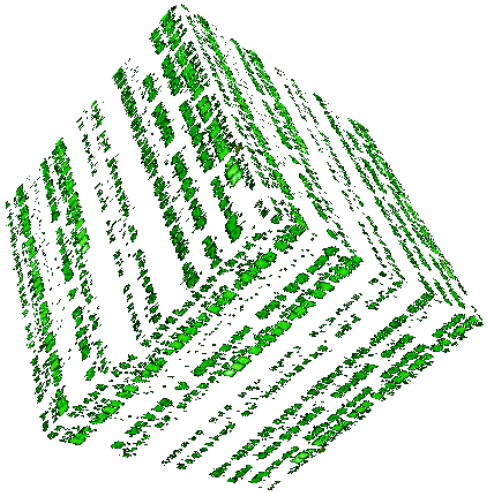


EarthServer @ RDA

RDA-DE/DINI Workshop, Karlsruhe, 2015-may-29

Peter Baumann

Jacobs University | rasdaman GmbH

baumann@rasdaman.com

RDA: Co-chairing
- Big Data IG
- Geospatial IG

OGC co-chairing:
- editor, „Big Geo Data“ stds
- BigData.DWG

ISO:
- SC 32 WG 3 SQL
-TC211 Geo Imagery

INSPIRE:
- Coverage Download Services

Co-funded by
the European Union

[gamingfeeds.com]

