OGC’s Big Data Initiative

Open-Source Park, INTERGEO 2013

Peter Baumann

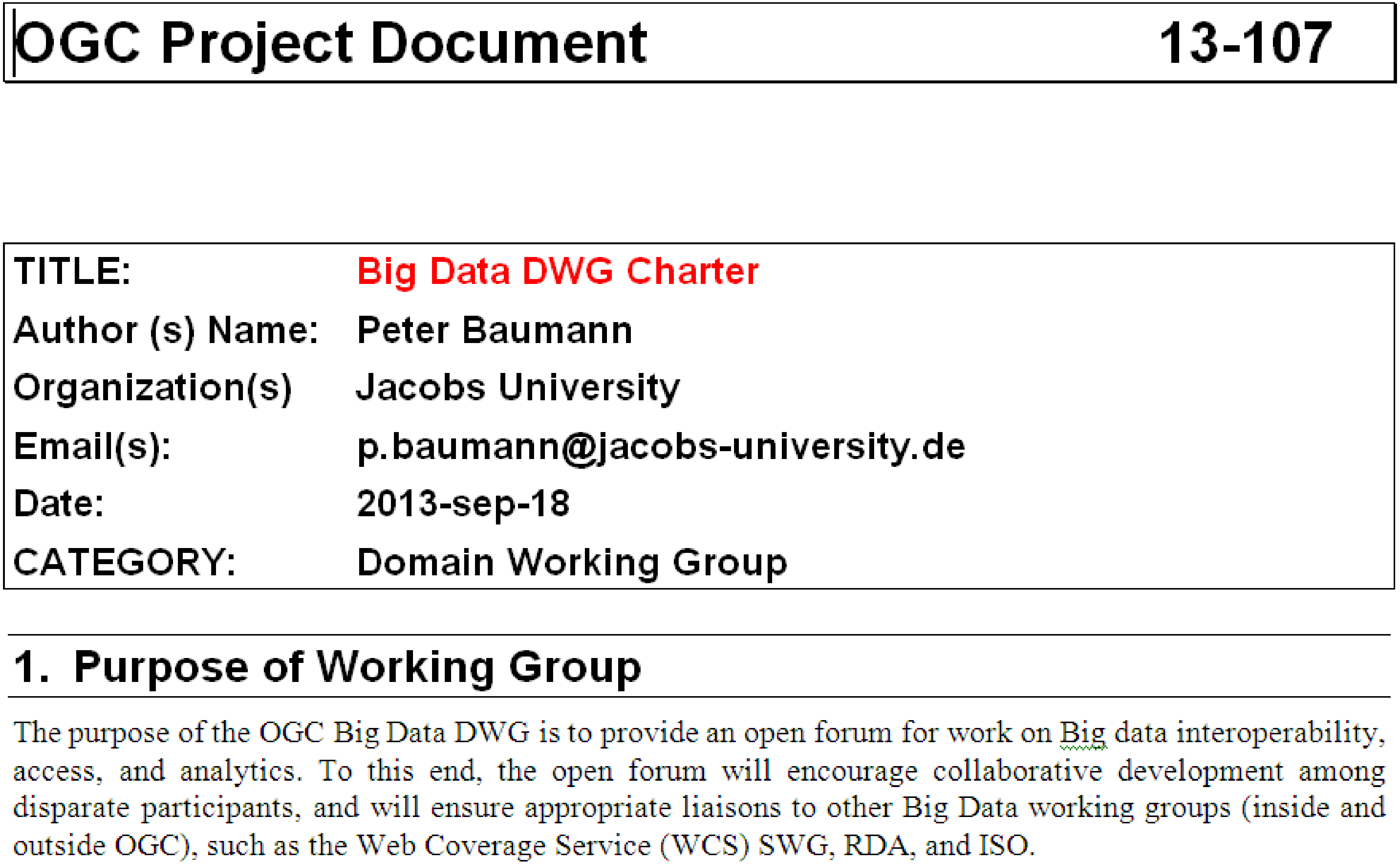
Jacobs University | rasdaman GmbH

Big Data Research @ Jacobs U

- Large-Scale Scientific Information Systems research group
  - www.jacobs-university.de/lsis
- Spin-off company: rasdaman GmbH
  - www.rasdaman.com
- Main results:
  - Array DBMS, rasdaman
  - OGC: unified coverage data & services, chair of 5 WGs, editor of 12 standards
  - INSPIRE invited expert
  - ISO: Array SQL
  - Research Data Alliance (RDA): Big Data co-chair

