Open Science goes Geo

Part III: Beyond Data and Software

European Geosciences Union
General Assembly 2015
Vienna | Austria | 12 – 17 April 2015

Conveners: Martin Hammitzsch and Samuel Illingworth

4-in-a-row

Short Course series: Open Science goes Geo

★ Part I: Research Data
Tuesday, 14 Apr, 17:30–19:00 / Room B1

★ Part II: Scientific Software
Wednesday, 15 Apr, 17:30–19:00 / Room B1

★ Part III: Beyond Data and Software
Thursday, 16 Apr, 17:30–19:00 / Room B5

★ Part IV: Winning Horizon 2020 with Open Science
Friday, 17 Apr, 08:30–10:00 / Room B4

Part III: Beyond Data and Software

Speakers

- ★ Rolf Sander
- ★ Jens Klump
- ★ Kerstin Lehnert
- ★ Dorit Kerschke represented by Joachim Wächter
- ★ Chris Herwig
- ★ Jochen Klar
- ★ Joachim Wächter

Part III: Beyond Data and Software

Today's menu

- ★ Geoscientific Model Development
- ★ International Geo Sample Number
- ★ Internet of Samples
- ★ Spatial Data Infrastructures
- ★ Research Infrastructures
- ★ Virtual Research Environments
- ★ Research Infrastructure Maturity Model

‘Mini’ Panel Discussion

Questions

- ★ save until the end
- ★ twitter hashtag #egu15sc25

Answers

- ★ at the end
- ★ follow-up later on

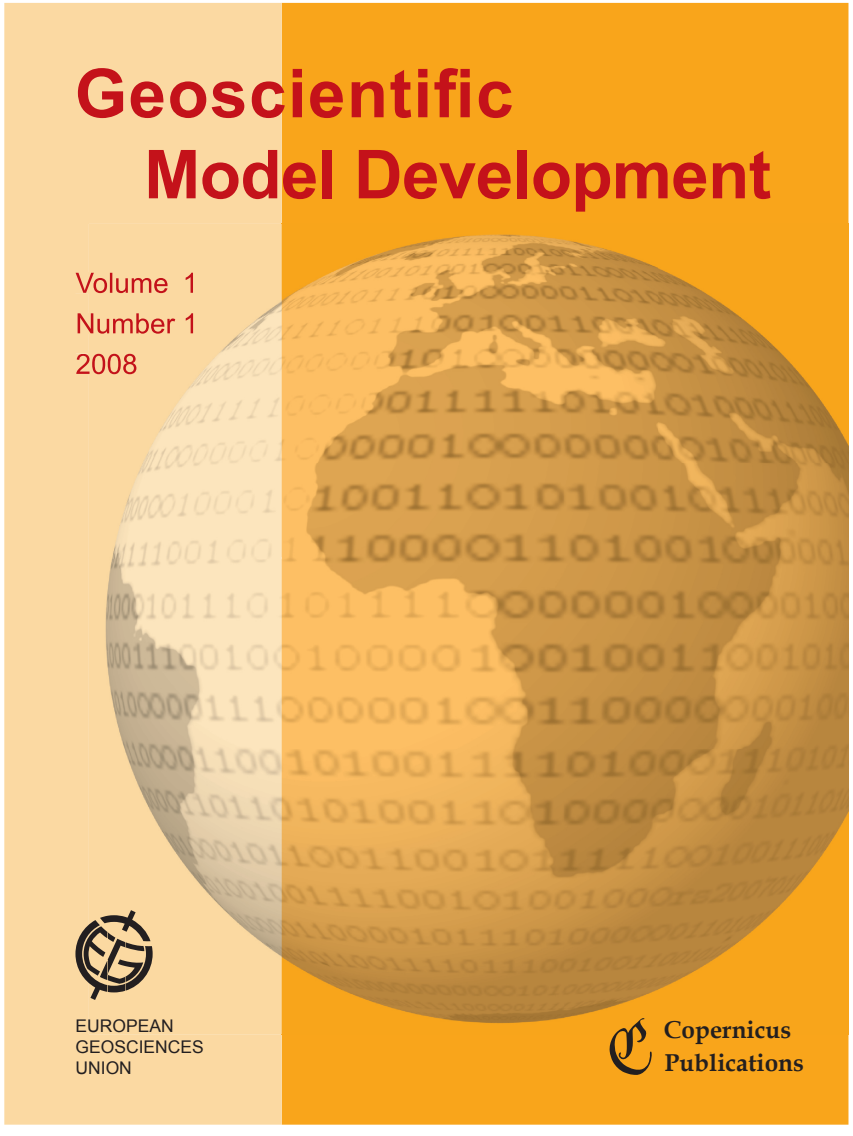
Geoscientific Model Development (GMD)

Julia Hargreaves, Astrid Kerkweg, Dan Lunt, Robert Marsh, Andy Ridgwell, Didier Roche, Ian Rutt & Rolf Sander[1]

[1]Max-Planck Institute for Chemistry, Mainz, Germany

2015-04-16

Geoscientific Model Development (GMD)



- What is GMD?
- Why do we need GMD?
- What is published in GMD?

What is GMD?

- International scientific journal

What is GMD?

- International scientific journal
- Description and evaluation of numerical models of the Earth System and its components

What is GMD?

- International scientific journal
- Description and evaluation of numerical models of the Earth System and its components
- Published since 2008 by Copernicus Publications on behalf of the European Geosciences Union (EGU)

What is GMD?

- International scientific journal
- Description and evaluation of numerical models of the Earth System and its components
- Published since 2008 by Copernicus Publications on behalf of the European Geosciences Union (EGU)
- Impact factor 6.086 (2013)

What is GMD?

- International scientific journal
- Description and evaluation of numerical models of the Earth System and its components
- Published since 2008 by Copernicus Publications on behalf of the European Geosciences Union (EGU)
- Impact factor 6.086 (2013)
- Open access

What is GMD?

- International scientific journal
- Description and evaluation of numerical models of the Earth System and its components
- Published since 2008 by Copernicus Publications on behalf of the European Geosciences Union (EGU)
- Impact factor 6.086 (2013)
- Open access
- Interactive discussion
  - reviews: public
  - referee: known or anonymous

Why do we need GMD?

Reproducibility: If model details are not fully described and published, then there is no way for other scientists to replicate the findings.

Why do we need GMD?

Reproducibility: If model details are not fully described and published, then there is no way for other scientists to replicate the findings.

Accessibility: Authors are (strongly!) encouraged to make code and user manuals available.

- Electronic supplement
- Data archive with DOI (e.g., zenodo)

Why do we need GMD?

Reproducibility: If model details are not fully described and published, then there is no way for other scientists to replicate the findings.

Accessibility: Authors are (strongly!) encouraged to make code and user manuals available.

- Electronic supplement
- Data archive with DOI (e.g., zenodo)

Traceability: Model versioning, collect papers about different versions in Special Issues

Why do we need GMD?

Reproducibility: If model details are not fully described and published, then there is no way for other scientists to replicate the findings.

Accessibility: Authors are (strongly!) encouraged to make code and user manuals available.

- Electronic supplement
- Data archive with DOI (e.g., zenodo)

Traceability: Model versioning, collect papers about different versions in Special Issues

Peer-review: Reviewers comment on

- Manuscript
- Model code (optional)

Why do we need GMD?

Reproducibility: If model details are not fully described and published, then there is no way for other scientists to replicate the findings.

Accessibility: Authors are (strongly!) encouraged to make code and user manuals available.

- Electronic supplement
- Data archive with DOI (e.g., zenodo)

Traceability: Model versioning, collect papers about different versions in Special Issues

Peer-review: Reviewers comment on

- Manuscript
- Model code (optional)

Recognition: Peer-reviewed publication for developer (developer $\neq$ user)

What is published in GMD?

Model Description papers:

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools
- model name and version number in the title

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools
- model name and version number in the title
- three parts: main paper, manual, source code

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools
- model name and version number in the title
- three parts: main paper, manual, source code
- describe underlying scientific basis and technical details

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools
- model name and version number in the title
- three parts: main paper, manual, source code
- describe underlying scientific basis and technical details
- model web page, hardware and software requirements, licence

What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools
- model name and version number in the title
- three parts: main paper, manual, source code
- describe underlying scientific basis and technical details
- model web page, hardware and software requirements, licence
- examples of model output

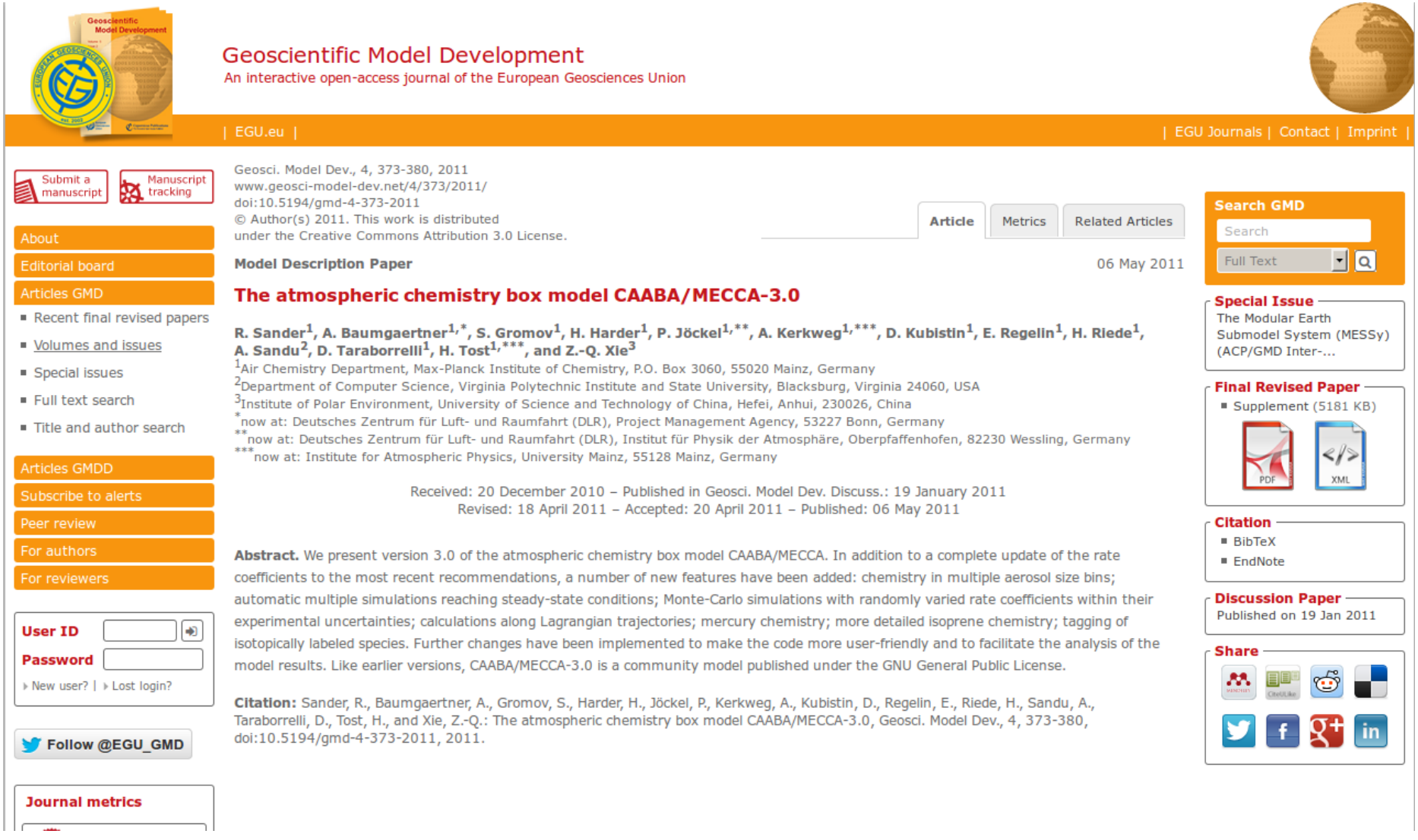
What is published in GMD?

Model Description papers:

- comprehensive descriptions of numerical models
- detailed, complete, rigorous, and accessible to a wide community of geoscientists
- complete models, components, modules, frameworks and utility tools
- model name and version number in the title
- three parts: main paper, manual, source code
- describe underlying scientific basis and technical details
- model web page, hardware and software requirements, licence
- examples of model output
- section "Code availability" (always available for editor and referees)

What is published in GMD?

Model Description paper example:



Geoscientific Model Development
An interactive open-access journal of the European Geosciences Union

| EGU.eu | | EGU Journals | Contact | Imprint |

Submit a manuscript | Manuscript tracking

About
Editorial board
Articles GMD
- Recent final revised papers
- Volumes and issues
- Special issues
- Full text search
- Title and author search

Articles GMDD
Subscribe to alerts
Peer review
For authors
For reviewers

User ID
Password
New user? | Lost login?

Follow @EGU_GMD

Journal metrics

Geosci. Model Dev., 4, 373-380, 2011
www.geosci-model-dev.net/4/373/2011/
doi:10.5194/gmd-4-373-2011

Article | Metrics | Related Articles

Model Description Paper 06 May 2011

The atmospheric chemistry box model CAABA/MECCA-3.0

R. Sander[1], A. Baumgaertner[1,*], S. Gromov[1], H. Harder[1], P. Jöckel[1,], A. Kerkweg[1,***], D. Kubistin[1], E. Regelin[1], H. Riede[1], A. Sandu[2], D. Taraborrelli[1], H. Tost[1,***], and Z.-Q. Xie[3]**

[1]Air Chemistry Department, Max-Planck Institute of Chemistry, P.O. Box 3060, 55020 Mainz, Germany
[2]Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060, USA
[3]Institute of Polar Environment, University of Science and Technology of China, Hefei, Anhui, 230026, China
*now at: Deutsches Zentrum für Luft- und Raumfahrt (DLR), Project Management Agency, 53227 Bonn, Germany
**now at: Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82230 Wessling, Germany
***now at: Institute for Atmospheric Physics, University Mainz, 55128 Mainz, Germany

Received: 20 December 2010 – Published in Geosci. Model Dev. Discuss.: 19 January 2011
Revised: 18 April 2011 – Accepted: 20 April 2011 – Published: 06 May 2011

Abstract. We present version 3.0 of the atmospheric chemistry box model CAABA/MECCA. In addition to a complete update of the rate coefficients to the most recent recommendations, a number of new features have been added: chemistry in multiple aerosol size bins; automatic multiple simulations reaching steady-state conditions; Monte-Carlo simulations with randomly varied rate coefficients within their experimental uncertainties; calculations along Lagrangian trajectories; mercury chemistry; more detailed isoprene chemistry; tagging of isotopically labeled species. Further changes have been implemented to make the code more user-friendly and to facilitate the analysis of the model results. Like earlier versions, CAABA/MECCA-3.0 is a community model published under the GNU General Public License.

