Serving Spatio-Temporal Grids: How Standards Help

Jinsongdi Yu[1]*, Peter Baumann[2]
[1] Spatial Information Research Center of Fujian, Fuzhou University, Fuzhou, China
[2] Large-Scale Scientific Information Systems, Jacobs University, Bremen, Germany
*Corresponding author, e-mail: yyx350@126.com

Abstract—**The global Earth Science Systems (ESS) cooperation requires both flexible and interoperable Web Service support built on large varieties of Earth Observation archives. Given the complexity and dynamics of each observation and the large number of disciplines involved, Open GIS Consortium (OGC) proposed a modular standardization approach to facilitate ESS data retrieval and analytics. Its latest Web Coverage Service (WCS) 2.0 standard follows the approach and presents as an entry point to build ESS information exchange blocks for the improvement of global Earth Science Systems cooperation in the coming Big Data Era**

Keywords-Earth Science Systems; interoperability; OGC WCS 2.0

I. INTRODUCTION

To achieve a better understanding of the Earth System, e.g. global warming and climate changes, humans since some time now are undertaking comprehensive, coordinated and sustained observations of the planet Earth's behavior Collection, analysis, and presentation of Earth information is being contributed from multiple disciplines, such as Atmospheric, Solid Earth, and Ocean research. In the Big Data [1][2] era, given the complexity and dynamics of each observation and the large number of disciplines involved, the Open Geospatial Consortium (OGC) proposes a modular standards approach to the interoperability of the global Earth Science Systems (ESS) by bringing together the diverse initiatives and archiving a common agreement in the involved communities. In particular, the Web Coverage Service (WCS) 2.0 standards suite [3] follows this modularity approach in the most rigorous way today, with one core data and service specification each, surrounded by more than a dozen extension specifications addressing specific facets öike format encodings and detail functionality like scaling.

WCS has been built specifically with Big Earth data in mind combining methods of best practices and insights from data repositories, high-performance analytics, and further areas with the domain expertise of Earth, Space, Life, and Social sciences and Computer Science. Following its adoption in 2012, WCS 2.0 meantime is widely adopted within open-source, commercial products, and user communities at large as a means to facilitate access, processing, and filtering of ESS data, not only as and when they are exchanging messages, but also during analytic function orchestration.

While the standard documents are valued for their conciseness and implementability it is an often heard demand that an explanatory overview is provided in addition so that both implementers and users can easily grasp concepts and best practices. Responding to this need, this paper summarizes the WCS data and service standards suite with emphasis on practical use. This includes coverage models in abstract and its implementation models, service models in core and extensions, the first WCS reference implementation, and the corresponding compliance testing initiative.

The remainder of this paper is organized as follows. Section 2 presents the coverage models. Section 3 introduces the core and extensions of WCS 2.0. Section 4 describes its implementation and shows the influence. Section 5 concludes the paper.

II. COVERAGE MODELS

A. Concept

There are several data categories in the GIS domain where a multidimensional grid is the natural representation. As one particular class of space/time varying data, such as 1D time-series, 2D remote sensing imagery, 3D x/y/t image time-series and x/y/z geophysical data, as well as 4D x/y/z/t atmospheric and ocean data, arrays [4] allow modeling of both regular and irregular grids. Remote sensing (RS) imagery arguably is the most prominent one: raster data acquired by a satellite come as a (series of) matrix-represented datasets and are then further processed into higher level products. The most straightforward fit with arrays are indeed products coming after several levels of processing, delivering highly regular GIS geometries such as ortho-rectified products. Dealing with remote sensing data, however, requires taking into account more than just the data array itself. The location information of the array contents is needed to properly relate such values to positions on Earth. Increasingly, elevation and time coordinates have to be considered as well. In general, there is a need to associate this data payload to their GIS domain, which leads the standard concept of a coverage, which is a generic data category defined by ISO 19123 [5]. A *coverage* in this abstract model encompasses regular and irregular grids, point clouds, and general meshes. OGC complements this with a coverage implementation model [6] which indeed is interoperable, as we will discuss below.