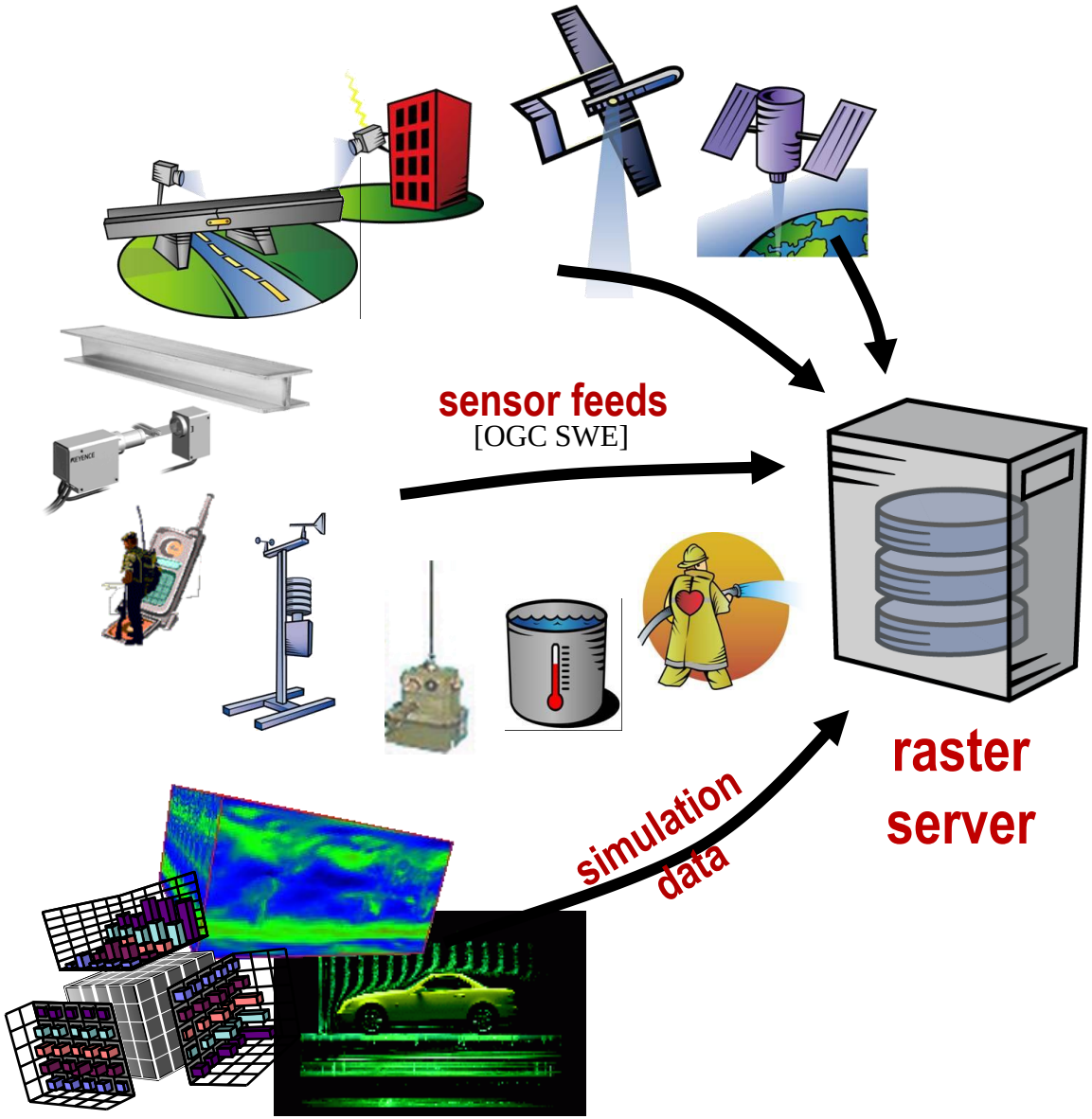
rasdaman raster data management

JACOBS UNIVERSITY

