First steps towards internationally integrating data and services in the solid Earth sciences and beyond

Lesley Wyborn[1], Ben Evans [1], Kerstin Lehnert[2], Tim Rawling[3], Jens Klump[4], Kirsten Elger[5], Simon Cox[6], Helen Glaves[7], Mohan Ramamurthy[8], Erin Robinson[9], and Shelley Stall[10]

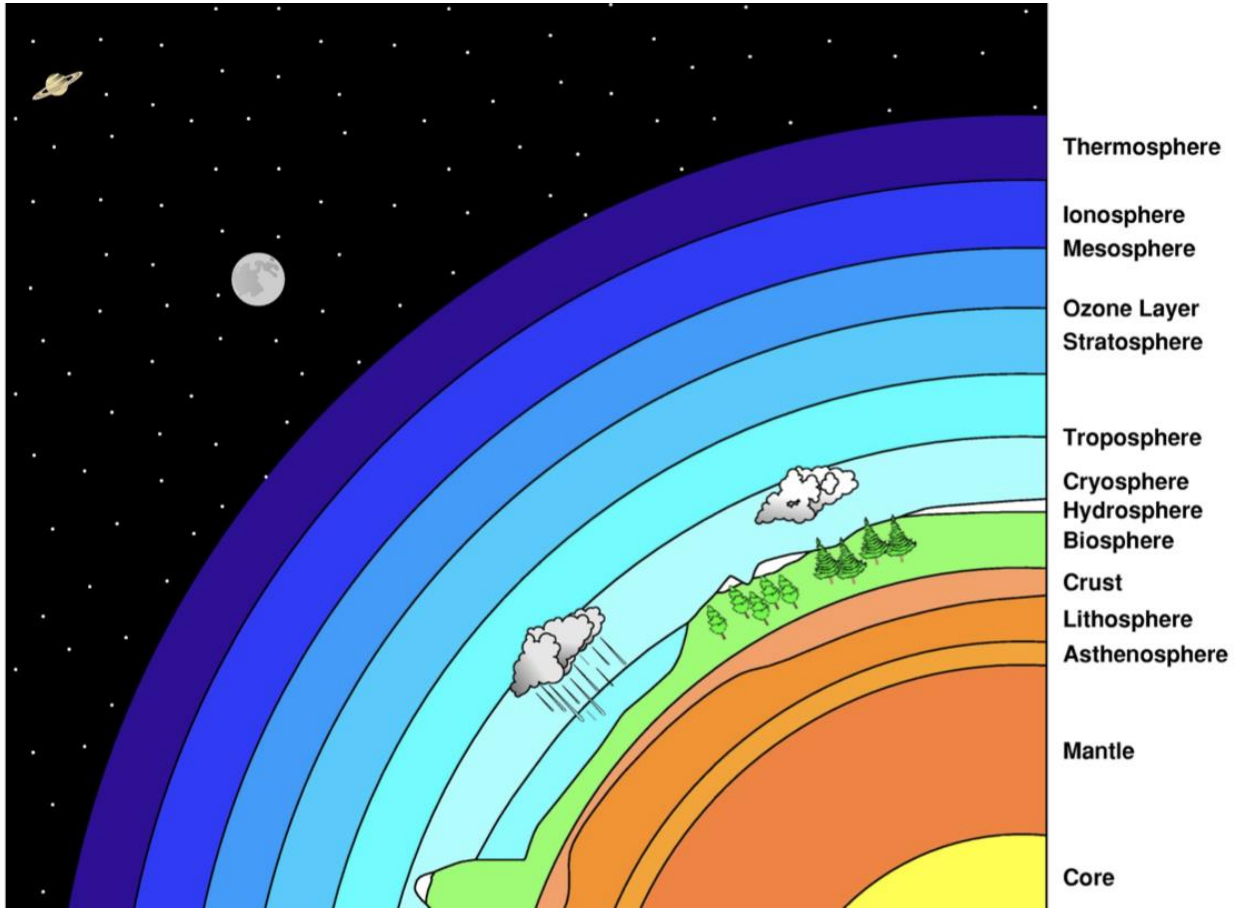
[1]National Computational Infrastructure, ANU, Australia
[2]Lamont-Doherty Earth Observatory, Columbia University, USA
[3]AuScope Ltd, Australia, [4]Mineral Resources, CSIRO, Australia
[5]GFZ German Research Centre for Geosciences, Germany
[6]Land and Water, CSIRO, Australia
[7]British Geological Survey, United Kingdom
[8]EarthCube, UCAR, USA
[9]Earth Science Information Partners, USA
[10]American Geophysical Union, USA

nci.org.au

Starting Premise

Evidence

1. Globally, solid Earth science data are collected by large numbers of organizations across the academic, government and industry sectors.
2. Spatially, the data collected covers multiple domains extending from the crust, through the lithosphere and mantle to the core.
3. Many observed phenomena cross national, if not continental, boundaries.