Let us look at the structure of a coverage in more detail. Mathematically resembling a function, a coverage establishes a mapping between a given multidimensional domain and the values associated to locations within such a domain. The

Supported by the National Natural Science Foundation of China (No. 41401454), Natural Science Foundation of Fujian Province of China (No.2014J05048), and the Scientific Research Starting Foundation of Fuzhou University (No.0102510073).

domain set of a coverage consists of all "direct points", i.e., locations where concrete values are available (as opposed to values obtained, for example, by interpolating between direct positions). The *range set* is the set of all values associated with direct positions. We call a direct position together with the value it carries a *cell*. For satellite imagery, the domain set is 2-D x/y and consists of the pixel locations which together fomr some (regular or irregular) grid; for image time series, the domain set is a 3-D x/y/t cube containing the pixel positions situated in space and, additionally, time. For ortho-rectified imagery obtained in regular time intervals this constitutes a quadrilateral, equidistant grid. This is different, however, for raw swath imagery that has not undergone ortho-rectification and also for time series where the time of acquisition is not always regular; in this case, the distance between locations of range value acquisition may be not regularly spaced on the spatial and temporal axes. Hence, when dealing which such kind of data domain geometries that might differ drastically from the regularly spaced world of array cells are investigated. The coverage model thus needs to address and support different geometries for its range values, encompassing array-style regular grids as well as irregular grids, or even meshes or point cloud datasets, such as laser scanning.

B. *The Coverage Implementation Standard*

ISO 19123, which is also Open GIS Consortium (OGC) Abstract Topic 6, establishes an abstract coverage model. It is abstract in the sense that many implementations are possible (and, in fact, exist) which are consistent with this model, but not interoperable among each other due to different implementation decisions made. To achieve interoperability, OGC has established, based on the abstract ISO 19123 model, a concrete coverage implementation model (CIV), the "GML 3.2.1 Application Schema – Coverages" standard [6] CIS. In this specification, a range type component has been added which describes, based on the Sensor Web Enablement definitions of SWE Common, the common data type (sometimes called attribute domain [7]) all cells in a coverage share. This is not confined to a data type description such as "8-bit unsigned integers", but may contain semantic links (such as URLs defining these values as representing radiance), value ranges, units of measure (such as W/cm^2), nil values, and more. This way a coverage is informationally complete enough to allow machine-based interpretation. Fig. 1 shows the high-level UML model of an OGC coverage.

Note that, despite the name GML in the title, UML and GML are only used to achieve some machine-testable interface specification; by no means is a coverage encoding tied exclusively to GML. In fact, a steadily increasing list of coverage encodings is standardized, such as GeoTIFF, JPEG2000, NetCDF, and others; see the OGC WCS page [8] for the current list. In recognition of this misperception OGC has renamed the specification in March 2015, so the document in future will be announced and distributed as "Coverage Implementation Schema", abbreviated as "CIS".

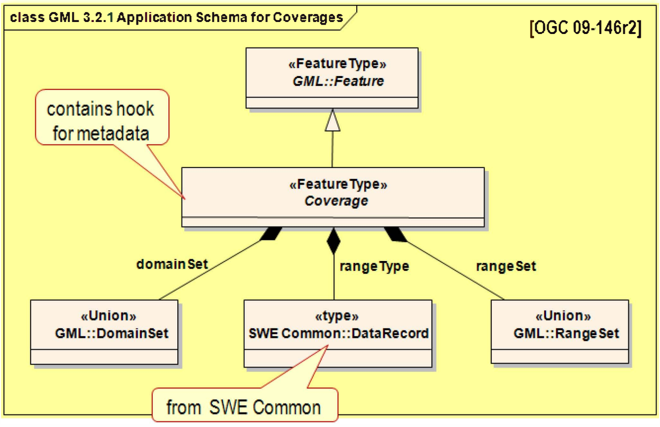

Figure 1. OGC coverage data model

The standard actually foresees two different encoding techniques, acknowledging storage efficiency considerations and also common best practices. First, a coverage can be encoded in some format that allows representation; for example, a 4D climate data cube can be represented in NetCDF, but not in JPEG as this is only capable of 2D. Some formats, like GML, will allow a comprehensive representation of all coverage elements defined in the standard. Other formats (such as GeoTIFF) are lossy in metadata, so the client has to make a wise choice on what to retrieve. Lossless formats will allow further processing while lossy formats like PNG or JPEG may deliver incomplete information, but still suitable for visualization in a browser.

