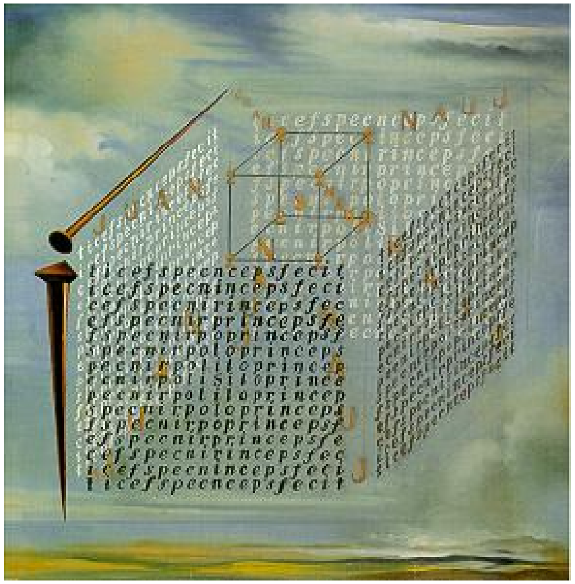


Preserving Geo-Scientific Data Assets Through Service Interoperability

PV 2009

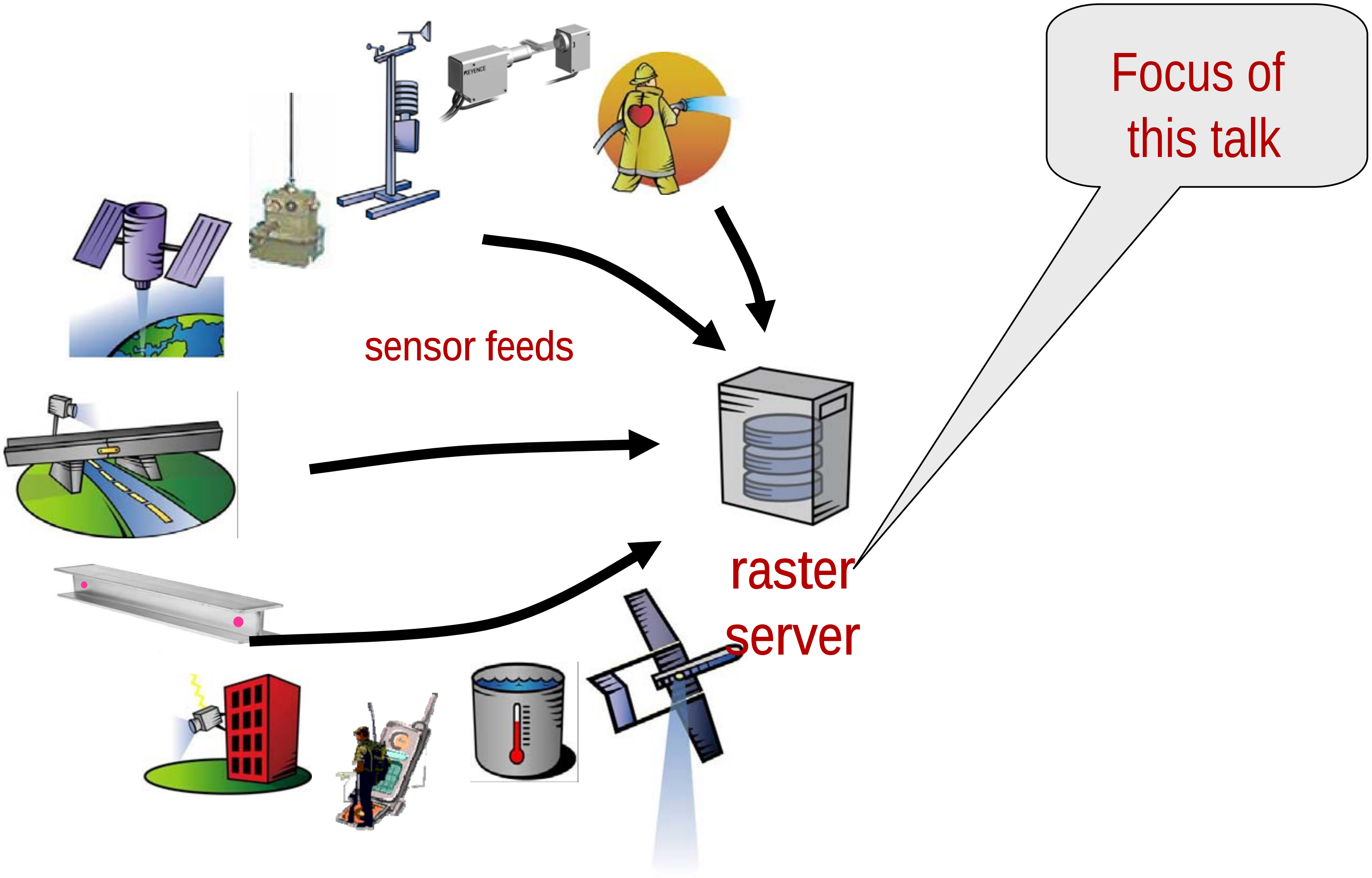
2009-dec-02, Villanueva de la Canada, Spain

Peter Baumann, Andrei Aiordachioaie
Jacobs University Bremen

Disclaimer:
our opinions, not necessarily those of OGC

- Motivation
- WCS & WCPS
- Implementation & optimization
- Summary & Outlook

Sensor Data Ev'ryWhere...

