.earthserver.eu>) is establishing agile data cube analytics on 3D x/y/t image timeseries and 4D x/y/z/t weather data, based on the rasdaman Array Database System (<http://www.rasdaman.org>). The largest installation, EO Data Service ([www.eodataservice.org](http://www.eodataservice.org)), recently has passed the 1 Petabyte [Strobl 2017].

Data cube assets of 2.5 PB have been federated intercontinentally, and single queries have been parallelized across more than 1,000 cloud nodes. As such, the EarthServer initiative is one of the large-scale proofs of practicality for WCS and, specifically, WCPS.

One success or activity, establishing a private/public datacube partnership, is the German funded project BigDataCube [Misev et al 2018]. Aside from routinely using the rasdaman federation capabilities it puts particular emphasis on versatile access control, down to the level of single pixels.

## B.4. rasdaman

---

Following its invention of data cube services [Baumann 1992], the rasdaman team since then has architected the rasdaman scalable datacube engine for cross-domain management and server-side processing of massive n-D arrays. The rasdaman declarative datacube analytics language allows any query, any time, on any size, ready for extraction, filtering, processing, analytics, distributed datacube fusion, and Artificial Intelligence. Through the OGC WMS, WCS, and WCPS standards rasdaman supports regular and irregular spatio-temporal grids. The open-source rasdaman engine is OGC WCS reference implementation, and the rasdaman query language has been standardized in the ISO SQL array extension. As the currently only tool rasdaman supports transparent federation on planetary scale. Individual nodes retain autonomy, with security configurable down to single pixel level.

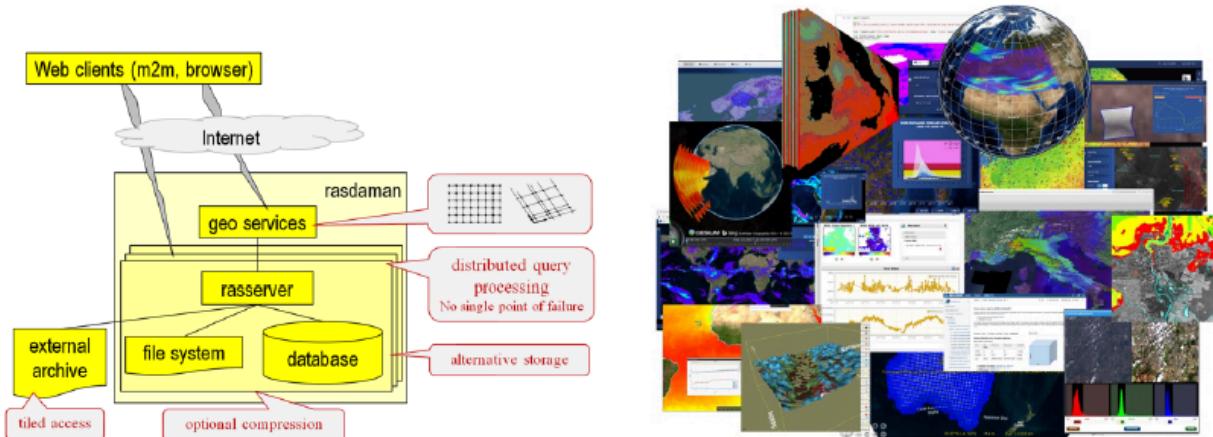
The rasdaman architecture is a full-stack implementation in fast C++, with every component optimized for arrays. A series of highly effective techniques boosts array query execution, including: adaptive partitioning, just-in-time compilation into machine code, heterogeneous CPU/GPU hardware support, intelligent query optimizations, parallelization and distribution (without single point of failure), and federated processing. Queries have been split across more than 1,000 Amazon cloud nodes.

Storage utilizes transparent tiling into sub-arrays maintained in some database (like PostgreSQL) or directly in the file system (about 2x faster), always using rasdaman's optimized storage methods. Any tiling can be configured by the admin, as one of the many tuning parameters in an easy-to-use, OGC based ETL tool. Additionally, rasdaman can tap into any pre-existing archive by just registering files rather than copying them.