To achieve lossless metadata while retaining an efficient encoding, a mixed representation is defined in addition. Based on the MIME definition of multipart messages (as used in emails with attachments) a coverage can consist of two parts, a GML header and a binary encoding of the range set, i.e., the pixel/voxel payload. This way the client always receives metadata in a canonical, well-defined, and machine-readable representation.

CIS is concrete in terms of testable data structure definitions and integrity constraints and, therefore, enables rigorous conformance testing down to single pixel level through the mechanisms established in OGC's CITE conformance testing program [9].

Furthermore, several efforts have been investigated to provide interoperable and standard-based solutions [8] for datasets up to 5D and bridge the gap between the atmospheric, oceanographic and GIS communities, based on the research of mapping existing archive models, e.g., the Common Data Model of the Unidata, to standardized coverage model [9]. By concerning access bit streams behind, relevant coverage information can be extracted with a proper data virtualization approach [10], regardless of physical formats.

A current limitation of GML is that the domain set of a coverage is meant to carry only numeric values. While this is acceptable for geographic coordinates (communities seem to accept floating point numbers in place of degree, minute, and second representations) this is a severe impedimentation when it comes to time coordinates. All our stakeholder discussions have shown that counting "number of seconds since epoch" is not acceptable, but common date strings as defined by ISO 8601 [11] need to be supported (such as "2013-04-20"). Currently an extension is under work to allow for different (and mixed) representations. A dedicated OGC working group [12] is currently elaborating a coherent handling of time and calendars – actually, beyond coverages as such, as temporal data handling is a cross-cutting issue.

III. WCS 2.0 Core and Extensions

The CIS coverage data model forms a pillar for coverage services, such as the OGC Web Coverage Service (WCS) standards suite. Starting with WCS 2.0 data and service model have been separated so as to allow for the coverage data model to be used independently from WCS. Today, coverages can be served by WFS, WCS, WCPS, WPS, and SOS, among others. WCS is distinguished, though, in that it offers the richest functionality – where a WFS can serve out a coverage only as a whole a WCS can extract from a coverage and manipulate its data (see later). At the heart of the coverage specifications are CIS for the core data model and WCS Core [13] establishing the service model. The core specification allows several extensions to be added for expanding server functionality.