OGC TC Meeting, ESA, Frascati, 2013-sep-26

- Big Data hot topic in science & markets
- Manifold Big Data in OGC's realm - coverages & others
- ...so OGC should have a say, establish a position
- „Big Data“ not just big; a main issue: analytics on variety of data
  - Therefore, overarching, cross-WG topic
- Kickoff meeting:
  - Some 30+ participants
  - Topic uniformly seen as relevant, participants want to see this group
  - No spec development, but position statements for OGC & possibly recommendations to SWGs



OGC Project Document 13-107

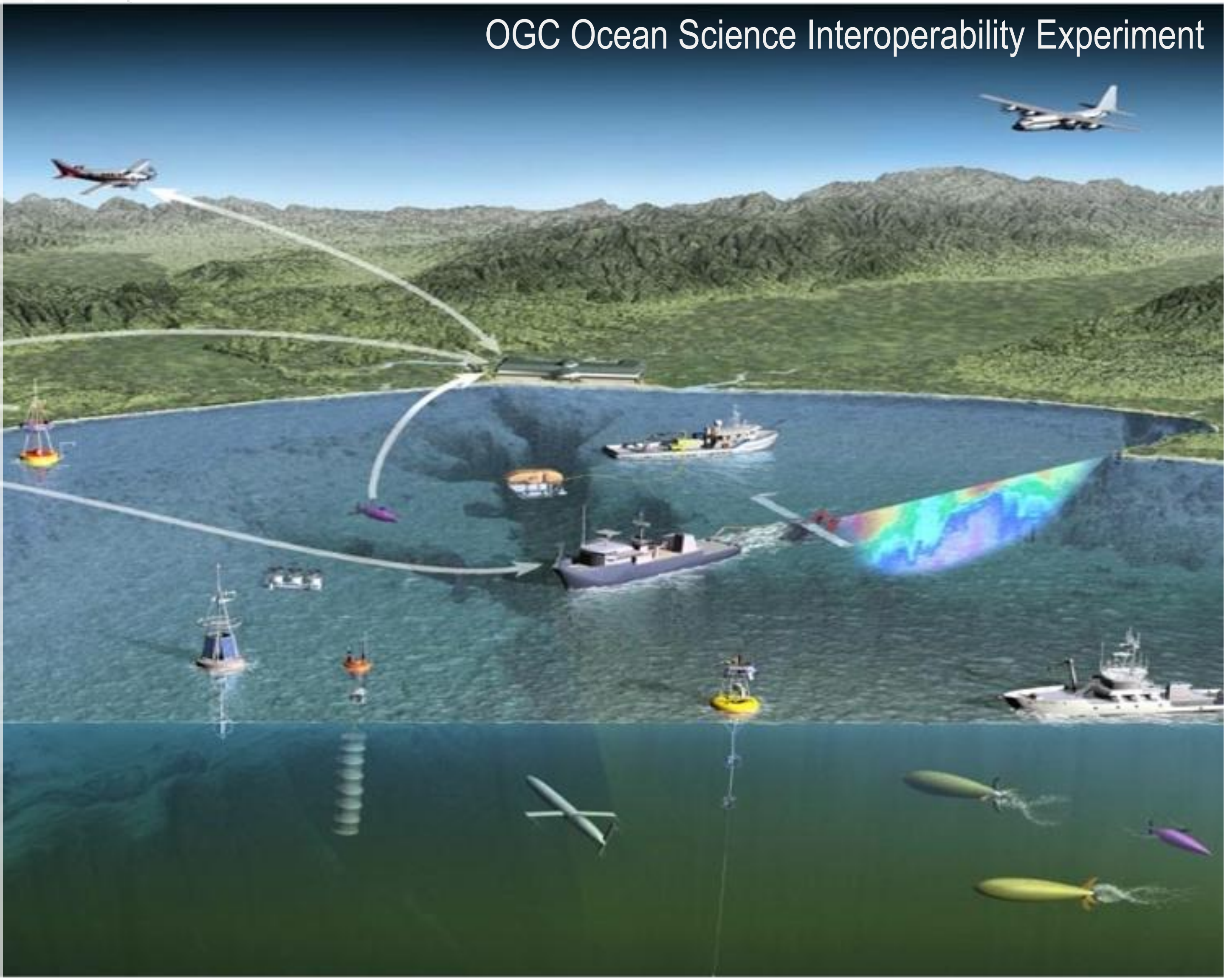
TITLE:	Big Data DWG Charter
Author (s) Name:	Peter Baumann
Organization(s)	Jacobs University
Email(s):	p.baumann@jacobs-university.de
Date:	2013-sep-18
CATEGORY:	Domain Working Group

