IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323424

BIG DATA STANDARDS AND ANALYSIS-READINESS: STATUS AND EVOLUTION

Peter Baumann

Jacobs University | rasdaman GmbH
Bremen, Germany
baumann@rasdaman.com

ABSTRACT

Flexible, scalable services on massive Earth data receive much attention today, in particular spatio-temporal data homogenized into datacubes. OGC has taken a lead in defining interoperable spatio-temporal datacubes through its coverage paradigm, together with tailored services, the Web Coverage Service (WCS) standards suite. This OGC Big Geo Data standard has attracted a who's who of major implementers of both open-source and proprietary servers and clients. We present status, uptake, and recent progress of the coverage datacube standards in OGC, ISO, and INSPIRE, in particular the ongoing work on the OpenAPI based OAPI-Coverages specification.

Index Terms— OGC, ISO, INSPIRE, Big Data, datacubes, coverage, WCS, WCPS, rasdaman

1. INTRODUCTION

Sensor, image, image timeseries, simulation, and statistics data contribute significantly to the Big Data challenge in the Earth sciences. Serving them at a high service quality and at high speed is one of the key challenges for modern technology. In particular, more and more it becomes clear that a zillion of single scenes is not the appropriate granularity for user-friendly offerings. A first step has been done some time back when seamlessly mosaicked maps enabled a smooth zoom and pan experience for users. Effectively, this has opened Earth Observation data a much larger, new community than just geo experts – we all embrace Google Maps and similar services today as part of our common Web experience. The same paradigm shift is now heralded by the datacube concept, but along the time axis: organizing all data from one satellite instrument into a single x/y/t datacube has the potential of simplifying access to multi-temporal data. Similar arguments hold for the vertical spatial axis.

Historically, OGC has established its Abstract Topic 6 for a high-level, abstract definition of coverage data. This has been adopted verbatim by ISO as 19123. Subsequently, both a concrete data and service model have been developed by OGC: the Coverage implementation Schema (CIS) and the Web Coverage Service (WCS). Both are organized in a modular way so as to keep the entry barrier for implementers low. CIS meantime has been adopted as ISO 19123-2, and 19123 is being modernized into 19123-2. "Concrete" in this context means: implementation guidance is concise enough to allow for interoperable implementations; OGC conformance tests actually allow assessing implementations down to the level of single pixels.

The WCS services spectrum ranges from simple extraction and reformatting over reprojection up to complex spatio-temporal analytics with the OGC Earth datacube language, Web Coverage Processing Service (WCPS) which allows ad-hoc queries following a paradigm of "any query, anytime". This way, WCS/WCPS as a data service suite complements the map navigation standard, WMS, and the generic, static processing service (with administrator-predefined fixed processes), WPS.

These coverage standards are proven on multi-Petabyte offerings, connected in intercontinental federations and with queries parallelized more than 1,000x in clouds; all this has been performed first in the intercontinental EarthServer initiative [7] which has taken a lead in advancing and promoting the datacube paradigm (Figure 1). Based on modern Array Database technology, rasdaman, large-scale data centers are establishing 3D and 4D datacubes for "any query, anytime" retrieval. All client/server interfaces of EarthServer rely exclusively on the OGC WCS and WCPS standards, plus WMS for flat map visualization; 3D visualization uses NASA WorldWind coupled again through WCS / WCPS.

In this contribution we present the current status as well as future directions for coverage standardization in OGC, ISO; and European INSPIRE, also putting into relation further, isolated initiatives such as by W3C. Observations and opinions are based on the author's continuous contribution to and leadership of coverage standards evolution.

2. OGC COVERAGES

OGC and ISO together have established coverages as a unifying concept for spatio-temporal regular and irregular grids, point clouds, and meshes. Their abstract definition is laid down in ISO 19123 [5][10]. Being abstract implies: incompatible implementations are possible (typically visible by the fact that such services use their own homegrown clients and cannot use, e.g., standard open-source clients).

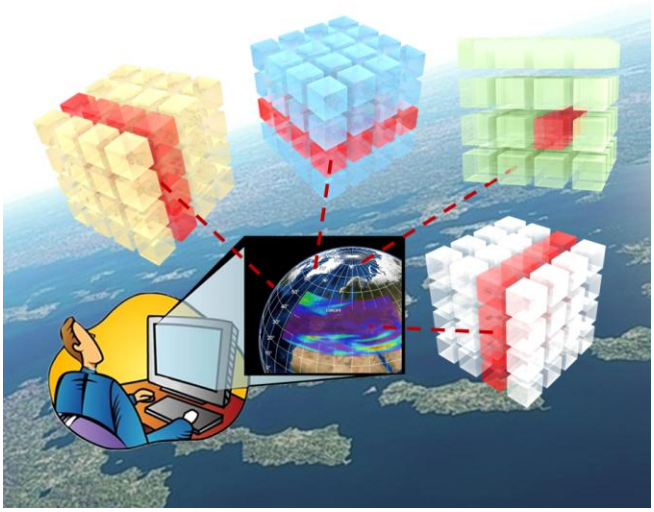

Figure 1: EarthServer datacube fusion based on OGC WCS and WCPS (source: EarthServer)

To remedy this, OGC has established a canonical coverage implementation model [2], derived from ISO 19123 and concrete, concise, and conformance testable down to pixel level. This coverage model supports both regular and irregular spatio-temporal grids, point clouds, and meshes.