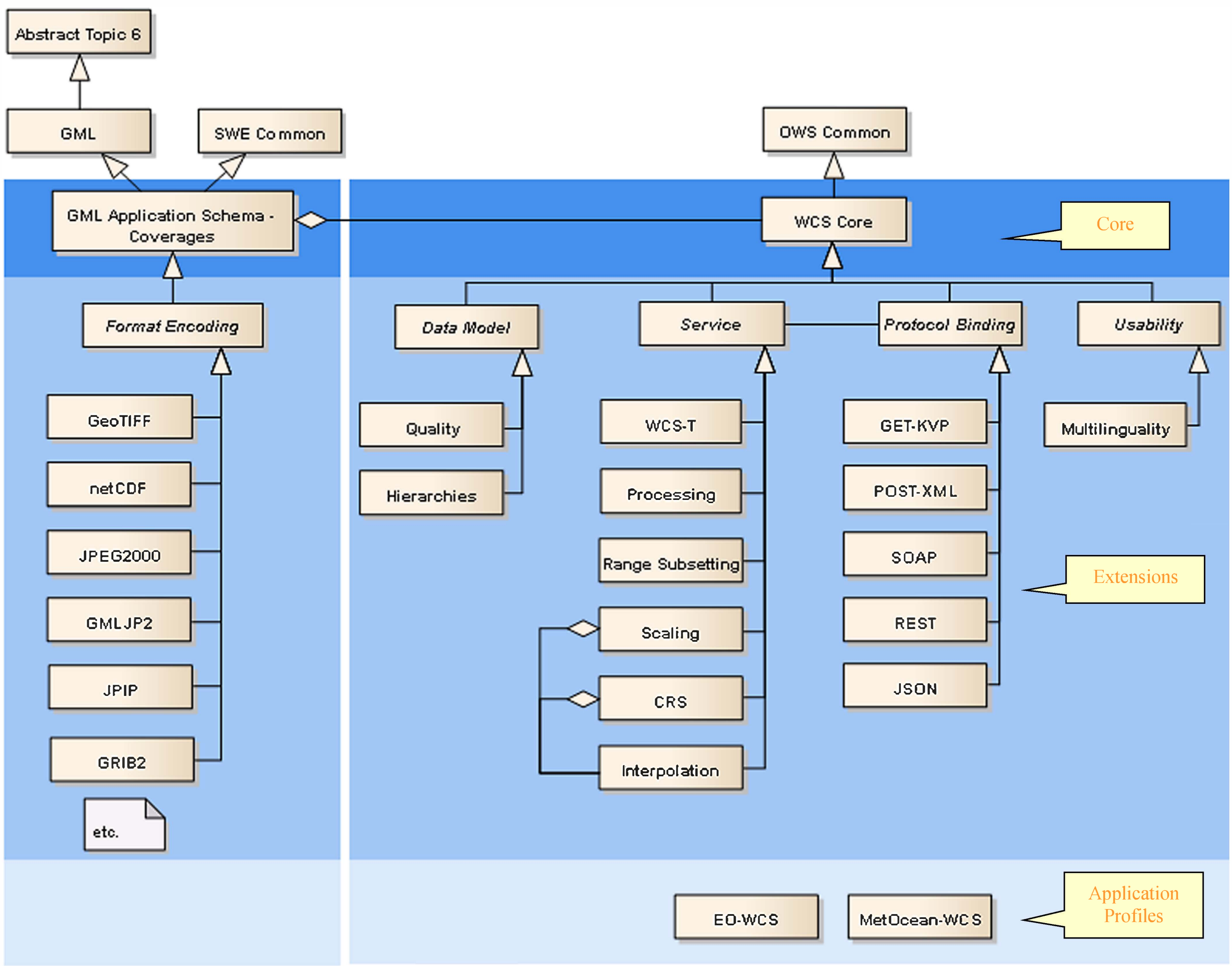

Figure 2. OGC coverage data and service model schematic overview

Fig. 2 provides an overall picture of the OGC unified coverage standards where a clear separation of concerns is visible by having the data model elements on the left side while the service model elements cover the center and right areas. On the service side, OGC OWS Common forms the common ground. Using these concepts, OGC WCS Core (the service model's central element) defines the basic coverage access functionality: downloading a whole coverage, or some spatio-temporal part (i.e., a subset) of it. The service extensions add further functionality facets and protocol bindings. We discuss these in turn.

A. WCS 2.0 core

WCS Core provides three main operations to be provided by any server implementing this standard:

- *GetCapabilities* allows a client to retrieve overall information about a server's capabilities as well as a list of offered coverages. In the GET/KVP protocol binding, such a request might look as follows:
  http://{*server*}?SERVICE=WCS&VERSION=2.0.0
  &REQUEST=GetCapabilities

- *DescribeCoverage* allows a client to retrieve detailed meta-data on particular offered coverages, such as their spatiotemporal extent and range cell data type. In the GET/KVP protocol binding, such a request might look as follows:
  http://{*server*}?SERVICE=WCS&VERSION=2.0.0&
  REQUEST=DescribeCoverage&COVERAGEID=*id*

- *GetCoverage*, the actual "workhorse", finally allows a client to retrieve actual coverage data for some selected set of locations, delivered in some coverage encoding format selected through its MIME identifier. In the GET/KVP protocol binding, such a request might look as follows:
  http://{*server*}?SERVICE=WCS&VERSION=2.0.0&
  REQUEST=GetCoverage&COVERAGEID=*id*&
  FORMAT=image/tiff

A request using only WCS Core is guaranteed to deliver all data unmodified (as long as a suitable lossless encoding format is chosen).