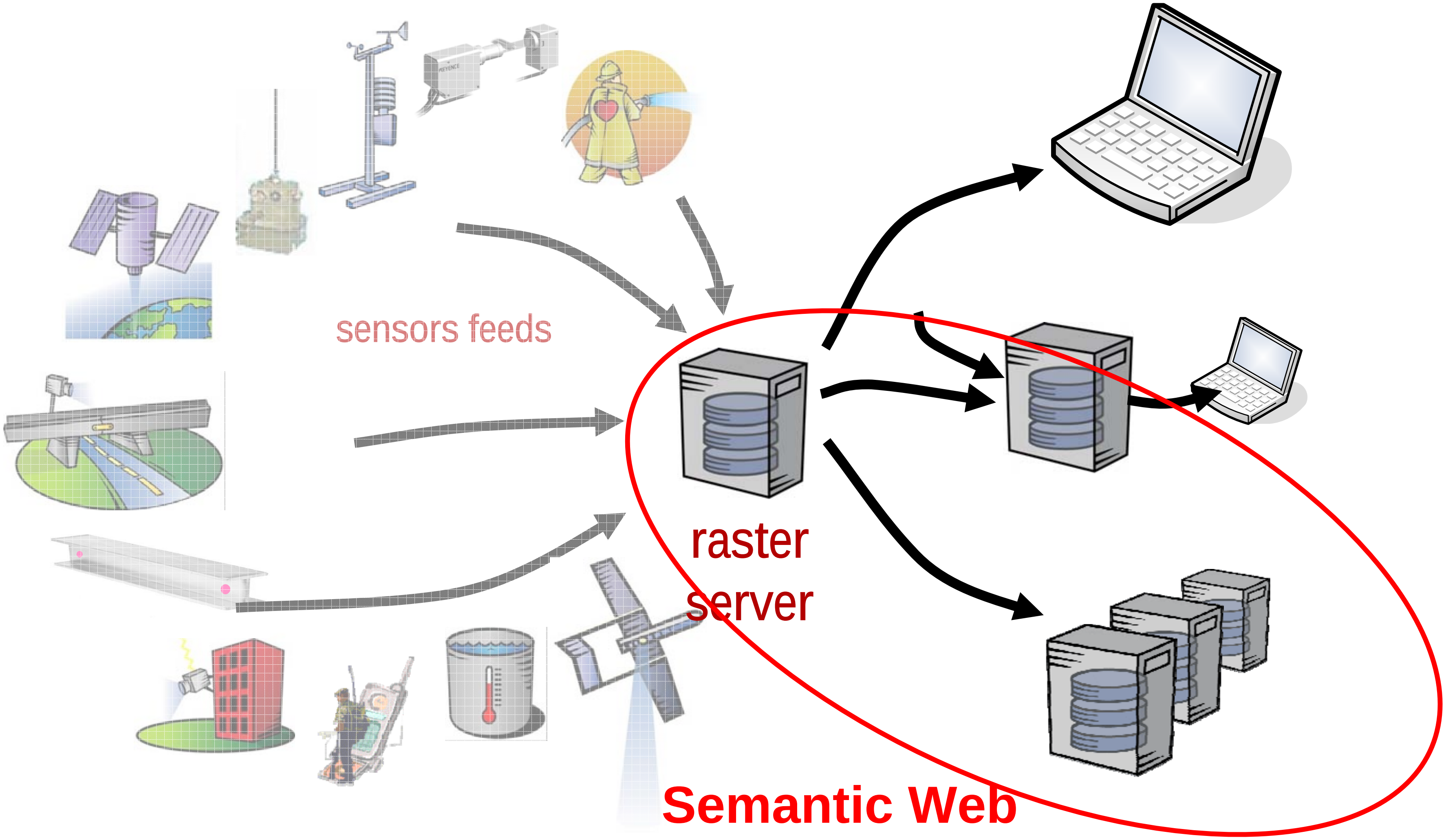


sensor feeds

raster server

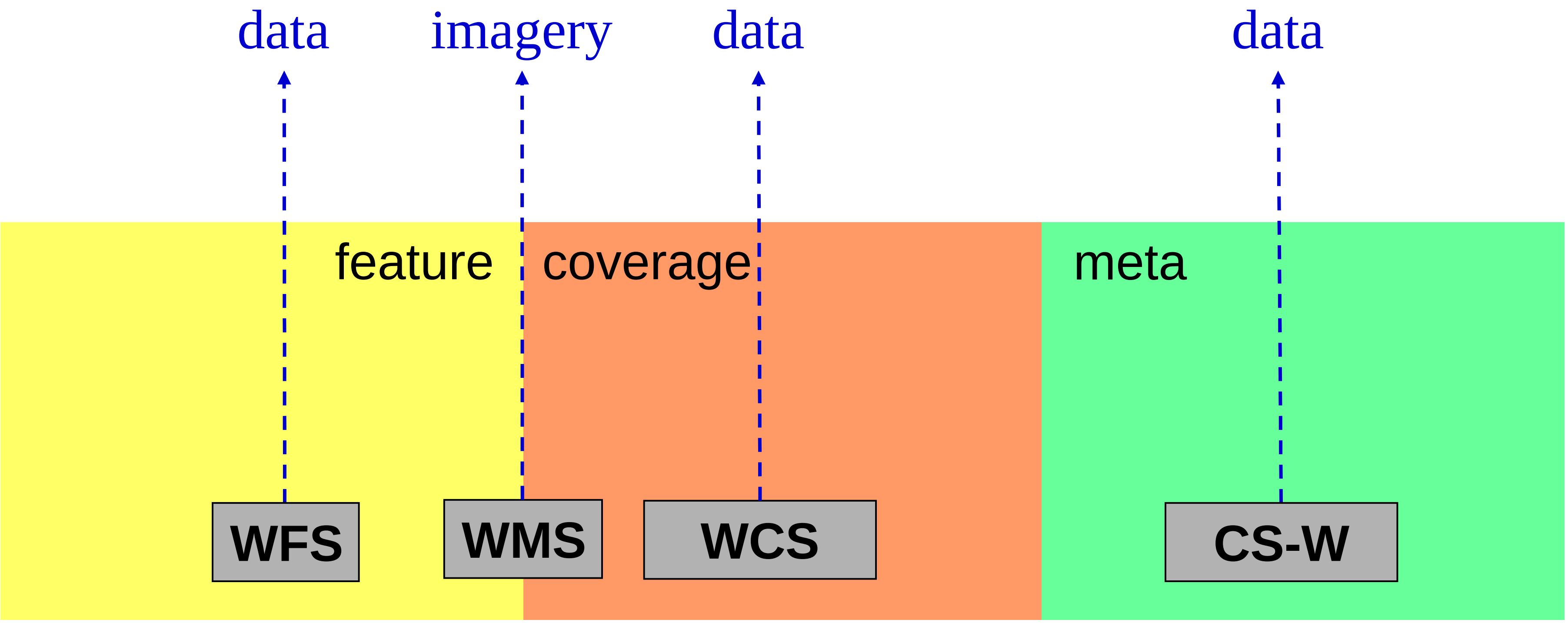
Focus of this talk

Sensor Data Ev'rywhere…



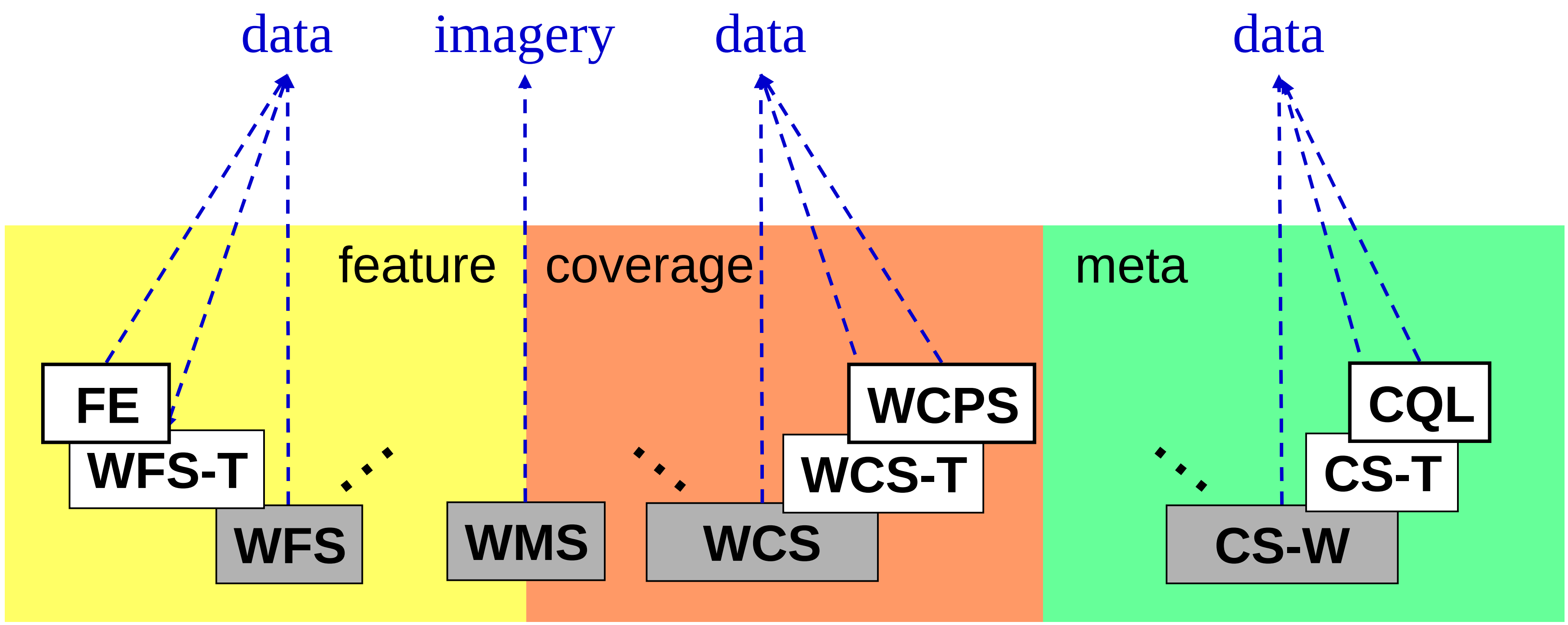
(Part of) The OGC Quilt

Open GeoSpatial Consortium, www.opengeospatial.org



(Part of) The OGC Quilt

JACOBS UNIVERSITY

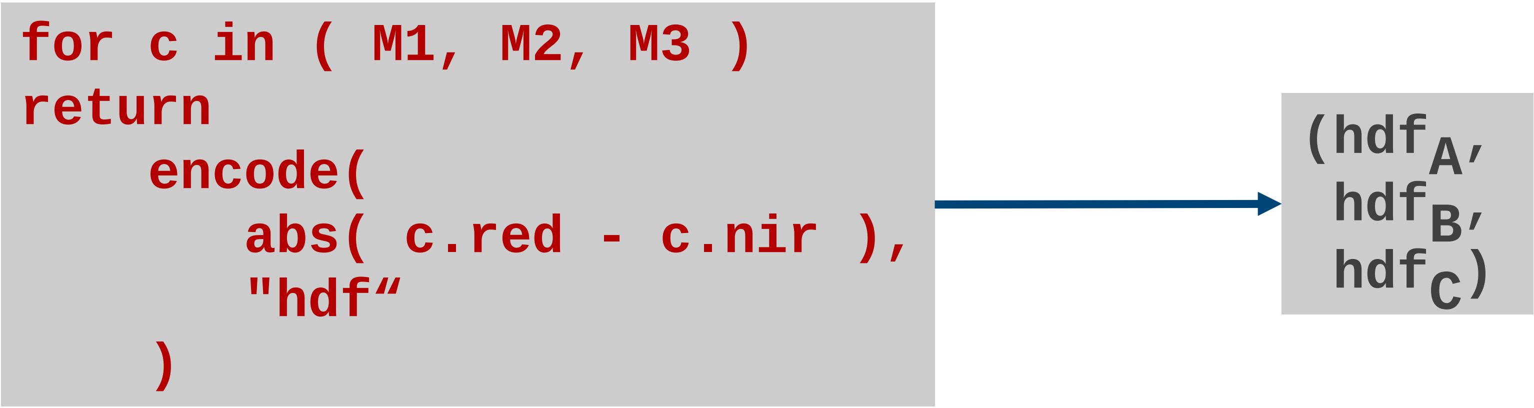


Web Coverage [Processing] Service

- Coverage = "space-time varying phenomenon"
  - ISO 19123, OGC Abstract Topic 6; GML
- WCS: simple access
  - subsetting, scaling, reprojection, encoding
- WCPS: raster language
  - Expressive power: image & signal processing, statistics
  - ...except recursion → safe
  - Formal evaluation model

WCPS By Example

- "From MODIS scenes **M1**, **M2**, and **M3**, the absolute of the difference between **red** and **nir**, in HDF-EOS"



```
for c in ( M1, M2, M3 )
return
    encode(
       abs( c.red - c.nir ),
       "hdf“
    )
```

(hdf_A, hdf_B, hdf_C)

WCPS By Example

- "From MODIS scenes **M1**, **M2**, and **M3**, the absolute of the difference between **red** and **nir**, in HDF-EOS"
  - ...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 ),
       "hdf“
    )
```

(hdf_A, hdf_C)

WCPS By Example

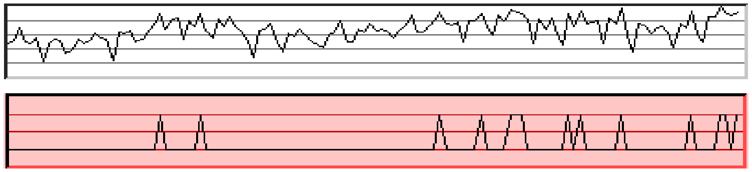
- "From MODIS scenes **M1**, **M2**, and **M3**, the absolute of the difference between **red** and **nir**, in HDF-EOS"
  - ...but only those where nir exceeds 127 somewhere
  - ...inside region R