Citation: Sander, R., Baumgaertner, A., Gromov, S., Harder, H., Jöckel, P., Kerkweg, A., Kubistin, D., Regelin, E., Riede, H., Sandu, A., Taraborrelli, D., Tost, H., and Xie, Z.-Q.: The atmospheric chemistry box model CAABA/MECCA-3.0, Geosci. Model Dev., 4, 373-380, doi:10.5194/gmd-4-373-2011, 2011.

Search GMD
Search
Full Text

Special Issue
The Modular Earth Submodel System (MESSy) (ACP/GMD Inter-...

Final Revised Paper
- Supplement (5181 KB)

PDF XML

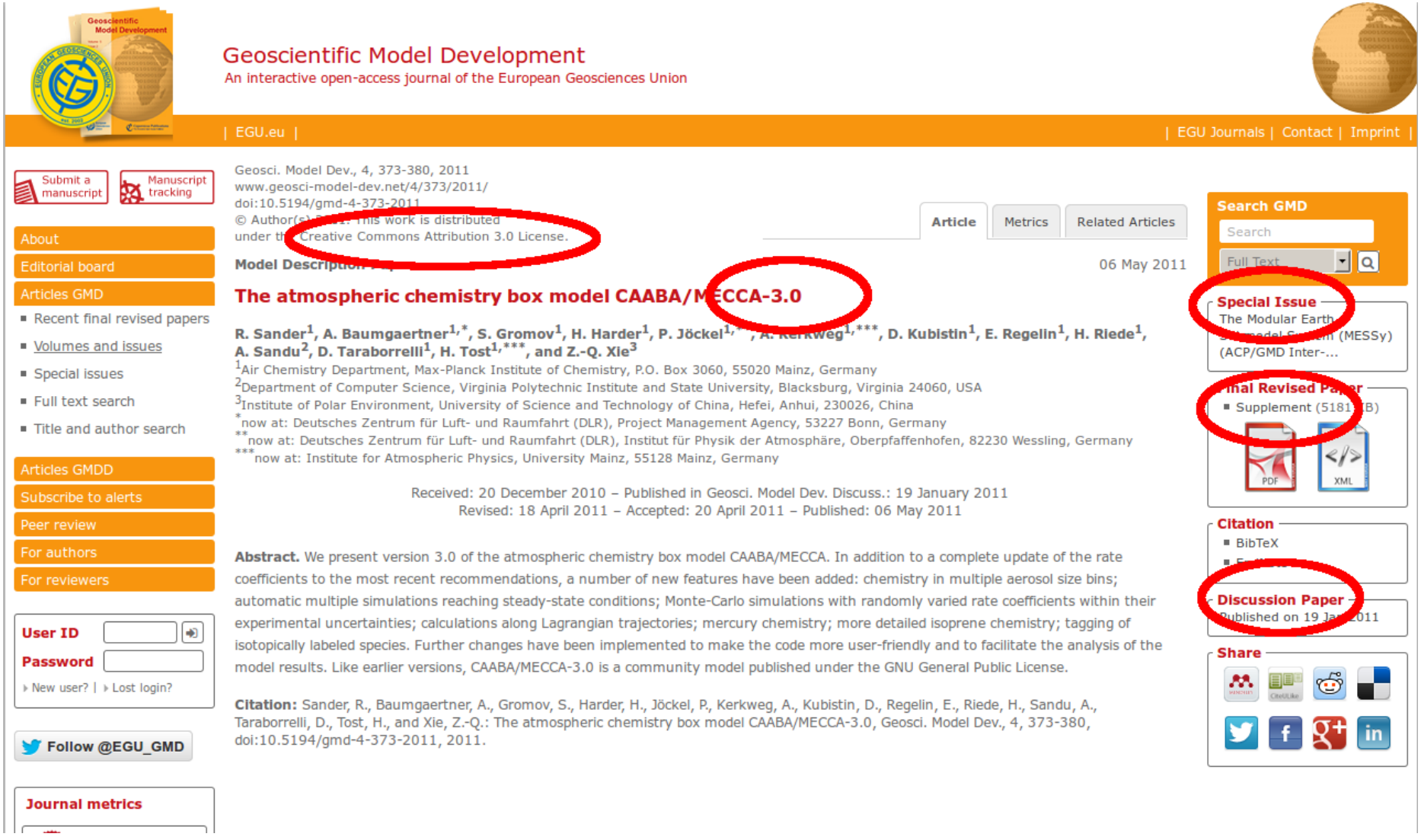
Citation
- BibTeX
- EndNote

Discussion Paper
Published on 19 Jan 2011

Share

What is published in GMD?

Model Description paper example:



Geoscientific Model Development
An interactive open-access journal of the European Geosciences Union

| EGU.eu | | EGU Journals | Contact | Imprint |

Submit a manuscript | Manuscript tracking

About
Editorial board
Articles GMD
- Recent final revised papers
- Volumes and issues
- Special issues
- Full text search
- Title and author search

Articles GMDD
Subscribe to alerts
Peer review
For authors
For reviewers

User ID
Password
New user? | Lost login?

Follow @EGU_GMD

Journal metrics

Geosci. Model Dev., 4, 373-380, 2011
www.geosci-model-dev.net/4/373/2011/
doi:10.5194/gmd-4-373-2011
© Author(s) 2011. This work is distributed
under the Creative Commons Attribution 3.0 License.

Model Description

Article | Metrics | Related Articles

06 May 2011

The atmospheric chemistry box model CAABA/MECCA-3.0

R. Sander[1], A. Baumgaertner[1,*], S. Gromov[1], H. Harder[1], P. Jöckel[1,], A. Kerkweg[1,***], D. Kubistin[1], E. Regelin[1], H. Riede[1], A. Sandu[2], D. Taraborrelli[1], H. Tost[1,***], and Z.-Q. Xie[3]**

[1]Air Chemistry Department, Max-Planck Institute of Chemistry, P.O. Box 3060, 55020 Mainz, Germany
[2]Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060, USA
[3]Institute of Polar Environment, University of Science and Technology of China, Hefei, Anhui, 230026, China
*now at: Deutsches Zentrum für Luft- und Raumfahrt (DLR), Project Management Agency, 53227 Bonn, Germany
**now at: Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82230 Wessling, Germany
***now at: Institute for Atmospheric Physics, University Mainz, 55128 Mainz, Germany

Received: 20 December 2010 – Published in Geosci. Model Dev. Discuss.: 19 January 2011
Revised: 18 April 2011 – Accepted: 20 April 2011 – Published: 06 May 2011

Abstract. We present version 3.0 of the atmospheric chemistry box model CAABA/MECCA. In addition to a complete update of the rate coefficients to the most recent recommendations, a number of new features have been added: chemistry in multiple aerosol size bins; automatic multiple simulations reaching steady-state conditions; Monte-Carlo simulations with randomly varied rate coefficients within their experimental uncertainties; calculations along Lagrangian trajectories; mercury chemistry; more detailed isoprene chemistry; tagging of isotopically labeled species. Further changes have been implemented to make the code more user-friendly and to facilitate the analysis of the model results. Like earlier versions, CAABA/MECCA-3.0 is a community model published under the GNU General Public License.

Citation: Sander, R., Baumgaertner, A., Gromov, S., Harder, H., Jöckel, P., Kerkweg, A., Kubistin, D., Regelin, E., Riede, H., Sandu, A., Taraborrelli, D., Tost, H., and Xie, Z.-Q.: The atmospheric chemistry box model CAABA/MECCA-3.0, Geosci. Model Dev., 4, 373-380, doi:10.5194/gmd-4-373-2011, 2011.

Search GMD
Search
Full Text

Special Issue
The Modular Earth ... (MESSy) (ACP/GMD Inter-...

Final Revised Paper
- Supplement (5181 KB)

PDF | XML

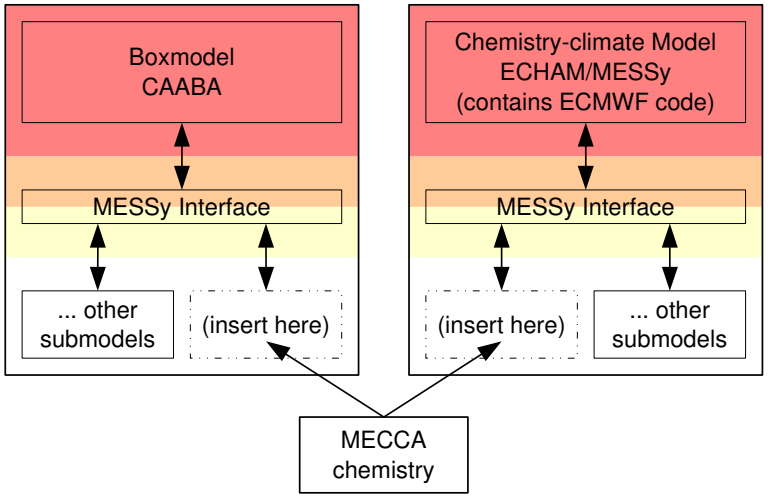
Citation
- BibTeX

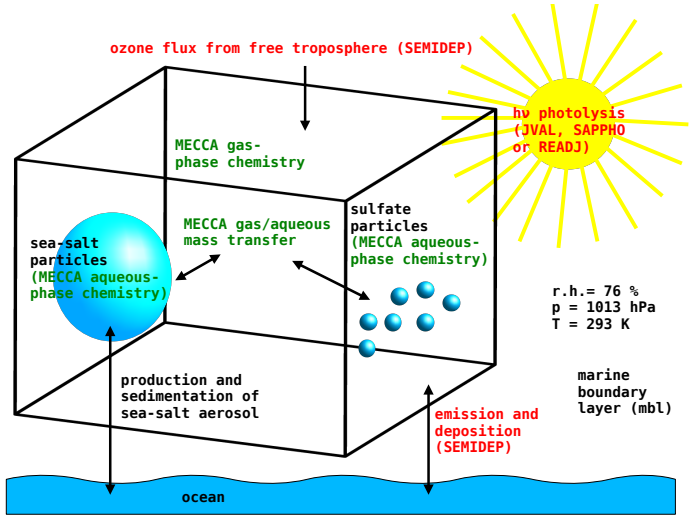
Discussion Paper
Published on 19 Jan 2011

Share

What is published in GMD?

Model Description paper example:

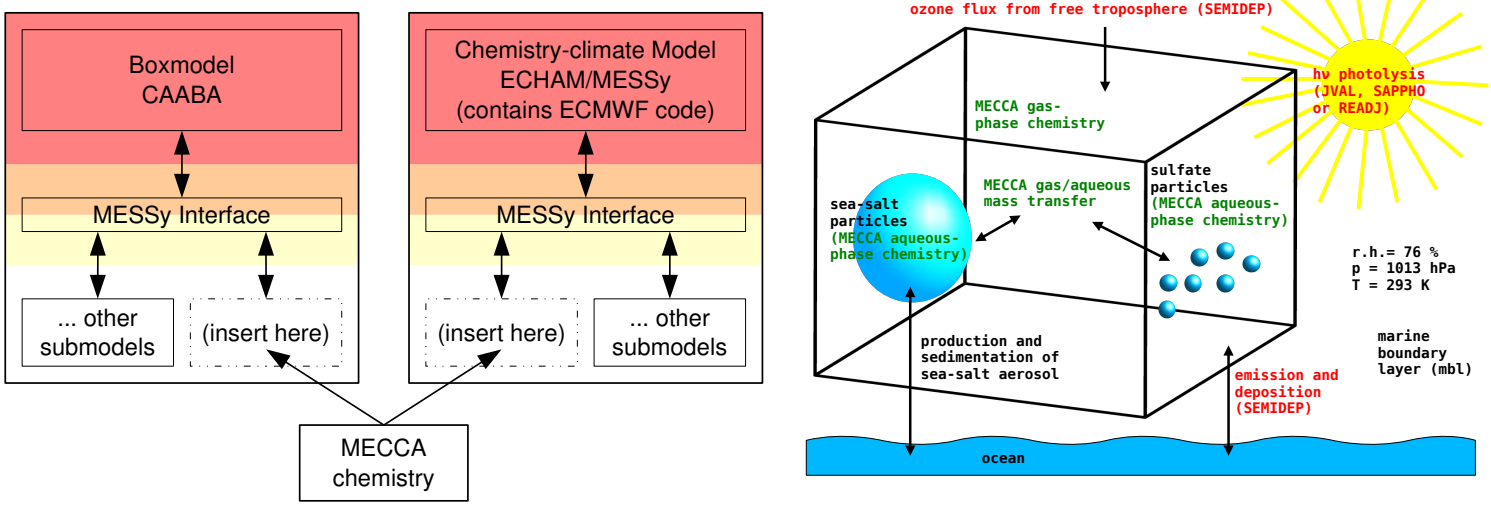




- MECCA chemistry module (open source)

What is published in GMD?

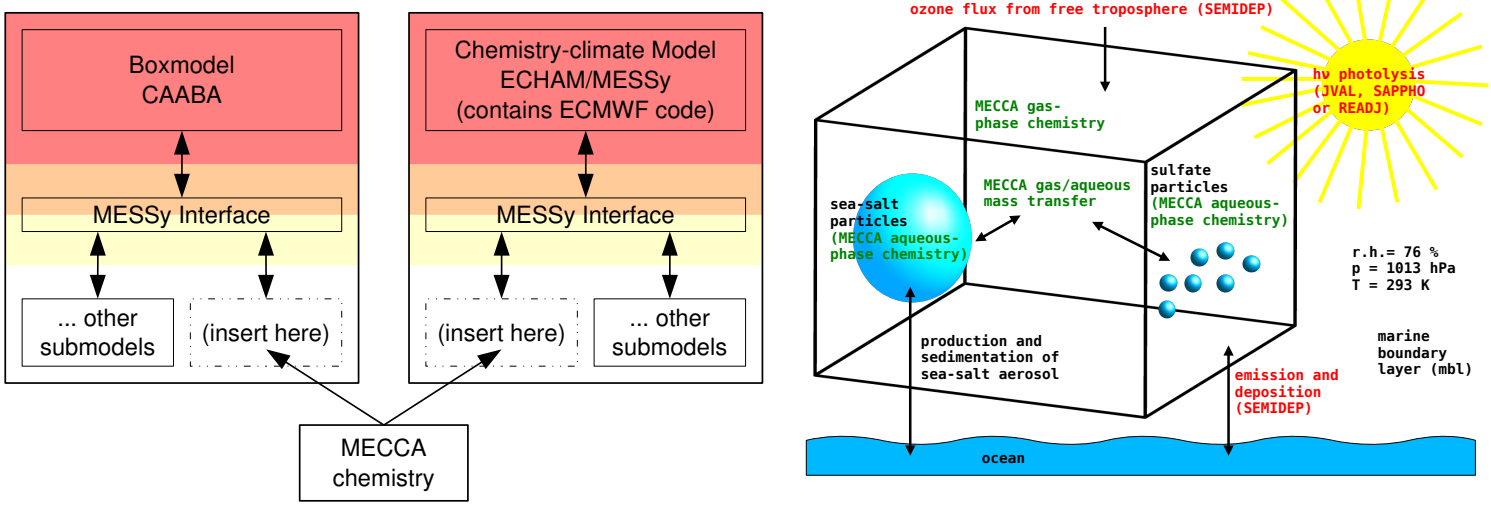
Model Description paper example:



- MECCA chemistry module (open source)
- chemistry-climate model ECHAM (contains ECMWF code, Software License Agreement necessary)

What is published in GMD?

