EarthServer: Federating Datacubes Intercontinentally

Geospatial IG, RDA P8, Denver, USA, 2016-sep-16

Peter Baumann
Jacobs University | rasdaman GmbH
baumann@rasdaman.com

[co-funded through EU EarthServer 1/2, PublicaMundi]
Why Standards?

- interoperability, no vendor lock-in
- construction guidance
- forward oriented
Features & Coverages

- The basis of all: geographic feature
- Special kind of feature: coverage
  - aka space-time varying phenomenon
  - regular & irregular grids, point clouds, meshes
- Usually, Big Geo Data are coverages
Modeling Coverages:

OGC Coverage Implementation Schema 1.1
Modeling Coverages:
OGC Coverage Implementation Schema 1.1

- OGC **abstract** coverage definition:
  OGC Abstract Topic 6 = ISO 19123
- OGC **concrete** coverage definition:
  Coverage Implementation Schema (CIS) = GMLCOV, renamed
Coverage Definition

«Feature Type»

GML::Feature

«Feature Type»

Coverage

«Data Type»

DomainSet

«Data Type»

SWE Common::DataRecord

«Data Type»

RangeSet

0..1 metadata

[OGC 09-146r2]
<generalGridCoverage ... gml:id="CIS_001">

  <domainSet>
    <generalGrid srsName="http://www.opengis.net/def/crs-compound?1=http://www.opengis.net/def/crs/EPSG/0/4979
                &amp;2=http://www.opengis.net/def/crs/OGC/0/Ansi1Date"
                 axisLabels="Lat Long h date">
      <regularAxis axisLabel="Lat" uomLabel="deg" lowerBound="40" upperBound="60" resolution="10"/>
      <regularAxis axisLabel="Long" uomLabel="deg" lowerBound="-10" upperBound="10" resolution="10"/>
      <irregularAxis axisLabel="h" uomLabel="m">
        <c> 0</c>
        <c>100</c>
      </irregularAxis>
      <irregularAxis axisLabel="date" uomLabel="d">
        <c>2015-12-01</c>
        <c>2015-12-02</c>
      </irregularAxis>
    </generalGrid>
  </domainSet>

  <gridLimits srsName="http://www.opengis.net/def/crs/OGC/0/Index4D" axisLabels="i j k l">
    <indexAxis axisLabel="i" lowerBound="0" upperBound="2"/>
    <indexAxis axisLabel="j" lowerBound="0" upperBound="2"/>
    <indexAxis axisLabel="k" lowerBound="0" upperBound="1"/>
    <indexAxis axisLabel="l" lowerBound="0" upperBound="1"/>
  </gridLimits>
</generalGridCoverage>

A Simple Coverage, in GML

also binary formats
CIS 1.1: Next-Gen Coverages

- Coverage Implementation Schema 1.1
  = bw-compatible evolution of CIS/GMLCOV 1.0 + GML 3.3
CIS 1.1: Next-Gen Coverages

- Coverage Implementation Schema 1.1

- General Grids
  - grid types → axis types
  - Regular + irregular + warped (generalizing GML 3.3), SensorML

```xml
<cis:domainSet>
  <cis:GeneralGrid srsName="http://www.opengis.net/def/crs/compound?1=http://www.opengis.net/def/crs/EPSG/0/4327&amp;2=http://www.opengis.net/def/crs/OGC/0/AnsiDate"
      uomLabels="deg deg m d" axisLabels="Lat Long h ansi" srsDimension="4">
    <cis:regularAxis axisLabel="Lat" lowerBound="-90" upperBound="-80" resolution="5"/>
    <cis:regularAxis axisLabel="Long" lowerBound="0" upperBound="10" resolution="5"/>
    <cis:irregularAxis axisLabel="h" lowerBound="0" upperBound="100" directPositions="0,100"/>
    <cis:irregularAxis axisLabel="ansi" lowerBound="2015-12-01" upperBound="2015-12-02"
      directPositions="2015-12-01, 2015-12-02"/>
  </cis:GeneralGrid>
</cis:domainSet>
```
CIS 1.1: Next-Gen Coverages

- Coverage Implementation Schema 1.1
- General Grids
- Interpolation added to rangeType
  - coverage knows desirable/allowed interpolations

```xml
<interpolationRestriction>
  <allowedInterpolation>http://www.opengis.net/def/interpolation/OGC/1/nearest-neighbor</allowedInterpolation>
  <allowedInterpolation>http://www.opengis.net/def/interpolation/OGC/1/linear</allowedInterpolation>
</interpolationRestriction>
```
CIS 1.1: Next-Gen Coverages

