OGC Project Document 17-071r1

Title:	**Datacube Domain Working Group Carter**
Author Names:	**Peter Baumann, Dimitar Misev, Patrick Hogan, Lucio Colaiacomo, Roger Brackin, Kent Thayer, Eric Hirschorn**
Organizations:	**rasdaman GmbH, Jacobs University, NASA, European Union Satellite Center, Envitia, Digital Globe, KEYW Corp.**
Emails:	baumann@rasdaman.com, misev@jacobs-university.de, Lucio.Colaiacomo@satcen.europa.eu, Patrick.Hogan@nasa.gov, roger.brackin@envitia.com, kthayer@digitalglobe.com, ehirschorn@keywcorp.com
Date:	**August 23, 2017**
Category:	**Domain Working Group**

1. Introduction

The purpose of this Datacube Domain Working Group (Datacube.DWG) is to maintain an open forum for the discussion and presentation of datacubes in all their aspects – including, but not limited to, theory, practice, technology, "analysis ready" data organization, and standardization in the datacube context.

As a DWG, this WG will not establish RFC submissions, candidate standards, or revisions to existing OGC Standards. However, the DWG can develop change requests as document interoperability requirements for submission as work items to a SWG or to TC for consideration.

As datacubes are being addresses in a variety of bodies recently, Datacube.DWG will maintain close liaison with at least the following bodies: ISO; INSPIRE; W3C; RDA maintaining mutual information exchange. In particular, the group will contribute to the Research Data Alliance (RDA) Array Database Assessment Working Group (ADA:WG) [8].

The Datacube.DWG, all of its resources and all of its meetings will be open to all interested parties. Consequently, the DWG plans to have public collaboration, such as in teleconference, email discussions, and a public Twiki; for the latter, the DWG shall make a motion to the TC to approve public participation in the DWG.

Voting in DWGs is by simple majority of OGC Members present at the WG meeting or telecon, not just Voting TC Members, with the caveat that no OGC Member organization may cast more than one vote in a WG vote.

Chairs will be elected in the new DWG's first meeting.

2. Purpose of Working Group

The purpose of the Datacube.DWG is to act as a discussion forum on datacube themes in the widest sense and to contribute to making the standards environment for geospatial application developers as effective, efficient, consistent, and user-friendly as possible, regardless of whose standards they use in implementation. Specifically, the Datacube.DWG will generate use cases and requirements for SWGs to address.

The Datacube DWG strives to maximize information flow about requirements, concepts, technology, best practice, and standards related to datacubes, thereby aiding stakeholders with the information necessary to find their best approach to coping with "Big Data" when arranged as datacubes.

3. Problem Statement

Never before in human history it was so easy to obtain such massive information, and it is getting more and more affordable for users at large. Datacubes represent an emerging paradigm which strives to arrange massive spatio-temporal data in an analysis-ready shape by stitching partial data (such as satellite scenes). What is particular to datacubes is the arrangement of data items along multiple dimensions – actually, any number of dimensions: 1-D sensor timeseries, 2-D satellite imagery, 3-D x/y/t image timeseries and x/y/z geophysics voxel data, 4-D x/y/z/t climate and ocean data comprise just a few representatives. Hence, datacubes can be defined as follows [7]:

Definition. A **datacube** is a massive multi-dimensional array, also called "raster data" or "gridded data"; "massive" entails that we talk about sizes significantly beyond the main memory resources of the server hardware. Data values, all of the same data type, sit at grid points as defined by the *d* axes of the *d*-dimensional datacube. Coordinates along these axes allow addressing data values unambiguously.

A *d*-dimensional **grid** is characterized by the fact that each inner grid point has exactly two neighbors along each direction; border grid points have just one (see Figure 1). Point clouds, e.g., are not grids. Note that each axis can be regular or irregular (see Figure 2).

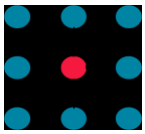

Figure 1: n-dimensional neighborhood in a grid

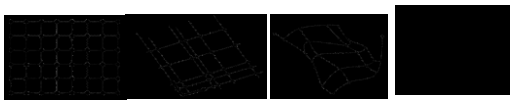

Figure 2: sample regular and irregular grid types (GML 3.3 and CIS 1.0)

The principle as such has been applied already earlier when data providers offered seamless maps for browsing, such as Google Maps, Bing, etc. Datacubes extend this principle to multi-dimensional spaces by performing exactly the same stitching along time, along elevation/bathymetry, etc. With suitable datacube APIs, users can reap benefits similar to those of seamless maps:

- Users see only one object (the cube) instead of a zillion files (such as satellite scenes), simplifying access, aggregation, and analysis in space and time; specifically, timeseries analysis of Big Data can get boosted massively.