Model Description paper example:



- MECCA chemistry module (open source)
- chemistry-climate model ECHAM (contains ECMWF code, Software License Agreement necessary)
- CAABA box model (open source)

What is published in GMD?

Technical, Development and Evaluation papers:

- Technical developments, e.g., related to speed and accuracy of numerical integration schemes
- New parameterizations for (subgrid) processes
- In-depth evaluations of already published models
- Assessments of performance with different compilers or under different computer architectures

What is published in GMD?

Technical, Development and Evaluation papers:

- Technical developments, e.g., related to speed and accuracy of numerical integration schemes
- New parameterizations for (subgrid) processes
- In-depth evaluations of already published models
- Assessments of performance with different compilers or under different computer architectures

Model Assessment Methods papers:

- Benchmarks for model performance
- Novel ways of comparing model results with observations
- Novel methods for data analysis or visualization

What is published in GMD?

Technical, Development and Evaluation papers:

- Technical developments, e.g., related to speed and accuracy of numerical integration schemes
- New parameterizations for (subgrid) processes
- In-depth evaluations of already published models
- Assessments of performance with different compilers or under different computer architectures

Model Assessment Methods papers:

- Benchmarks for model performance
- Novel ways of comparing model results with observations
- Novel methods for data analysis or visualization

Model Experiment Description papers:

- Descriptions of model intercomparison projects (MIP)
- Configurations, overview results, project protocols

What is published in GMD?

Technical, Development and Evaluation papers:

- Technical developments, e.g., related to speed and accuracy of numerical integration schemes
- New parameterizations for (subgrid) processes
- In-depth evaluations of already published models
- Assessments of performance with different compilers or under different computer architectures

Model Assessment Methods papers:

- Benchmarks for model performance
- Novel ways of comparing model results with observations
- Novel methods for data analysis or visualization

Model Experiment Description papers:

- Descriptions of model intercomparison projects (MIP)
- Configurations, overview results, project protocols

Corrigenda

THE END

www.geoscientific-model-development.net

The IGSN Implementation Organisation IGSN e.V.

Jens Klump

EGU General Assembly, SC25, 16 April 2015

MINERAL RESOURCES FLAGSHIP
www.csiro.au

CSIRO

IGSN e.V.

www.igsn.org

IGSN e.V.

- Non-profit organization to implement & govern the IGSN
- Registered in Germany (Potsdam)
- Managing Agent at LDEO
- Currently 14 members

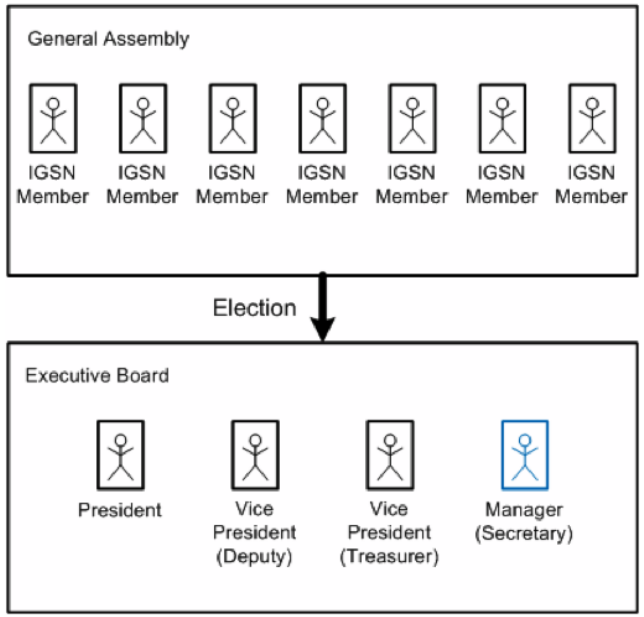
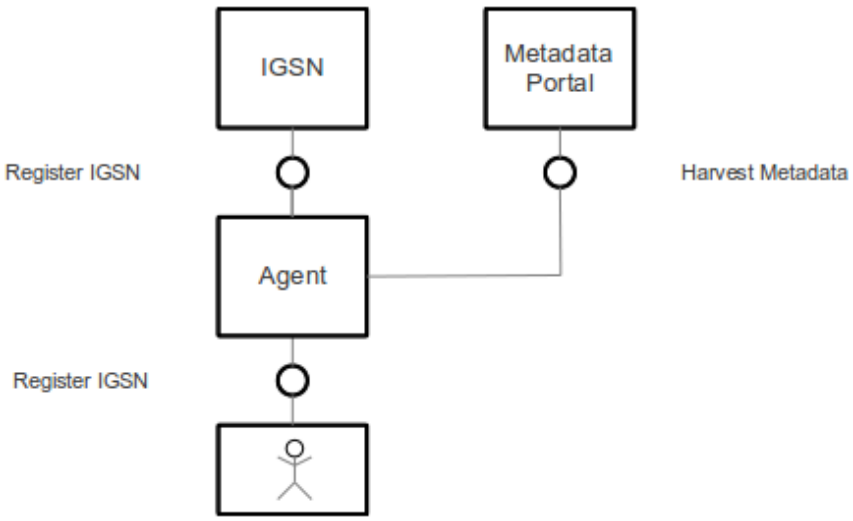


Figure: Constituent bodies of the Association.

IGSN Objectives

- provide identifiers that are guaranteed to be unique via a centralised control mechanism
- ensure preservation of sample metadata
- facilitate internet-based discovery and access to physical samples
  - web application and programmatic access to sample metadata catalogues
  - network with repositories
- Work with publishers to embed IGSN in publications.
  - Example: **IGSN: SSH000SUA** In: Dere, A. L., T. S. White, R. H. April, B. Reynolds, T. E. Miller, E. P. Knapp, L. D. McKay, and S. L. Brantley (2013), Climate dependence of feldspar weathering in shale soils along a latitudinal gradient, Geochimica et Cosmochimica Acta, 122, 101–126, doi:10.1016/j.gca.2013.08.001.

How IGSN works (tech)



- The sample owner/curator registers the IGSN and the URL of the landing page with an Allocating Agent.
- The Allocating Agent registers the IGSN/URL pair and administrative metadata with the IGSN Registry.
- The Allocating Agent provides a discipline specific metadata catalogue for harvesting by OAI-PMH.

How to resolve an IGSN

Example: SSH000SUA

HTTP URI

- http://hdl.handle.net/10273/SSH000SUA
- http://dx.doi.org/10273/SSH000SUA

IGSN metadata can also record parent-child relationships between samples:

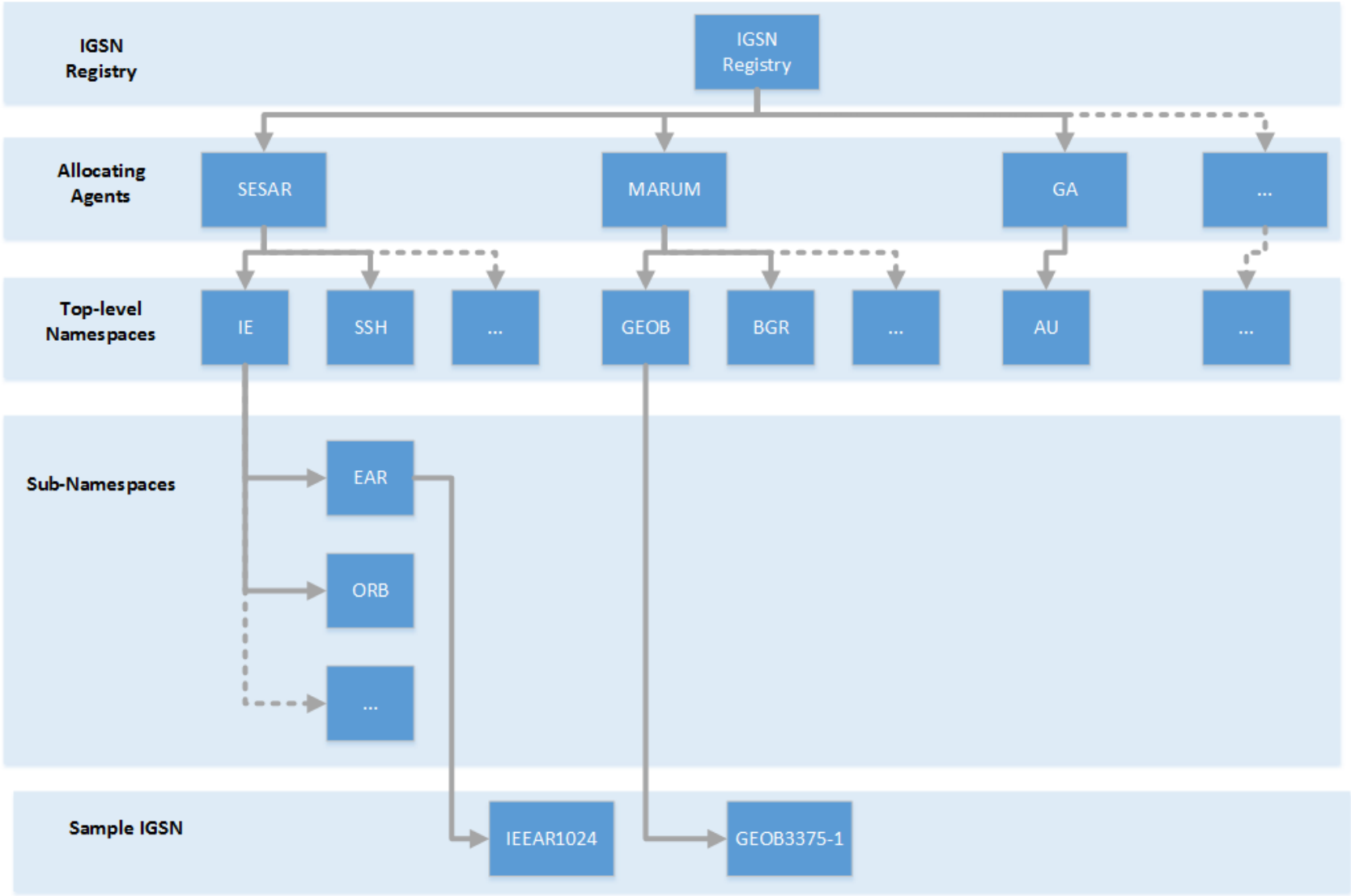
- drill hole (parent) and drill core sections (children)
- subsamples (children) of a larger sample (parent)

IGSN: SSH000STR

IGSN: SSH000STR
Sample Name: ald-10-01
Other Name(s):
Sample Type: Individual Sample
Parent IGSN: Not Provided

Description	
Material:	Rock
Classification:	Not Provided
Field Name:	shale
Description:	rock outcrop sample
Age (min):	Not Provided
Age (max):	Not Provided
Collection Method:	rock hammer
Collection Method Description:	Not Provided
Size:	Not Provided
Geological Age:	Not Provided
Geological Unit:	Not Provided
Comment:	Not Provided
Purpose:	CZO Shale Transect
Geolocation	
Latitude:	52.470683
Longitude:	-3.69255
Elevation:	323.088

IGSN Hierarchical Delegation



How to Participate

- As an individual
  - register your samples through an Allocating Agent
  - use the IGSN
    - sample labels, data tables on your computer and in publications
- As an organisation
  - become a member of the IGSN e.V.
    - as a regular member at act as an Allocating Agent
    - as an affiliate member to promote and support the IGSN

www.igsn.org

Thank you

Mineral Resources Flagship
Jens Klump

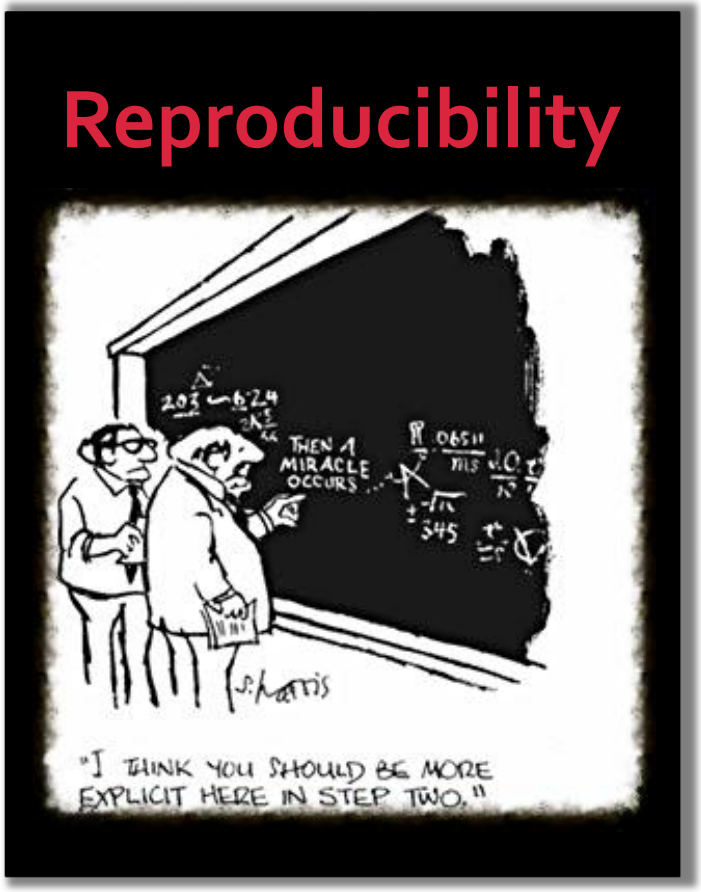
t +61 8 6436 8828

e jens.klump@csiro.au
w www.csiro.au
w www.igsn.org

The IGSN in Action:
Building the Internet of Samples

Kerstin Lehnert, Interdisciplinary Earth Data Alliance

Internet of Samples: Why?