With Athena Research's FeMME extension, rasdaman allows integrated data and metadata search, thereby establishing a seamless integration of catalogs with datacubes. Users always remain in the comfort zone of their well-known tools – rasdaman supports a large and

growing number of visual and API clients, including OpenLayers, Leaflet, QGIS, ArcGIS, NASA WebWorldWind, Cesium, C++, Java, python, and R.

User-friendly, flexible, fast, scalable, and standards-conformant datacube services become a commodity with rasdaman: providers enjoy simple deployment and maintenance, users enjoy the massively simplified access, visualization, and analysis capabilities not requiring any programming skills. A series of innovation awards, most recently the NATO Defence Innovation Award 2018, acknowledges the rasdaman innovation.



**Figure B.1** – Left: Rasdaman architecture [Misev et al 2018];  
right: Kaleidoscope of rasdaman-based datacube portals

## B.5. EOx Sentinel-2 cloudless

Sentinel-2 cloudless (<https://s2maps.eu>) is a yearly series of global sunny homogeneous 10m satellite maps. It combines trillions of pixels collected during differing weather conditions over a whole year and merges them into a mosaic, almost free from atmospheric impacts. The mosaics are provided via Web Map Tile Service (WMTS) as well as Web Map Service (WMS) under a Creative Commons license (2016 CC-BY; newer CC-BY-NC-SA). A WCS interface is provided.

Sentinel-2 cloudless is the global visualization product of the EOxCloudless product family (<https://cloudless.eox.at>) that provides off-the-shelf or custom tailored mosaics as Data Cubes. In its simplest form the mosaics are provided as numerous GeoTIFFs in a pyramid structure following the WMTS tilematrixset definition (see this blog post for details <https://eox.at/2018/01/visualizing-geotiff-tiles-with-openlayers/>). Off-the-shelf mosaics contain 4 bands (8- or 16-bits) but custom mosaics may include any number of bands as well as potentially a metadata band allowing tracing back to individual input satellite images on a per-pixel basis. These GeoTIFFs can be loaded in a simple HTTP server or a more complex server software like the Open Data Cube. The Open Source library geotiff.js (<https://geotiffjs.github.io>) for example allows exploration and analysis of GeoTIFFs directly in the browser like for example shown in the COG-Explorer (<https://geotiffjs.github.io/cog-explorer>) demonstration.



## B.6. Advanced geospatial Data Management platform (ADAM)

---

The Advanced geospatial Data Management platform (ADAM) is a platform that implements the Digital Earth concept: ADAM allows accessing to large variety of multiyear global geospatial collections allowing data discovery, visualization, combination, processing and download. ADAM allows exploiting data from global to local scale.

The core of ADAM is a Data Access System (DAS), a software module that manages a large variety of geospatial information featuring different data formats, geographic / geometric and time resolution. The DAS exposes OGC standardized Open Search and Web Coverage Service (WCS 2.x) interfaces that allow discovering available collections and subset them in any dimension with a single query. Thus, ADAM provides an effective subsetting functionality that accesses the data only when requested and serves to the client only the data amount that is really needed.

ADAM is a very modular platform: various DAS are deployed on different data sources (e.g. different DIAS, AWS, MEO Data Facility, etc.), allowing accessing and subsetting the available collections without downloading / duplicating the data.

ADAM, referred also as a virtual datacube technology, enables seamless access to all available data collections through a single interface; cutting the access barriers to a large variety of data (no need to know the exact format of any of the available products) it saves data preparation time and facilitates the simultaneous use of data coming from different sources.