- The internal data organization of the data provider remains hidden; this simplifies life for the user and opens up opportunities for implementations to develop (and evolve) efficient implementations, such as adaptive data partitioning as well as parallel and distributed processing.
- With a suitable service API, datacubes implement the "ship code to data" paradigm; standardized query languages exist already for interoperable access to and analysis of datacubes [3][5].

Based on such datacube data and service models, implementers can devise efficient and user-friendly tools on both client and server side. The Datacube.DWG aims at proliferating knowledge on datacubes by bringing together stakeholders in the widest sense, but also to spot underexplored areas.

For example, the following technical questions are not entirely (sometimes barely) addressed today:

- How to ensure consistency of datasets within the grids (geometric, temporal, etc.) so the data can be effectively used.
- How to ensure that data within a grid can be translated accurately to the location of the Earth.
- How to combine datacubes with novel CRS approaches, such as DGGS.
- Harmonization of raster standards, such as WMS, WMTS, and WCS.
- How to model tile systems as coverages, in particular given the tiling capabilities of CIS 1.1.

4. Charter

The initial membership of the Datacube.DWG will consist of the following members and individuals with extensive education and experience in datacube issues, namely:

NAME	AFFILIATION
Peter Baumann	rasdaman GmbH
Dimitar Misev	Jacobs University
Patrick Hogan	NASA
Lucio Colaiacomo	European Union Satellite Center
Roger Brackin	Envitia
Kent Thayer	Digital Globe
Eric Hirschorn	KEYW Corp.

5. Organizational Approach and Scope of Work

5.1 DWG Business Goals

Datacubes are an important approach towards making gridded data analysis ready – and, incidentally, such gridded data make up for a large part of today's "Big Data". As such, datacubes have the potential of increasing use of a key part of the "Big Data" on which industry, academia, open-source projects, data centers, and users worldwide are working intensively. Obviously, the topic enjoys a massive potential for innovation the market can benefit from.

As such, the Datacube.DWG has a potential of

- Showing new directions of research and development, to both industry and open-source communities;
- Showing novel technologies and best practices to data centers and data users
- Ultimately, fostering a new market segment on the emerging domain of datacubes.

5.2 Mission and Role

The Datacube DWG has a mission in acting as a focal point and information turnpike on datacubes within and beyond OGC. Its role, therefore, is to provide the necessary forum that collects and disseminates datacube relevant information from within OGC and beyond.

5.3 Activities planned

The Datacube.DWG will

- provide a forum for collecting, discussing, and disseminating information about datacubes, with a particular emphasis on interoperability requirements;
- generate use cases and requirements for SWGs to address;
- conduct regular meetings at OGC TC events and exceptionally possibly also at other venues agreed by the DWG;
- where appropriate, conduct joint meetings (in particular with Coverages.DWG due to the overlap)
- liaise with other OGC WGs on datacube related issues, in particular – but not limited to –Coverages.DWG, WCS.SWG, and BigData.DWG;
- liaise with at least the following bodies: ISO; INSPIRE; W3C; RDA for mutual information exchange. In particular, the group will exchange with the Research Data Alliance (RDA) Array Database Assessment Working Group (ADA:WG).
- develop Change Requests Proposals (CRPs) for existing OGC Standards;
- develop engineering reports with the intent seeking approval by the TC for release of these documents as OGC White Papers, Discussion Papers or Best Practices Papers;
- develop input for interoperability program planning;
- Perform informational presentations and discussions about the market use of datacube APIs and both client and server side technology.

Further activities may be pursued, in line with OGC's R&Ps on DWGs.

6. References

[1] OGC: OGC Coverage Implementation Schema. OGC 09-146r3, http://www.opengeospatial.org/standards/wcs
[2] OGC: OGC Web Coverage Service - Core. OGC 09-110r4, http://www.opengeospatial.org/standards/wcs
[3] OGC: OGC Web Coverage Processing Service (WCPS) Interface Language. OGC 08-068r2, http://www.opengeospatial.org/standards/wcs
[4] ISO: ISO 19123 Coverage Implementation Schema CD
[5] ISO: ISO 9075 SQL Part 15: Multi-Dimensional Arrays (MDA)
[6] INSPIRE: WCS as Coverage Download Service. http://inspire.ec.europa.eu/id/document/tg/download-wcs
[7] P. Baumann: The Datacube Manifesto. http://earthserver.eu/tech/datacube-manifesto
[8] RDA: Array Database Assessment Working Group. https://www.rd-alliance.org/group/array-database-assessmentac-wg/wiki/array-database-assessment-working-group