1. Purpose of Working Group

The purpose of the OGC Big Data DWG is to provide an open forum for work on Big data interoperability, access, and analytics. To this end, the open forum will encourage collaborative development among disparate participants, and will ensure appropriate liaisons to other Big Data working groups (inside and outside OGC), such as the Web Coverage Service (WCS) SWG, RDA, and ISO.

Big Data in Geo

- Volume
- Velocity
- Variety
- Veracity


OGC Ocean Science Interoperability Experiment

[Doug Laney / Gartner & IBM; image: OGC]

Big Data: Volume

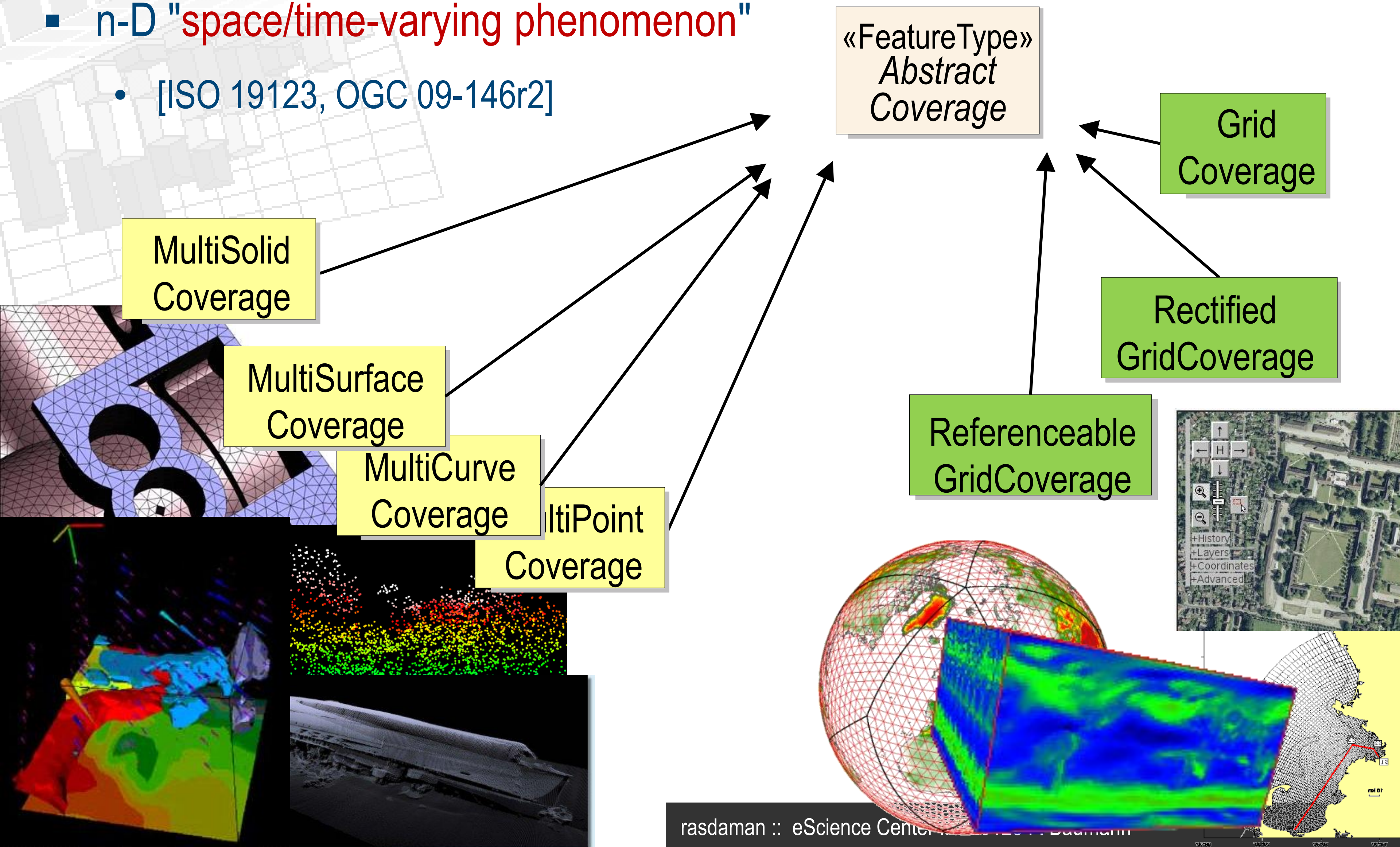
- Social Networks
  - Incidence matrix of size 10^8 x 10^8 ...now do linear algebra!
- Satellite Imagery
  - ESA planning for 1,000,000,000,000 images
- HPC
  - *„Even with multi-terabyte local disk sub-systems and multi-petabyte archives, I/O can become a bottleneck in HPC.“*
    - *-- Jeanette Jenness, LLNL, ASCI-Project, 1998*
  - „Users download 10x more data than needed“
    - *-- Kerstin Kleese van Dam, 2002*