B. *WCS 2.0 extensions*

Extension specifications add further functionality facets which a server may or may not implement, depending on the vendor's choice. The only mandatory selection is to have at least one protocol binding, as otherwise no communication is possible at all. In the *GetCapabilities* response, every server is obliged to list the extensions it supports.

For convenience – and without being normative – extensions are grouped into several categories. *Data model* groups extensions intended to extend or refine data structures such as the support to uncertainty in measurement (quality) or hierarchical data structures (such as nested range types); *Service model* groups extensions providing additional service capabilities, hence they add new operations or functional behaviors (these are detailed hereafter); *Protocol binding* group extensions dealing with request/response shipping defining the client/server communication protocols, at the time of writing (July 2013) the approved protocols are the HTTP based GET-KVP [14], XML/POST [15] plus XML/SOAP [16] with the REST binding being finalized; *Usability* collects extensions that generally improve service use, such as multilingual names and error messages. Finally, *Format encoding* is a noticeable change in recent coverage specifications with respect to its predecessors: once being adopted WCS extensions, with the upcoming/current standards, these belong to the coverage data model, as described in the previous section.

Service model extensions, adding functionality, are detailed hereafter:

- *Range subsetting*: extracts range components (commonly known as "bands" or "variables") from a coverage. A practical example is the extraction of the red channel from an RGB image.
- *Scaling*: Different methods for resampling a grid coverage are provided herein, effectively changing resolution of the grid.
- *Interpolation*: this specifies the interpolation method to be applied whenever interpolation has to be done by the server; common examples are scaling and CRS transforms. Embedded in a common framework, implementations can make their choice between several interpolation methods taken from ISO 19123. Note that if this extension is not supported then the server will silently make its own implementation dependent choice.
- *CRS*: this extension enables the provision of coverages where either the input subsets and/or the output coverage require on-the-fly re-projection between the native CRS and some specified one. EPSG is an established authority that defines CRS for GIS data while a general CRS framework aims at supporting the OGC-defined nD spatiotemporal CRSs by means of CRS composition and parametrization. The OGC CRS Name Type Specification allows composing CRSs on the fly based on existing spatial and temporal axes and CRSs.
- *Transactional service (WCS-T)*: the "transactional" extension allows modification a server's offering by inserting, deleting, or updating coverages [17]. Especially the latter functionality is important when large coverages are being built piecewise, such as image mosaics. WCS-T has between designed in a way that its input is compatible with a GetCoverage output. This enables mashups where coverages can be exchanged between computes and data centers. At the time of this writing this extension is under adoption vote.
- *Web Coverage Processing Service (WCPS):* WCPS [18] establishes a high-level spatiotemporal raster query language which enables ad-hoc server-side processing and filtering on coverages. It summarizes and extends the functionality of the other retrieval extensions and embeds it into a language concept for flexible ad-hoc composition of complex requests. Similarly as database query languages allow to flexibly search meta-data, WCPS allows versatile navigation, extraction, aggregation, and analytics on coverage data that can consist of a single coverage or sets thereof. The query language includes n-dimensional map algebra operators and functions, aggregations, plus encoding of the resulting data in different formats for download. This is a step forward in the current landscape of ad-hoc, exploratory data analysis over large data archives – in this sense, WCPS is a Big Earth Data standard. The WCS *Processing* Extension links WCPS into WCS world; a binding into

WPS world has been established in parallel to provide WPCS in both environments.

The service model presented, paired with its coverage data model counterpart, allows for multiple processing steps to be applied in a service oriented architecture. For example, a coverage might be generated through a Web Processing Service (WPS) [19] or a Sensor Observation Service (SOS) [20], then it can get imported into a coverage server through WCS-Transaction (WCS-T) and subsequently be served through a WCS, a WCPS, or a WPS to a computing or visualization client.