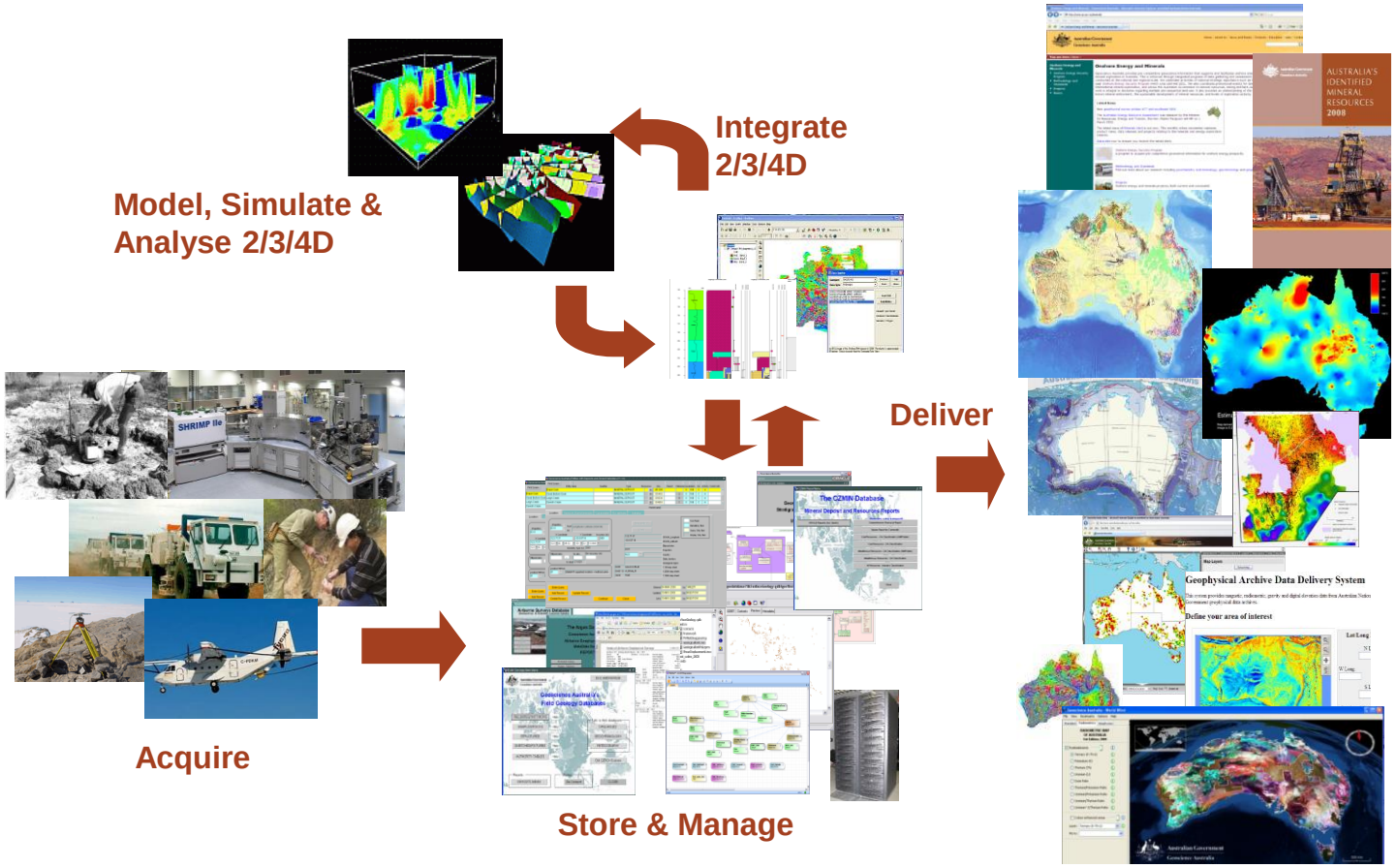
Question:

Why can't we work together to develop international networks of Earth science researchers to contribute to growing global challenges such as:

A. scarce non-renewable resources
B. risk reduction for natural hazards
C. fundamental research on the nature of the planet

© NCI Australia 2018

The last decade has seen a dramatic growth in online



- Earth science datasets
- Computational power, particularly utilising Cloud or HPC hosted data and compute resources.

But

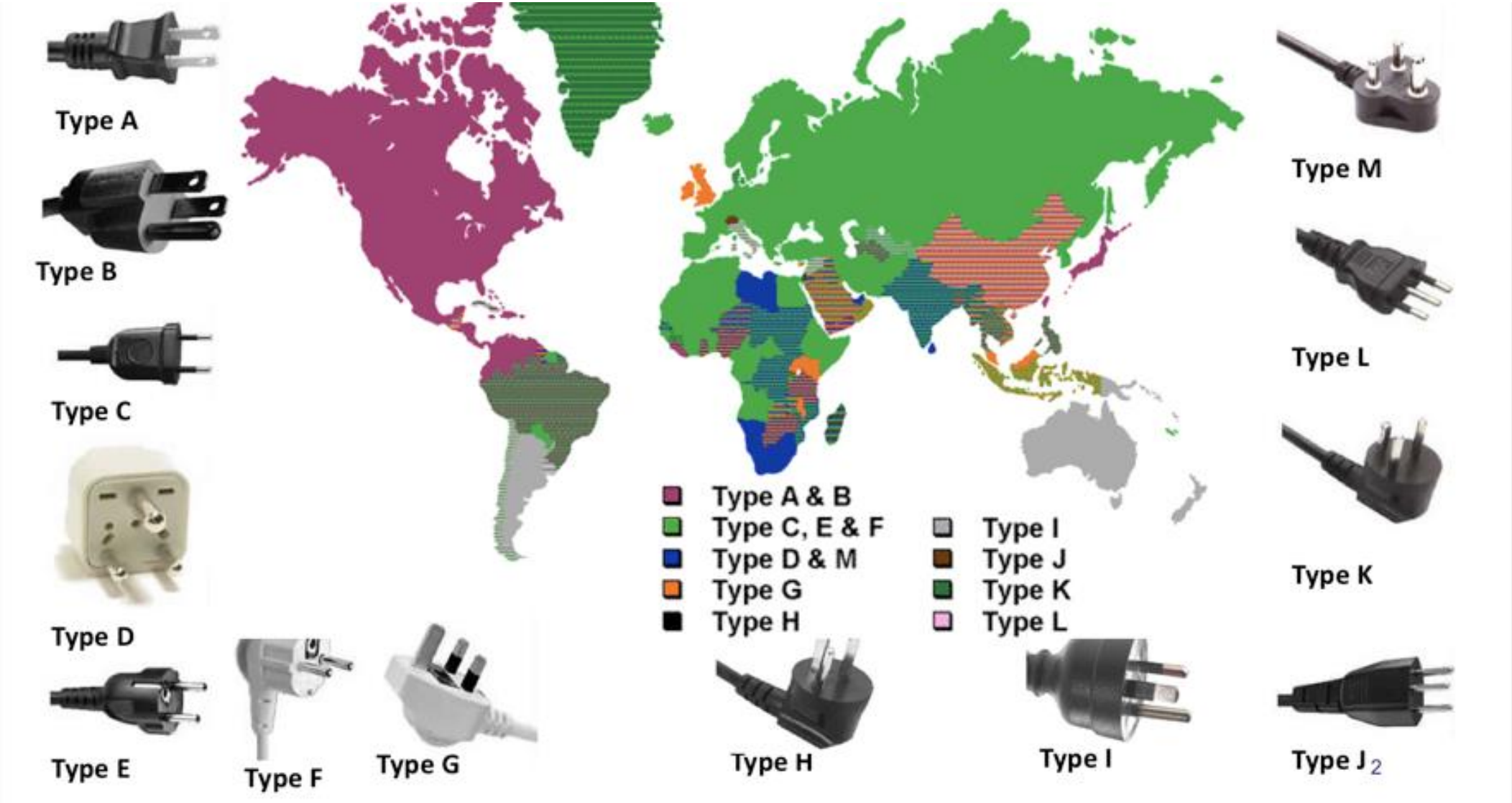
- There are inconsistent and incompatible data descriptions and formats
- Software is developed locally around specific applications and data sources
- There is a multiplicity of software providing similar and overlapping functions.