Internet of Samples: How?

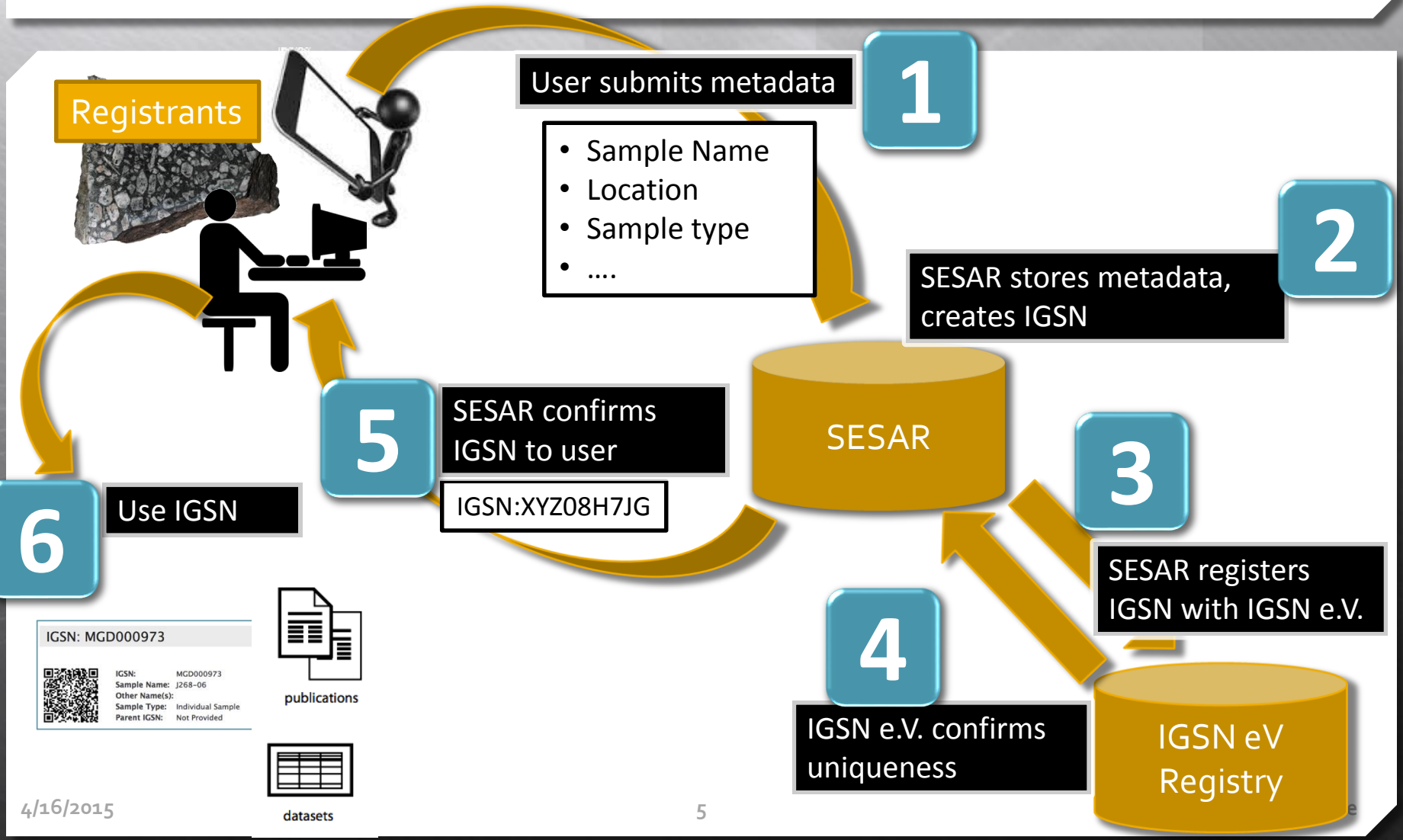
- Access to digital representations of physical samples
- Use of resolvable unique & persistent identifiers

- **IGSN**: Architecture & governance for sample PIDs
- **SESAR** (System for Earth Sample Registration): Software tools for sample registration, sample metadata management, etc.
- **DESC**: Shared cyberinfrastructure for sample and collection management
- **iSamples**, CODATA TG, Belmont Forum, etc.: Communities of Practice

SESAR (www.geosamples.org) System for Earth Sample Registration

- Allocating Agent for individual investigators, sample repositories, and science programs
  - tools and services for users to catalog and manage sample metadata (MySESAR)
    - personal (authenticated) workspace
    - metadata template creator
    - label creation & printing (including QR code)
    - transfer of sample ownership
  - web services for client systems
    - register sample metadata & obtain IGSNs
    - access to IGSN metadata
  - preservation & persistent access of sample metadata
- Global Sample Catalog (harvest metadata from other AAs

IGSN Registration Workflow



app.geosamples.org/index.php

MySESAR

Back to SESAR Home | My Home | My Samples | My Groups | Sample Reg | Logout

My Home

Welcome, Megan Carter

REGISTRATION

> Register an individual sample
> Download batch registration template
> Upload my batch samples

SAMPLES

> Search sample catalog
> View/Edit my samples
> View/Edit my groups

Pending Batch Registratio

There is a total of 1 batch registrati

- batch_MCT_315_1427904897.xls su

Registered Samples Summ

You have a total of 795 registered s

- 1 Hole
- 9 Core
- 1 CoreSection
- 1 CoreHalfRound
- 1 CoreQuarterRound
- 780 IndividualSample
- 2 Site

Related Samples

Parents: ODP000230 Hole 3-22*

Siblings:
- ODP000231 Core 3-22*-1R
- ODP000232 Core 3-22*-2R
- ODP000233 Core 3-22*-3R
- ODP000235 Core 3-22*-5R

Children:
- ODP0115YA Section 3-22*-4R-1
- ODP0115YE Section 3-22*-4R-2
- ODP0115YI Section 3-22*-4R-3
- ODP0115YM Section 3-22*-4R-4
- ODP0115YQ Section 3-22*-4R-5
- ODP0115YU Section 3-22*-4R-6
- ODP0115YY Section 3-22*-4R-7

IGSN:	GRO000186
Sample Name:	12PCT04
Other Name(s):	Pawcatuck at water wells
Sample Type:	Individual Sample
Parent IGSN:	Not Provided

Description

Material:	Liquid>aqueous
Classification:	Not Provided
Field Name:	Not Provided
Description:	Not Provided
Age (min):	Not Provided
Age (max):	Not Provided
Collection Method:	Not Provided
Collection Method Description:	Not Provided
Size:	Not Provided
Geological Age:	Not Provided
Geological Unit:	Not Provided
Comment:	Not Provided
Purpose:	Not Provided

Geolocation

Latitude:	41.39177
Longitude:	-71.83908

wells

02543-1541, USA

Current Archive:	WHOI
Current Archive Contact Details:	Bernhard Peucker-Ehrenbrink
Original Archive:	Not Provided
Original Archive Contact Details:	Not Provided

Relation To Parent

GUIs

Institutions
Collection Mgmt

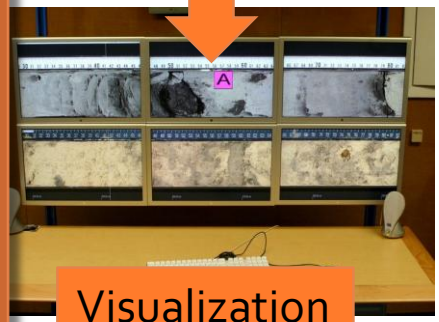
Investigators
Sample Mgmt

Public
'Virtual Museum'

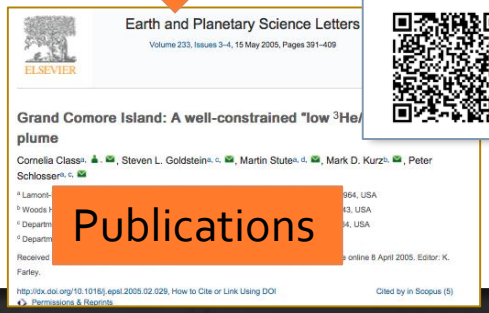
DESC: Digital Environment for Sample Curation

(storage, software solutions, & services)

APIs



Visualization



Publications

Sample Registries

EarthChem

mindat.org

USGIN
United States Geoscience Information Network

Data Systems

6/2015

iSamples

EARTHCUBE

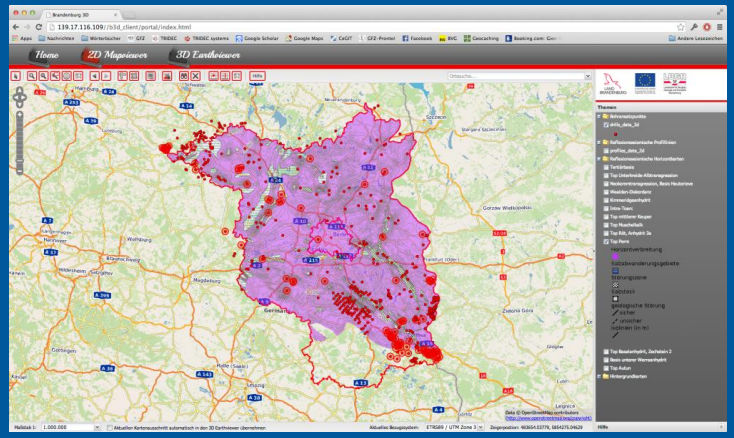
- Research Coordination Network to advance access and re-use of physical samples through use of innovative cyberinfrastructure
  - advance best practices, standards, & policies for sample curation, distribution, attribution, and citation
  - plan a "Digital Environment for Sample Curation" (DESC)
- Cross-disciplinary coordination (BIO, archeology, etc.)
- International coordination

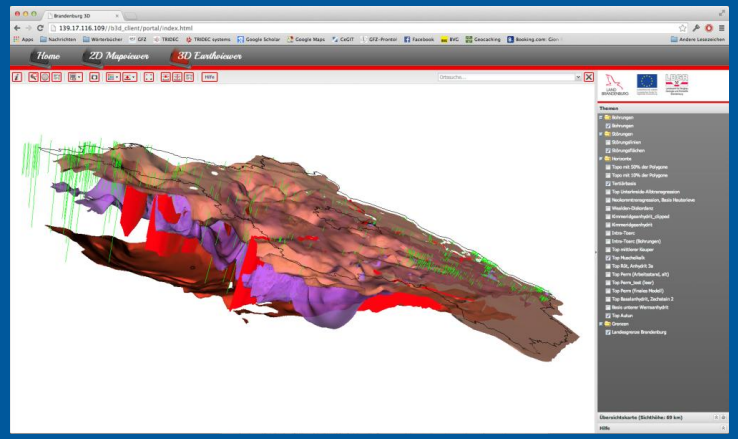
International Initiatives

- **CODATA Task Group** "Physical Samples in the Digital Era"
  - http://www.codata.org/task-groups/management-of-physical-objects
- **SciColl:** Scientific Collections International (Consortium)
  - http://www.scicoll.org/
- ("Samples of Planet Earth": proposal as Belmont Forum Action of WP4 under development)

Brandenburg 3D
a Contribution to Open Science

Dorit Kerschke & Joachim Wächter





Potential of Spatial Data Infrastructures for Open Science

- Scientific research infrastructures (RIs) have the urging task to integrate and federate scientific data; along with sensor platforms, satellites and other instruments or repositories; from highly heterogeneous sources.
- SDIs (INSPIRE) have already been implemented with major efforts and high costs, over a long-term period.
- The interoperability for disciplinary and domain applications highly depends on the adoption of generally agreed technologies and standards (OGC, ISO…) originating from SDI-related efforts (e.g., INSPIRE).
- SDI integration concepts based on standardized services platforms will leverage the construction of efficient RIs.
- SDI thus can facilitate the sustainable and cost-efficient utilization, exchange, and re-use of data and software.

B3D as an infrastructure node

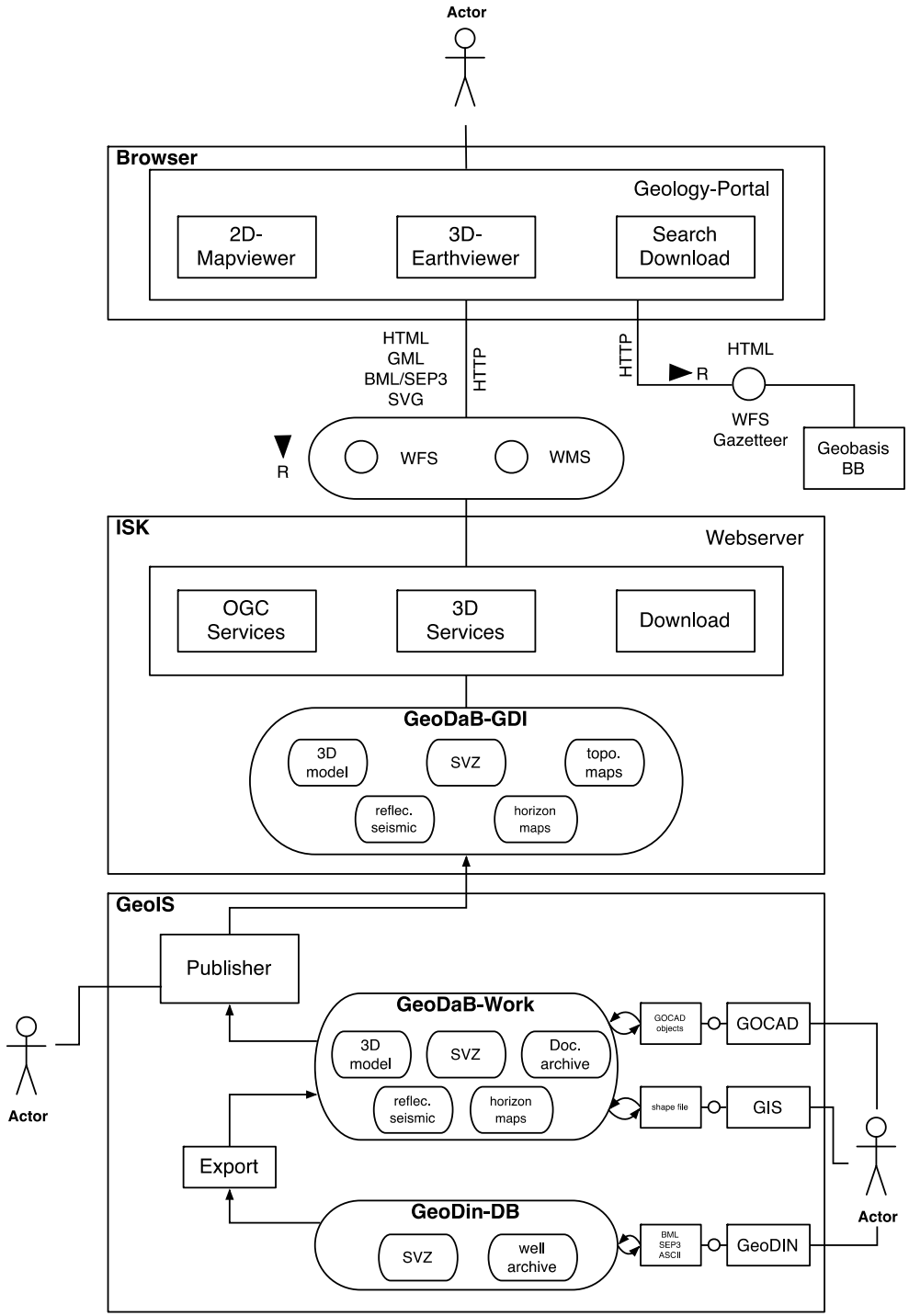
- ‘Brandenburg 3D’ (B3D) – EFRE project.
- Digitalization of available subsurface data.
- Development of a comprehensive 3D subsurface model for Brandenburg.
- Central storage of all data, regardless of format.
- OGC-conformable integration into the ‘Spatial Data Infrastructure Brandenburg’ (SDI-BB) ➤ Infrastructure node.
- Implementation of a plugin free web application ‘B3D’ for interactive real-time visualization and manipulation of data.