- Coverage Implementation Schema 1.1
- General Grids
- Interpolation added to rangeType
- Additional representations
  - domain/range
    + "geometry/value pair" lists
    + partitions ("tiles")
  - More containers: multipart/MIME, zip, GMLJP2, GeoPackage, ...
  - More encodings: JSON + RDF
A Simple Coverage, in JSON

```
{   "type": "CoverageByDomainAndRangeType",
    "domainSet":{ 
        "type": "DomainSetType",
        "generalGrid":{ 
            "type": "GeneralGridCoverageType",
            "srsName": "http://www.opengis.net/def/crs/OGC/0/Index2D",
            "axisLabels": ["i", "j"],
            "axis": [{       "type": "IndexAxisType", "axisLabel": "i", "lowerBound": 0, "upperBound": 2 },
            {       "type": "IndexAxisType", "axisLabel": "j", "lowerBound": 0, "upperBound": 2 }]
        }
    },
    "rangeSet": {      "type": "RangeSetType",
        "dataBlock": { "type": "VDataBlockType", "values": [1,2,3,4,5,6,7,8,9] } },
    "rangeType": {       "type": "DataRecordType",
        "field": [{       "type": "QuantityType",
            "definition": "ogcType:unsignedInt",
            "uom": { "type": "UnitReference", "code": "10^0" } } ]
    }
}
```
A Simple Coverage, in RDF

<http://www.opengis.net/cis/1.1/examples/CIS_05_2D>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.opengis.net/cis/1.1/CoverageByDomainAndRangeType>.

<http://www.opengis.net/cis/1.1/examples/CIS_05_2D>
<http://www.opengis.net/cis/1.1/domainSet>
<http://www.opengis.net/cis/1.1/examples/CIS_DS_05_2D>.
<http://www.opengis.net/cis/1.1/examples/CIS_DS_05_2D>
<http://www.opengis.net/cis/1.1/generalGrid>
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_05_2D>.
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_05_2D>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.opengis.net/cis/1.1/DomainSetType>.
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_05_2D>
<http://www.opengis.net/cis/1.1/axis>
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_I_05_2D>.
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_05_2D>
<http://www.opengis.net/cis/1.1/axis>
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_J_05_2D>.
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_05_2D>
<http://www.opengis.net/cis/1.1/axisLabels>
<http://www.opengis.net/cis/1.1/axisLabels0>.
<http://www.opengis.net/cis/1.1/axisLabels0> <http://www.w3.org/1999/02/22-rdf-syntax-ns#first> "i".
<http://www.opengis.net/cis/1.1/axisLabels1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#first> "j".
<http://www.opengis.net/cis/1.1/axisLabels1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#rest> <http://www.w3.org/1999/02/22-rdf-syntax-ns#nil>.
<http://www.opengis.net/cis/1.1/examples/CIS_DS_GG_05_2D> <http://www.opengis.net/cis/1.1/srsName> <http://www.opengis.net/def/crs/OGC/0/Index2D>.

A Simple Coverage, in RDF also binary formats
Encoding Coverages

- **Single file encoding:**
  - Informationally complete: GML, JSON, RDF, …
  - Other Format: GeoTIFF, NetCDF, JPEG2000, …

- **Multipart: container( “header” + file1 + file2 + … )**
  - Multipart/MIME, zip, GMLJP2, …
Serving Coverages:
Web Coverage Service (WCS),
Web Coverage Processing Service (WCPS)
Web Coverage Service (WCS)

- **WCS Core**: access to spatio-temporal coverages & subsets
  - subset = trim | slice

- **WCS Extensions**: optional facets
  - Functionality
  - Protocols: KVP, POST, SOAP, (REST), ..

Large, growing implementation basis: rasdaman, GDAL, QGIS, OpenLayers, OPeNDAP, MapServer, GeoServer, GMU, NASA WorldWind, EOxServer; Pyxis, ERDAS, ArcGIS, ...
WCS Core GetCoverage

- Download a coverage (or a subset thereof), values guaranteed unaltered
  - Ex: "download coverage c001"
    - http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0 & REQUEST=GetCoverage & COVERAGEID=c001

- Ex: "coverage c001, lat/long cutout, time slice t=2009-11-06T23:20:52"
  - http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0 & REQUEST=GetCoverage & COVERAGEID=c001 & SUBSET=Long(100,120) & SUBSET=Lat(50,60) & SUBSET=time("2009-11-06T23:20:52")