```
for c in ( M1, M2, M3 ),
    r in ( R )
where
    some( nir > 127 and R )
return
    encode
       abs( c.red - c.nir ),
       "hdf“
    )
```

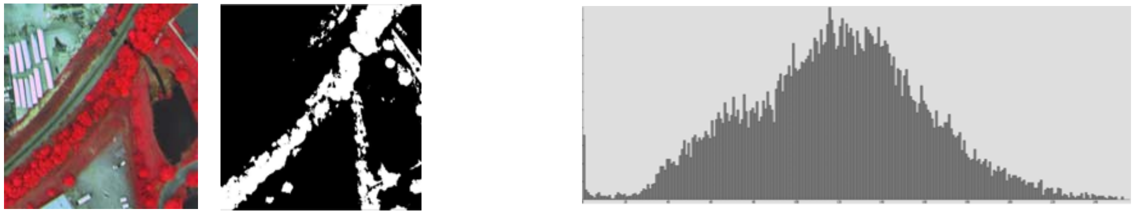
(hdf$_A$)

WCPS: Use Cases

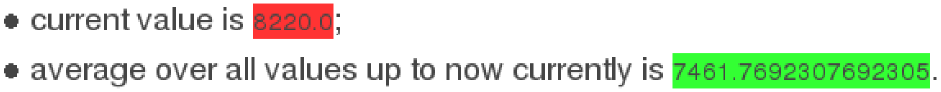
- „SQL for coverages“: ad-hoc navigation, extraction, aggregation, analysis
- Time series



- Image processing

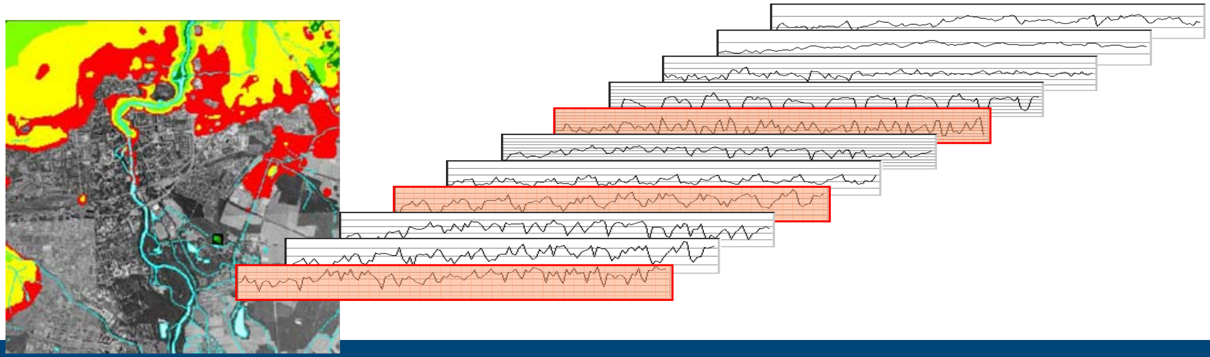


- Summary data



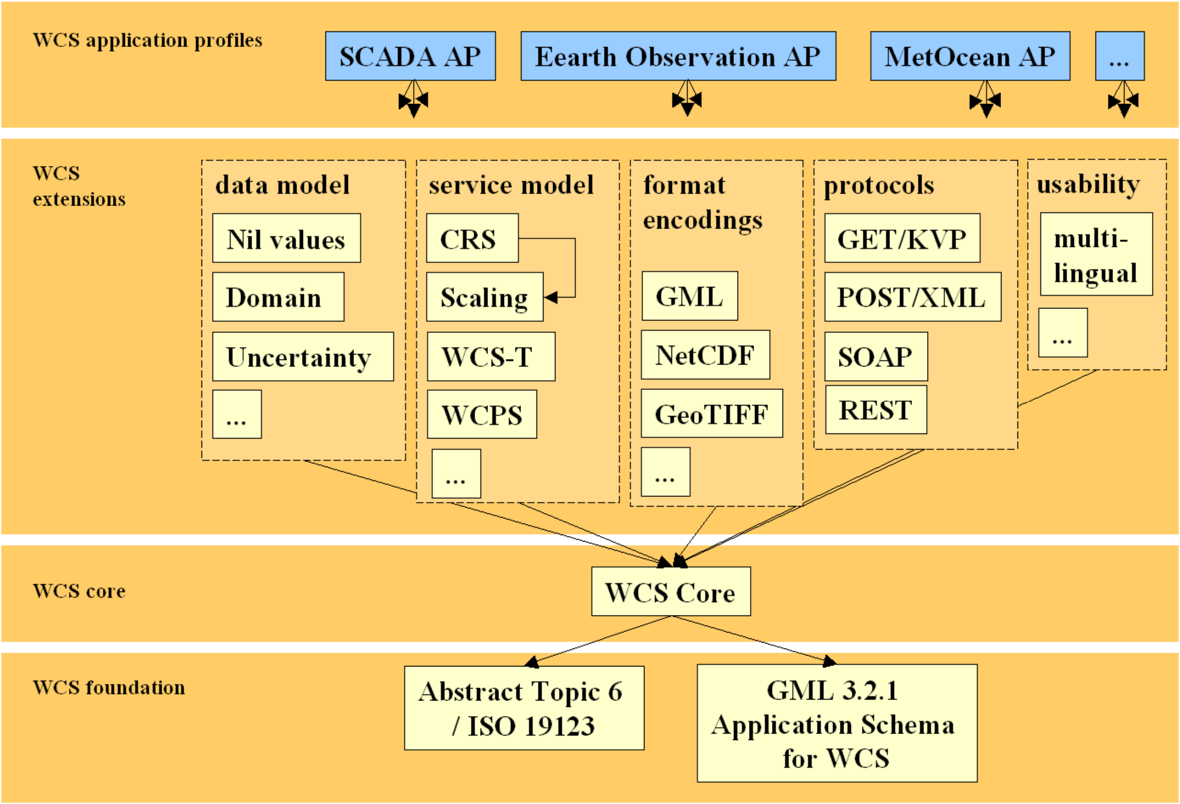
- current value is 8220.0;
- average over all values up to now currently is 7461.7692307692305.