Sencha Ext JS

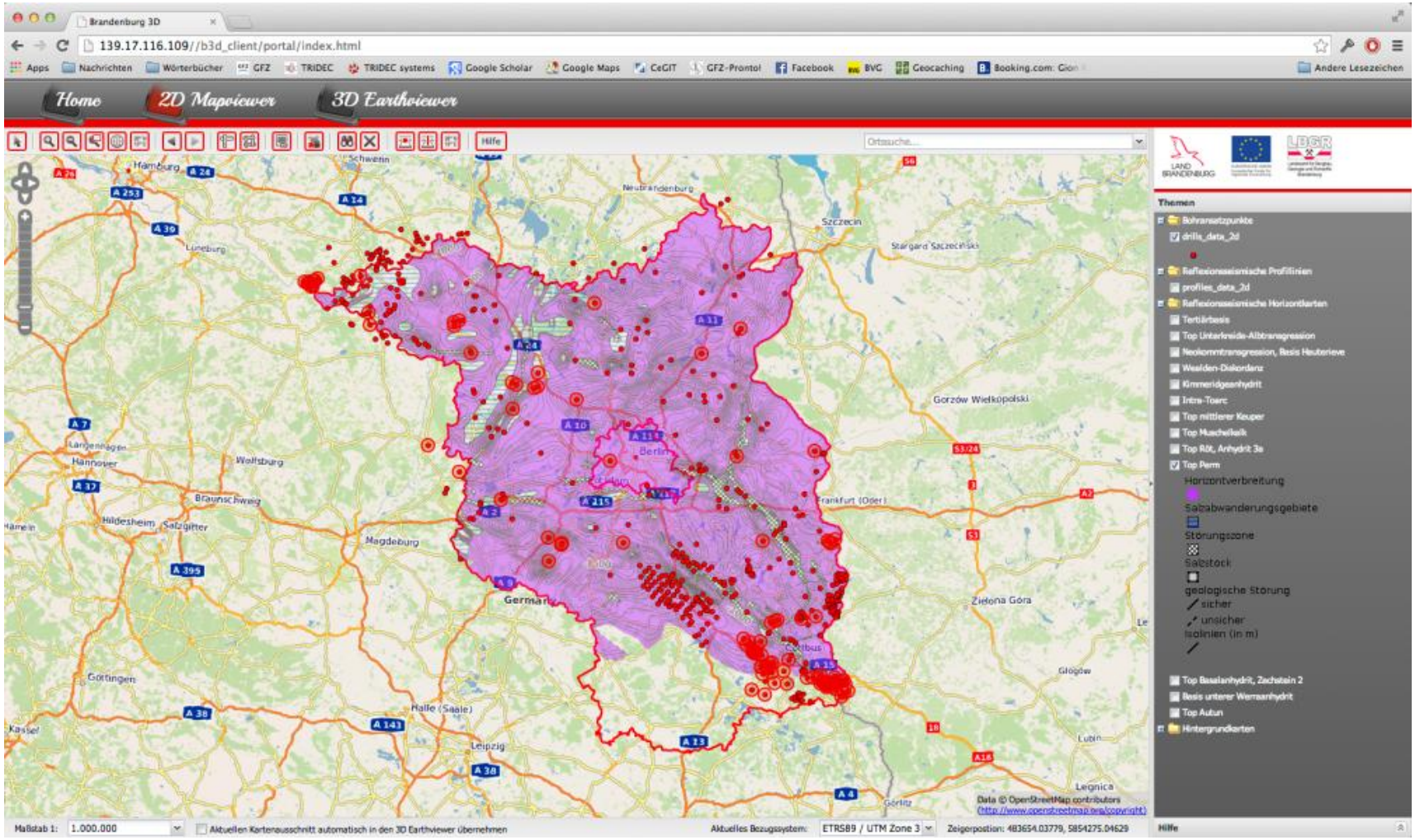
GeoExt

x3dom

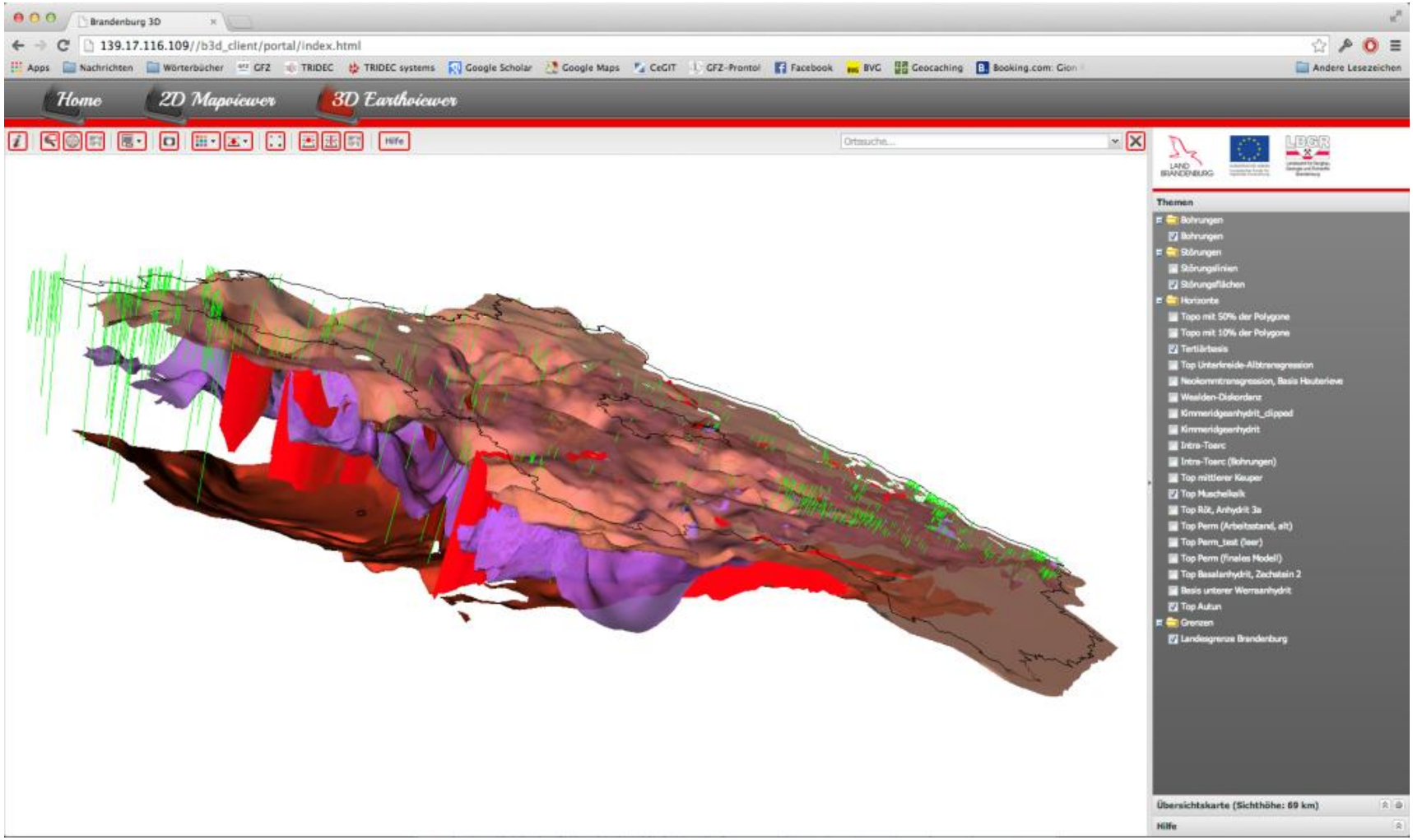
B3D architecture



B3D User Interface



B3D User Interface



Upscaling the concept

- It is now applied for the implementation of an interoperable data management platform for the Helmholtz Observatory in Chile (IPOC – Integrated Plate boundary Observatory Chile).
- New approach for project data management, interfaces for tools used by the researchers, and a web-based user interfaces for the sharing and reusing of data.
- The use of standards also enables the integration into research infrastructures.
- The B3D design serves as an architectural blueprint and framework for geological /drilling data management
- Framework software is Open Source can be transferred to other geoscientific application fields
- … and helps to collect and provide valuable scientific data
- **One step towards Open Science for scientific geological data**

Google Earth Engine

Chris Herwig, Google

The Fourth Paradigm

“Often it turns out to be more efficient to move the questions than to move the data.”

- Jim Gray

What is Google Earth Engine?

A cloud-based geospatial processing platform

Goals

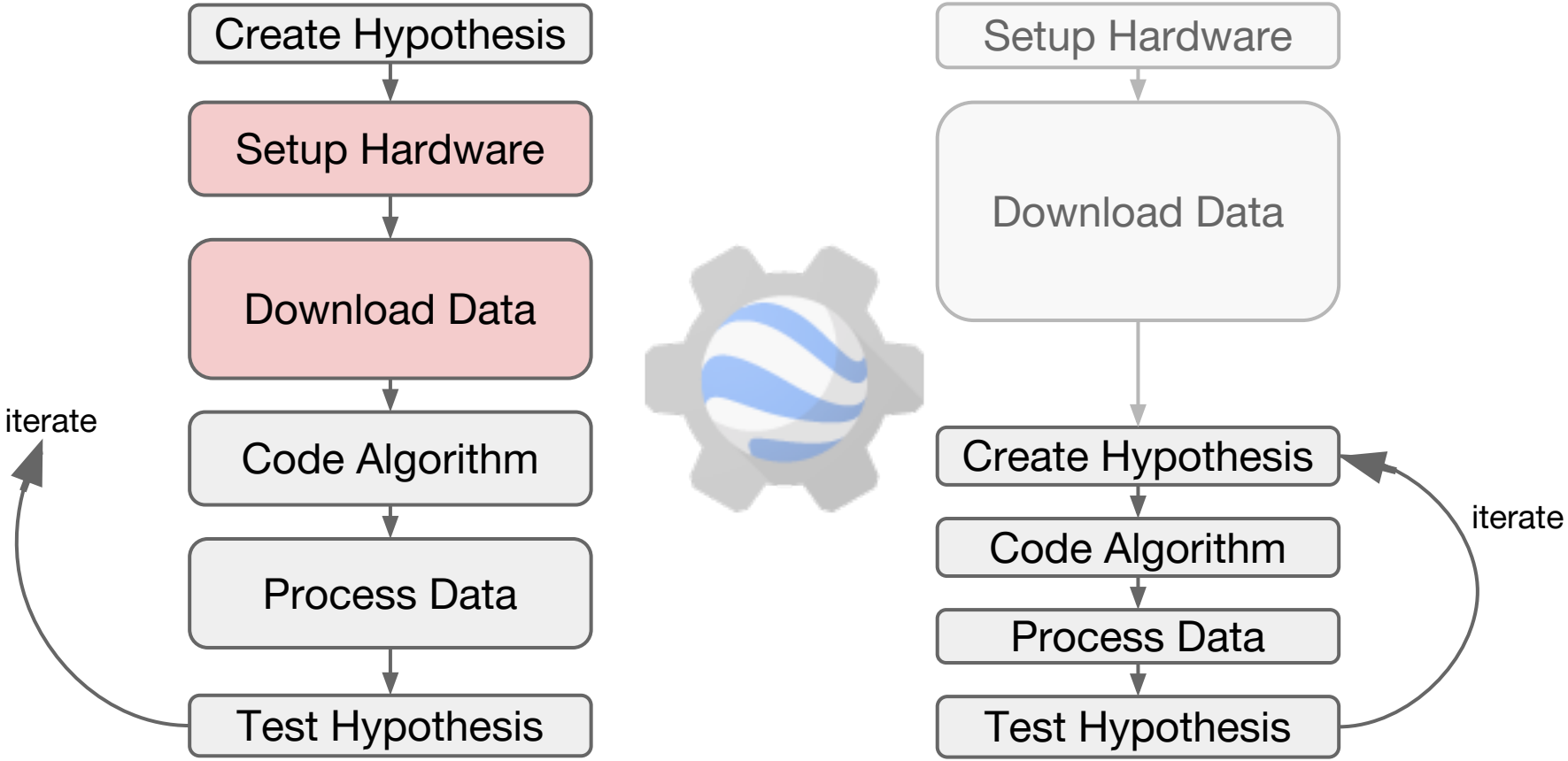
- Make substantive progress on global challenges that involve large geospatial datasets. Do the things no one else can do.
- Push the edge of the envelope for big data in remote sensing.
- Enable high-impact, data-driven science.

