Integration of Statistical and Geospatial Information: Metadata Standards and More

United Nations

Global Forum on Integration of Geospatial-Statistics

4-5 August 2014

Mark E. Reichardt

President and CEO, Open Geospatial Consortium

mreichardt@opengeospatial.org

OGC®
Making location count.

Geospatial Information: Essential to Address Social, Environmental & Economic Issues

Logistics & Transport

Education & Research

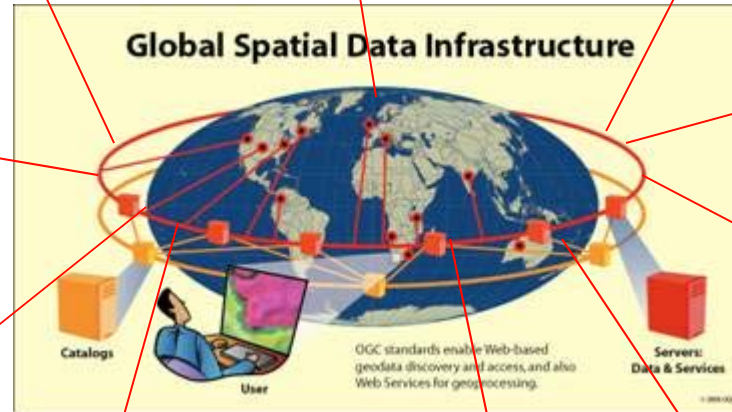Sustainable Development

Utilities

Health

Global Spatial Data Infrastructure

OGC standards enable Web-based geodata discovery and access, and also Web Services for geoprocessing.

Catalogs

User

Servers: Data & Services

E -Government

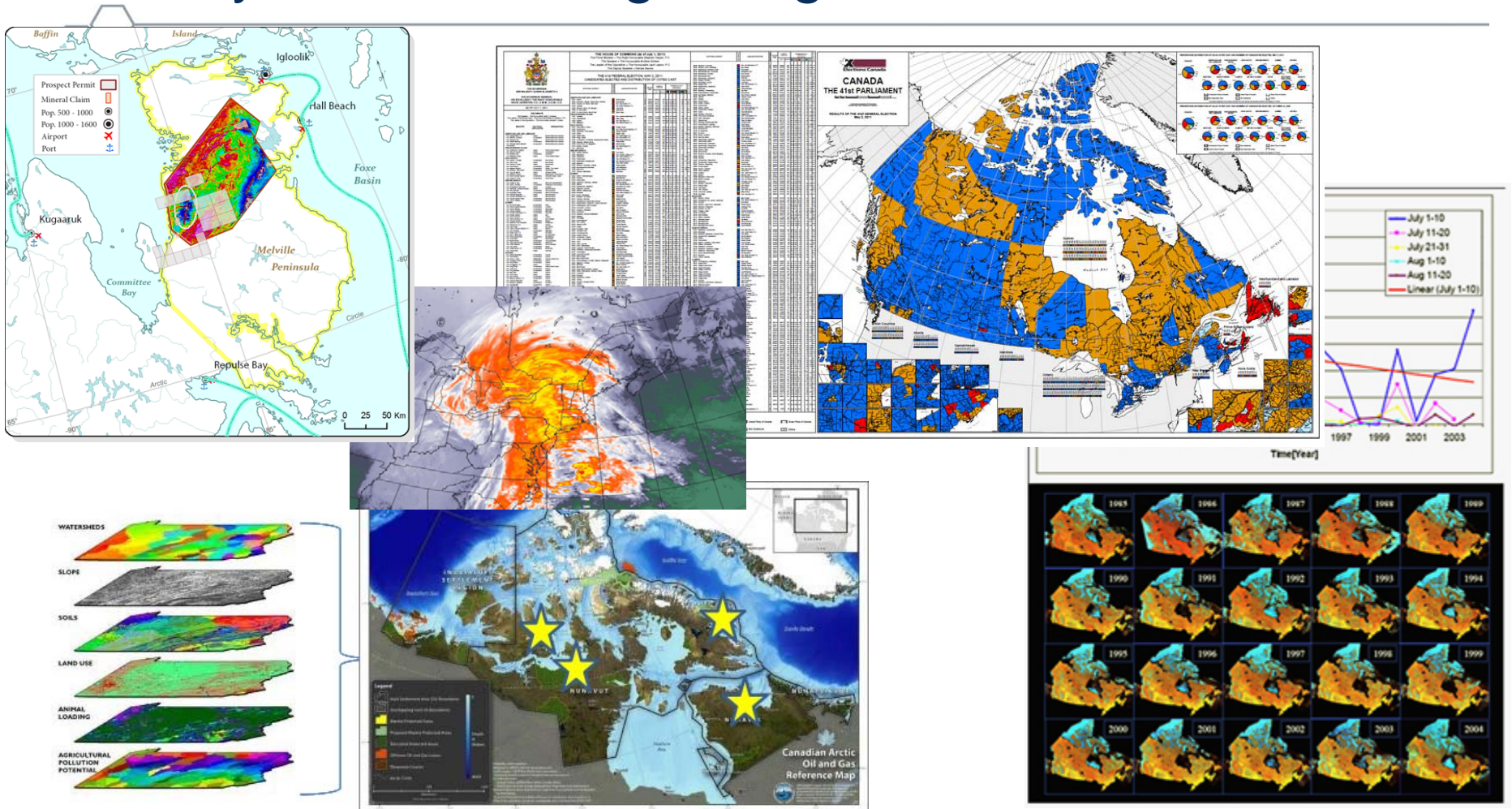Emergency and Disaster Mgt / Response

Consumer Services

Energy

Geosciences

OGC®

# Geospatial information via Spatial Data Infrastructures is widely utilized throughout government…

…to support broad national objectives such as economic growth, social cohesion and well-being, and environmental management

Source: GeoConnections Canada

The Open Geospatial Consortium

Not-for-profit, international voluntary consensus standards organization; leading development of geospatial standards