- Ex: “coverage c001, in GeoTIFF”
  - http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0 & REQUEST=GetCoverage & COVERAGEID=c001 & FORMAT="image/tiff"
WCS Extension – Range Subsetting

- Extract range components

Ex:  
http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0
& REQUEST=GetCoverage & COVERAGEID=c001
& FORMAT=image/jpeg
& RANGESUBSET=red

- or:  ...
- or:  ...
- or:  ...
- or:  ...
- or:  ...

http://standards.rasdaman.com
WCS Extension – Transaction

- **WCS-T**: Modify coverage offerings on a server via Web

- New requests: *InsertCoverage* + *DeleteCoverage* + *UpdateCoverage*
  - incl partial update

- Ex: `http://www.acme.com/wcs
  ? SERVICE=WCS & VERSION = 2.0
  & REQUEST=InsertCoverage
  & COVERAGEREF=http://bcme.com/archive/hurricane.nc`

[http://standards.rasdaman.com](http://standards.rasdaman.com)
Web Coverage Processing Service (WCPS)

= spatio-temporal datacube analytics language

- "From MODIS scenes M1, M2, M3: difference between red & nir, as TIFF"
  - …but only those where nir exceeds 127 somewhere

```python
for $c$ in ( M1, M2, M3 )
  where
    some( $c$.nir > 127 )
  return
    encode( $c$.red - $c$.nir, "image/tiff" )
```

[JacobsU, FhG; NASA; data courtesy BGS, ESA]

http://standards.rasdaman.com
Outlook: Integrated Data + Metadata Queries

- WCPS 2.0 (draft): coverage expressions + XQuery

- Ex1: „difference of red bands for all coverages on Hubei“

```xml
for $c in doc("http://acme.com/wcs")//coverage,
  $d in doc("http://bcme.com/wcs")//coverage
where some( $c.nir > 127 ) and $d/metadata/@city = "Praha"
return encode( $c.red - $d.red, "image/tiff" )
```

- Ex2: „name & location of coverages showing some phenomenon“

```xml
for $c in doc("WCPS")//coverage[some( $c.nir - $c.red > 0 )]
return
  <id> { $c/@id } </id>
  <area> { $c/boundedBy } </area>
```
Implementation: Grid Coverages
= „raster data manager“: SQL+ n-D arrays
- pioneer Array Database System

- Scalable parallel “tile streaming” architecture

- Mature, in operational use
  - OGC WCS Core Reference Implementation
  - 130+ TB databases
  - 1 query → 1000+ cloud nodes
Adaptive Partitioning

- **Any partitioning ("tiling")**
  - 130+ TB datacubes
  [IJDE 2015]

- **Why irregular tiling?**

[OpenStreetMap]
Parallel / Distributed Query Processing

```
select
    max((A.nir - A.red) / (A.nir + A.red))
- max((B.nir - B.red) / (B.nir + B.red))
- max((C.nir - C.red) / (C.nir + C.red))
- max((D.nir - D.red) / (D.nir + D.red))
from A, B, C, D
```

1 query → 1,000+ cloud nodes

[SIGMOD DANAC 2014]
Hadoop? One size does not fit all

- “Since it was not originally designed to leverage the structure its performance is suboptimal” [Daniel Abadi]
- U Madison / GMU benchmark [AGU 2015]

[Common Sense]
Just because you can, doesn't mean you should.

EarthServer: Datacubes At Your Fingertips

- Intercontinental initiative, 3+3 years: EU + US + AUS

- Agile Analytics on Earth & Planetary datacubes
  - Rigorously standards: OGC WMS + WCS + WCPS
  - EU rasdaman + NASA WorldWind
  - 100s of TB sites now, next: 1+ PB

- Goal: global federation
  - single common information space, location transparent
  - unified standards-based access
Conclusion
Standardization Convergence

- OGC “Big Earth Data” suite:
  - Coverage Implementation Schema & WCS with WCPS

- ISO TC211:
  - OGC CIS 1.1 → 19123-2
  - 19123, modernized → 19123-1
  - OGC WCS 2 → WCS

- ISO SC32: SQL/MDA („Multi-Dimensional Arrays“)

- INSPIRE:
  - WCS as Coverage Download Service
  - rasdaman as reference implementation
Conclusion

„One cube says more than a million images“

- OGC Coverages for pixel-level interoperability
  - consensus: stds bodies, major tools & vendors
  - From simple access (WCS Core) to agile analytics (WCPS)
- rasdaman: scalable array engine, blueprint for datacube standards
- EarthServer: WCS-based datacube federation

[gamingfeeds.com]