ADAM exposes three user interfaces to satisfy the needs (and match with the skills) of a large variety of users:

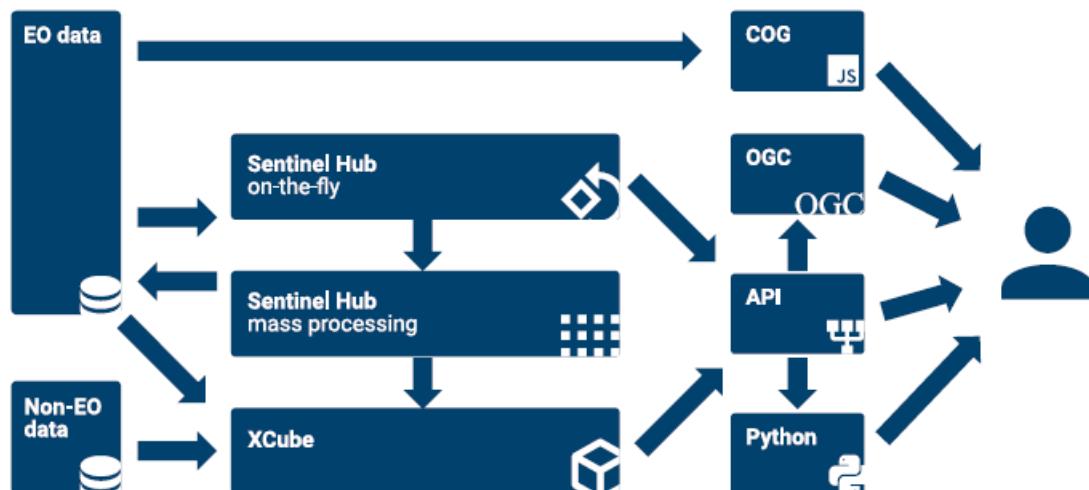
- A web based graphic user interface (Explorer) and a mobile application (ADAM Mobile);
- A web-based notebook (Jupyter Notebook); and
- OGC standardized REST interfaces: OpenSearch for data discovery, WCS 2.x for data access).

## B.7. Euro Data Cube

---

The Euro Data Cube (EDC, <https://eurodatacube.com>) is the cube service for global Earth Observation (EO) data exploitation. The objective of EDC is to provide users with functionality to query and exploit large volumes of EO data and information resources, with a scheme that promotes open data sharing and reselling. The EDC ensures the storage and online availability of data in the form of multi-dimensional arrays suitable for retrieving, analyzing, and correlating EO information resources, based on innovative interfaces and interoperability standards.

Figure B.2 shows a high-level architecture of the EDC offerings highlighting the data and information flow from the EO and Non-EO data offerings on the left over the data cube building blocks and service interfaces to the user on the right.



**Figure B.2 – Euro Data Cube**

The top arrow shows EDC's Client Data Processing Engine (CDPE), supporting clientside processing of Cloud-optimized GeoTiffs (COGs) directly in the web browser.

The on-the-fly data cube access service as well as the asynchronous mass processing service are powered by Sentinel Hub (<http://www.sentinel-hub.com>).

Sentinel Hub is a satellite imagery processing service, which is capable of on-the-fly gridding, re-projection, mosaicking, compositing and other actions required for efficient fetching of data for end-users, either for integration in web-applications, where pictures are mostly served, or in machine learning and similar analysis processes, where pixel values and statistics are essential. The service works with original satellite data files and does not require replication or pre-processing. It uses cloud infrastructure like CreoDIAS, Mundi Web Services, ONDA, or AWS and innovative methods to efficiently process and distribute data in a matter of seconds. It provides catalog service, data processing and rendering service, custom scripting engine including open-source custom scripts library on top of Sentinel-1 GRD, Sentinel-2 (L1C and L2A), Sentinel-3 (OLCI & SLSTR), Sentinel-5P, Landsat-5, -7, -8, Envisat MERIS, and MODIS data as well as Planet and SPOT data on demand.

Versatile pre-generated data cubes are provided by the Open Source Python package `xcube`. It is based on `xarray` just like the CEOS Open Data Cube, Pangeo, and other leading data cube initiatives.