Big Data: Velocity

- NASA EOSDIS
  - ~ 5 TB per day
- LOFAR: distributed sensor array farms for radio astronomy
  - 2 Gb per second per station, consolidated into 2 – 3 PB per year
- M. Stonebraker: „drinking from the firehose“

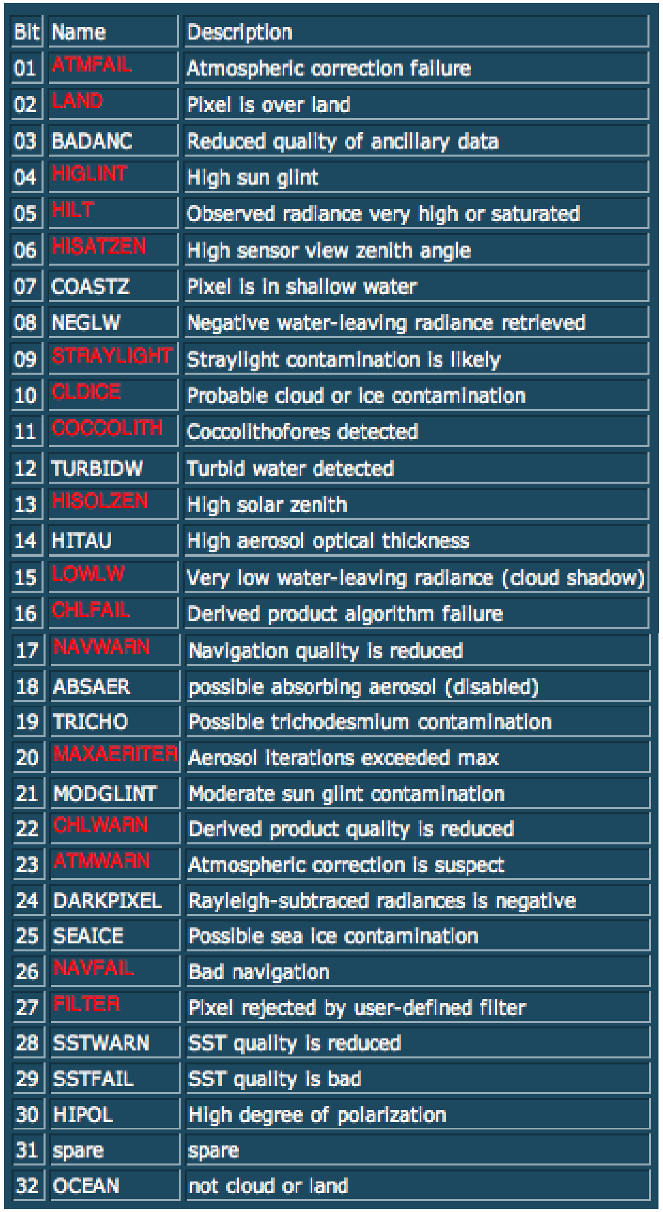
Big Data Variety: Coverages

- n-D "space/time-varying phenomenon"
  - [ISO 19123, OGC 09-146r2]