IV. Adoption & Conformance Testing

The WCS standard series meantime has won widespread attention and support by both industry-leading vendors (such as ESRI, ERDAS, PCI, Pyxis, etc.) and open-source projects (such as rasdaman, MapServer, GeoServer, etc.). A non-authoritative list of tools that has been reported to support WCS 2.0 as client and/or server is provided at [21]. Specifically, rasdaman, which is the first official OGC WCS 2.0 Core Reference Implementation. The OGC compliance program - Compliance Interoperability Testing & Evaluation (CITE) determine whether the software complies with the requirements of the OGC standard [22]. In this way, different application domains can set up specific domain services with the same retrieval interfaces. WCS Core and almost all extensions (with the exception of WCS CRS) are supplied with extensive conformance testing utilities available from OGC.

The EarthServer initiative [23] has established large-scale services in all Earth sciences. In this project, six services have been set up for integrated data/metadata retrieval and distributed query processing. The EarthServer platform is based on the rasdaman Array DBMS which is in operational use on coverage holdings in excess of 130 TB and successfully parallelized over more than 1,000 cloud nodes while serving coverages through WMS, WCS, WCPS, and WPS. WCPS has been applied successfully in climate anomaly search and Meteorological data services [24] [25].

V. Conclusions

Standardized Web service frameworks have the potential to solve a large range of questions today in the context of global Earth System. The OGC coverage data and service models contribute greatly in terms of interoperability on regularly and irregularly gridded data. With the WCS suite, OGC's Big Geo Data specifications, a flexible and interoperable service offering can be provided, ranging from simple sub-setting with WCS Core to ad-hoc queries and analytics with WCPS, possibly combined with a WPS binding for asynchronous processing. Manifold implementations have demonstrated convenience of the specification techniques for developers, and both flexibility and scalability have been proven as well.

However, research communities like to use different technologies like R, Matlab, Array Databases, etc., and best in mixed mode to address global analytic challenges. Research on seamless integration of different data and processing models, therefore, is a topic receiving high attention in research today. In terms of interoperability and standards, the OGC core and extension framework appears flexible to integrate emerging data and technologies, such as *in-situ* data, semantic search, and cloud paradigms [26] [27], under common interface umbrellas. On implementation side, services with their high-level data modeling and querying capabilities paired with their extensibility and scalability mechanisms seem to have a say in the thrive for solving interoperability problems of rich observation data and systems. With WCS the community-driven standardization in the OGC has established interfaces proven, accepted value based on stakeholder experience and in operational use on Big Earth Data.

Acknowledgment

The authors would like to express appreciations to the OGC WCS.SWG colleagues for their valuable comments and support.

References

[1] D. Laney, 2001. "3D Data Management: Controlling Data Volume, Velocity and Variety," MetaGroup, 2001, http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf, seen on 2015-mar-15

[2] D. Snow, "Adding a 4th V to BIG Data – Veracity." 2012, http://dsnowondb2.blogspot.de/2012/07/adding-4th-v-to-big-data-veracity.html, "last seen on 2015-mar-15"

[3] P. Baumann (ed.), "OGC® Web Coverage Service 2.0 Primer: Core and Extensions Overview," OGC 09-153r1, 2012, p. 15, URL: https://portal.opengeospatial.org/files/?artifact_id=46442

[4] P. Baumann, S. Holsten, "A comparative analysis of array models for databases," International Journal of Database Theory and Application, vol. 5(1), 2012, pp. 89-120

[5] ISO, "Geographic information -- Schema for coverage geometry and functions," ISO 19123:2005, 2009, p. 65, URL: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=40121

[6] P. Baumann (ed.), "OGC® GML Application Schema - Coverages," OGC 09-146r2, 2012, p. 43, URL: https://portal.opengeospatial.org/files/?artifact_id=48553

[7] OGC SWE Common Standard Page, http://http://www.opengeospatial.org/standards/swecommon, seen on March 15, 2015

[8] OGC WCS standards page. http://www.opengeospatial.org/standards/wcs, seen on March 15, 2015

[9] OGC Compliance Program Page, http://http://cite.opengeospatial.org/, seen on March 15, 2015

