Supported by EU FP7 eInfrastructure, contract 283610 EarthServer



Ingredients for an Integrated Dinner: Parsley, Sage, Rosemary and Thyme EGU 2013

Vienna, Austria, 2013-apr-12

Peter Baumann¹, Marina Sahakyan, Giorgios Kakaletris, Panagiota Koltsidas, Thanasis Perperis

Jacobs University, rasdaman GmbH, Athena Research



[Wikipedia]



Integration :: EGU 2013, Vienna :: ©2013 Peter Baumann



Our Kontrapunktic Melodies

Data:

- Large in volume "Big Data"
- semantic-poor
- difficult to interpret, at best statistics
- only for clumsy download
- Search only through previously extracted metadata

Metadata:

- Small in volume
- rich in content & semantics
- Manifold techniques for interpretation, such as ontologies
- flexibly queryable
- rich body of search technologies
 - SQL, OpenSearch, ...

...so much about history; can we do better?





EarthServer: Big Earth Data Analytics

- Scalable On-Demand Processing for the Earth Sciences
 - EU funded, 3 years, 5.85 mEUR
 - Platform: rasdaman (Array Analytics server)
 - Distributed query processing, integrated data/metadata search 3D clients 4
- Strictly open standards: OGC WMS+WCS+WCPS; W3C Xquery; X3D
- 6 * 100+ TB databases for all Earth sciences + planetary science







Big Earth Data Have a Name

Core element in OGC: geographic feature

- = abstraction of some real-world phenomenon
- associated with a location relative to Earth
- Special kind of feature: coverage
 - space-time varying multi-dimensional phenomenon
 - Typical representative: raster image
 - ...but there is more!
- Typically, coverages are Big Data







Integration :: EGU 2013, Vienna :: ©2013 Peter Baumann



Web Coverage Service (WCS)

Core: Simple & efficient access to multi-dimensional coverages



see later

- WCS Extensions for additional functionality facets
 - "band extraction", scaling, reprojection, interpolation, query language,...
- Application Profiles define domain-oriented bundling
 - EO, MetOcean, ...





Ex: Spatio-Temporal Image Services

rasdaman / Oracle, Solaris



[Diedrich et al 2001]



Integration :: EGU 2013, Vienna :: ©2013 Peter Baumann



WCS Service Model: Structure





WCS Service Model: Operations

- GetCapabilities
 - what service extensions?
 - What formats? CRSs?
 - What coverages?



- DescribeCoverage
 - coverage metadata



- GetCoverage
 - coverage, or subset thereof





Let's Try Some XPath!





"Identifiers of all coverages offered"

/CoverageOfferings/OfferedCoverage/coverage/@id

"All formats supported by this server"

/CoverageOfferings/ServiceIdentification/ServiceMetadata/
formatSupported/text()

"spatial extent of coverage X"

/CoverageOfferings/OfferedCoverage/coverage[@id="X"]/domainSet

"...and its pixel values"

/CoverageOfferings/OfferedCoverage/coverage[@id="X"]/rangeSet

- "...a histogram" ? "...percentile aggregation" ? ...etc.
 - Multi-dimensional arrays not well supported by XML / XQuery





OGC Web Coverage Processing Service (WCPS)

- High-level query language on n-D spatio-temporal gridded data
 - Ex: Time series, image/signal processing, summary data, sensor fusion, pattern mining
 - Demo: www.earthlook.org
- OGC standard since 2010 [OGC 08-068r2]





WCPS By Example

 "From MODIS scenes M1, M2, and M3, the absolute of the difference between red and nir, in GeoTIFF"

• ...but only those where nir exceeds 127 somewhere

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





...and now: Integration

merge WCPS with XQuery

```
let $r := doc("WCS")
    //coverage/metadata/region/[ @name = "Austria" ]
for $c in doc("WCS")
    //coverage/[ some( $c.nir > 127 and $r ) ] )
```

return

```
encode( abs( $c.red - $c.nir ), "image/tiff" )
```

for \$c in doc("WCPS")//coverage/[some(\$c.nir > \$c.red)]
return

```
<id> { $c/@id } </id>
<area> { $c/boundedBy } </area>
```

- Implementation: federation of eXist + rasdaman in progress
 - Jacobs University & Athena Research Lab





The rasdaman Raster Analytics Server

- Array DBMS for massive n-D raster data
 - Stored in standard database or preexisting file archive
 - Scalable through massive optimization / parallelization
- In operational use
 - Has pioneered the research field of Array Databases

 http://en.wikipedia.org/wiki/Array_DBMS
 - GEOSS component for Big Data Analytics
 - OGC WCS Core, OGC WCPS reference implementation



www.rasdaman.org







3D Database Visualization

- Problem: coupling DB / visualization
- Approach:
 - deliver RGBA image to X3D client, transparency as height
 - Feed directly into client GPU

```
select
    encode(
        { red: (char) s.b7[x0:x1,x0:x1],
        green: (char) s.b5[x0:x1,x0:x1],
        blue: (char) s.b0[x0:x1,x0:x1],
        alpha: (char) scale( d, 20 )
        },
        "png"
    )
from SatImage as s, DEM as d
```







Conclusion

- Data / metadata separation is history, overcoming it a core underlying challenge of "Big Data"
- Coverages = sensor, image, simulation, & statistics data

 a main source of *Big Data* in Earth Sciences
 Petrol industry has "more bytes than barrels"
- Integration of data means integration of operations
- For flexibility, no way around high-level query languages
 - Semantic interoperability
 - Visual clients can hide QL, servers can parallelize
- **EarthServer** working on it...