Quote from Industry supporter of a multi-client project in 2004:

"The Minerals Industry spends 80% of its time finding and reformatting data – what if that 80% could be used to develop better and smarter algorithms to process the data"

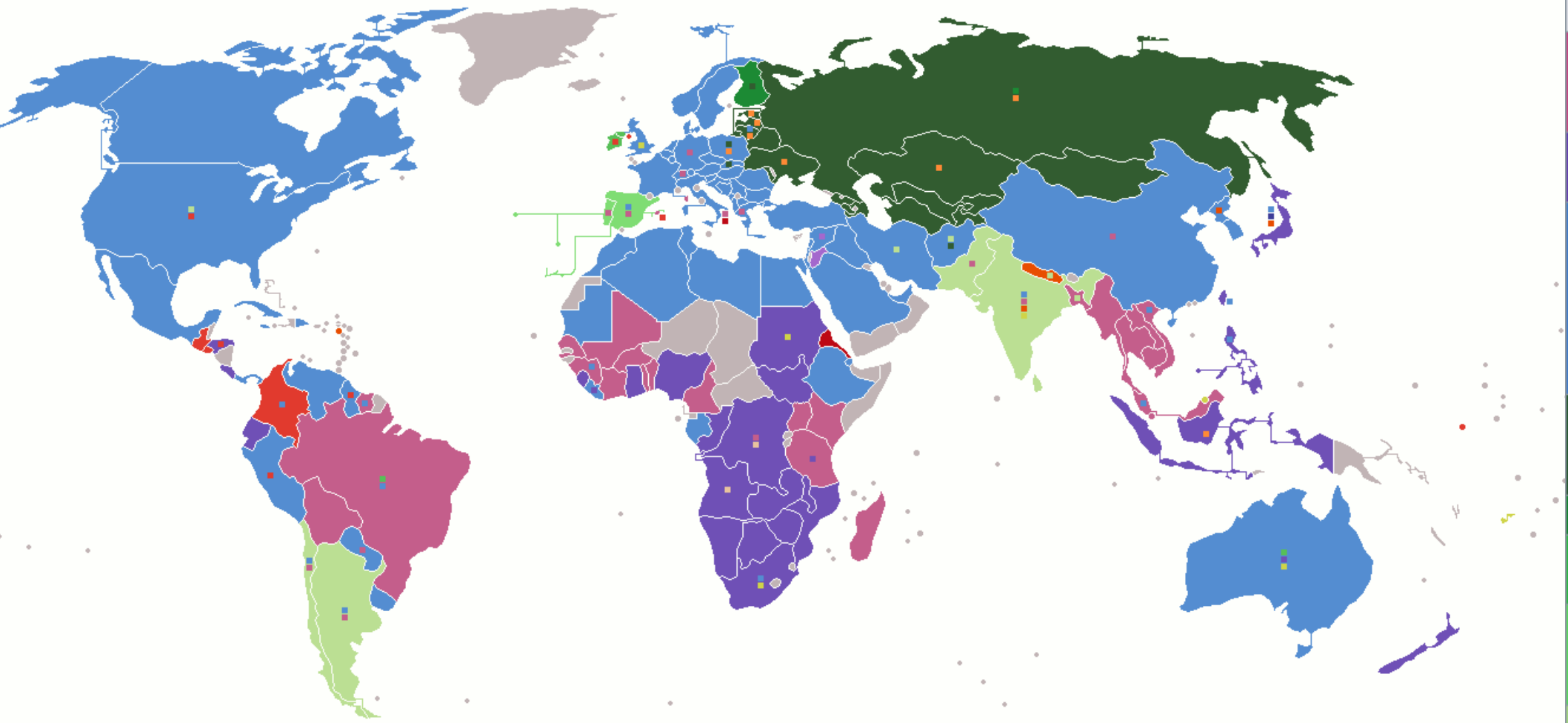
© NCI Australia 2018

Where are the community agreed standards for Interfacing?



© NCI Australia 2018

But is it also about infrastructure

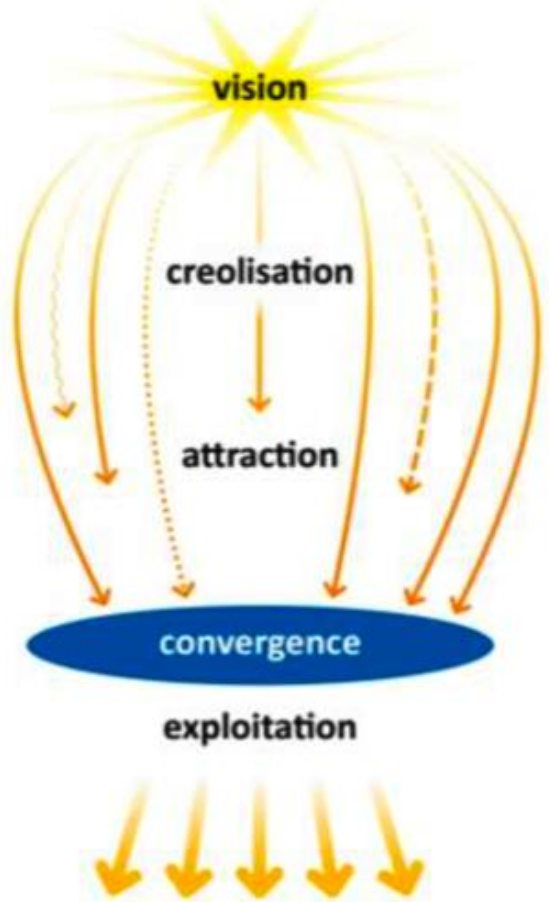


Gauge	Name
1,000 mm (3 ft 3⅜ in)	Metre gauge
1,067 mm (3 ft 6 in)	Three foot six inch gauge
1,435 mm (4 ft 8½ in)	Standard gauge
1,520 mm (4 ft 11 27/32 in)	Five foot and 1520 mm gauge
1,524 mm (5 ft)	Finnish gauge
1,600 mm (5 ft 3 in)	Five foot three inch gauge
1,668 mm (5 ft 5 21/32 in)	Iberian gauge
1,676 mm (5 ft 6 in)	Five foot six inch gauge