- **20th Anniversary**

- **470+ members worldwide**

- **Over 40 standards and best practices**

- **Hundreds of product implementations**

- **Broad user community implementation worldwide**

- **Alliances and collaborative activities with many other organizations**

University 24%

Commercial 41%

Research 7%

NGO 10%

Government 18%

OGC®
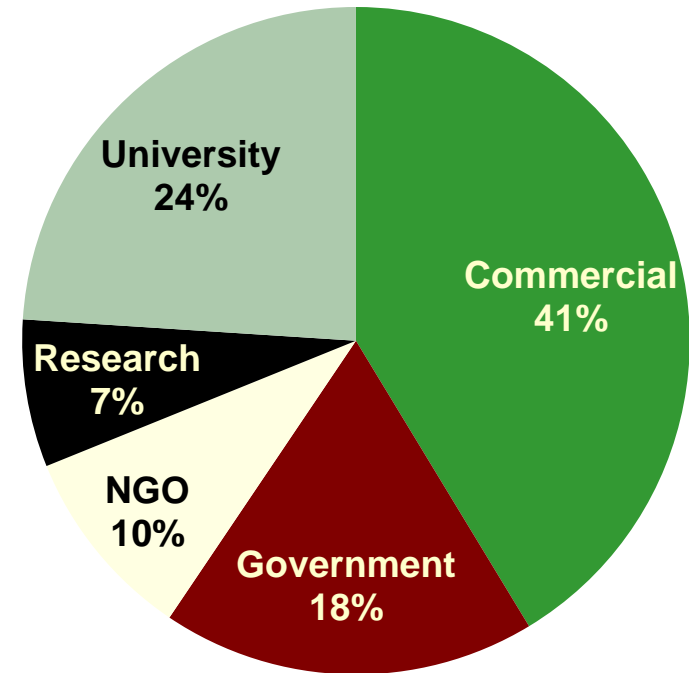
Metadata

- ## Data about Data

    "…provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data."
   US Federal Geographic Data Committee

- ## Helps guide use of correct data for the appropriate purpose

- ## Well defined metadata constructs (ISO 19115), implemented worldwide, national profiles, many tools to support metadata creation and maintenance

OGC®

OGC/ISO Web Services Standards

Rapid discovery, access, fusion and application of location information for:
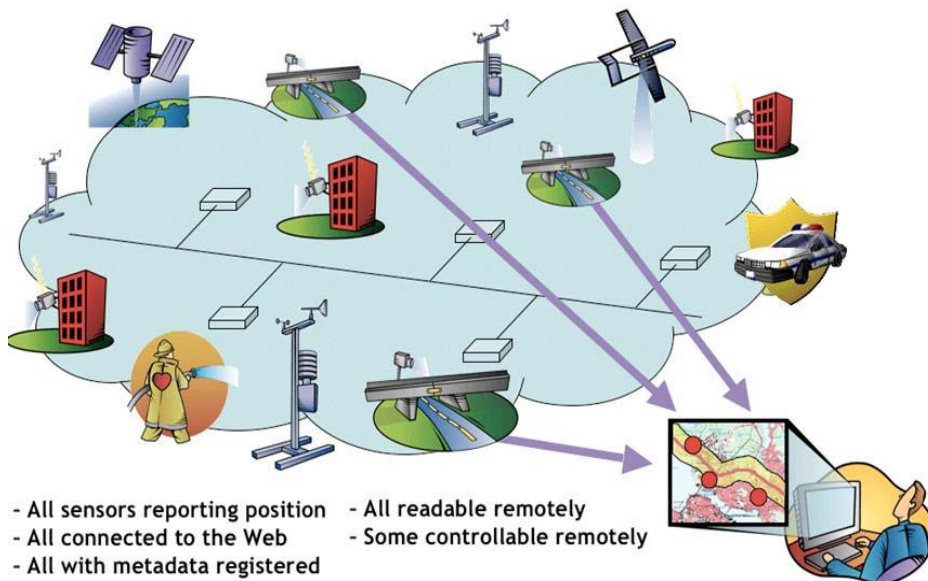
- Catalogue (leverages ISO 19115)
- Geography Markup Language(ISO 19136)
- KML
- Observations and Measurements (ISO 19156)
- SensorML
- Sensor Observation Service
- Sensor Planning Service
- Simple Features (ISO 19125)
- Web Coverage Service
- Web Feature Service (ISO 19142)
- Web Map Service (ISO 19128)
- Web Map Tile Service
- Web Processing Service

OGC®

Complete OGC Standards List: http://www.opengeospatial.org/standards
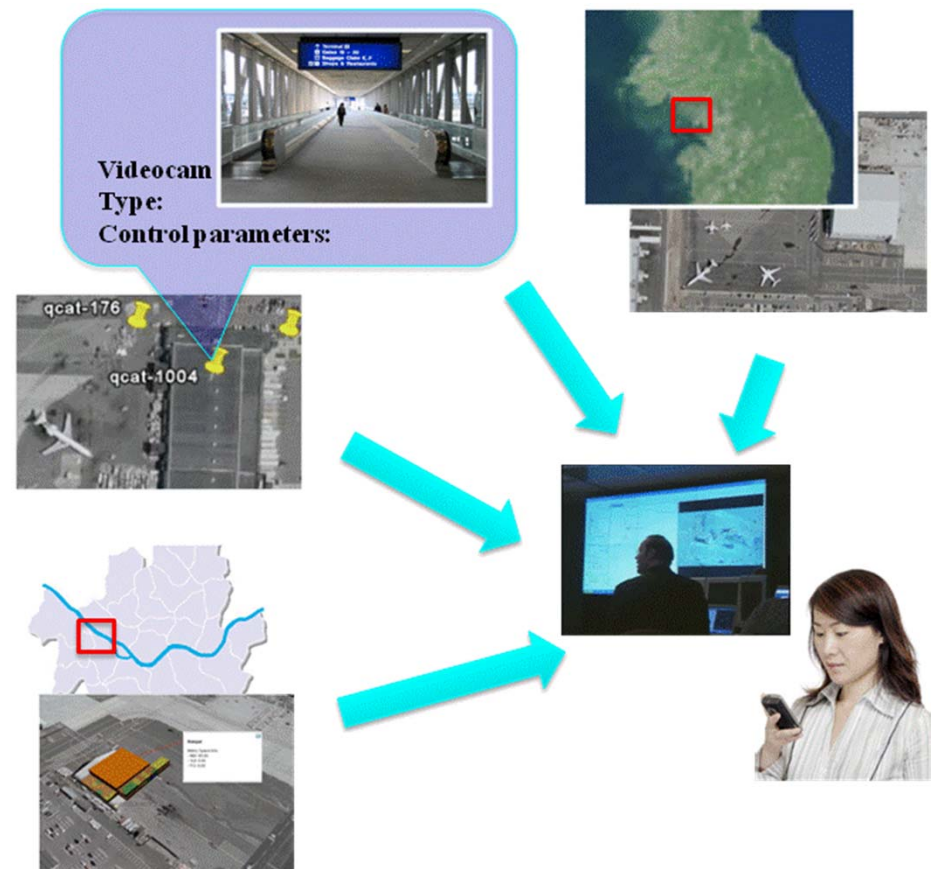
OGC Sensor Web Enablement Standards

Enable discovery and tasking of sensor assets, and the access and application of sensor observations for enhanced situational awareness


- All sensors reporting position
- All connected to the Web
- All with metadata registered

- All readable remotely
- Some controllable remotely

✓Sensor Model Language (SensorML)