- Sensor fusion & pattern mining



Outlook: WCS 2.0

- public comment & e-vote starting Dec-10, 2009
- GML coverage model
- n-D coverages, incl. non-space/time axes
- Formal semantics definition (GML, XPath)

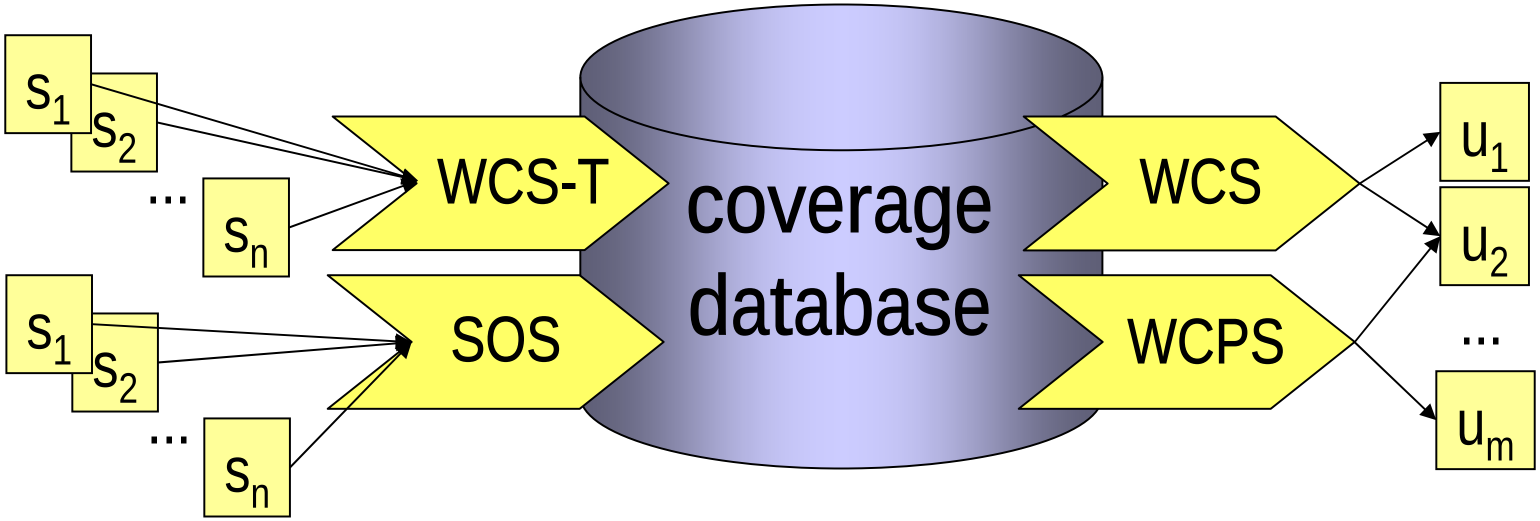


source: [OGC 09-153]

- Crisp: Core 17 p, KVP extension 3 p, ...
- Harmonization WCS – GML – SWE

Multi-Source Merge & Delivery

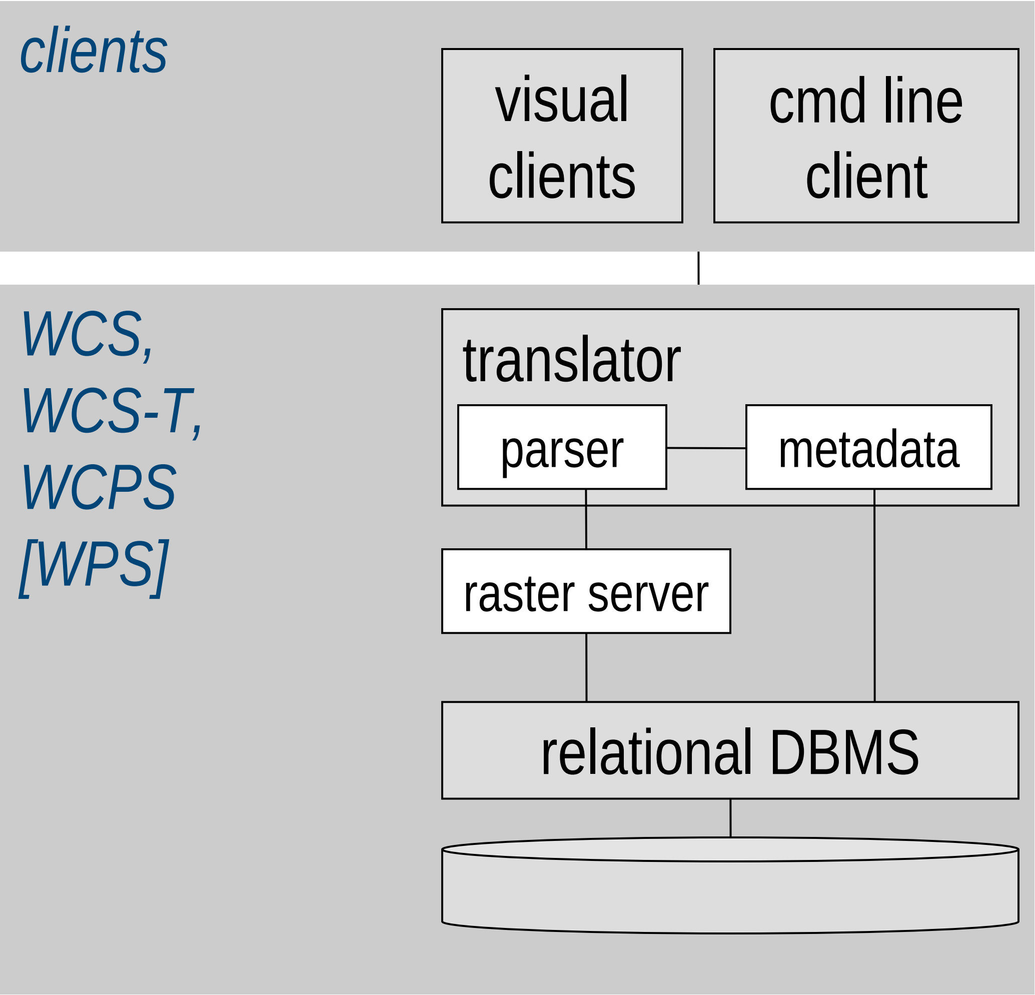
- Scenario 1: ground segment: integrated WCS + WCS-T + WCPS
  - ESA Heterogeneous Mission Access Follow-On (HMA-FO)
- Scenario 2: on board: Deploy WCPS on satellite
  - NASA