© NCI Australia 2018

Patterns of Infrastructure Development

Kerstin Lehnert: EGU McHarg
Lecture April 2018



Wittenburg & Strawn 2018

1. Inventions and development of start-up systems
2. Technology transfer between regions and also society (creolization)
3. Planning for system growth where "reverse salients" need to be tackled
4. Substantial momentum (mass, velocity, direction)

System Building

Growth

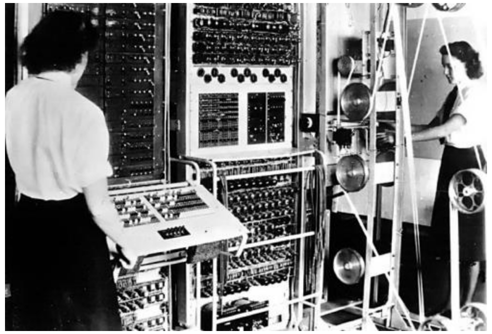
Consolidation

April 10, 2018 17

Wittenburg, P., and Strawn, G., 2018. Common Patterns in Revolutionary Infrastructures and Data
https://www.rd-alliance.org/sites/default/files/Common_Patterns_in_Revolutionising_Infrastructures-final.pdf

The lessons of history: **Industrial Revolution vs the Geoinformatics Revolution**

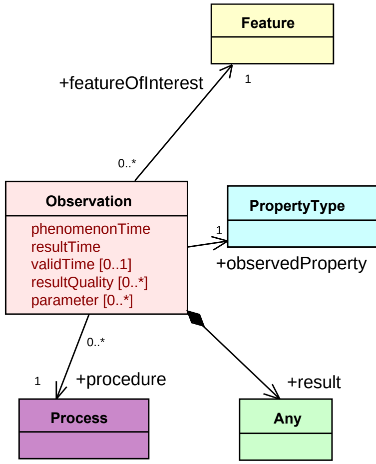
James Watt



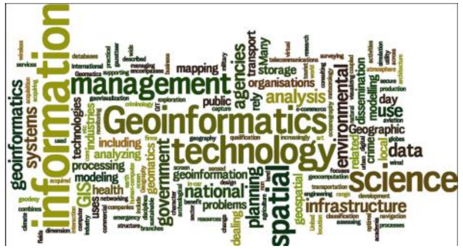
Simon Cox

- **1776 - Invention of steam engine**
- **1829 - Invention of railways**
- **1834 - First rail-networks**
- **1880 - First Standards Association for individual components**
- **1890 - Manufacturing age**

- **1940 - Invention of the computer**
- **1989 - First Generation Internet**
- **1996 - First Grid networks**
- **1996 - First Standards for components: W3C & XML appear**
- **2007 - OneGeology & Geonformatics Age**



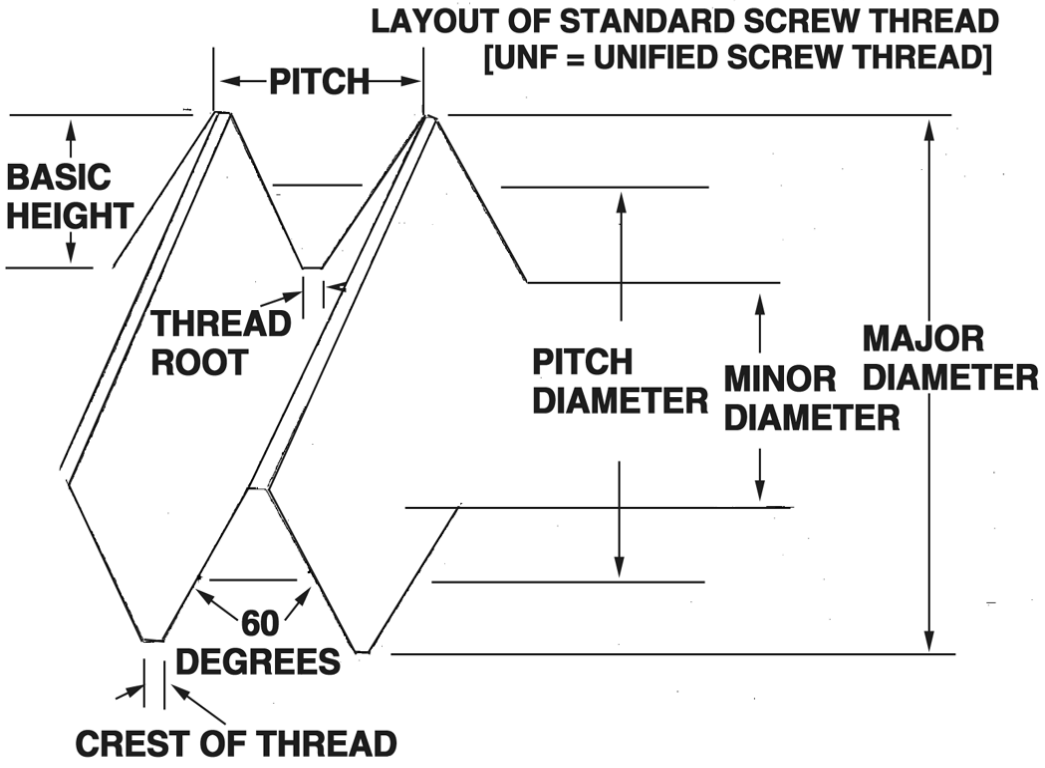
Observation & Measurement Model



Peter Drucker - Beyond the Information Revolution: http://www.theatlantic.com/issues/99oct/9910drucker.htm

The Screw – the Minimum Component of the Industrial Revolution

The anatomy of the thread of a screw



Standards of the industrial age were developed at the level of the lowest common component

© NCI Australia 2018

Containers, Content and Packaging

Technical Metadata
(Computer Scientists)

Content Metadata
(Domain Specialists)

Bibliographic Metadata
(Data Librarian Specialists)

https://www.notonthehighstreet.com/auntiemims/product/24-personalised-gift-labels-from-the-north-pole http://arnoldpackaging.com/products-services/the-container/ https://packagingsales.com/tamper-evident-packaging/#!prettyPhoto/0/

© NCI Australia 2018

Competition vs Collaboration

https://www.walthers.com/products/trains/

Competition

Collaboration

We need collaborative informatics and computational platforms on which competitive research can be undertaken

But where do we put the boundary?

© NCI Australia 2018

What do we have?

NCI AUSTRALIA

ICSU CODATA

ICSU WORLD DATA SYSTEM

NSF ARCTIC DATA CENTER

EGU European Geosciences Union

GEOSS Global Earth Observation System of Systems

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

IEEE

AGU American Geophysical Union

GEO GROUP ON EARTH OBSERVATIONS

AuScope

OGC Making location count.