A cloud-based geospatial processing platform

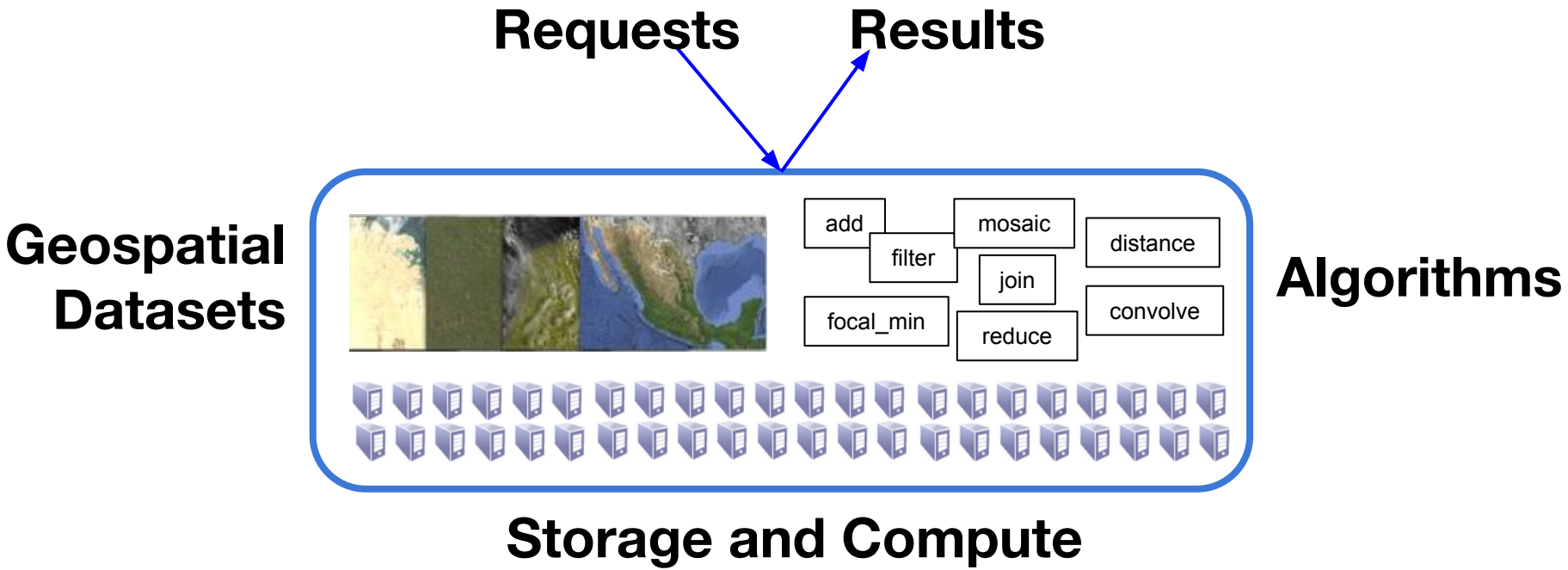
Approach

- Build a geospatial analysis platform that allows both highly-interactive algorithm development and global-scale analysis.

Changing Large Data Analysis Workflows


Create Hypothesis
Setup Hardware
Download Data
Code Algorithm
Process Data
Test Hypothesis
iterate
Setup Hardware
Download Data
Create Hypothesis
Code Algorithm
Process Data
Test Hypothesis
iterate

Web-based Geospatial API


Requests
Results
Geospatial Datasets
add
filter
mosaic
distance
join
focal_min
reduce
convolve
Algorithms
Storage and Compute

Who uses Google Earth Engine?


JRC
EUROPEAN COMMISSION

THE UNIVERSITY OF CALIFORNIA BERKELEY
1868

Imazon

THE UNIVERSITY OF IDAHO
1889


CONAFOR
COMISIÓN NACIONAL FORESTAL

NASA

Yale University

University of
Nebraska
Lincoln

DRI
Desert Research Institute

WORLD
RESOURCES
INSTITUTE

Stanford
University

BOSTON
UNIVERSITY

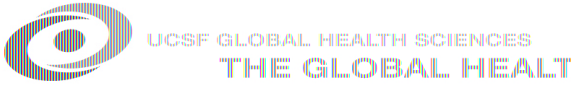
UNIVERSITY OF
MARYLAND

USGS
science for a changing world

CSIRO

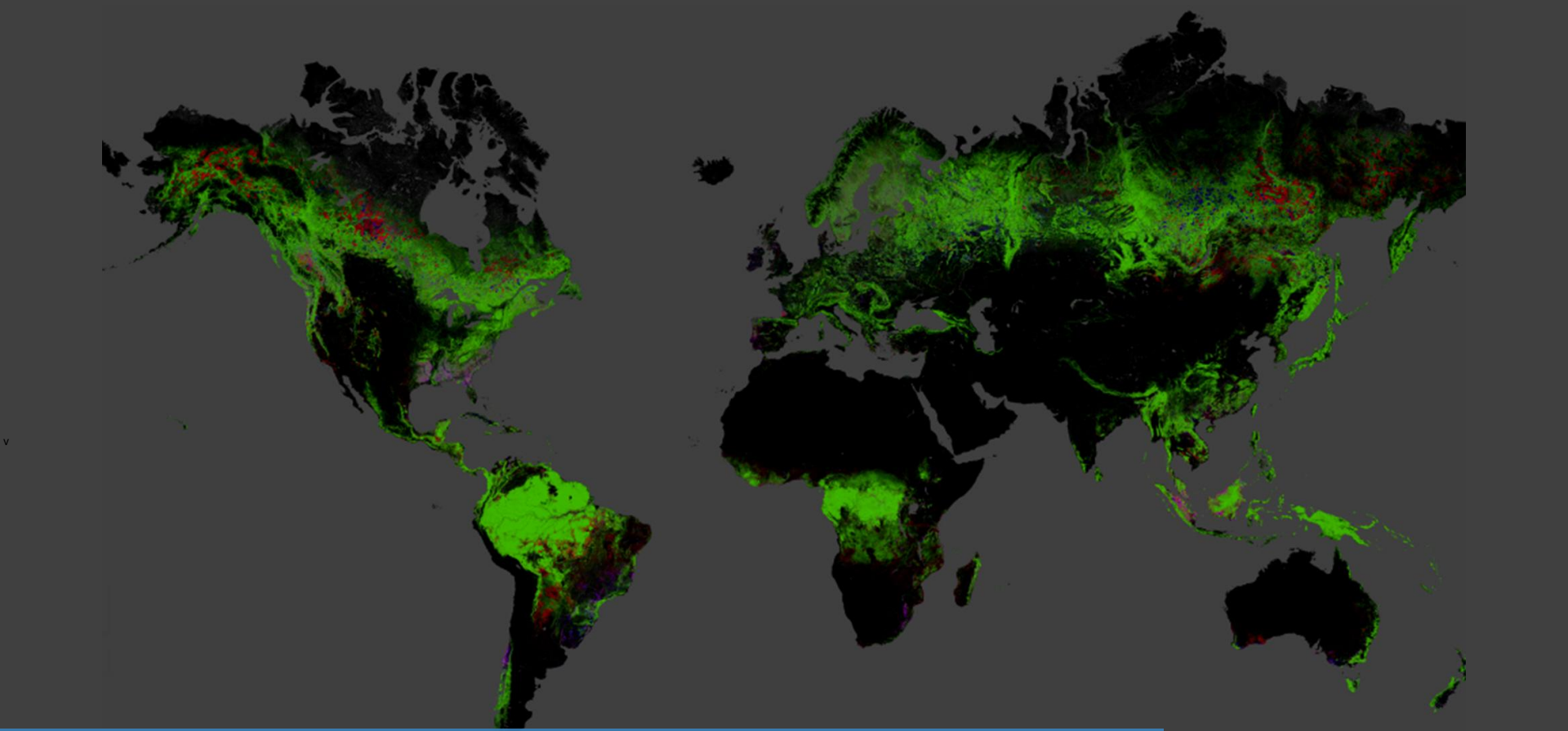
crc•si

CARNEGIE
INSTITUTION FOR
SCIENCE

University of
BRISTOL

UCSF GLOBAL HEALTH SCIENCES
THE GLOBAL HEALT
From evi

the Jane Goodall Institute

WAGENINGEN UR
For quality of life



Global Forest Extent and Change 2000–2012

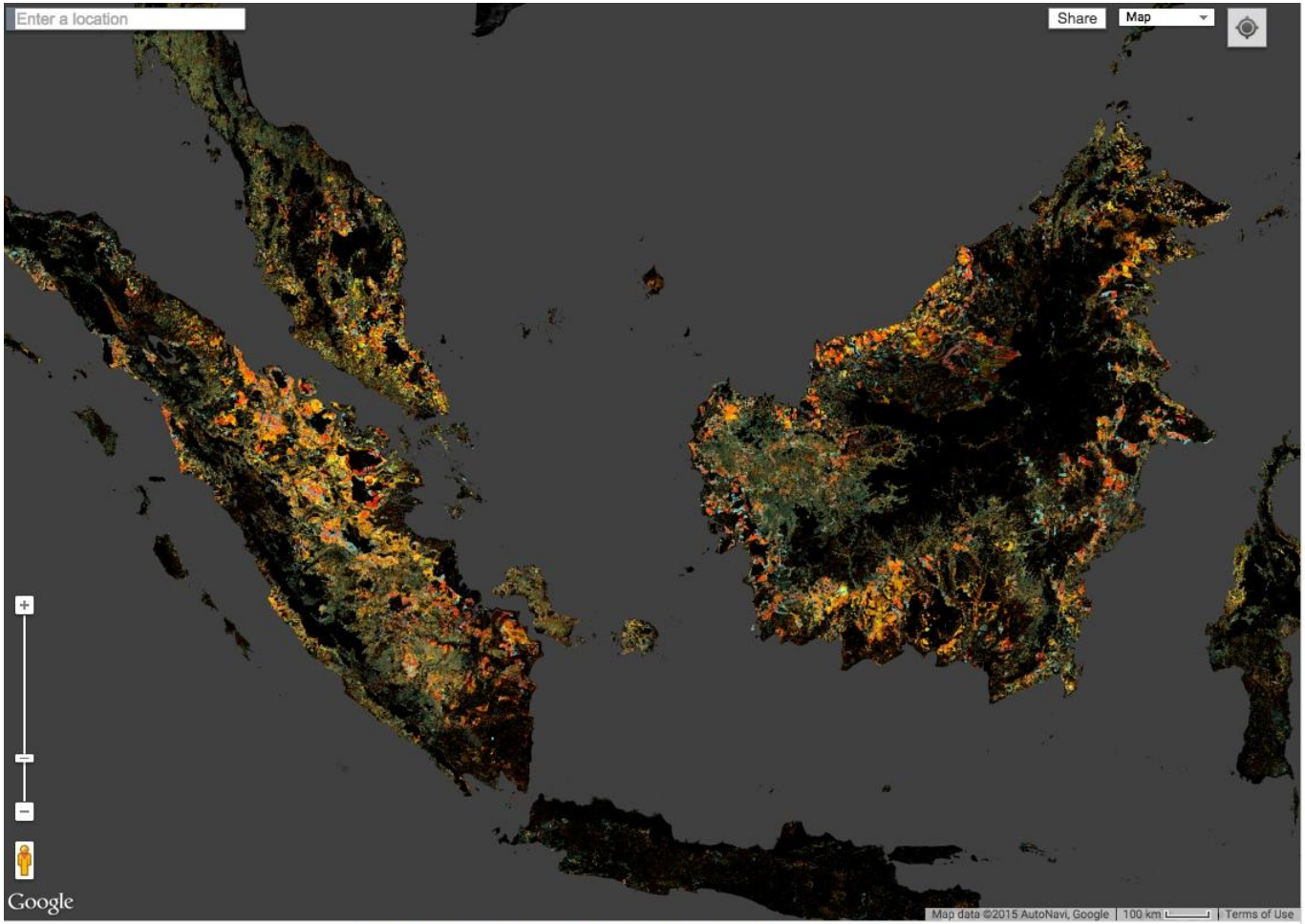
Hansen, Potapov, Moore, Hancher et al. , Science, 15 November 2013