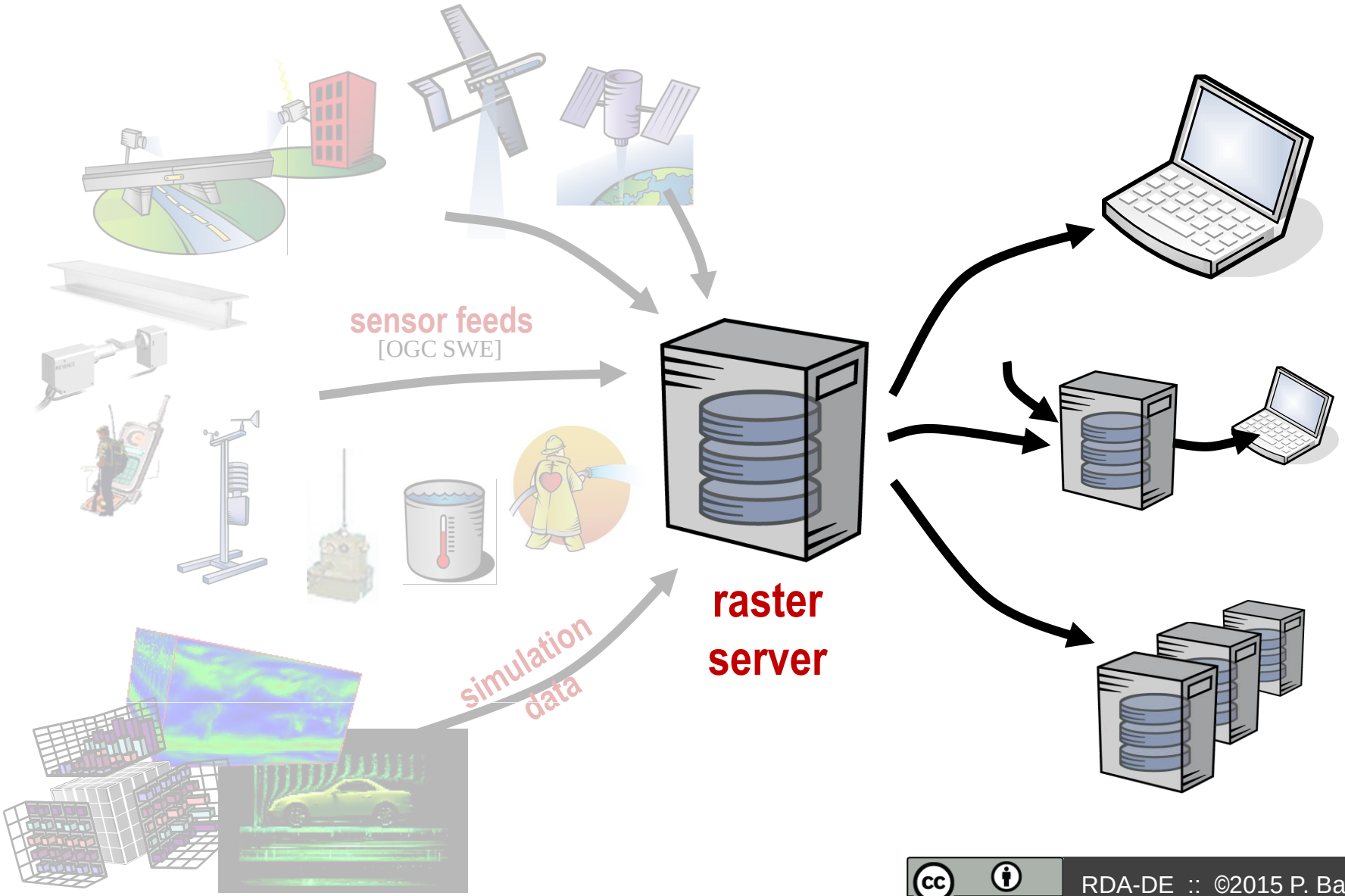
Big Data (not only) in Geo

Ocean Science Interoperability Experiment [OGC]