✓Sensor Planning Service (SPS)

✓Sensor Observation Service (SOS)

✓PUCK

✓Sensor Alert Service (SAS)

--Complementary Standards--

✓OASIS (alert) standards

✓IEEE 1451 smart sensor standard

OGC®

Basic Geospatial Integration Challenge Solved
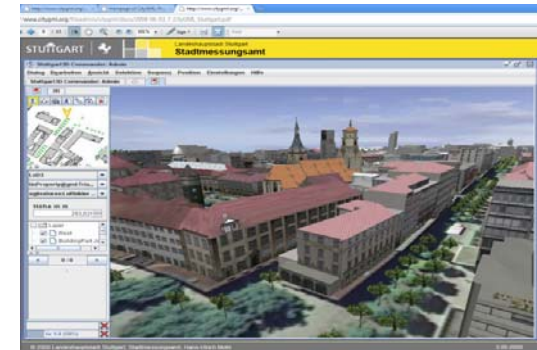## Standards-based Technologies and Information Sources Abound
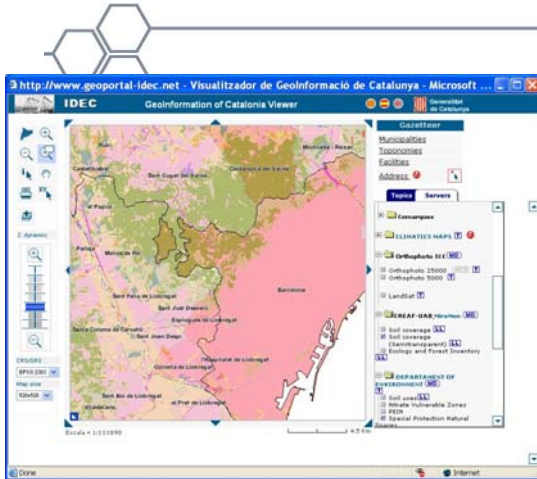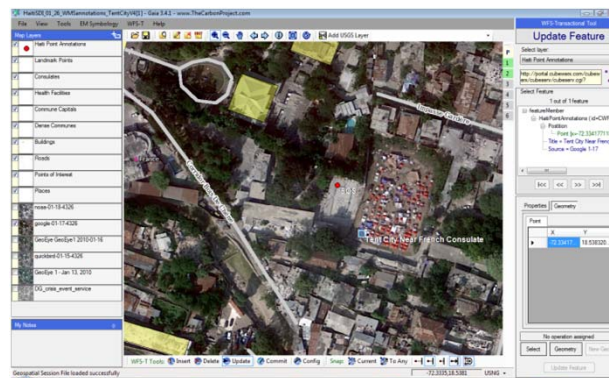

Source: Geoportal of the Catalonia SDI

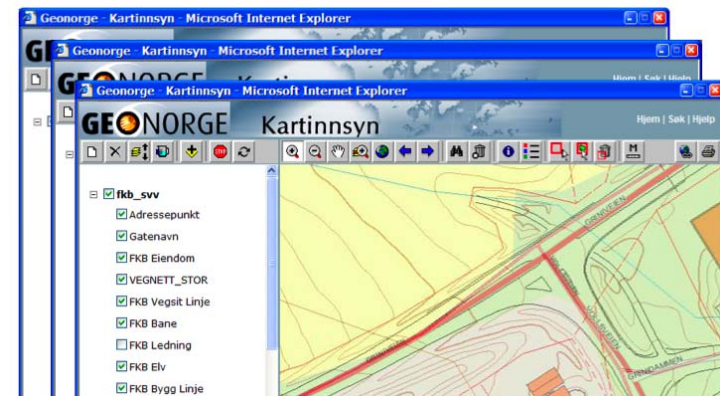**Source: OpenIOOS.org**
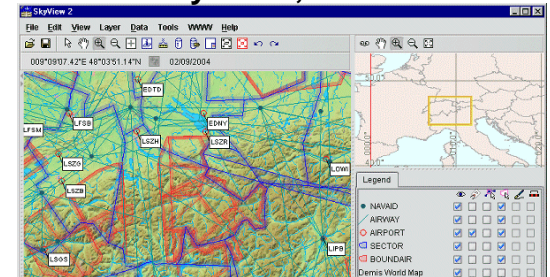

Source: Landeshauptstadt Stuttgart

Source: The Carbon Project

Source: onegeology.org

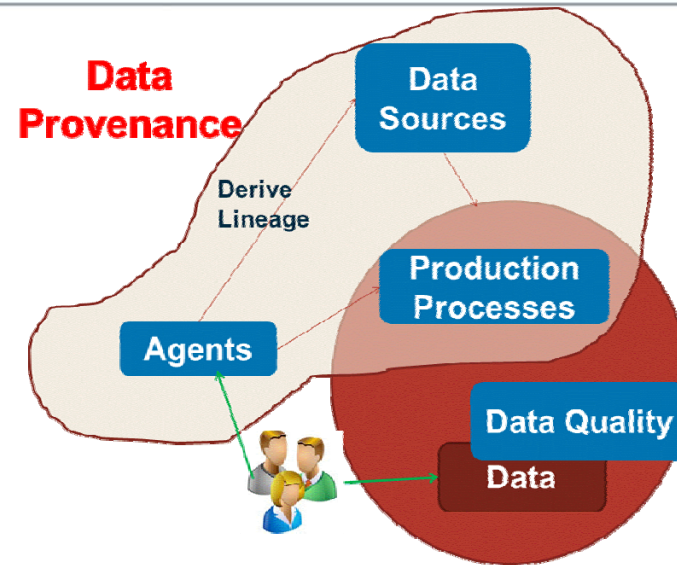**Source: Skyview2, Eurocontrol**

OGC®

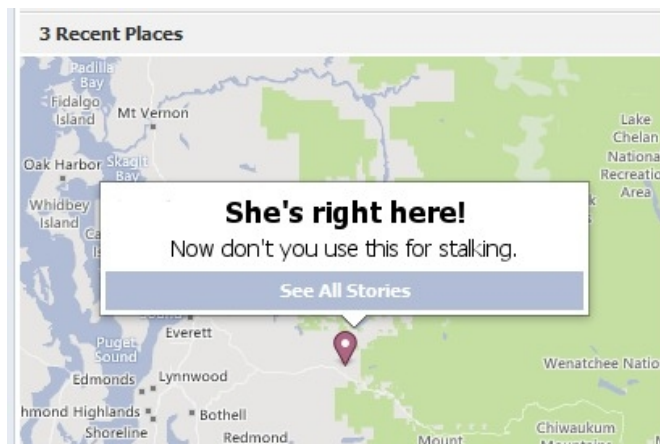# Some Common Challenges
All Roads Lead to Metadata

- **Data Quality**
  - Uncertainty
  - Provenance
  - Conflation, Semantic Mediation
- **Privacy**
- **Security**

OGC Web Services Phase 9 Testbed
http://www.opengeospatial.org/pub/www/ows9/cci.html

Source: http://www.change.org/petitions/facebook-provide-global-privacy-settings-for-location-data

Data Quality
Metadata, provenance, and uncertainty measures important!