GeoScience DWG

W3C

BELMONT FORUM

e-INFRASTRUCTURES DATA MANAGEMENT

Earth System Grid Federation ESGF

CEOS Committee on Earth Observation Satellites

EARTHCUBE

RDA RESEARCH DATA ALLIANCE

Earth, Space, and Environmental Sciences Interest Group

ESIP The Federation of Earth Science Information Partners Fostering connections to make data matter

buildingSMART International home of openBIM

EUROPEAN SCIENTIFIC ASSOCIATION EISCAT

GÉANT

Environmental Data Initiative

GMSG

INSPIRE Infrastructure for Spatial Information in Europe

DataONE Data Observation Network for Earth

enes EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

TERN

EPOS EUROPEANPLATEOBSERVINGSYSTEM

plus ENVRI

SWIM DRINK FISH

Make Data FAIR AGU

LTER Long-Term Ecosystem Research in Europe

IMOS

SeaDataNet

IGSN

ONE Geology

energistics Energy Standards

CC BY

© NCI Australia 2018

EGU2018-12639: Wyborn et al. (lesley.wyborn@anu.edu.au)

European Plate Observing System (EPOS)

What is EPOS?

Overarching European Research infrastructure for the solid Earth

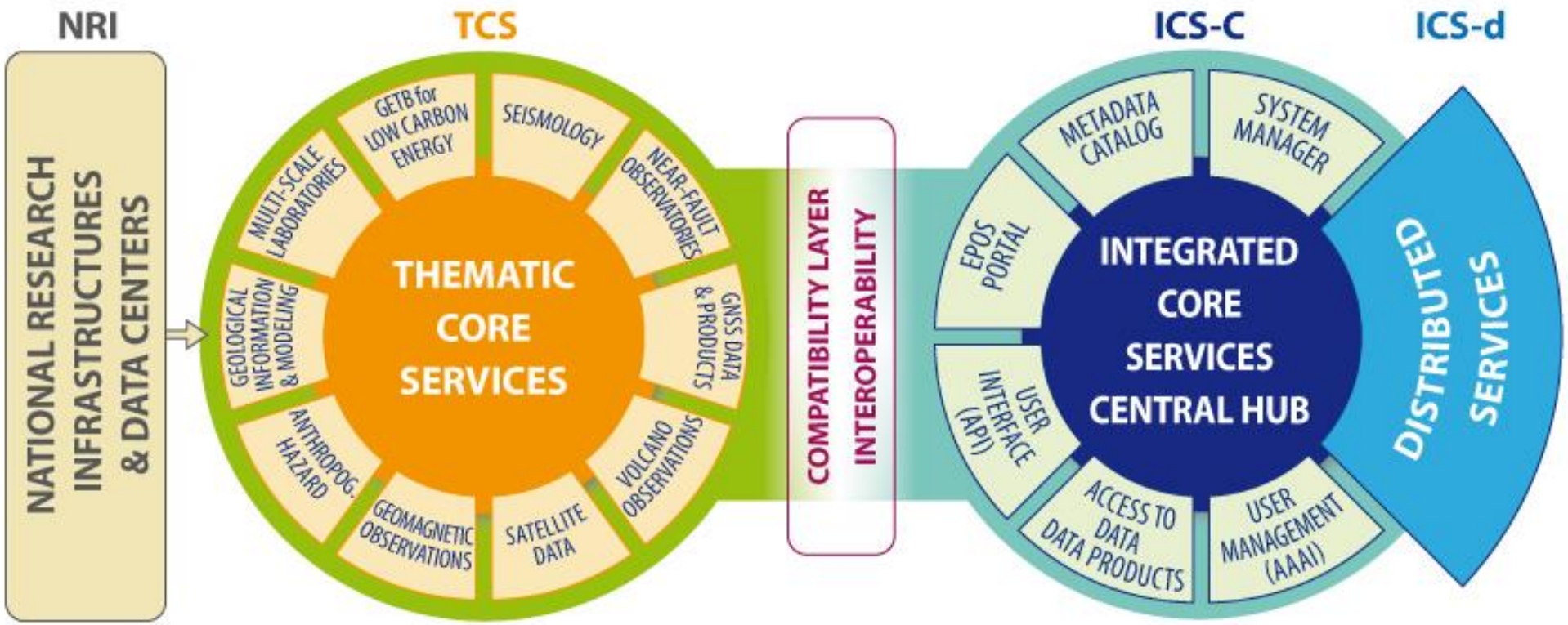
Integrating diverse range of European Research Infrastructures for solid Earth science

On the ESFRI Roadmap (Landmark?)

EU-funded H2020 project

EPOS ERIC (European Research Infrastructure Consortium): legal entity expected in 2018