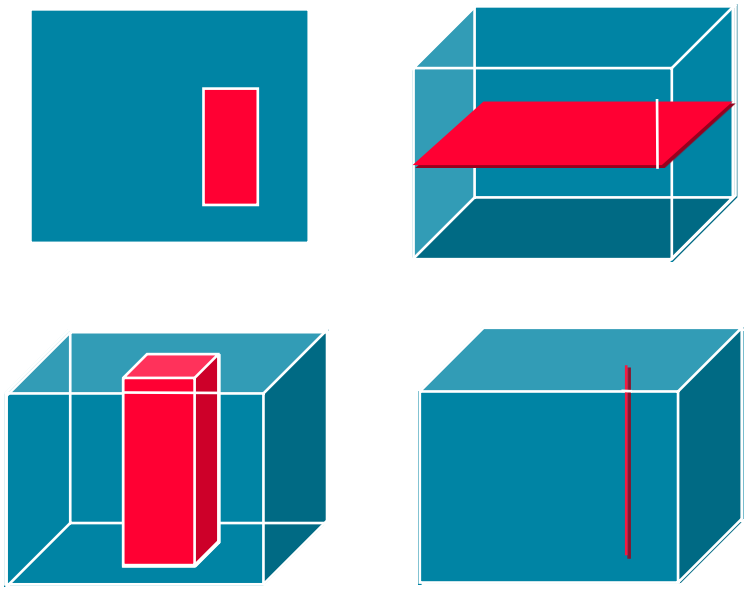
From Data Collecting…



...to Data Offerings



Common Operations on Data Cubes

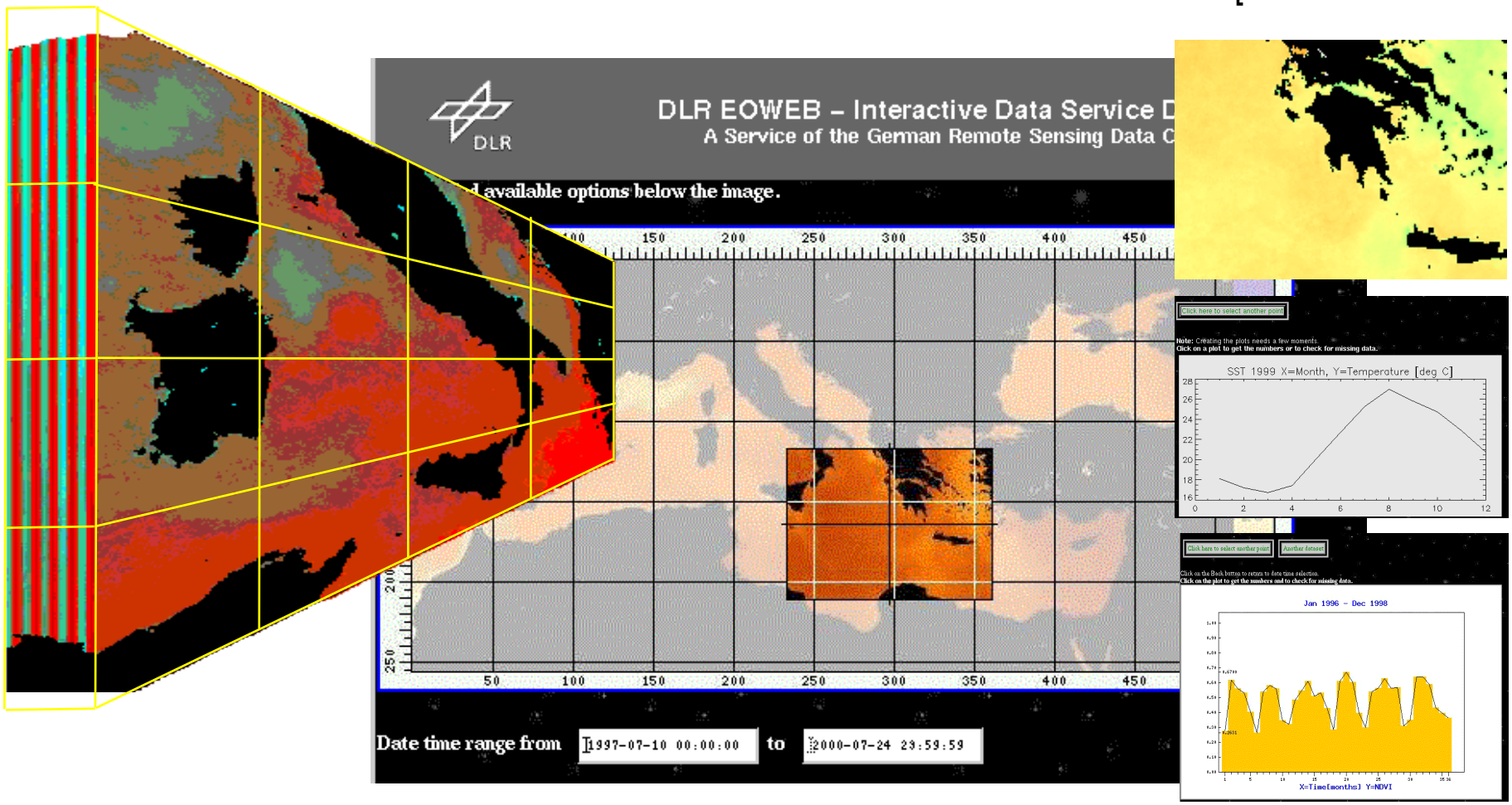


- plus data formats
- plus processing

→ „Coverage“ and „Web Coverage Service“ in OGC

Use Case: Satellite Image Time Series

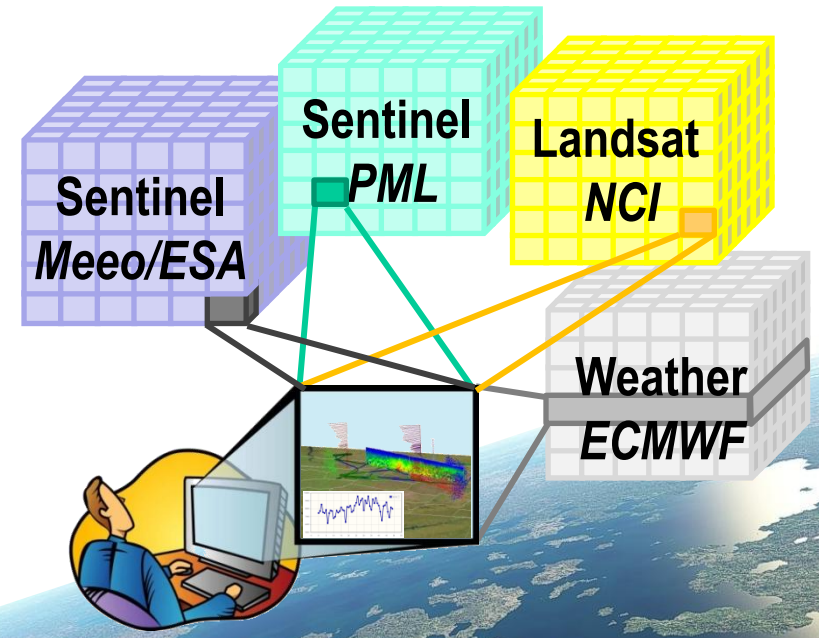
[Diedrich et al 2001]



rasdaman raster data management | JACOBS UNIVERSITY

EarthServer: Datacubes At Your Fingertips

- Standards on Steroids:
  Real-time mix&match on 1+ Petabyte space/time datacubes
  - 3D sat image timeseries, 4D weather data
  - Any query, anytime – with integrated data/metadata search
  - Strictly open standards
  - 1D...4D Web visualization, incl NASA WorldWind
- Intercontinental initiative
  - EarthServer-1: 2011-2014, EU+US
  - EarthServer-2: 2015-2018, EU+US+AUS
- www.earthserver.eu