Are (the data) fit for their intended uses in operations, decision making and planning?" ([J. M. Juran](#)).

http://geospatial.blogs.com/geospatial/2014/05/canadas-open-geospatial-data-initiative.html

OGC®

Location Data Quality

- **Committee of Experts UN GGIM, Report of the Secretariat: Establishment and implementation of standards for the global geospatial information community**

    – …"quality is a key element towards the goal of producing authoritative/official/credible information as an institutional asset; (ii) in a world of abundant information the communication of quality (metadata) becomes increasingly important; and (iii) an international framework is needed to establish a 'language' of quality"

- **OGC Challenge: Develop Location Data Quality standard**

    – Data Quality DWG: surveyed ~1000 Geospatial professionals, researched ISO specs, reviewed data quality use cases
    – Need more focus on quality and uncertainty

OGC®

Addressing Uncertainty

- ISO/IEC Guide to expression of Uncertainty in Measurement (GUM)

> **"…it is obligatory that some quantitative indication of the quality of the result be given so that those who use it can assess its reliability."**
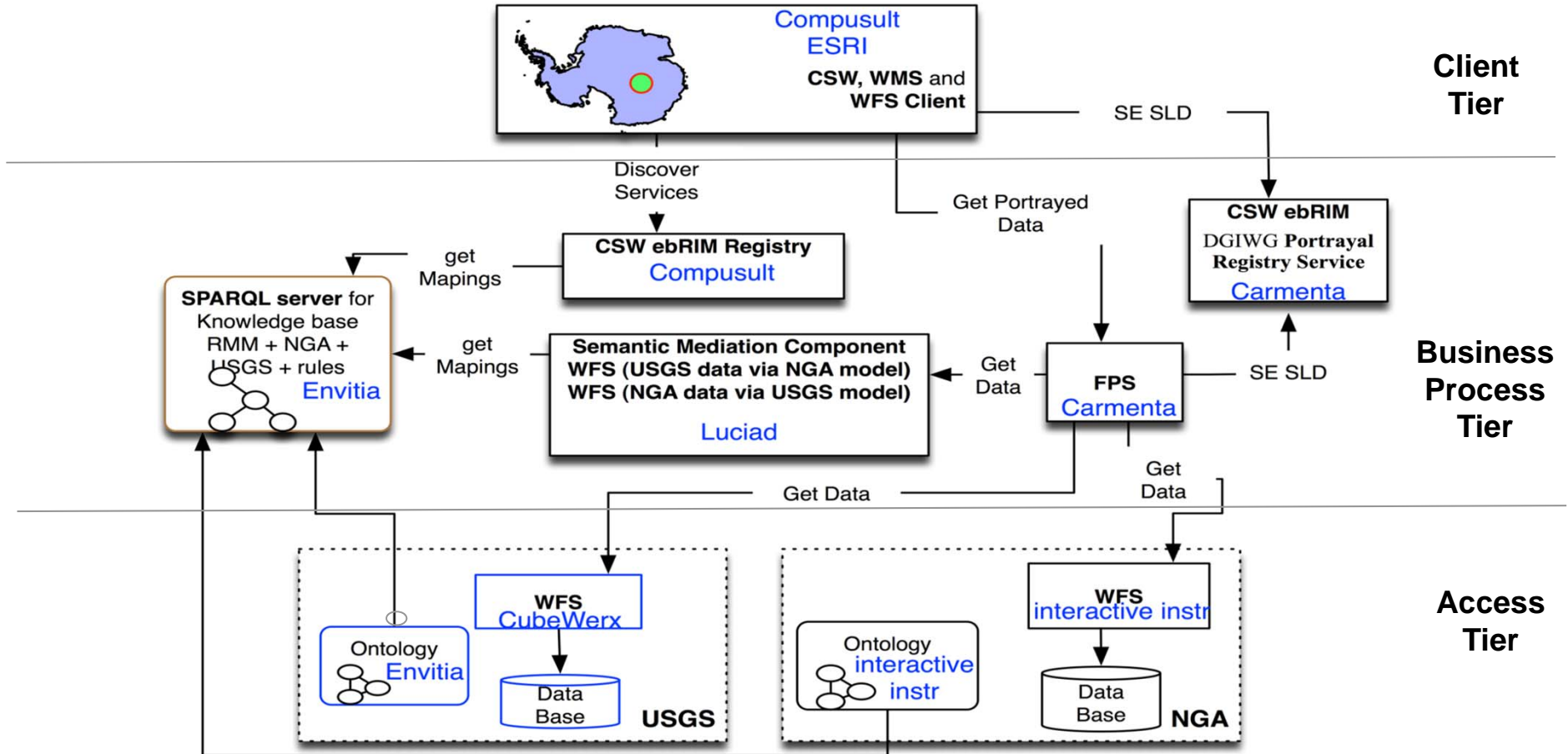>
> http://www.iso.org/sites/JCGM/GUM/JCGM100/C045315e-html/C045315e_FILES/MAIN_C045315e/Intro_e.html

- UncertML is a conceptual model, with accompanying XML schema, that may be used to **quantify** and **exchange complex uncertainties** in data.

- OGC collaborating with www.uncertml.org
  - Leverage UncertML as a way to express uncertainty across a web services workflow
    - Source error, integration, processing…
    - Described as samples, simple statistics, probabilistic distributions
  - OGC Discussion Paper: Uncertainty Markup Language (UnCertML)