So Why "Semantic"?

- Formal semantics for language allows machine-machine communication, no human intervention required
  - Clients (other services?) can compose requests
- Ex:
  - Client: "Let's see, which server can handle reprojection / exponentials / ... ?"
  - In a cloud: "hm, this subexpression I better pass on to node X"
  - "Evaluating this request will take an estimated 3.5min, over 500 objects match."
  - "Sorry, this request's complexity exceeds your CPU quota"

PetaScope:
WCPS Reference Implementation



clients

visual clients

cmd line client

WCS,
WCS-T,
WCPS
[WPS]

translator

parser

metadata

raster server

relational DBMS

Java 1.6;
Swing, Jgraph, GWT

Java 1.6;
Tomcat, Xerces, ANTLR

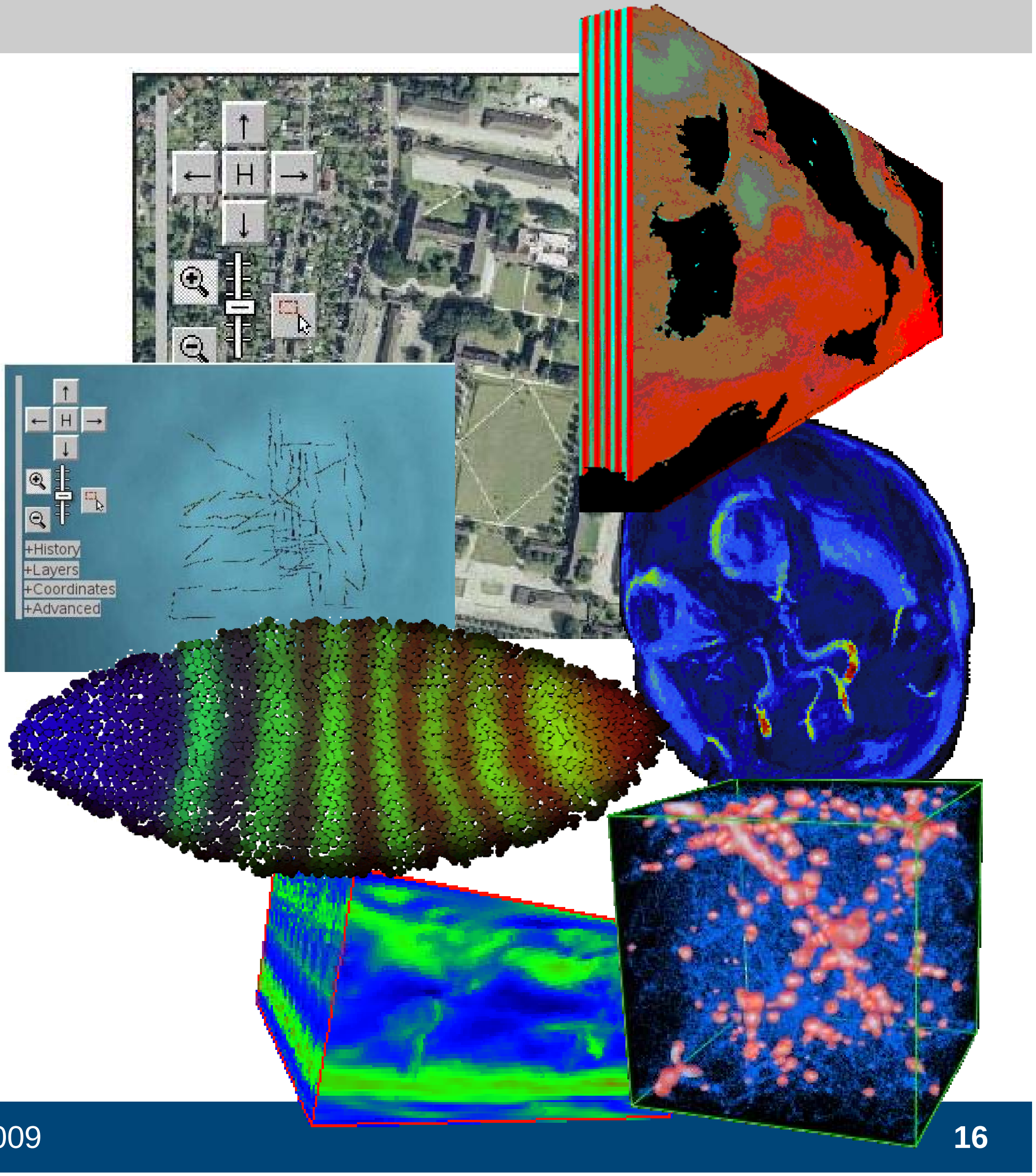
rasdaman

PostgreSQL

www.petascope.org

Optimization

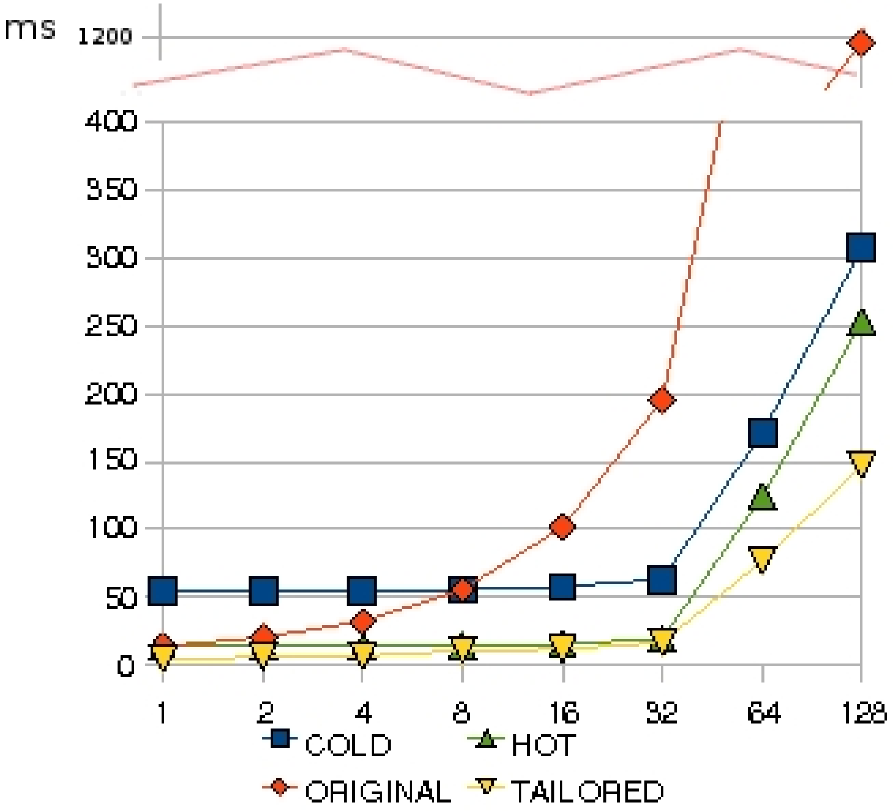
- Adaptive tiling
- Adaptive compression
- Multi-dimensional indexing
- Distributed query processing
- Query rewriting
- Pre-aggregation
- Physical operator clustering
- Transparent tape integration
- Just-in-time compilation
- GPU processing
- Tile caching
- ...



Optimization Example:
Just-In-Time Compilation