Co-funded by the European Union

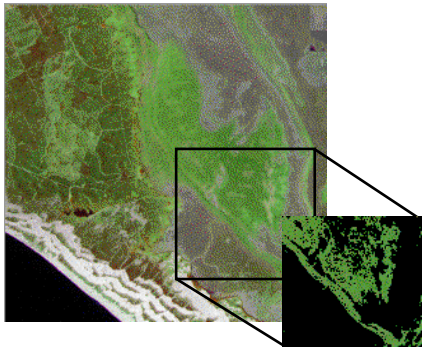
rasdaman raster data management

JACOBS UNIVERSITY

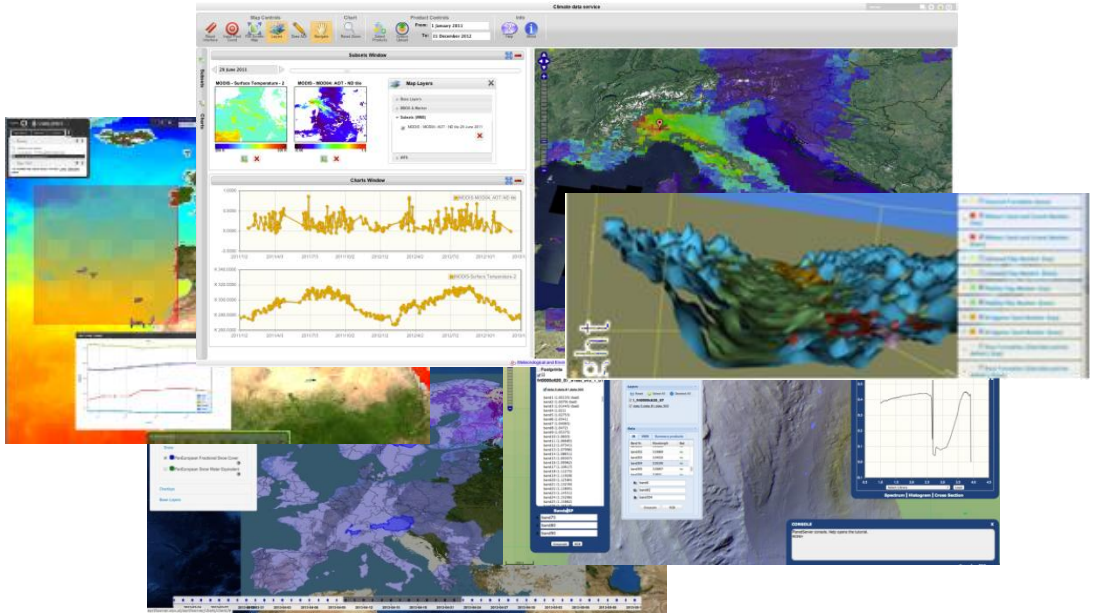
Underlying Technology: Array Database

- „raster data manager“: SQL + n-D arrays

```
select ls.img.green[x0:x1,y0:y1] > 130
from   LandsatArchive as ls
where  avg_cells( ls.img.nir ) < 17
```