[10] J. Yu, P. Baumann, S. Crompton, S. Wu, Y. Lu, "Facilitating earth science data interoperability using the SCIDIP-ES data virtualisation toolkit", Earth Science Informatics, 2014, doi: 10.1007/s12145-014-0189-8

[11] ISO, "Data elements and interchange formats -- Information interchange -- Representation of dates and times", ISO 8601:2004, 2004

[12] OGC, "Temporal Domain Working Group public wiki", http://external.opengeospatial.org/twiki_public/TemporalDWG/WebHome, seen on March 15, 2015

[13] P. Baumann, (ed.), "OGC® WCS 2.0 Interface Standard - Core," OGC 09-110r4, 2012, p. 57, URL: https://portal.opengeospatial.org/files/?artifact_id=48428

[14] P. Baumann, (ed.), "OGC® Web Coverage Service 2.0 Interface Standard - KVP Protocol Binding Extension," OGC 09-147r3, 2013, p. 17, URL: https://portal.opengeospatial.org/files/?artifact_id=50140

[15] P. Baumann(ed.), "OGC® Web Coverage Service 2.0 Interface Standard - XML/POST Protocol Binding Extension," OGC 09-149r1, 2010, p. 13, URL: http://portal.opengeospatial.org/files/?artifact_id=41441

[16] P. Baumann(ed.), "OGC® Web Coverage Service 2.0 Interface Standard - XML/SOAP Protocol Binding Extension," OGC 09-148r1, 2010, p. 14, URL: http://portal.opengeospatial.org/files/?artifact_id=41440

[17] P. Baumann: OGC WCS Extension – Transaction. OGC 13-057, 2015

[18] P. Baumann(ed.), “Web Coverage Processing Service (WCPS) Implementation Specification,” OGC 08-068, 2008, p. 75, URL: http://portal.opengeospatial.org/files/?artifact_id=32319

[19] OGC WPS Standard Page, http://www.opengeospatial.org/standards/wps, seen on March 15, 2015

[20] OGC SOS Standard Page, http://www.opengeospatial.org/standards/wps, seen on March 15, 2015

[21] OGC, “WCS implementations and services”, http://www.ogcnetwork.net/node/1673, seen on March 15, 2015

[22] J. Yu, “A Model for Tracing Cross-referenced Statement Validity”, Computer Standards & Interfaces, 2015, vol. 41, pp. 10-16

[23] P. Baumann, P. Mazzetti, J. Ungar, R. Barbera, D. Barboni, A. Beccati, L. Bigagli, E. Boldrini, R. Bruno, A. Calanducci, P. Campalani, O. Clement, A. Dumitru, M. Grant, P. Herzig, G. Kakaletris, J. Laxton, P. Koltsida, K. Lipskoch, A.R. Mahdiraji, S. Mantovani, V. Merticariu, A. Messina, D. Misev, S. Natali, S. Nativi, J. Oosthoek, J. Passmore, M. Pappalardo, A.P. Rossi, F. Rundo, M. Sen, V. Sorbera, D. Sullivan, M. Torrisi, L. Trovato, M.G. Veratelli, S. Wagner, “Big Data Analytics for Earth Sciences: the EarthServer Approach”, International Journal of Digital Earth, 0(0)2015, Taylor & Francis Group 2015, pp 1 – 27, http://www.tandfonline.com/10.1080/17538947.2014.1003106

[24] P. Baumann, and D. Misev, 2012. “Towards Scalable Ad-Hoc Climate Anomalies Search”, ACM SIGSPATIAL BIGSPATIAL Workshop, 2012, pp. 101-110

[25] D. Misev, P. Baumann, J. Seib, “Towards Large-Scale Meteorological Data Services: A Case Study,” Datenbank-Spektrum, 2012, pp. 1-10

[26] J. D. Blower, “GIS in the Cloud: Implementing a Web Map Service on Google App Engine,” Proceedings of the First International Conference on Computing for Geospatial Research and Application, 2010, pp. 1-4

[27] M. Owonibi, “Dynamic Resource-Aware Decomposition of Geoprocessing Services Based on Declarative Request Languages,” PhD Thesis, 2012