OGC®

Semantic Mediation and Conflation

This diagram shows the main components and interactions developed. The thread overview video provides context and examples of how this architecture was used.
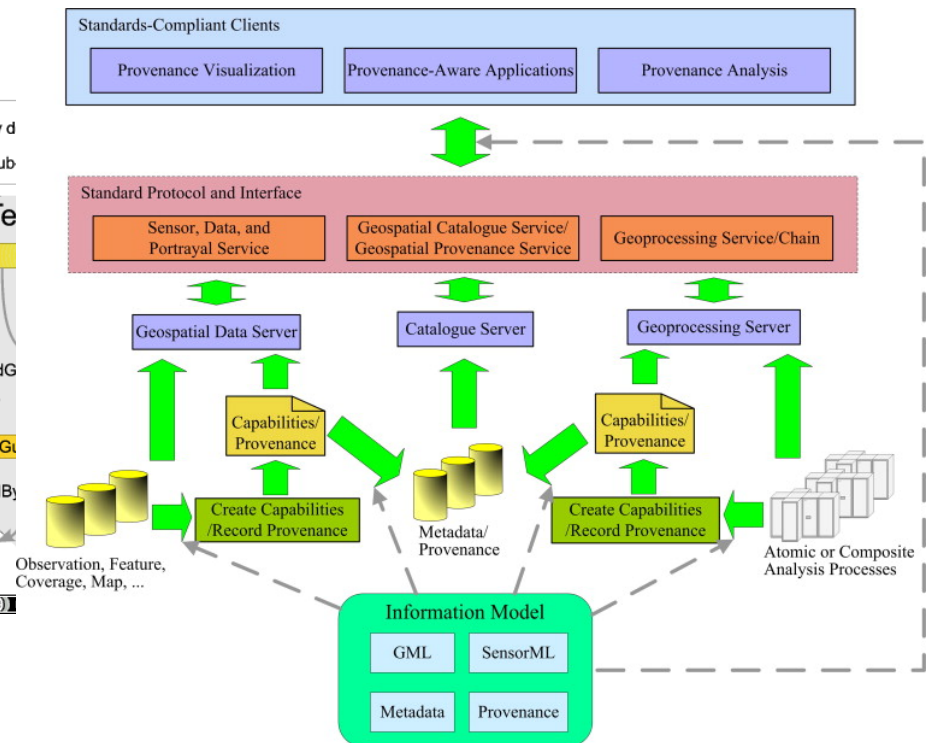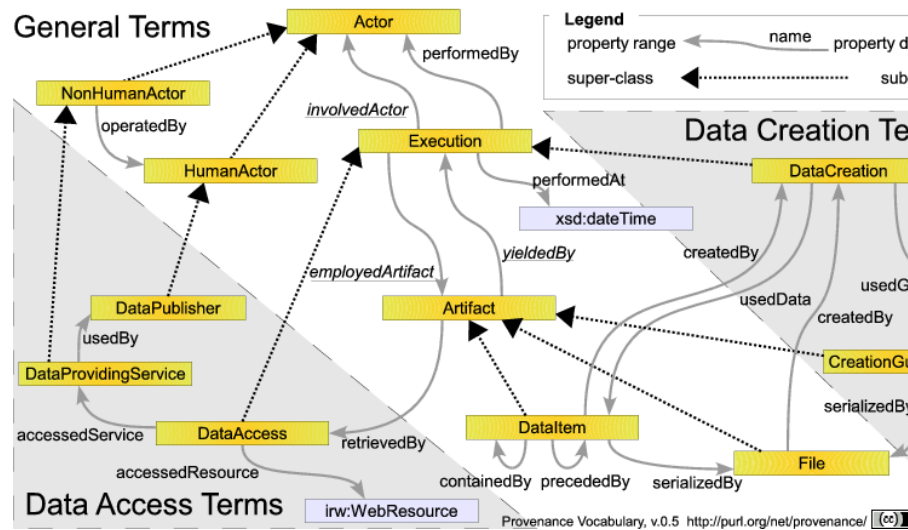

See OGC Web Services Phase 8 and Phase 9 Cross Community Interoperability threads
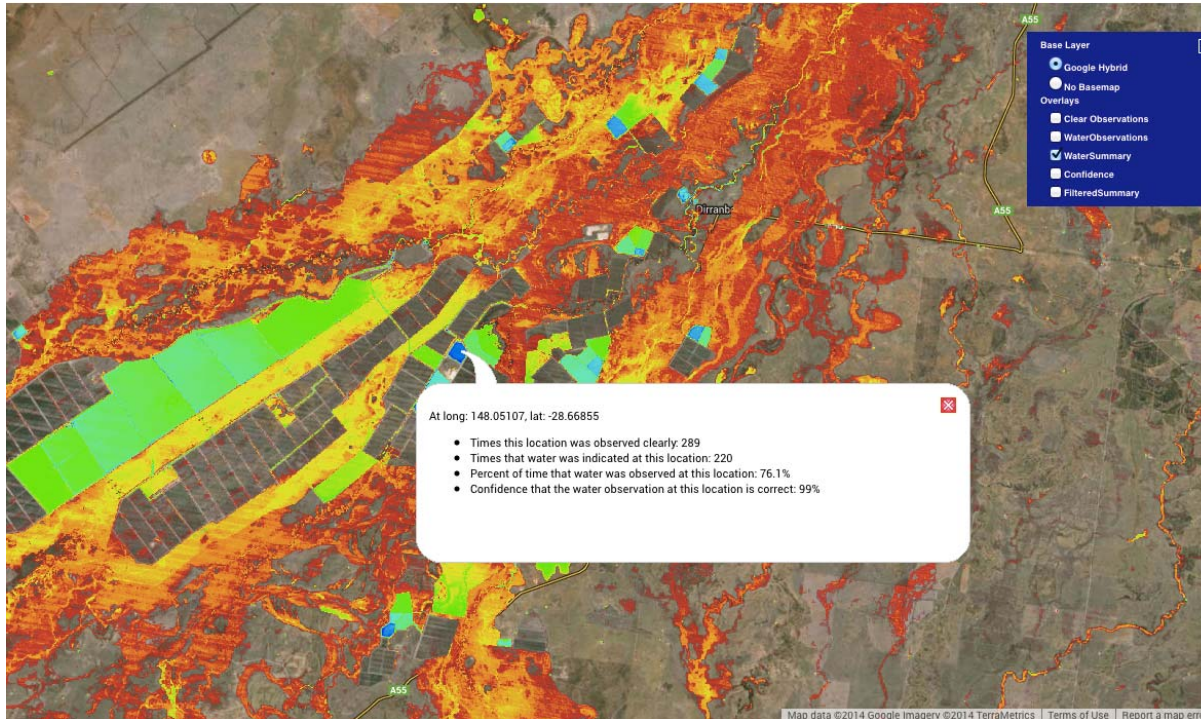http://www.opengeospatial.org/pub/www/ows8/cci.html
http://www.opengeospatial.org/pub/www/ows9/cci.html

OGC®

Provenance

- "to come from", refers to the chronology of the ownership, custody or *location* of a historical object. A type of metadata.

OGC®

Australia: Water Observations From Space
Addresses Provenance, Uncertainty, Data Quality

- 30+ Years of Landsat data, water observations correlated at pixel level
- 130,000 individual Landsat scenes
- Served via OGC/ISO standards based architecture

Supporting assessment of flood risk, water recharge, farm dam assessment, floodplain analysis

OGC®

Privacy

- In the context of the location data collected by so many mobile apps these days, anonymization generally refers to the decoupling of the location data from identifiers such as the user's name, address, location or phone number.