- Scalable, distributed “tile streaming” architecture
- In operational use



Key Conceptual Challenges in Geo Services

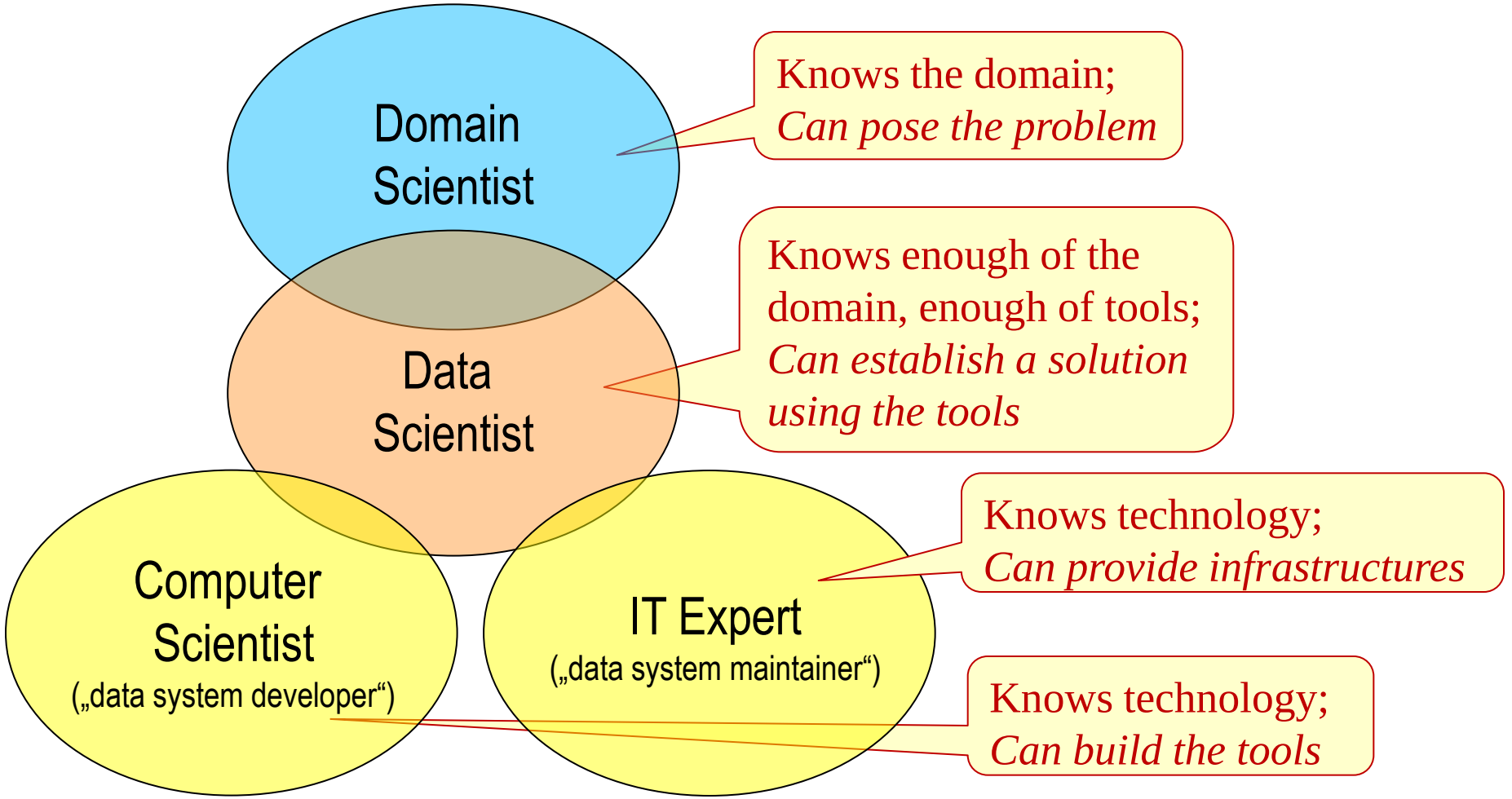
- Internet of (georeferenced) things
- Time coordinates
- Client/server interface design
  - Powerful, scalable, easy, safe
  - Server-side filtering & processing
- Integration of heterogeneous paradigms: UML, XML, JSON, RDF, ...
- ...plus many more