EPOS Functional Architecture



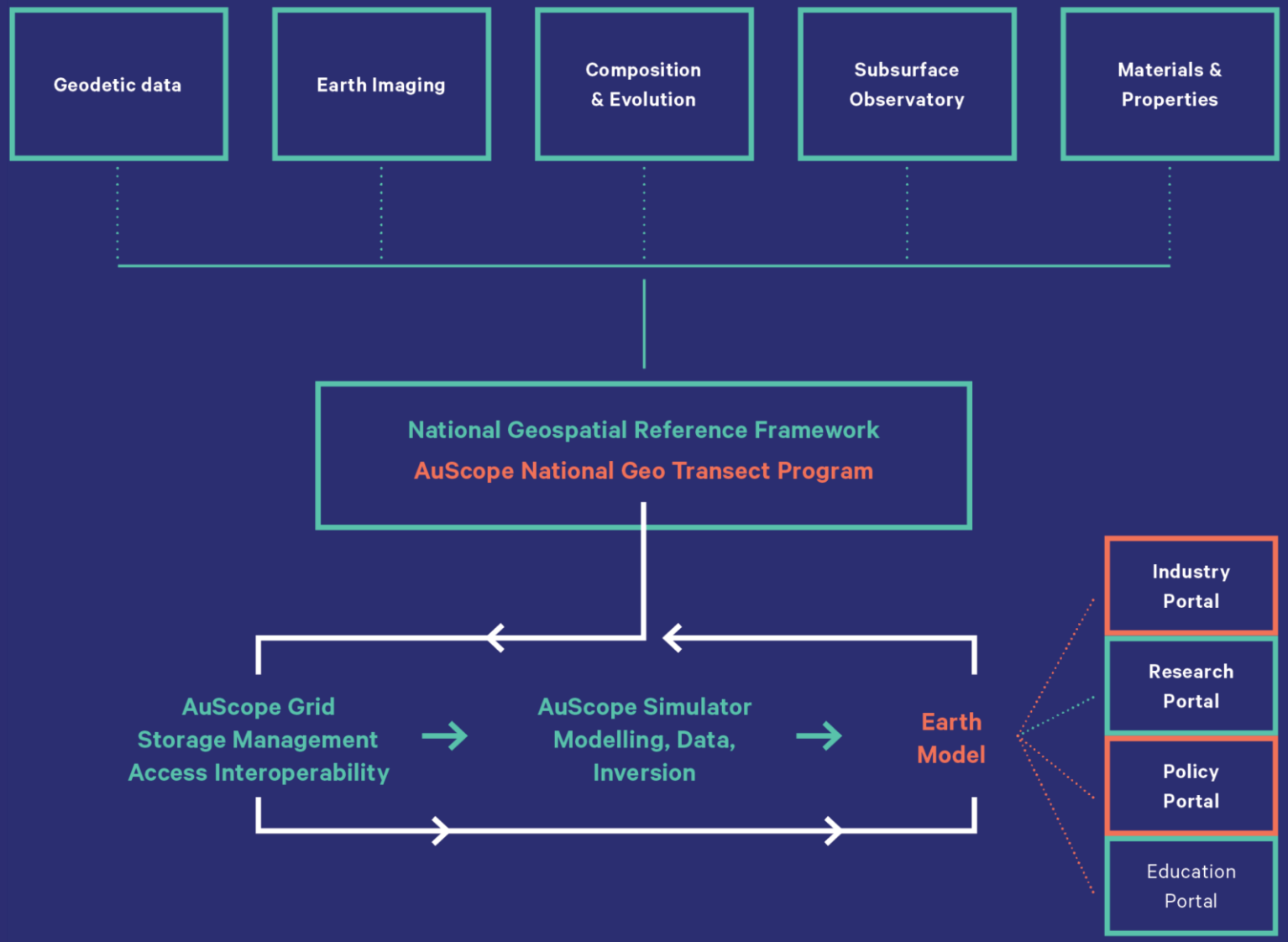
Data collection
Data generation

Data curation
Metadata
Registration
Community Services

Interoperability
Brokerage

Metadata registry
Processing
Aggregation
Integrated analyses
Visualization

AuScope Model



Introduction

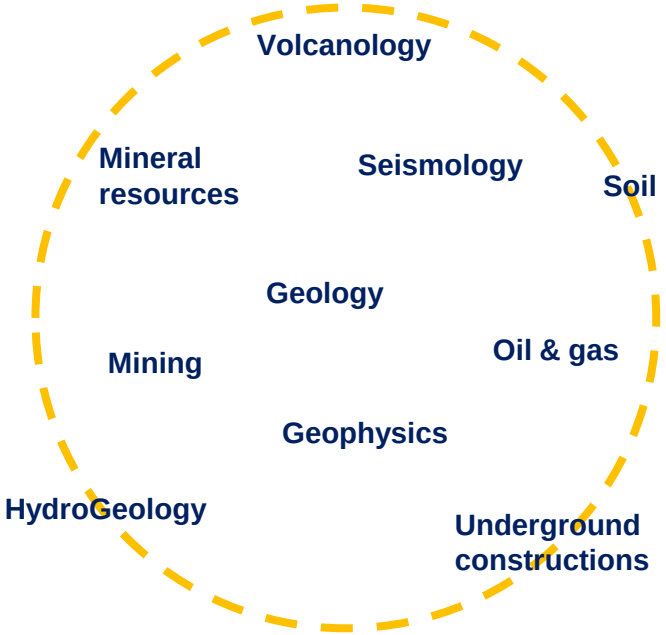
ENVRIplus

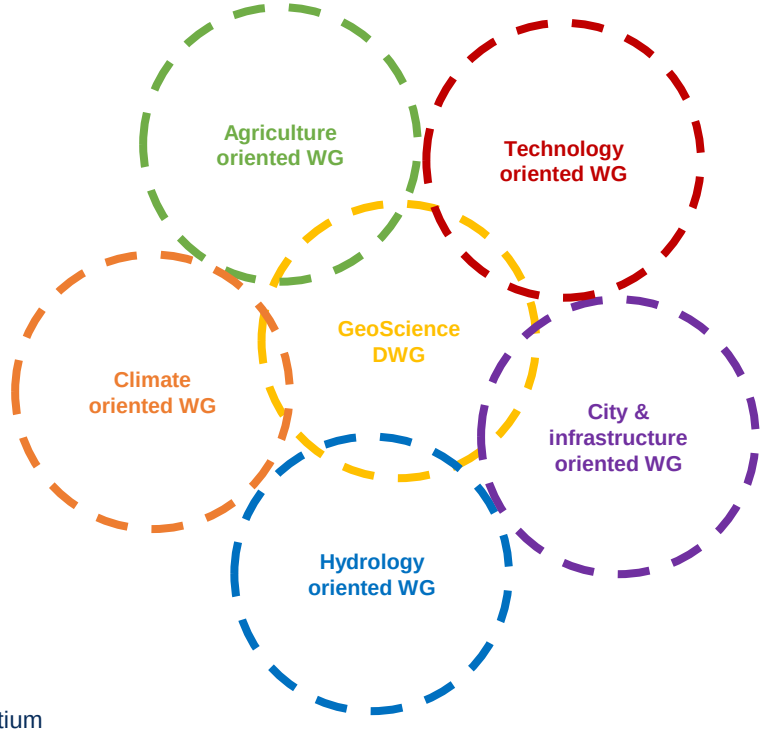
Participation

- 20 research infrastructures
- 7 associated RIs
- 4 domains (biosphere, marine, atmosphere and solid Earth)

GeoScience DWG position

- Main targets:
  - Harmonize geoscience data expression and facilitate usage
  - Stay connected and even propose enhancements to technologies / solutions
  - Connect people interested in the geoscience topic
  - Ensure proper connections with other groups / communities inside and outside OGC





