Unleashing The Sensor Deep Web: Next-Generation Sensor Data Services

Peter Baumann[1]

[1] Jacobs-University Bremen, Campus Ring 12, 28759 Bremen, Germany
p.baumann@jacobs-university.de

1 Motivation

"Google Earth" – and enough is said for both experts and home Internet users when it comes to fast and user-friendly access to satellite imagery. An open standard allowing such map navigation is the Web Map Service (WMS) [6] of the Open GeoSpatial Consortium (OGC, www.opengeospatial.org) which develops, publishes, and maintains interface specifications for interoperable geo services.

Yet, still in 2002 a technical expert from a state mapping agency told the author: "Airborne images? Why should we serve them to the public? Nobody understands them anyways!" Then Google Earth and similar services popped up. Said another expert, "Google offers all we need; I cannot imagine anything beyond".

So what *is* beyond Google Earth that communities might need? Actually, 2-D remote sensing imagery is only the most well-known species - the general concept of multi-dimensional spatio-temporal raster data covers 1-D sensor time series, 2-D imagery, 3-D CAT scans (x/y/t) and exploration data (x/y/z), 4-D climate, ocean, and cosmological simulation results as well as life science microarray data (x/y/z/t), and many more. Sizes frequently range well into multi-Terabyte object sizes, in future: multi-Petabyte, resulting in a "Deep Web" of raster data waiting to be surfaced.

The term "coverage" denotes some "space-varying phenomenon" [4,7], in practice today: raster data. WCS, hence, defines an open access protocol for multi-dimensional raster data. Among the extensions that WCS offers beyond its basic service level is the Web Coverage Processing Service (WCPS) [2]. WCPS defines a raster expression language allowing clients to phrase requests of arbitrary complexity for evaluation by the server. This allows for manifold ad-hoc processing of raster data, such as deriving the vegetation index, determining statistical evaluations, and generating different kinds of plots, such as classification, histograms, etc. A hands-on demonstration website with real-life 1-D to 4-D data is being established at www.earthlook.org.

Use cases for WCPS are manifold: For sensor and streaming data it allows subsetting, on-demand processing and summarisation, as well as standing queries for alerting. Hyperspectral satellite imagery will not just be served as is, but derived products like vegetation index or snow index are be computed on the fly and without redundant storage. Human brain imaging benefits from analyzing thousands of brain activity maps simultaneously. Multi-Petabyte datacubes can be leveraged for online analysis.

In this poster and demo we present the state of the art in semantic Web services for large-scale, multi-dimensional raster data by showcasing the emerging WCPS standard and its implementation, based on real-life use cases. We hope to inspire creativity of the research community and, at the same time, get back valuable stimuli.

2 WCPS

WCPS is an optional component ("extension") for the basic WCS ("core") adding a coverage expression language to the static retrieval functionality of WCS. In this section, following a brief sketch of WCS, we give an overview of concept and implementation of the WCPS language, the object of demonstration.

2.1 WCS Coverage Model

Technically, a coverage mainly consists of a multi-dimensional array, a title (i.e., a name acting as identifier in requests), an optional geo reference plus the coordinate reference system name in which geo coordinates are expressed, and information about how to interpret the array data. The array cell type (the coverage *range description*) consists of one or more components, which themselves can be atomic or composite. For example, a Landsat satellite image would contain a range list of five to seven components, which are integer values, hence atomic.

While currently restricted to spatio-temporal coordinates WCS in future is expected to allow arbitrary combinations of spatial, temporal, and "subtract (i.e., non-spatiotemporal) axes. Examples for an abstract axis are pressure in climate simulation or sales figures in data warehousing / OLAP.

2.2 WCPS Request Language

We present the WCPS language's flavour by means of an example rather than giving a complete presentation which can be found in [2]. The following code piece assumes three coverage objects `LandsatA`, `LandsatB`, and `LandsatC` on the server:

```
for c in ( LandsatA, LandsatB, LandsatC )
where   max( c.red ) > 127
return  encode( c.red + c.nir, "tiff" )
```

These three coverages are inspected in sequence, binding it to variable `c`. Coverages failing the `where` predicate will be dropped; in this case, the maximum value of the `red` component must exceed 127. For the coverages selected, their `red` and `nir` ("near infrared") components are added pixelwise. The result image is encoded in TIFF and shipped back to the client together with accompanying metadata.

Among the WCPS operators available are *induced operations* to apply a cell operation simultaneously to all cell values; *subsetting operations*: cutout ("trim") and slicing ("section"); *condensers*, for deriving summary information; *array constructors* which allow to construct, e.g., image transformations; *reprojection*, i.e., transformation into another coordinate reference system; and metadata functions.

The expressive power of this language allows statistics, imaging, and signal processing operations, such as aggregation/roll-up, modes, convolutions and filter kernels. Recursion tentatively has been omitted to keep with a safe language. These characteristics motivate why WCPS has been dubbed “SQL for raster data”.

2.3 WCPS Implementation

The WCPS reference implementation consists of a service stack as shown in Fig. 1. A Java servlet accepts XML requests, which must conform to the WCPS schema, and returns coverage results. The rasdaman array database system [1] receives queries translated from the WCPS requests and generates the response. Array data are stored in some relational database, in the reference implementation is open-source PostgreSQL. Coverage results consist of an XML summary plus binary coverage data, all within a multipart/ mixed HTTP response.

3 Demonstration

The demonstration will present the reference implementation of the WCPS standard through interactive navigation, extraction, and ad-hoc analysis. A number of embedded clients will be used, each one tailored to different data categories, such as 1-D time series, 2-D maps, and 3-D/4-D geo-scientific data sets. Data sets used encompass all spatio-temporal dimensions: 1-D time series; 2-D satellite and seafloor maps; 3-D geophysics data; 4-D climate simulation data.

The demonstration’s objective is to present Internet services for the emerging field of open, large-scale, multi-dimensional raster services, and give a condensed overview on techniques, issues, standards, and applications.

References

1. Baumann, P.: Large-Scale Raster Services: A Case for Databases. Invited keynote, 3rd Intl Workshop on Conceptual Modeling for Geographic Information Systems (CoMoGIS), Tucson, USA, 6 - 9 November 2006. In: Roddick, J. et al (eds): Advances in Conceptual Modeling - Theory and Practice, Springer 2006, pp. 75 – 84.
2. Baumann, P. (ed.): Web Coverage Processing Service (WCPS) Implementation Standard. OGC document 08-068, 2008
3. Botts, M., Robin, A., Davidson, J., Simonis, I. (eds.): Sensor Web Enablement Architecture. OGC document 06-021r1, 2006
4. N.n.: Abstract Specification Topic 6: Schema for coverage geometry and functions, OGC 07-011 (ISO 19123), 2007
5. Whiteside, A., Evans, J. (eds.): Web Coverage Service (WCS) Implementation Specification, version 1.1.2. OGC document 07-067r5, 2008
6. de La Beaujardiere, J. (ed.): Web Map Service. OGC document 03-109r1, 2003
7. N.n.: Geographic information – Schema for coverage geometry and functions. ISO DIS 19123, TC 211/SC/WG 2, 2002