http://earthenginepartners.appspot.com/science-2013-global-forest

654,178

L7 Scenes

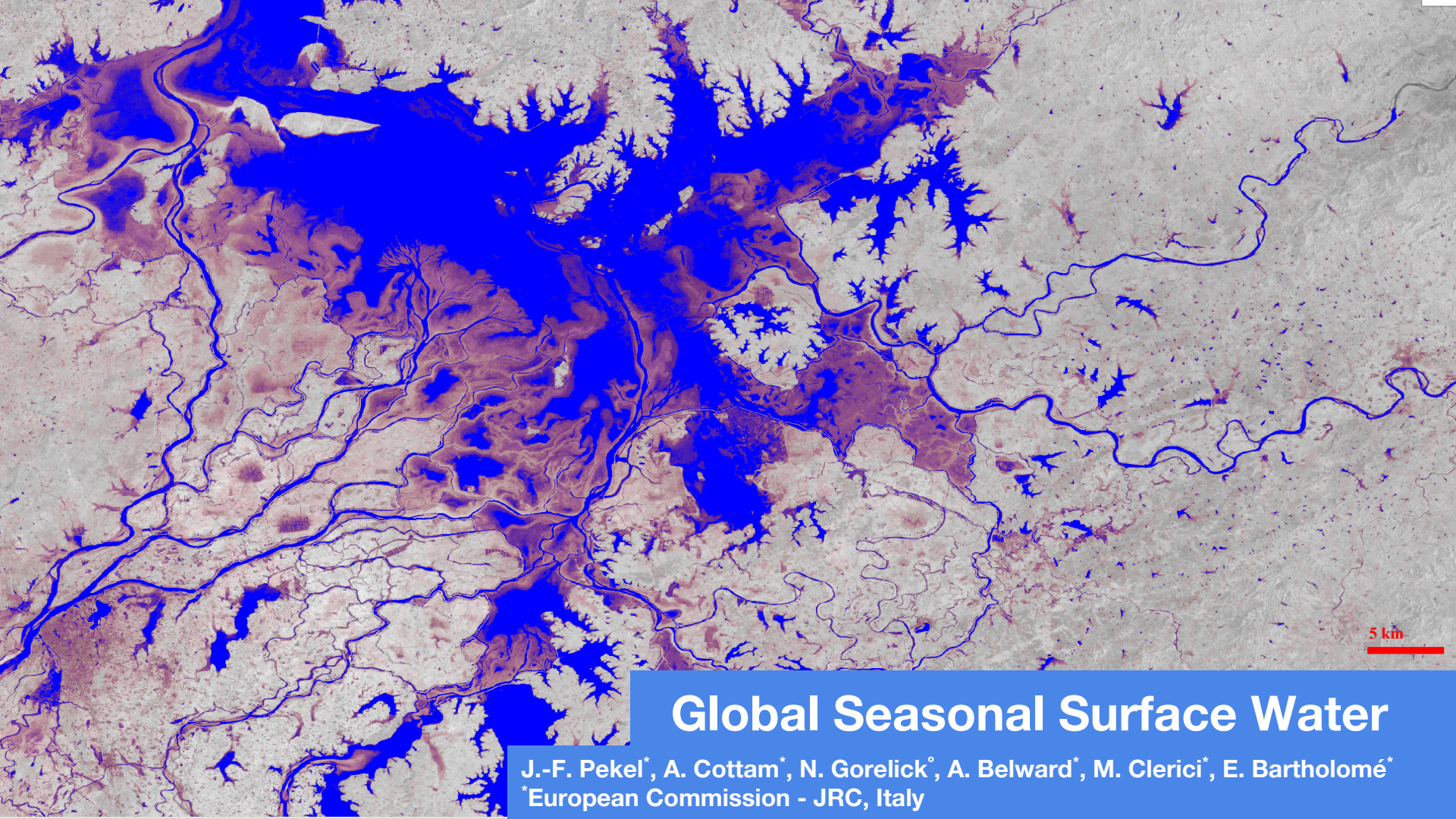
370 x 185 feet



http://earthenginepartners.appspot.com/science-2013-global-forest

333

citations

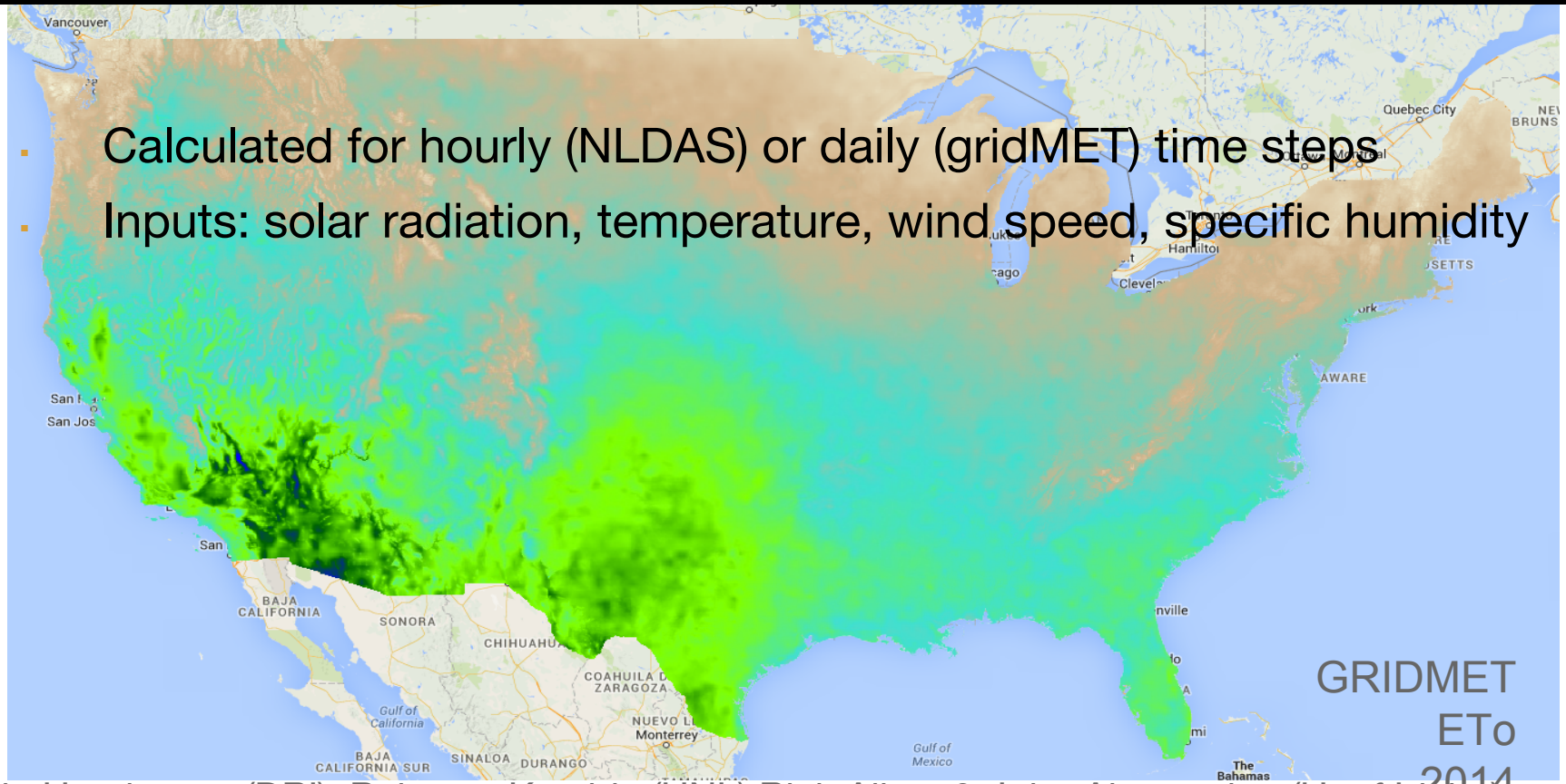


Global Seasonal Surface Water

J.-F. Pekel*, A. Cottam*, N. Gorelick°, A. Belward*, M. Clerici*, E. Bartholomé*
*European Commission - JRC, Italy

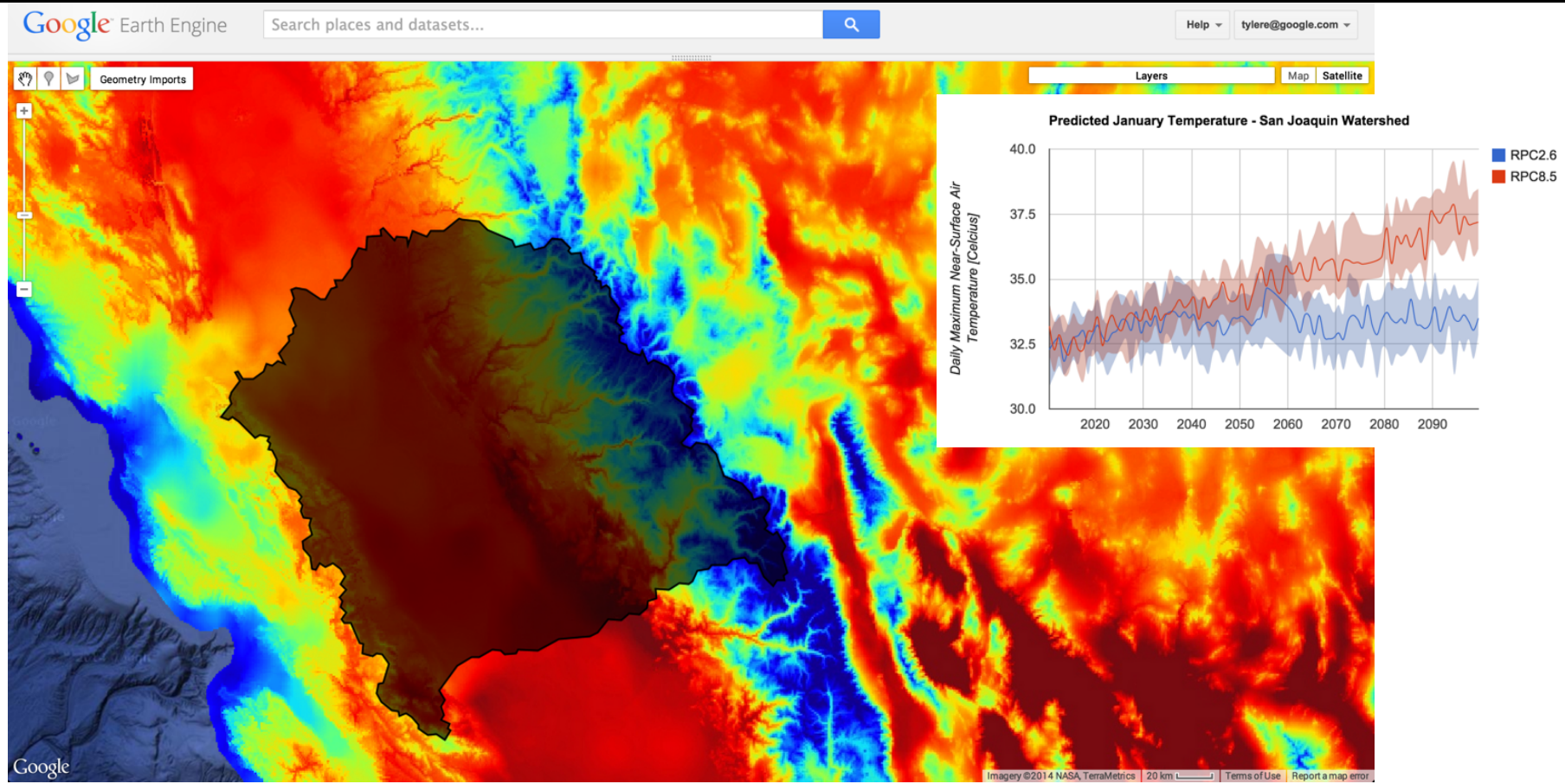
ASCE Standardized Reference ET (ETo/ETr)

- Calculated for hourly (NLDAS) or daily (gridMET) time steps
- Inputs: solar radiation, temperature, wind speed, specific humidity

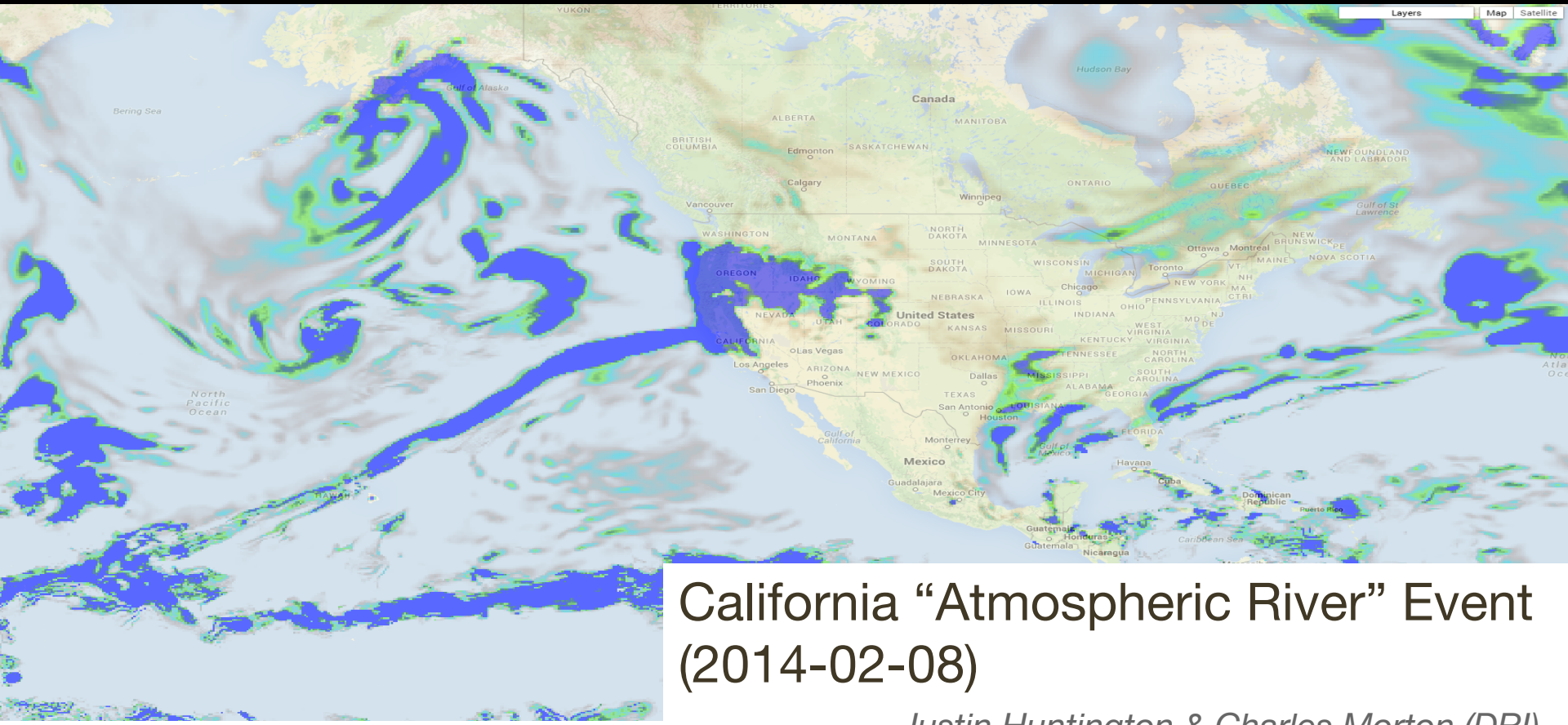


Justin Huntington (DRI), Baburao Kamble (UNL) Rick Allen, & John Abatzoglou (U. of Idaho)

Climate Model Forecast Data (NEX-DCP30)



NCEP Climate Forecast System (CFSR & CFSv2)



California “Atmospheric River” Event (2014-02-08)