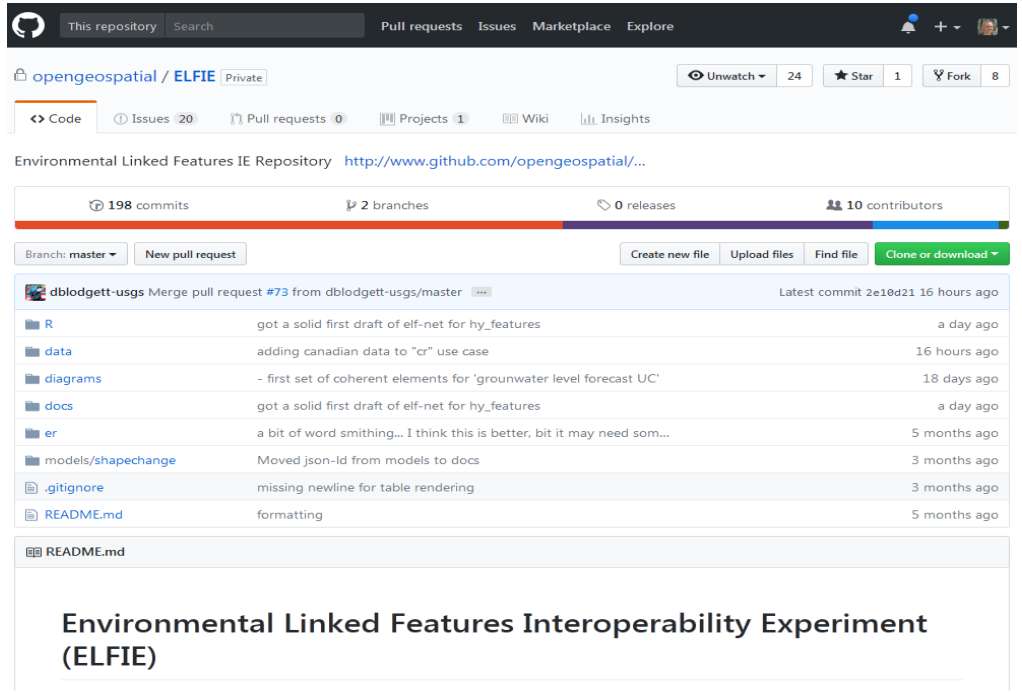
Copyright © 2018 Open Geospatial Consortium

Another IE driven by the GeoScience community

- Environmental Linked Features IE (ELFIE)
- ‘Demonstrate the use of existing and pending OGC standards for the encoding of environmental observation data in an integrated dataset of features linked according to ReSTful and Linked Data principles.’

- Initiators:
  - U.S. Geological Survey (US)
  - Land Information New Zealand (NZ)
  - BRGM (FR)
- Participants
  - Tumbling Walls and Dewberry (US)
  - Meta-linkage (AU)
  - INSPIRE (EU)
  - Natural Resources Canada (CA)
  - Manaaki Whenua and Horizons Regional Council (NZ)

From: https://github.com/opengeospatial/ELFIE



Copyright © 2018 Open Geospatial Consortium

ESIP GROUPS

Standing Committees

- Data Stewardship
- Education
- Information Technology and Interoperability
- Products and Services (Retired)
- Semantic Technologies

Administrative Committees

- Constitution and Bylaws (Renamed Governance)
- Finance and Appropriations
- Partnership
- Nominations

Working groups

- Visioneers
- Energy & Climate
- Data Management Training

Clusters

- Ag & Climate
- CLEAN Network
- Cloud
- Data Coordination
- Data Mgmt Training
- Data to Decisions*
- Disaster Lifecycle
- Discovery
- Documentation
- Drones
- Earth Science Data Analytics
- Energy & Climate
- EnviroSensing
- Information Quality
- Pre-prints/EarthArXiv*
- Science Communication
- Science Software
- Semantic Technologies
- Software and Services Citations**
- Sustainable Data Mgmt
- Toolmatch
- Usability
- VR/AR
- Web Services

*New Cluster

** Starting 2018

… And yours?

http://esipfed.org/collaboration-areas

ESIP The Federation of Earth Science Information Partners

Enabling FAIR Data Project

Align publishers and repositories in following best practices to enable FAIR and open data and to create workflows so that researchers will have a simplified, common experience when submitting their paper to Earth and space science journals

1. ESS publishers will follow consistent and rigorous policies and guidelines for sharing and citing data used in scholarly literature;

2. Open ESS repositories will enable those policies and other data applications by providing persistent identifiers, rich metadata, and related services for the data they hold; and