Key Questions

Thesis 1: Technology Thrust

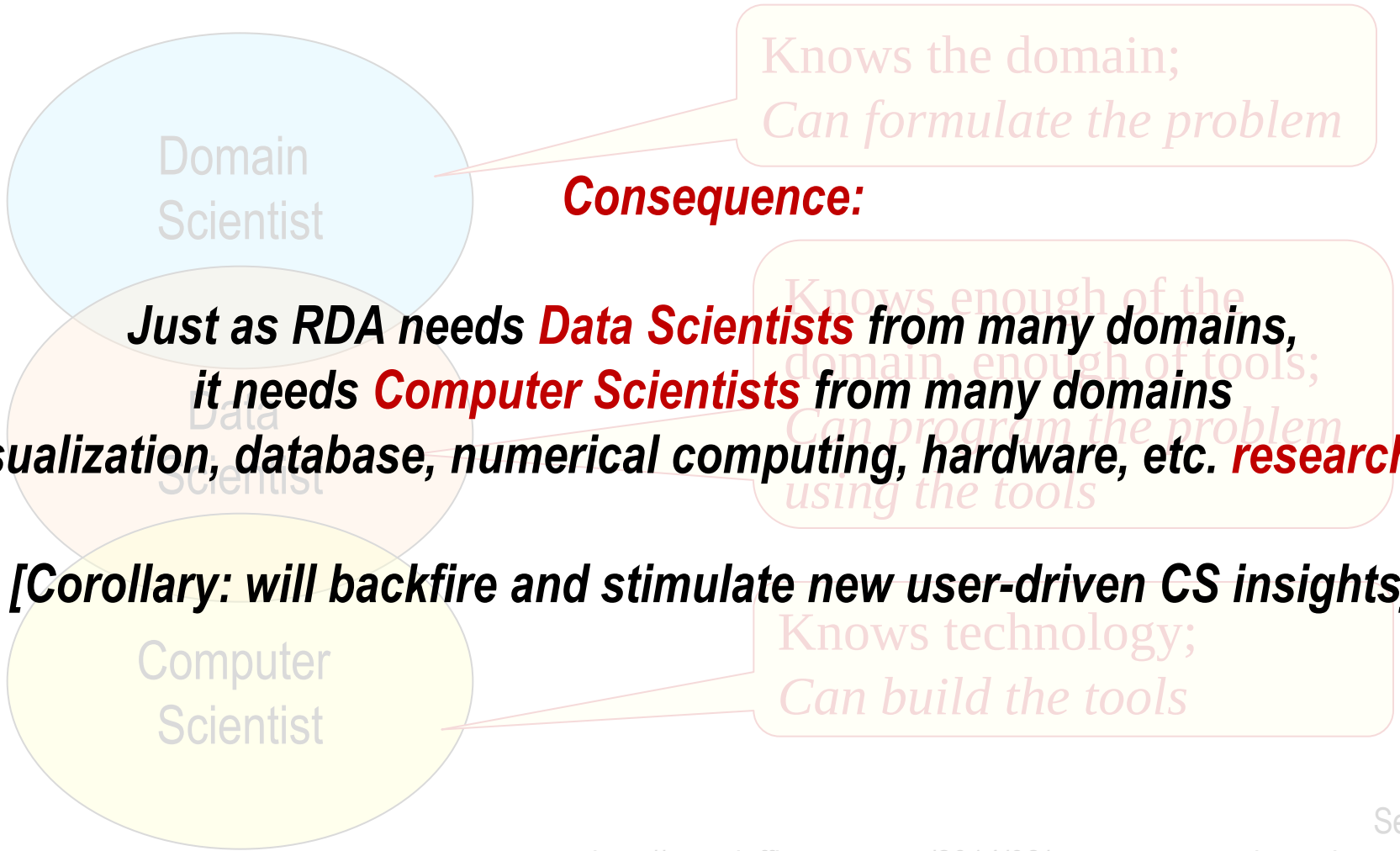
- Main upcoming challenge: common information space for users
  - data based on few data types: sets/tables; trees; graphs; arrays; text; ...
  - Technologies & stds present or coming: (New)SQL, SQL/MDA, XML, SPARQL, ...
  - Integration of models: an excellent interdisciplinary cooperation theme!
- Deep references into objects, dynamic references on virtual products
- Overcome data/metadata divide!
- Future services based on formalized languages
  - Role model: SQL
  - Not end user interface (except for experts), but powerful client/server API
  - Client side: Flexibility, server side: scalability
- Trust, more generally: „fitness for purpose, relative to task“
  - as (discoverable) property of data & services
  - Ex: weather forecast & climate change

Thesis 2: Data Science as Team Work



See also:
http://www.jeffheaton.com/2014/02/so-you-want-to-be-a-data-scientist/

Thesis 2: Data Science as Team Work



Consequence:

Just as RDA needs Data Scientists from many domains,
it needs Computer Scientists from many domains
(visualization, database, numerical computing, hardware, etc. researchers)

[Corollary: will backfire and stimulate new user-driven CS insights]

See also:
http://www.jeffheaton.com/2014/02/so-you-want-to-be-a-data-scientist/

Thesis 3: ...therefore, Do‘s and Dont‘s

☹ Don‘t:

- add more and more metadata just for discovery
- Establish *centralized* registries, brokers, etc.
- Focus exclusively on *data* stewardship
- Answer user needs with ad-hoc technology solutions

☺ Do:

- Make data itself searchable & smart
- Take *federation* serious, think peer
- Consider *service* stewardship
- Team up with CS experts
- Array Databases (Europe leads)