- Except, according to research published in *Scientific Reports* on Monday, people's day-to-day movement is usually so predictable that even anonymized location data can be linked to individuals with relative ease if correlated with a piece of outside information. Why? Because our movement patterns give us away.

  » David Meyer Mar. 25, 2013 (http://gigaom.com/2013/03/25/why-the-collision-of-big-data-and-privacy-will-require-a-new-realpolitik/)
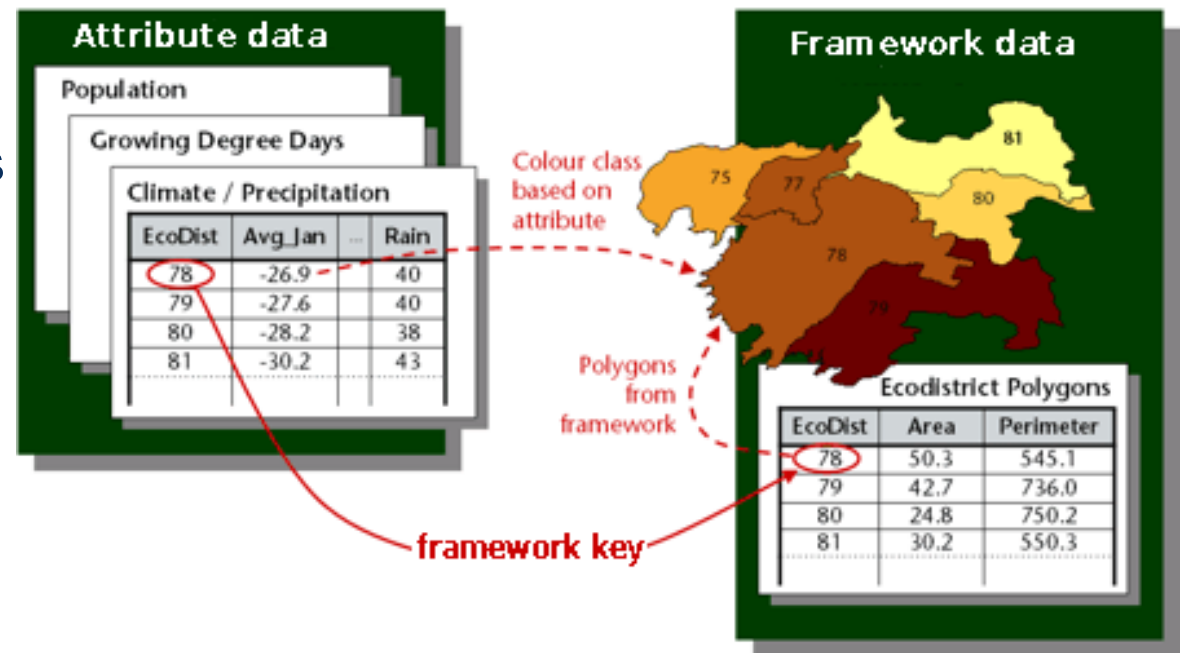
OGC®

Other Relevant Standards Work

OGC®
Making location count.

Statistics and Geospatial

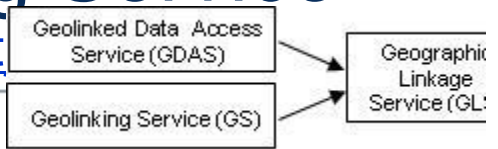- Common Links - Geographic Identifiers

  - Postal codes

  - School districts

  - Congressional districts

  - Economic zones

  - Incorporated Places

  - Telephone area codes

  - Country codes

  - Point location

Source: http://geoprocessing.info

OGC®

© 2014 Open Geospatial Consortium

OGC Georeferenced Table Joining Service

Geolinked Data Access Service (GDAS)

Geolinking Service (GS)

Geographic Linkage Service (GL

Geolinked Data Access Service (GDAS) : OGC 04-010r1
http://portal.opengeospatial.org/files/?artifact_id=5858

Geolinking Services: OGC 04-011r1
http://portal.opengeospatial.org/files/?artifact_id=5859

OGC Discussion Papers - 2004

Attribute data – Source A

Geographic data – Source B

Joined data

OGC TJS

http://www.geonovum.nl/sites/default/files/Report%20Geonovum-Table%20Joining%20Service%20v1.1.pdf

---

Open Geospatial Consortium

Date: 2010-11-22

Reference number of this document: OGC 10-070r2

Version: **1.0.0**

Category: OpenGIS® Implementation Standard

Editor: Peter Schut

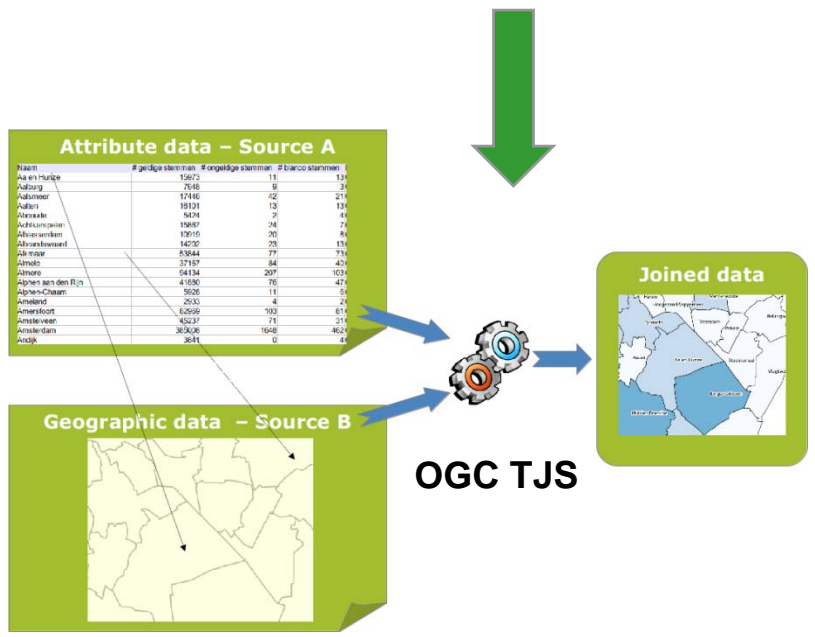External identifier of this OGC® document: http://www.opengis.net/doc/IS/tjs

OpenGIS®

Georeferenced Table Joining Service (TJS)

Implementation Standard

Copyright © 2010 Open Geospatial Consortium.
To obtain additional rights of use, visit http://www.opengeospatial.org/legal/.
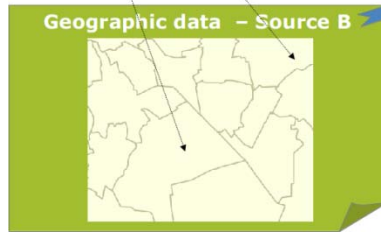
Warning

This document is an OGC Member approved international standard. This document is
available on a royalty free, non-discriminatory basis. Recipients of this document are
invited to submit, with their comments, notification of any relevant patent rights of
which they are aware and to provide supporting documentation.