3. ESS researchers will understand how to consistently share, document, and reference the data they collect and use

https://eos.org/agu-news/enabling-fair-data-across-the-earth-and-space-sciences

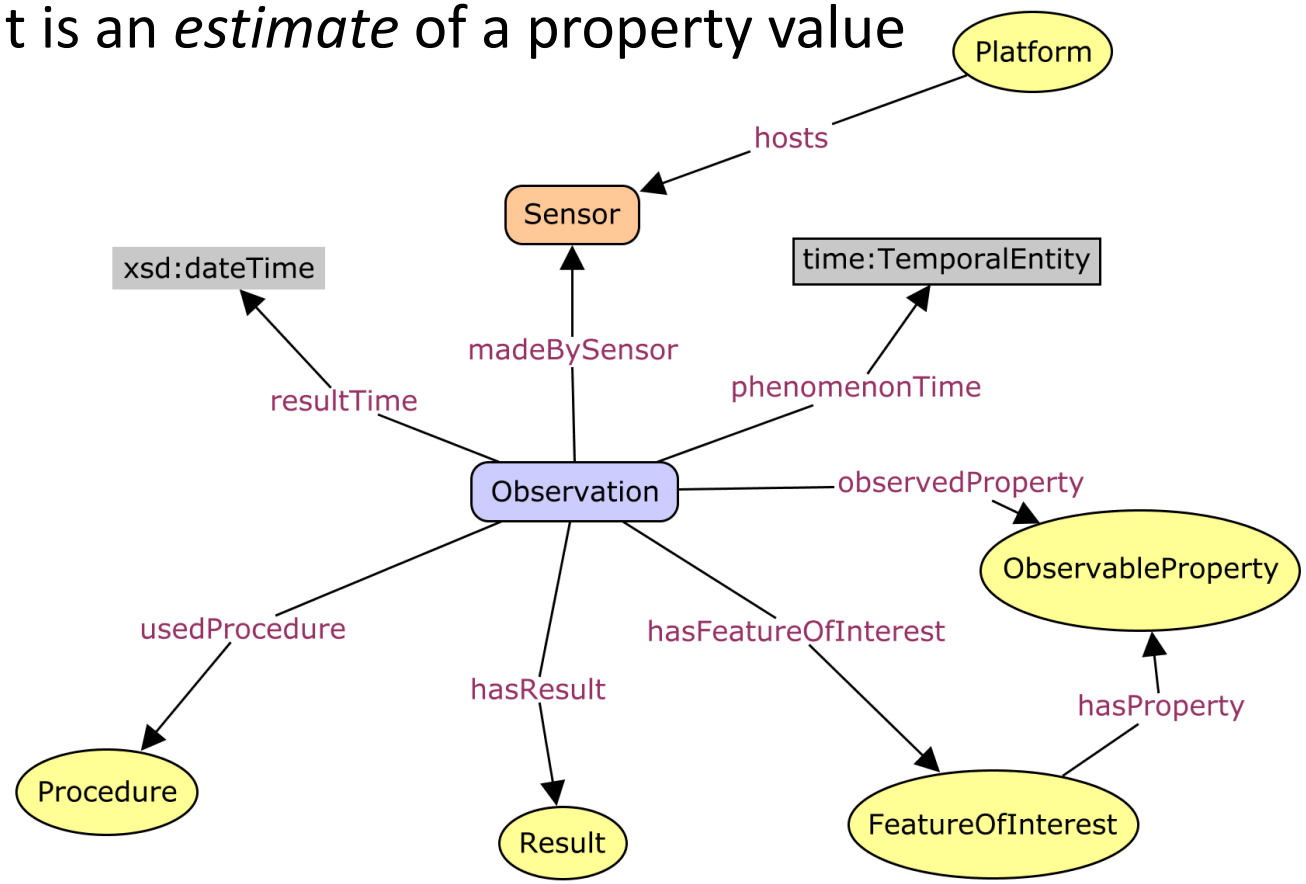
SSN/SOSA Observation

(SSN= Semantic Sensor Network, SoSA = Sensor Observation, Sample & Actuatir)

Universal Pattern in Science:
An observation **is** any action whose result is an *estimate* of a property value

Scope

- In situ observations
- Remote sensing
- Ex-situ (laboratory) observations
- Numerical models/simulations
- Forecasts
- Interpretations, classifications



https://www.w3.org/TR/vocab-ssn/

© NCI Australia 2018

Motivation for a common model

- Integrated analysis and modelling
  - Discovery & data integration a significant challenge
  - Different disciplines use different words for the same things
- O&M provides a standard, domain-neutral terminology
  - Reduces ambiguity
  - Increases interoperability
- **X-disciplinary terminology**

  ~~Many private contracts~~
  one public agreement!

Remote sensing
Sensor
Value
Parameter
Scene

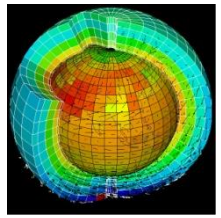
Earth science
Algorithm, code, simulator
Model, field
Variable
Volume, grid

Metrology
Instrument
Value
Measurand
Sample

Chemistry
Instrument, analytical process
Analysis
Analyte
Sample

Environmental monitoring
Gauge, sensor
Value, time-series
Parameter
Station

Observations & Measurements
procedure
result
observed property
feature of interest



Slide courtesy of Simon Cox

© NCI Australia 2018

What do we need to progress

What we need now are mechanisms to internationally link these major infrastructures to provide:

1. efficiencies in funding (stop reinventing the wheel!)
2. an environment where the research efforts can create globally interoperable networks of solid Earth science data, information systems, software and researchers

But how?

© NCI Australia 2018

Ocean Data Interoperability Platform (ODIP/ODIP II)

ODIP: October 2012 – September 2015

ODIP II: April 2015 – March 2018

Collaborative project:

- Europe, USA, Australia, Canada



Partners

Europe: 19 EU-funded partners (9 countries)

- NERC-BGS/BODC, MARIS, OGS, IFREMER, HCMR, ENEA, ULG, CNR, RBINS, TNO, AWI, BSH, RIHMI-WDC, VLIZ, UniHB, CSIC, 52ONorth, IEEE, SOCIB

USA: 11 Organisations

- Scripps Institution of Oceanography (SIO), Woods Hole Oceanographic Institute (WHOI), Lamont-Doherty Earth Observatory (LDEO), Florida State University (FSU): Center for Ocean-Atmospheric Prediction Studies, NOAA, US-IOOS, UNIDATA, MMI, ESRI

Australia: 5 organisations

- University of Tasmania (IMOS), CSIRO, Geoscience Australia (GA), NCI, ANDS

International: 4 organisations

- UNESCO IOC-IODE, GEO/GEOSS, POGO, ICSU – WDS

ODIP/ODIP II Objectives and Outputs

- Development of a series of prototype interoperability solutions demonstrating coordinated approach to marine data management on a global scale
- Promote development of a common global framework for marine data management
- Output 1: Metadata from regional data discovery systems accessible via global portals GEOSS portal and IODE Ocean Data Portal (ODP)
- Output 2: ISO Cruise Summary Reports harvested from regional nodes and exposed in the POGO portal
- Output 3: Establishing a global SWE community of practice and working towards OGC Sensor Web enablement standards for ocean sensors

First Steps

Formation of the ESIP/RDA Earth, Space and Environmental Sciences Interest Group
https://rd-alliance.org/groups/esiprda-earth-space-and-environmental-sciences-ig

Objective: Focus on awareness, and coordination where applicable, of independent efforts across the international Earth, space, and environmental science communities.

© NCI Australia 2018

Who are we?

Lesley Wyborn: National Computational Infrastructure, ANU, Australia (lesley.wyborn@anu.edu.au)

Ben Evans: National Computational Infrastructure, ANU, Australia (ben.evans@anu.edu.au)

Kerstin Lehnert: Lamont-Doherty Earth Observatory, Columbvia University, New York, USA (lehnert@ldeo.columbia.edu)

Tim Rawling: AuScope Ltd, Melbourne, Australia (tim.rawling@unimelb.edu.au)

Jens Klump: Mineral Resources, CSIRO, Kensington, WA, Australia (jens.klump@csiro.au)

Kirsten Elger: GFZ German Research Centre for Geosciences, Potsdam, Germany (kelger@gfz-potsdam.de)

Simon Cox: Land and Water, CSIRO, Clayton, Vic, Australia (simon.cox@csiro.au)

Helen Glaves: British Geological Survey, Nottingham, UK (hmg@bgs.ac.uk)

Mohan Ramamurthy: EarthCube, UCAR, Boulder, USA (mohan@ucar.edu)

Erin Robinson: Earth Science Information Partners, Boulder, CO, United States (erinrobinson@esipfed.org)

Shelley Stall: American Geophysical Union, Washington, D.C., United States (sstall@agu.org)

© NCI Australia 2018