3. OGC WEB COVERAGE SERVICE (WCS)

The *Web Coverage Service* (WCS), whose mission is to provide streamlined functionality for spatio-temporal coverage, technically is subdivided into a common core and several extensions. WCS Core [2] consists of only the most fundamental functionality: extracting a coverage or a subset thereof (Figure 2). The coverage result can be delivered in any suitable encoding. In a vanilla WCS request data are guaranteed to be delivered unaltered.

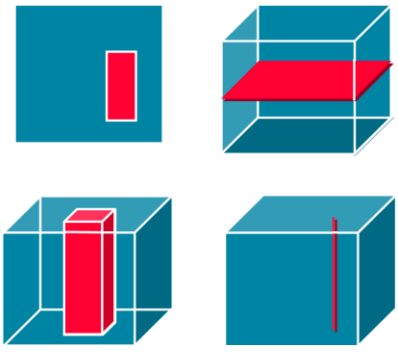

Figure 2: Subsetting of a 3-D coverage: Trimming (left) extracts a cutout which retains the number of dimensions, slicing (right) cuts out hyperplanes of reduced dimension

WCS extensions add further functionality facets to WCS, which an implementer may freely choose to support. Web Coverage Processing Service (WCPS) [1], for example, adds a declarative n-D geo raster query language into WCS.

4. IMPLEMENTING BIG DATACUBES

WCS Core Reference Implementation is *rasdaman*, the pioneer Array Database [8]. It offers an n-D array query language which on server-side is supported by a massively scalable array engine based on the "tile streaming" paradigm. Incoming queries undergo effective optimization, parallelization across node and heterogeneous hardware. Databases have reached multi-Petabyte **Error! Reference source not found.**, and single queries have been split over more than 1,000 cloud nodes [12].

The same principle, query splitting, that enables heterogeneous cloud parallelization can also be applied between data centers, leading to datacube federations of rasdaman installations, such as the EarthServer federation.

5. STANDARDIZATION OUTLOOK

Since its adoption in 2010, coverages and WCS have experienced manifold, continuously growing take-up from both open-source and proprietary implementers, and recently from further standardization bodies. OGC CIS 1.1, adopted in 2018, has unified regular and irregular grids, into a common conceptual framework which massively simplifies coverage handling. ISO TC211 has adopted CIS 1.0 as ISO 19123-2; in 2020, an upgrade to OGC CIS 1.1 is planned. Simultaneously, OGC is working on a new generation of protocol bindings for coverage, OAPI-Coverages. While still under development, it is obvious that OAPI-Coverages will constitute a WCS compatible extension, thereby preserving existing assets.

INSPIRE in 2016 has adopted OGC WCS as Coverage Download Service.

Based on the rasdaman query language, in 2019 domain-independent datacubes have been included into ISO SQL, forming Part 15: Multi-Dimensional Arrays (MDA).

The World-Wide Web Consortium (W3C) has been discussing coverages, although independently from and not synchronized with OGC, ISO, and INSPIRE standardization.

6. CONCLUSION

OGC coverages represent a concrete, interoperable data model which unifies n-D regular and irregular grids, point clouds, and meshes – obviously main contributors to today's Big Geo Data. The OGC WCS standards suite offers a wide range of functionality, from simple extraction and download up to complex ad-hoc analytics. Flexibility and scalability of

the WCS suite has been demonstrated in practice through large-scale services. Particularly instrumental in demonstrating scalability and federation of datacubes was pioneered by the EarthServer initiative, offering access and fusion on datacubes meantime by far exceeding the Petabyte frontier.

Obviously, coverage standardization is rather agile and progressing the state of the art. Fortunately, the main standardization bodies agree on the coverages definition, thereby achieving a stable platform for implementers, service providers, and users alike.

ACKNOWLEDGEMENT

Research supported by EU H2020 EarthServer-1/-2, LandSupport, CopHub.AC, PARSEC, German BigDataCube, BigPicture, ORBiDANSe, DeepRain; and several more.

8. REFERENCES

[1] P. Baumann: OGC Web Coverage Processing Service (WCPS) Language Standard, version 1.0. OGC 08-068r2

[2] P. Baumann: Web Coverage Service (WCS) Interface Standard – Core, version 2.1. OGC 17-089r1

[3] P. Baumann, E. Hirschorn, J. Maso: OGC Coverage Implementation Schema, version 1.1. OGC 09-146r6

[4] P. Baumann: The OGC Web Coverage Processing Service (WCPS) Standard. Geoinformatica, 14(4)2010, pp. 447 - 449

[5] ISO: 19123 Geographic information – Schema for coverage geometry and functions. IS 19123:2004

[6] N.n.: Earthlook. http://earthlook.org, seen 2020-01-20

[7] N.n.: EarthServer – Datacubes at Your Fingertips. www.earthserver.eu, seen 2015-11-06

[8] N.n.: rasdaman. www.rasdaman.org, seen 2020-01-20

[9] OGC: WCS standards suite. www.opengeospatial.org/standards/wcs; 2020-01-20

[10] OGC: The OpenGIS Abstract Specification Topic 6: Schema for coverage geometry and functions, version 7.0. OGC 07-011

[11] 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

[12] A. Dumitru, V. Merticariu, P. Baumann: Exploring Cloud Opportunities from an Array Database Perspective. Proc ACM SIGMOD Workshop on Data analytics in the Cloud (DanaC'2014), June 22 - 27, 2014, Snowbird, USA