Big Data Veracity

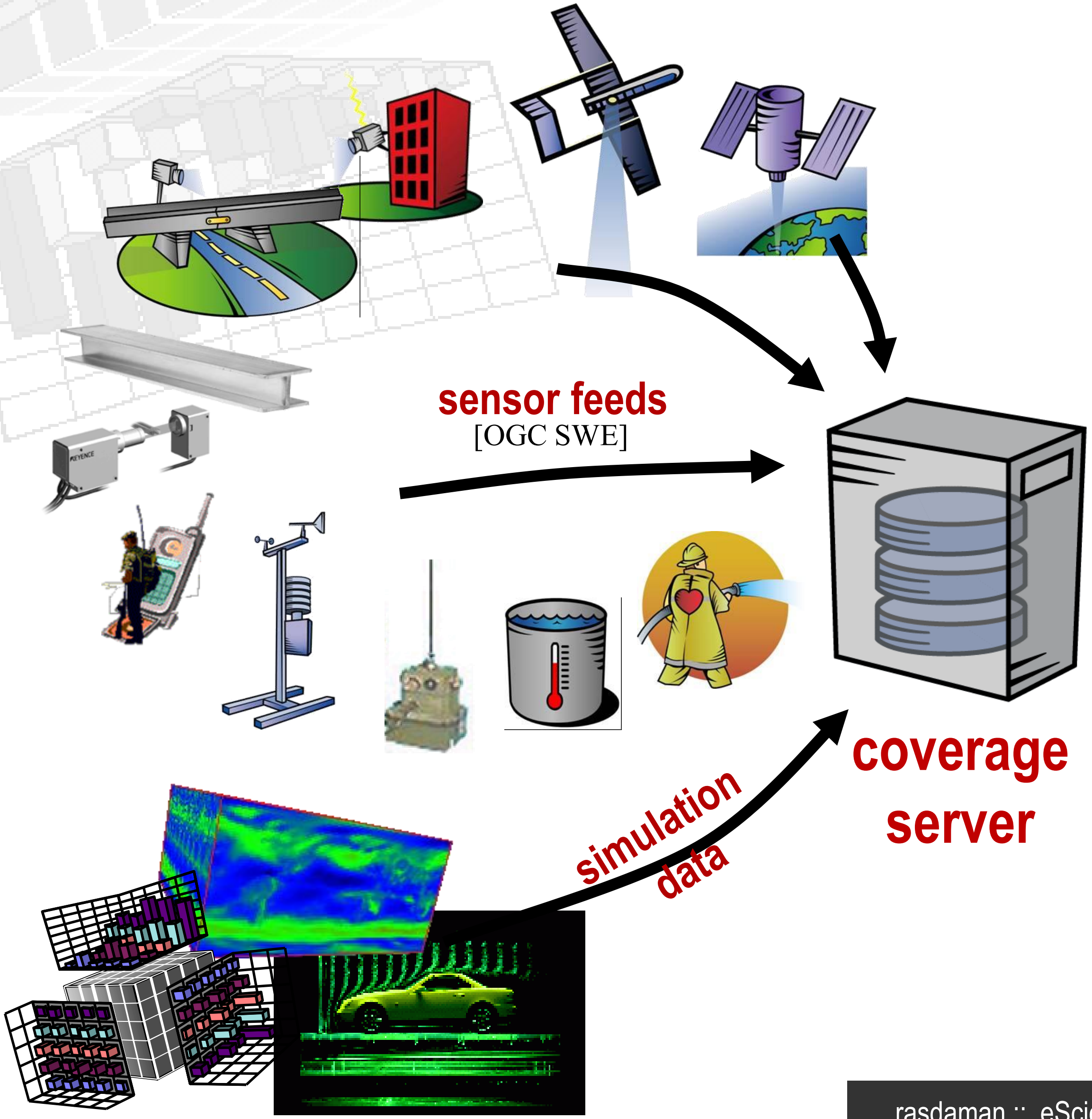
- More sources, more data = less trust
  - Cannot verify individual item any longer
  - measured & computed data:
    quality information as part of provenance
  - Crowdsourcing!
- Sometimes already well established procedures in scientific domains
- Complicates life of data consumer