```
for x in (float_matrix)
return x*x*...*x
```

- Observation: interpreted mode slows down
- Approach:
  - cluster suitable operations
  - compile & dynamically bind
- Benefit:
  - Speed up complex, repeated operations
- Variation:
  - compile code for GPU

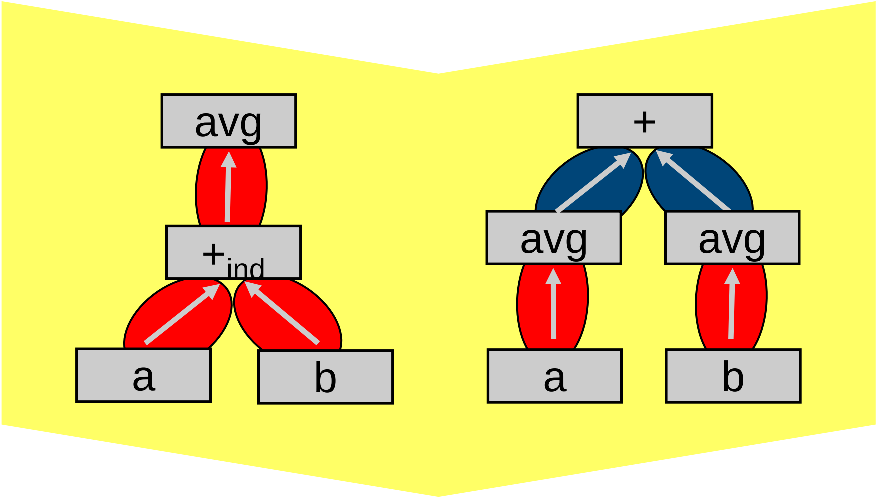


Times [ms] for 512^2 * n ops

[Jucovschi, Stancu-Mara 2008]

Optimization 2: Query Rewriting

```
for a in ( A ), b in ( B )
return avg( a + b )
```



```
for a in ( A ), b in ( B )
return avg( a ) + avg( b )
```

🔴 Tile stream
high traffic

🔵 Scalar stream
low traffic

Summary

- OGC coverage standards for open, interoperable access
  - Starting WCS 2.0 GML based → harmonization
- Language-based interfaces add flexibility + optimizability + scalability
  - Data interopability → service interoperability
- Our research: flexible, scalable raster services
- Websites:
  - www.petascope.org
  - www.earthlook.org