Document type: OpenGIS® Standard
Document sub type: Interface
Document stage: Approved 1.0
Document language: English

OGC®
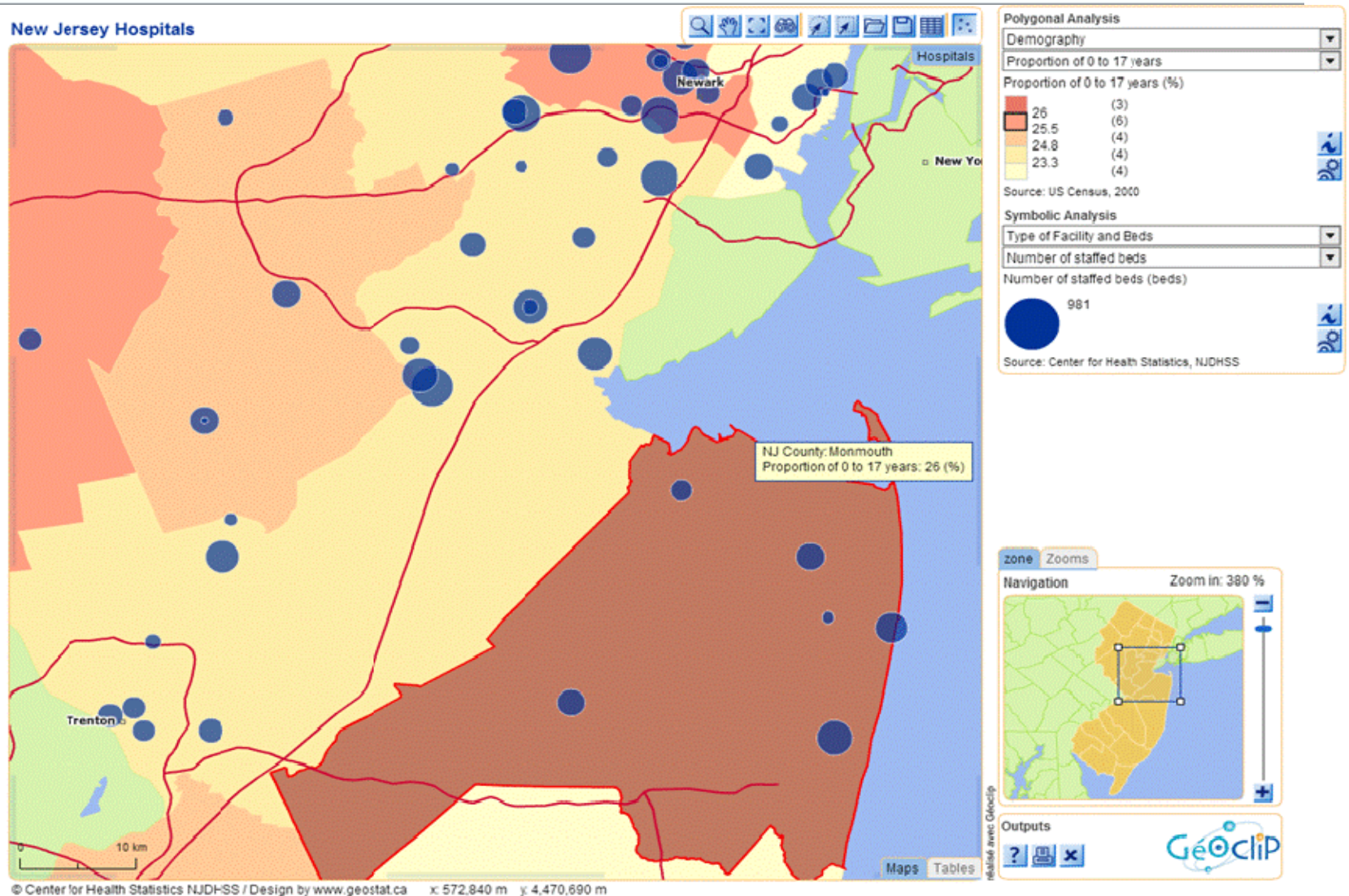
TJS Example: New Jersey Hospitals
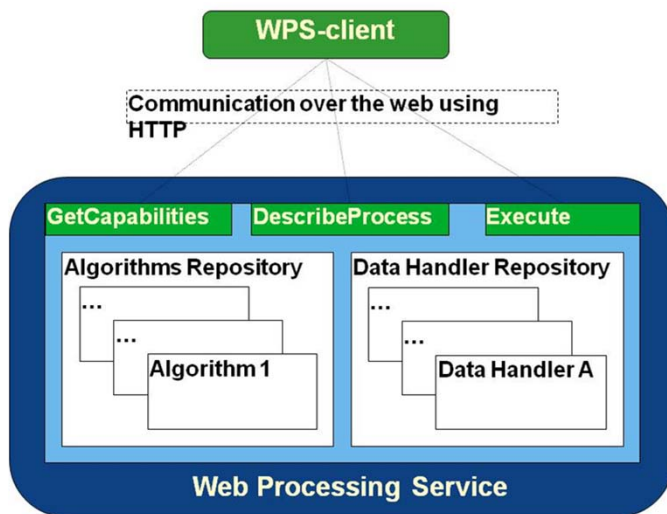
http://www.geostat.ca/realisation/sqlnj/carto.php?curCodeThemeSB=type&typindSB=R&curCodeIndSB=beds&couchesymb=ch

OGC®

Web Processing Service

- Provides developers with an agile and interoperable approach to deploying information processing, analysis, and dissemination capabilities as open web services

- Builds on the open geospatial standards framework

- Gaining use for geoanalytics, geoprocessing in the cloud, web service work flow (orchestration)

OGC®

OGC Web Processing Service

Example: StatisticsServer

- Implements popular statistical tools in a standards based WPS

- Deploys tools as web service

- Improves agility in deployment of statistical tools

- Avoids cost and time of hard coding

StatisticsServer

Apache Tomcat

Apache Axis2

StatisticsService

Controller

ApplicationConnector

config files

methods.wsdl

statisticsServer.xml

Java Apps

R

other Apps
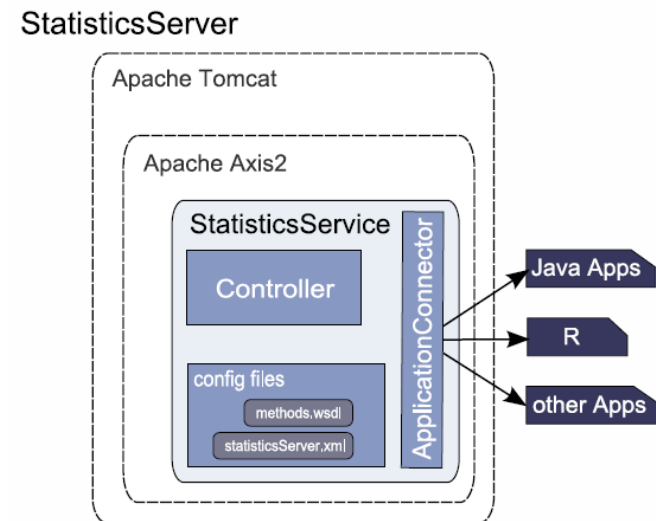
Figure 1: Statistical Web Service Architecture