Bit	Name	Description
01	ATMFAIL	Atmospheric correction failure
02	LAND	Pixel is over land
03	BADANC	Reduced quality of ancillary data
04	HIGLINT	High sun glint
05	HILT	Observed radiance very high or saturated
06	HISATZEN	High sensor view zenith angle
07	COASTZ	Pixel is in shallow water
08	NEGLW	Negative water-leaving radiance retrieved
09	STRAYLIGHT	Straylight contamination is likely
10	CLDICE	Probable cloud or ice contamination
11	COCCOLITH	Coccolithofores detected
12	TURBIDW	Turbid water detected
13	HISOLZEN	High solar zenith
14	HITAU	High aerosol optical thickness
15	LOWLW	Very low water-leaving radiance (cloud shadow)
16	CHLFAIL	Derived product algorithm failure
17	NAVWARN	Navigation quality is reduced
18	ABSAER	possible absorbing aerosol (disabled)
19	TRICHO	Possible trichodesmium contamination
20	MAXAERITER	Aerosol iterations exceeded max
21	MODGLINT	Moderate sun glint contamination
22	CHLWARN	Derived product quality is reduced
23	ATMWARN	Atmospheric correction is suspect
24	DARKPIXEL	Rayleigh-subtraced radiances is negative
25	SEAICE	Possible sea ice contamination
26	NAVFAIL	Bad navigation
27	FILTER	Pixel rejected by user-defined filter
28	SSTWARN	SST quality is reduced
29	SSTFAIL	SST quality is bad
30	HIPOL	High degree of polarization
31	spare	spare
32	OCEAN	not cloud or land

[l2gen, bitmask for ocean color]