The EDC provides the integrated power of Sentinel Hub and `xcube` via one API in an enhanced end-to-end service. The services are provided via an API following the OpenAPI approach (process, statistics) as well as OGC standards to the extent feasible (mainly WMS, WMTS, WCS, WPS).

The EDC aims to contribute to adoption and further evolution of open interfaces particularly of OGC standards, mainly the Web Coverage Service (WCS) but also the forthcoming OGC API Coverages. The EDC further intends to be interoperable with the CEOS supported “Open Data Cube” Open Source initiative by contributing to the OGC web services interoperability layer.

The Euro Data Cube is a joint service offering by Sinergise, Brockmann Consult, EOX, Gisat, Planet, and Sentinel Hub VAS. ESA has awarded a contract to European industry leaders to build the EDC facility service, ensuring its openness through a standardization path with OGC and will act initially as “anchor tenant” being also a main customer of the service.

## B.8. PCI Geomatics Tools for OpenDataCube implementation and data preparation

---

PCI Geomatics provides tools for preparing EO data for use in the OpenDataCube, and loading these data into an OpenDataCube. PCI has also prototyped analytics for OpenDataCube that leverage (for example) the Continuous Change Detection algorithm, and Object Based Image Analysis.

Specifically, the following tools exist:

- `GXL` (Geolmaging Accelerator), a software framework for processing large volumes of EO image data very quickly via parallelized code and load balancing across multiple

computers. Usable in cloud computing environments and local clusters. Includes workflows for:

- Image orthocorrection
- DSM extraction
- Analysis Ready Data processing; image geometry (fine alignment) and image radiometric correction and calibration.
- Object Based Image Analysis
- Pan sharpening using MRA Fusion (preserves radiometric character of the lower resolution inputs)
- Spectral pre-classification, including cloud and shadow detection and masking
- Data quality assessment and attribution
- Custom workflows including the addition of user-generated algorithms
- OpenDataCube data indexing and loading (loading data TO a cube).
- Image quality and history meta-data handling to support CARD4L certification
- Algorithm library, including all of the above noted functions, and several hundred more; accessible via Python
- Tools are generic; can include sensors other than Landsat and Sentinel-2. Supported sensors (to date) include Radarsat-2, Spot, Sentinel 1 and RapidEye.
- Data cube visualizations and analytics that build on example algorithms such Water Observations from Space and Continuous Change Detection. Enhancements to these algorithms include Object Based Image Analysis to improve speed and interpretability.
- Demonstration data cubes have been set up for Forestry Canada and Environment and Climate Change Canada. A data cube is in operational use by Natural Resources Canada (Emergency Geomatics Services) using RADAR and optical EO data.



C

# ANNEX C (INFORMATIVE) REVISION HISTORY

---

## ANNEX C (INFORMATIVE) REVISION HISTORY

---

DATE	RELEASED	AUTHOR	PARAGRAPH MODIFIED	DESCRIPTION
2019-09-14	v0.5	G. Percivall, (Editor) with Submitters	First public version	Published for Public Request for Comments
2019-10-24	V0.6	G. Percivall, (Editor) with Submitters	1, 6 , 7, 11.5, Annex C	Edits based on comments received from public review