Source: http://journal.imbio.de/articles/pdf/jib-140.pdf

OGC®

Web Processing Service Example
Ehabitat
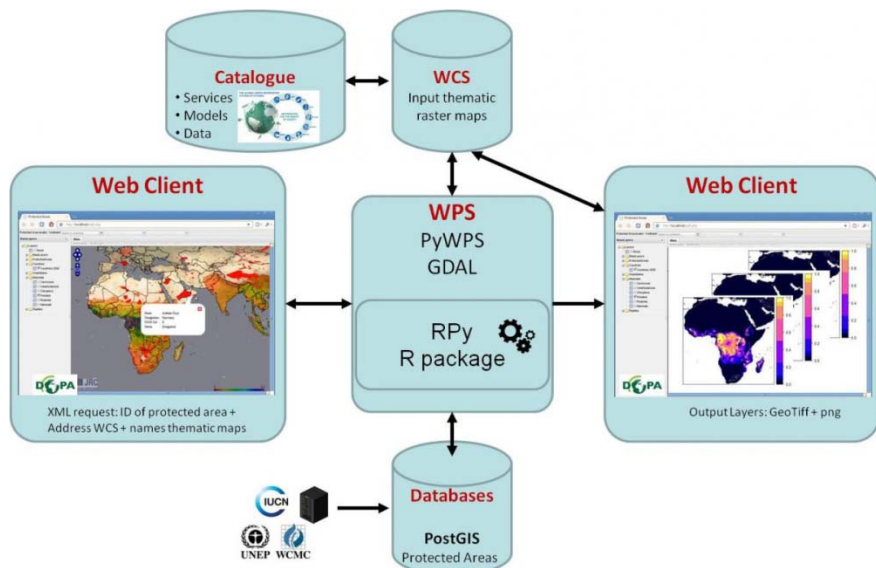
Catalogue
• Services
• Models
• Data

WCS
Input thematic
raster maps

Web Client

WPS
PyWPS
GDAL

RPy
R package

Web Client

XML request: ID of protected area +
Address WCS + names thematic maps

Output Layers: GeoTiff + png

IUCN

UNEP WCMC

Databases
PostGIS
Protected Areas

http://ehabitat.jrc.ec.europa.eu/content/developing-web-modeling-services-ecological-forecasting

See also:
http://www.ogcnetwork.net/pub/ogcnetwork/GEOSS/AIP3/documents/CCBio-eHabitat-ER-v2.0-FINAL.pdf

- Provides developers with an agile and interoperable approach to deploying information processing, analysis, and dissemination capabilities as open web services

- Builds on the OGC geospatial standards framework

- Gaining use for geospatial analytics, geoprocessing in the cloud, web service work flow (orchestration)

OGC®

Example - INTAMAP

- **Implemented as an OGC Web Processing Service**

- **Introduced as a** "**project for development of an interoperable framework for real time automatic mapping of critical environmental variables by extending spatial statistical methods and employing open, web-based, data exchange and visualisation tools**"

http://www.intamap.org/index.php

Try INTAMAP

Applications
▸ Radioactivity
▸ Air quality
▸ Video examples

INTAMAP system
▸ Send a WPS Request
▸ R intamap package
▸ R intamap examples
▸ R intamapInteractive
▸ psgp R package
▸ Adding new methods

INTAMAP Clients
▸ Generic INTAMAP client
▸ INTAMAP Java API
▸ SeeSharp
▸ Mobile client

Sample Data
▸ Radiation data
▸ NO2 Diffusion tube data
▸ NO2 Satellite data

Crossvalidation service
▸ CrossValidation request
▸ CrossValidation results

Using INTAMAP

INTAMAP sample datasets

This page supplies several datasets which can be used to test the INTAMAP service. Some of the csv files have no header with column names, but take the form **'x-coordinate', 'y-coordinate', 'value', 'error [optional]'**.

Using the simple web client, datasets with no error column will currently be handled using the 'automap' function of the INTAMAP WPS. Datasets with an error column will be handled using the 'psgp' function. These files may also be used as test cases for the INTAMAP API, where users may explicitly control the kriging option used.

Nitrogen Dioxide

Ground measurements: diffusion tube data for nitrogen dioxide.

Mean NO2 for 2007

•	0 - 8
•	9 - 19
•	20 - 28
•	29 - 36
•	37 - 44
•	45 - 52
•	53 - 62
•	63 - 74
•	75 - 91
•	92 - 123
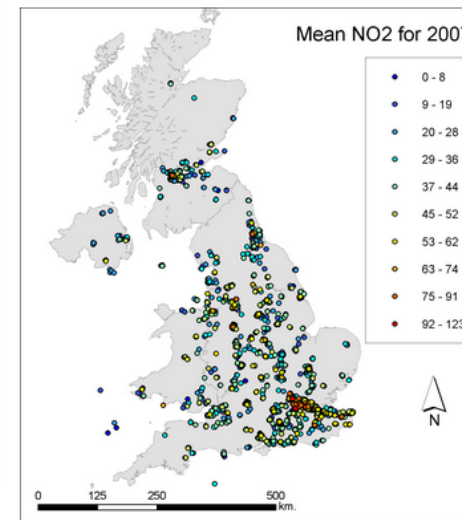
This dataset consists of 2213 point measurements, averaged over the course of the year 2007 from diffusion tube monitoring stations around the UK. The data is supplied by the UK National Air Quality Archive, and is Crown Copyright, but may be re-used if its source is acknowledged.

OGC®

Future Challenges

- Seamless integration of geospatial and statistics information is a critically important goal
  - Addressing metadata in this context is key - "fitness for use"

- Standards can help to ease this integration

- Some progress made, but there's more work / opportunity to achieve greater interoperability, and to address continuing challenges (definition/semantics, quality, provenance…)

- Let's identify ways in which we can work collaboratively to discuss, prototype and advance an interoperable global Statistical Geospatial information Framework

OGC®

Thank You

Mark Reichardt

mreichardt@opengeospatial.org

+1 301 840-1361

www.opengeospatial.org

www.myogc.org

OGC ®