Collecting Coverages

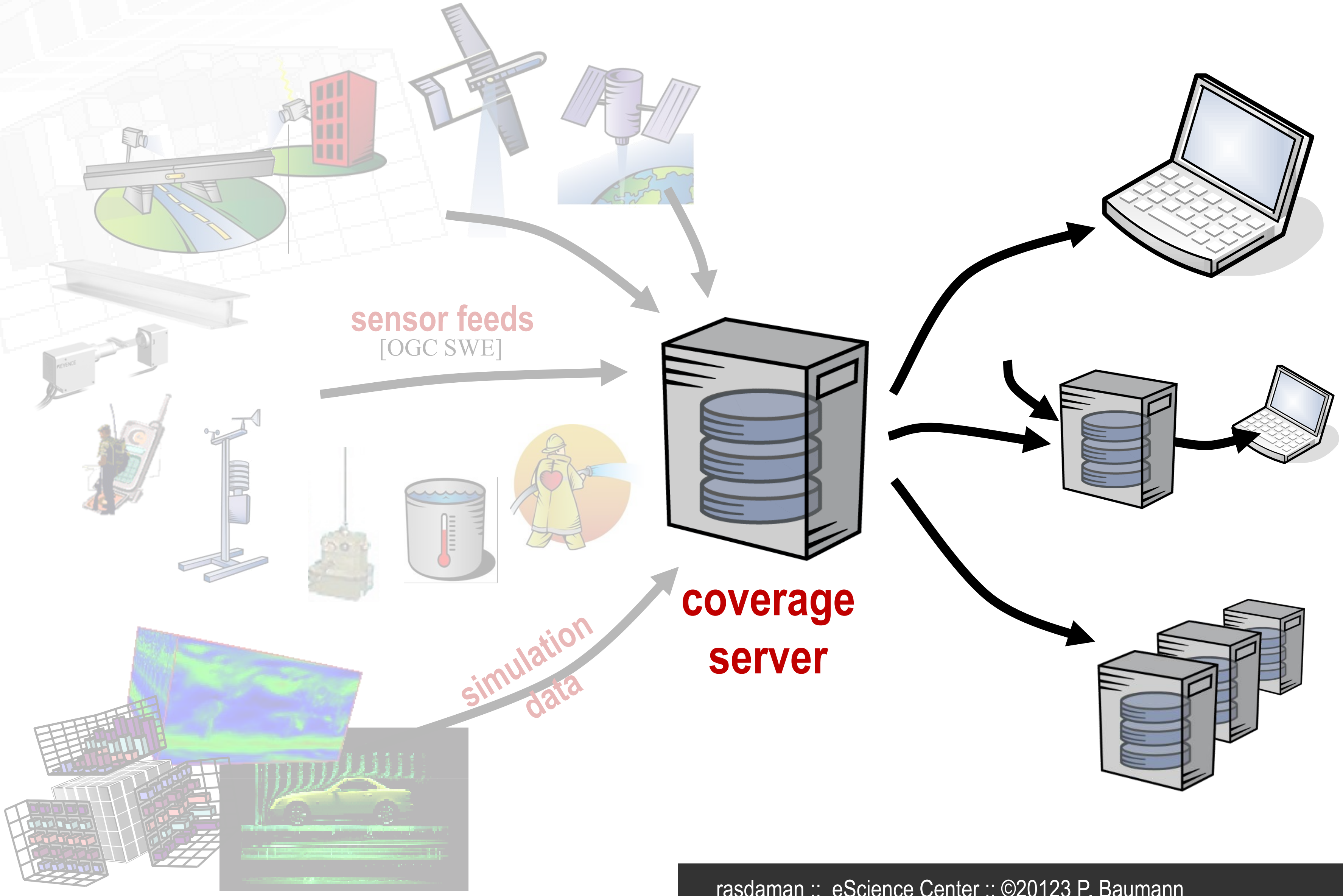


sensor feeds
[OGC SWE]

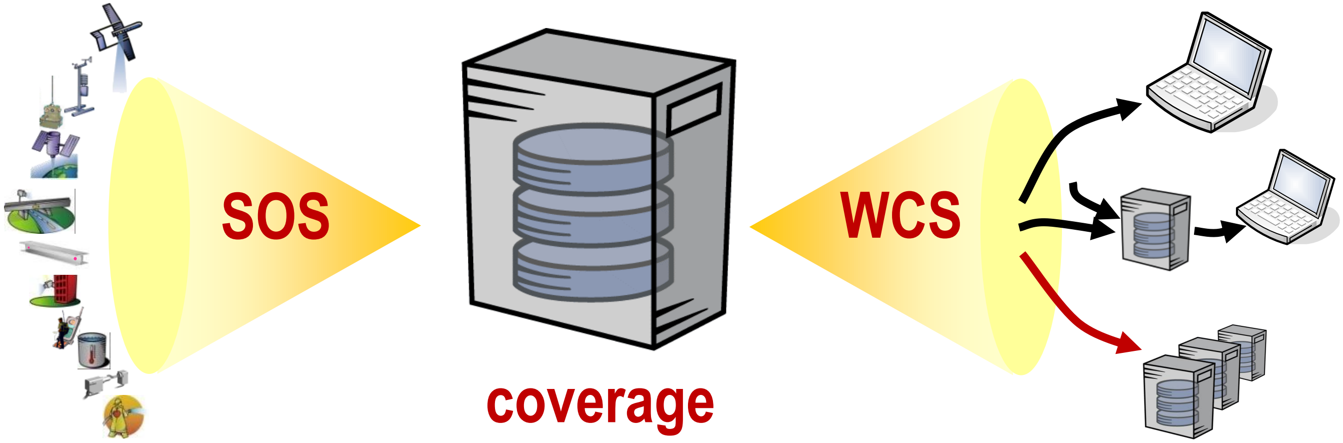
simulation data

coverage server

Serving Coverages



Serving Coverages



coverage server

WCS:
download & processing

SWE SOS:
data capturing

BigData.DWG Focus

- spatio-temporal data, in line with OGC's mission
  - What does Big Earth Data mean in an OGC context? What characterizes them?
  - What are the challenges, if any, of Big Earth Data for OGC's data and service interface specifications?
  - What is the market value of Big Earth Data, and how can OGC support leveraging it?
- will aim to clarify some foundational terminologies in the context of data analytics
  - differences/overlaps with terms like data analysis, data mining, etc.
- systematic classification of analysis algorithms, analytics tools, data and resource characteristics, and scientific queries

BigData.DWG: Planned Activities

- Establish a working communication infrastructure, including a public wiki.
- Meet regularly at TC meetings and through telecons.
- Establish liaisons with relevant OGC WGs, such as WCS.SWG, and maintain exchange.
- Establish liaison with relevant OGC-external entities, such as RDA, ISO TC211 and ISO JTC1/SC32/WG3 SQL, and maintain exchange.
- Foster an agile, member-driven agenda of topics and facilitate information sharing and consolidation.
- Proactively publish discussion and findings through wiki and other appropriate channels.
- The WG will identify additional activities as it sees fit.
- But not do standards.

Next Steps, Actions

- Action item:
  - OGC staff, please set up public BigData.DWG twiki + mailing list
  - URL most likely: http://external.opengeospatial.org/twiki_public/BigDataDWG/WebHome
  - On that page, will establish instructions on how to subscribe to list
- Next steps (mainly via twiki + list)
  - Refine charter
  - Find charter members, chairs
  - Start work

Join us!