# BIBLIOGRAPHY

---



## BIBLIOGRAPHY

---

1. P. Baumann: Language Support for Raster Image Manipulation in Databases. Proc. Int. Workshop on Graphics Modeling, Visualization in Science & Technology, Darmstadt/Germany, April 13 – 14, 1992, Springer 1993, pp. 236 – 245
2. P. Baumann, S. Feyzabadi, C. Jucovschi: Putting Pixels in Place: A Storage Layout Language for Scientific Data. Proc. IEEE ICDM Workshop on Spatial and Spatiotemporal Data Mining (SSTD'M'10), December 14, 2010, Sydney, Australia
3. CEOS and the Australian Geoscience DataCube – Towards Integrated Earth Environmental Information Systems, Presentation by Dr. Alex Held, CSIRO and CEOS Chair Representative, DIAS Symposium, Tokyo, August 2016
4. Vinhas, Lubia, et.al, Web Services for Big Earth Observation Data, Proceedings XVII GEOINFO, November 27-30, 2016, Campos do Jordão, Brazil
5. Baumann, P., The Datacube Manifesto, EarthServer Project, 2017. Retrieved from <http://www.earthserver.eu/tech/datacube-manifesto>
6. OGC Discrete Global Grid Systems (DGGS) Abstract Specification Topic 21, OGC Document 15-104r5, Publication Date: 2017-08-01
7. P. Baumann, A.P. Rossi, B. Bell, O. Clements, B. Evans, H. Hoenig, P. Hogan, G. Kakaletris, P. Koltsida, S. Mantovani, R. Marco Figuera, V. Merticariu, D. Misev, B. Pham Huu, S. Siemen, J. Wagemann: Fostering Cross- Disciplinary Earth Science Through Datacube Analytics. In. P.P. Mathieu, C. Aubrecht (eds.): Earth Observation Open Science and Innovation – Changing the World One Pixel at a Time, International Space Science Institute (ISSI), 2017, pp. 91 – 119
8. Nativi, Stefano Nativi, et.al., A view-based model of data-cube to support big earth data systems interoperability, Big Earth Data, 1:1-2, 75-99, Published online: 04 Dec 2017. DOI: 10.1080/20964471.2017.1404232
9. Strobl, Peter, et.al., The Six Faces of the Data Cube, Proc. of the 2017 conference on Big Data from Space (BiDS'17) Toulouse, France 28–30 November 2017 doi: 10.2760/383579
10. Tan, Zhenyu, et.al. An Array Database Approach for Earth Observation Data Management and Processing, ISPRS Int. J. Geo-Inf. 2017, 6, 220; Published: 19 July 2017. doi:10.3390/ijgi6070220
11. Wagemann, Julia, et.al., Geospatial web services pave new ways for server- based on-demand access and processing of Big Earth Data, International Journal of Digital Earth, 11:1, 7-25, Published online: 17 Jul 2017. DOI: 10.1080/17538947.2017.1351583
12. P. Baumann, D. Misev, V. Merticariu, B. Pham Huu: Datacubes: Towards Space/Time Analysis-Ready Data.. In: J. Doellner, M. Jobst, P. Schmitz (eds.): Service Oriented

Mapping – Changing Paradigm in Map Production and Geoinformation Management, Springer Lecture Notes in Geoinformation and Cartography, 2018

13. Open Data Cube Manual, CEOS. Accessed June 27, 2018. <http://datacube-core.readthedocs.io/en/latest/>
14. Invitation to Tender: “Data Cube Facility Service 2018–2020” <https://eo4society.esa.int/2018/09/06/data-cube-facility-service-2018-2020/>
15. D. Misev, P. Baumann, V. Merticariu, D. Bellos, S. Wieghe: BigDataCube: Making Big Data a Commodity. Proc. 69th International Astronautical Congress (IAC), Bremen, Germany, 1-5 October 2018
16. The Feasibility and Utility of Implementing Temporal Data Cubes to support Projection or “Forecast” models and land change trends. A Report of the National Geospatial Advisory Committee (NGAC) Landsat Advisory Group, April 2018
17. RDA: Array Database Assessment Working Group Report. <https://www.rd-alliance.org/groups/array-database-working-group.html>
18. M Appel and E Pebesma, “On-Demand Processing of Data Cubes from Satellite Image Collections with the gdalcubes Library”, Data 4 (3), 92, 2019.
19. Purss, Matt, et.al., Datacubes: A Discrete Global Grid Systems Perspective, Cartographica, Volume 54 Issue 1, Spring 2019, pp. 63-71. DOI: 10.3138/cart.54.1.2018-0017