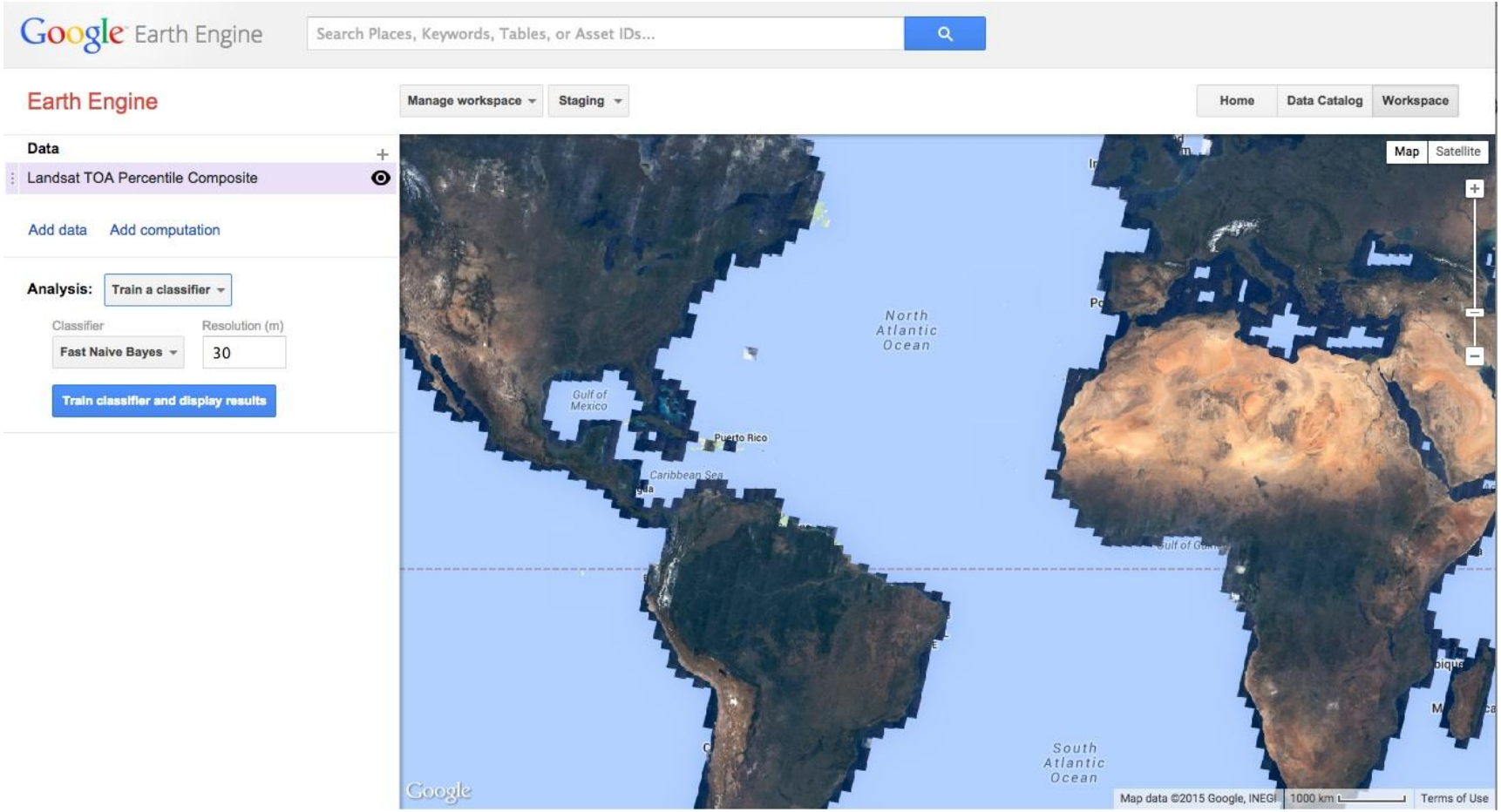
Justin Huntington & Charles Morton (DRI)

Data Types and Algorithm Primitives

- **Image** - band math, clip, convolution, neighborhood, selection ...
- **Image Collection** - map, aggregate, filter, mosaic, sort ...
- **Feature** - buffer, centroid, intersection, union, transform ...
- **Feature Collection** - aggregate, filter, flatten, merge, sort …
- **Filter** - by bounds, within distance, date, day-of-year, metadata ...
- **Reducer** - mosaic, mean, linearFit, percentile, histogram ….
- **Join** - simple, inner, outer, inverted ...
- **Kernel** - square, circle, gaussian, sobel, kirsch ...
- **Projection** - transform, translate, scale …

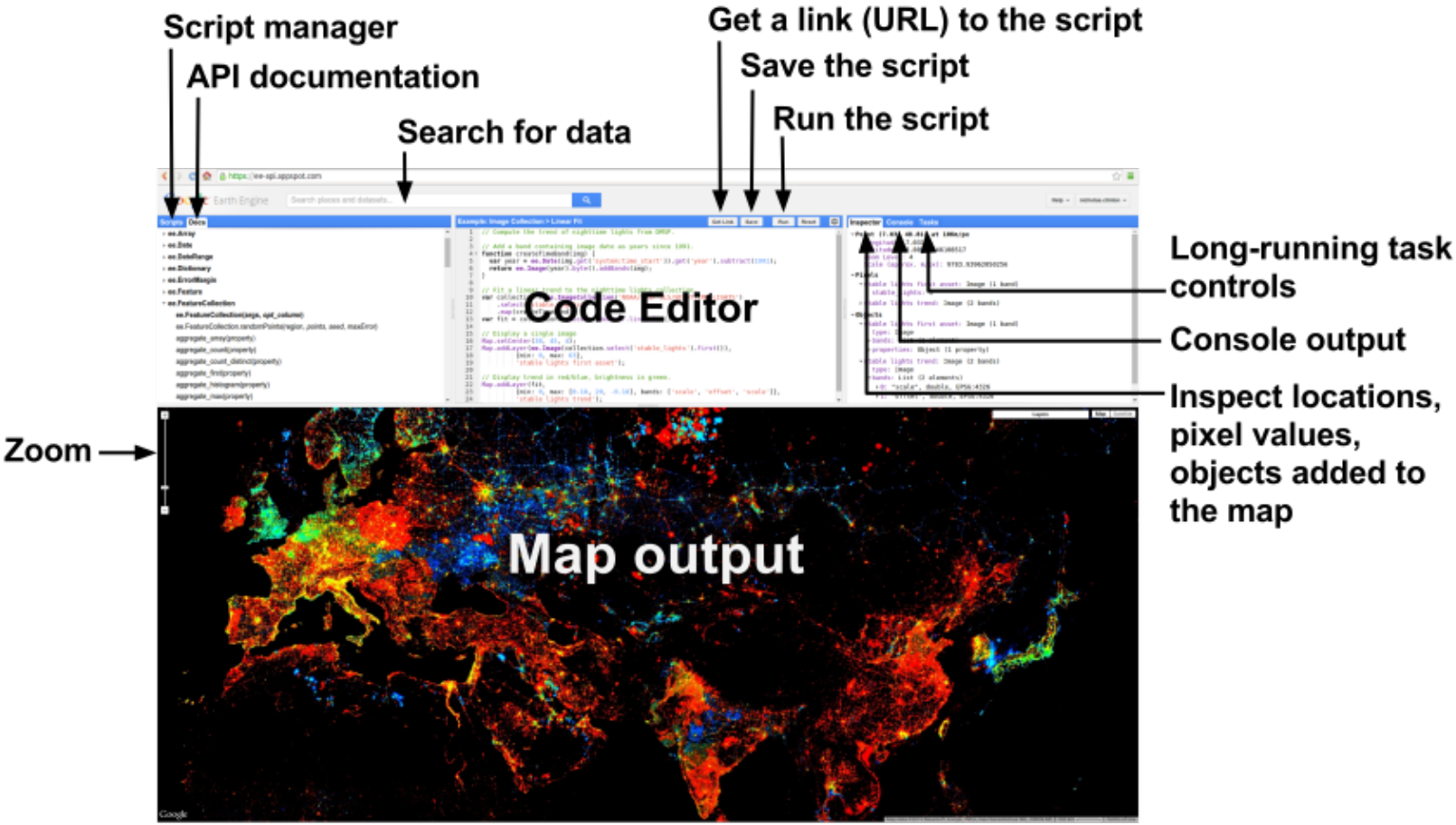
List, Dictionary, Array… over 500 primitives and growing!

Earth Engine Workspace



https://earthengine.google.org/#workspace

JavaScript Playground



http://ee-api.appspot.com

10 Second JavaScript Overview

```
print('Hello world!'); // print something

print(ee.Image('LANDSAT/LC8_L1T/LC80440342014077LGN00')); // image metadata

var image3 = image1.add(image2); // variables and objects

var terrainImage = ee.Algorithms.Terrain(dem); // Earth Engine algorithms

var myFunction = function(arguments) {
     // do something
     return something;
};

var collection2 = collection1.map(function); // Map

var statistic = collection1.reduce(ee.Reducer.something()) // Reduce
```

https://ee-api.appspot.com/708571296a2ea5a1d8fff64108b5d4be

Example 1
Finding, loading and displaying images

```
var collection = ee.ImageCollection('LANDSAT/LT5_L1T_TOA')
  .filterDate('2011-01-01', '2011-12-31')
  .filterBounds(ee.Geometry.Point(-122.2627, 37.8735));
print(collection);

var image = ee.Image('LANDSAT/LT5_L1T_TOA/LT50440342011261PAC01');

var vizParams = {bands: ['B4', 'B3', 'B2'], min: 0, max: 0.5};
Map.setCenter(-122.2627, 37.8735, 11);
Map.addLayer(image, vizParams, 'Landsat 5 false color composite');
```

https://ee-api.appspot.com/f809a8d1f36f394078a15b5e17e6be4b

Example 2 - Mapping

```
var collection = ee.ImageCollection('LANDSAT/LT5_L1T_TOA')
  .filterDate('2011-01-01', '2011-12-31')
  .filterBounds(ee.Geometry.Point(-122.2627, 37.8735));

var image = ee.Image('LANDSAT/LT5_L1T_TOA/LT50440342011261PAC01');

var addNDVI = function(image) {
  return image.addBands(image.normalizedDifference(['B4', 'B3']));
};

var ndvi = addNDVI(image);

var vizParams = {bands: ['nd'], min: -0.5, max: 1, palette: ['FF0000', '00FF00']};
Map.addLayer(ndvi, vizParams, 'NDVI');

var ndviCollection = collection.map(addNDVI);

print(ee.Image(ndviCollection.first()));
```

https://ee-api.appspot.com/6e8019d26ca50731abe40f78bb9a339e

Example 3 - Reducing

```
var collection = ee.ImageCollection('LANDSAT/LT5_L1T_TOA')
  .filterDate('2011-01-01', '2011-12-31')
  .filterBounds(ee.Geometry.Point(-122.2627, 37.8735));

var addNDVI = function(image) {
  return image.addBands(image.normalizedDifference(['B4', 'B3']));
};

var ndviCollection = collection.map(addNDVI);

var median = ndviCollection.reduce(ee.Reducer.median());

var vizParams = {bands: ['nd_median'], min: -0.5, max: 1, palette: ['FF0000', '00FF00']};
Map.setCenter(-122.2627, 37.8735, 11);
Map.addLayer(median, vizParams, 'NDVI');
```

https://ee-api.appspot.com/b4724841aa1358dd3bcf478d2b0f3448

What will you create with Earth Engine?

earthengine.google.org/signup

Leibniz-Institut für
Astrophysik Potsdam

Virtual Research Environments and Open Science

Open Science goes Geo - Part III: Beyond Data and Software

Jochen Klar

Astronomy!

Astronomy!

Leibniz-Institute for Astrophysics Potsdam (AIP)

- Cosmic magnetic fields
- Extragalactic Astrophysics
- Development of Research Technology and Infrastructure

E-Science @ AIP

- Data management
- Data publication

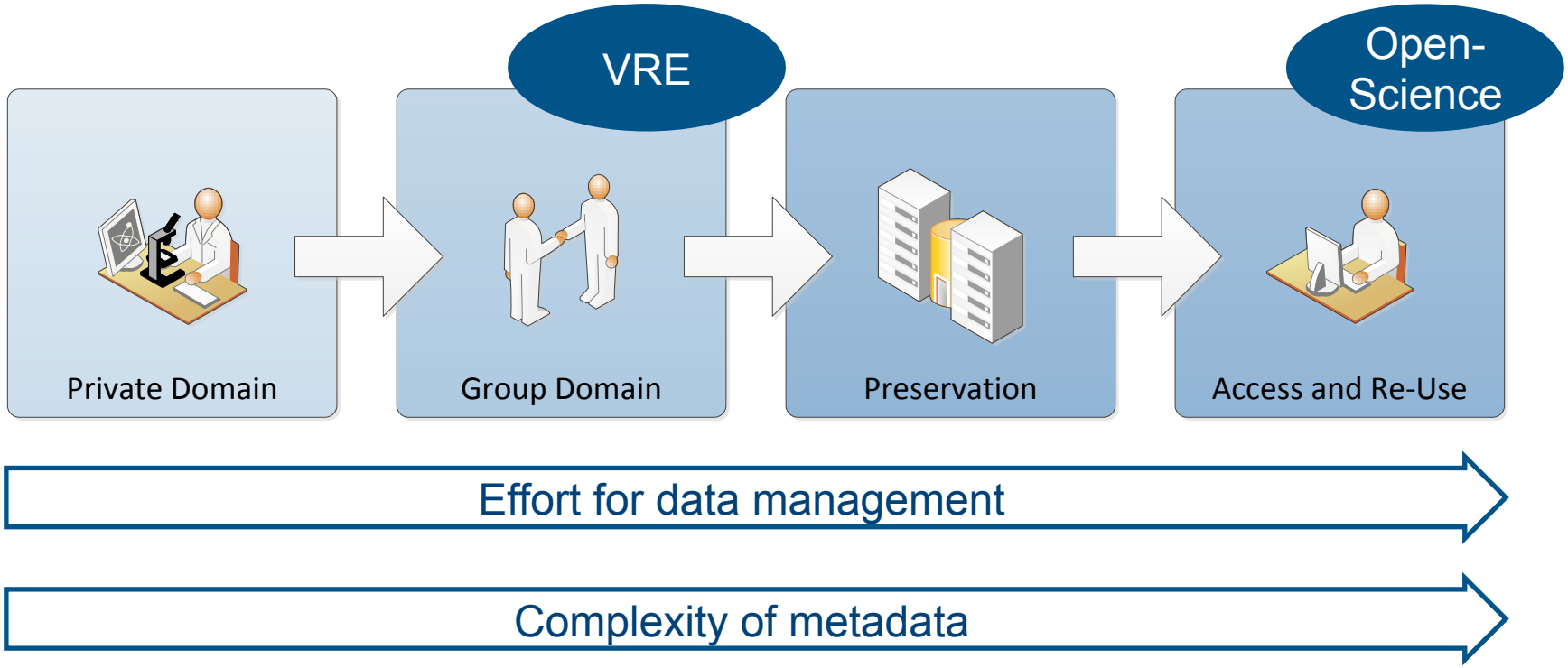
Virtual Research Environments

Virtual? Research? Environment?

- Cooperative research platform
- Available from different institutes, countries, continents
- *Potentially* comprises the whole research process
- Software services and communication networks
- Access to data, tools, resources, infrastructure
- Virtual organizations

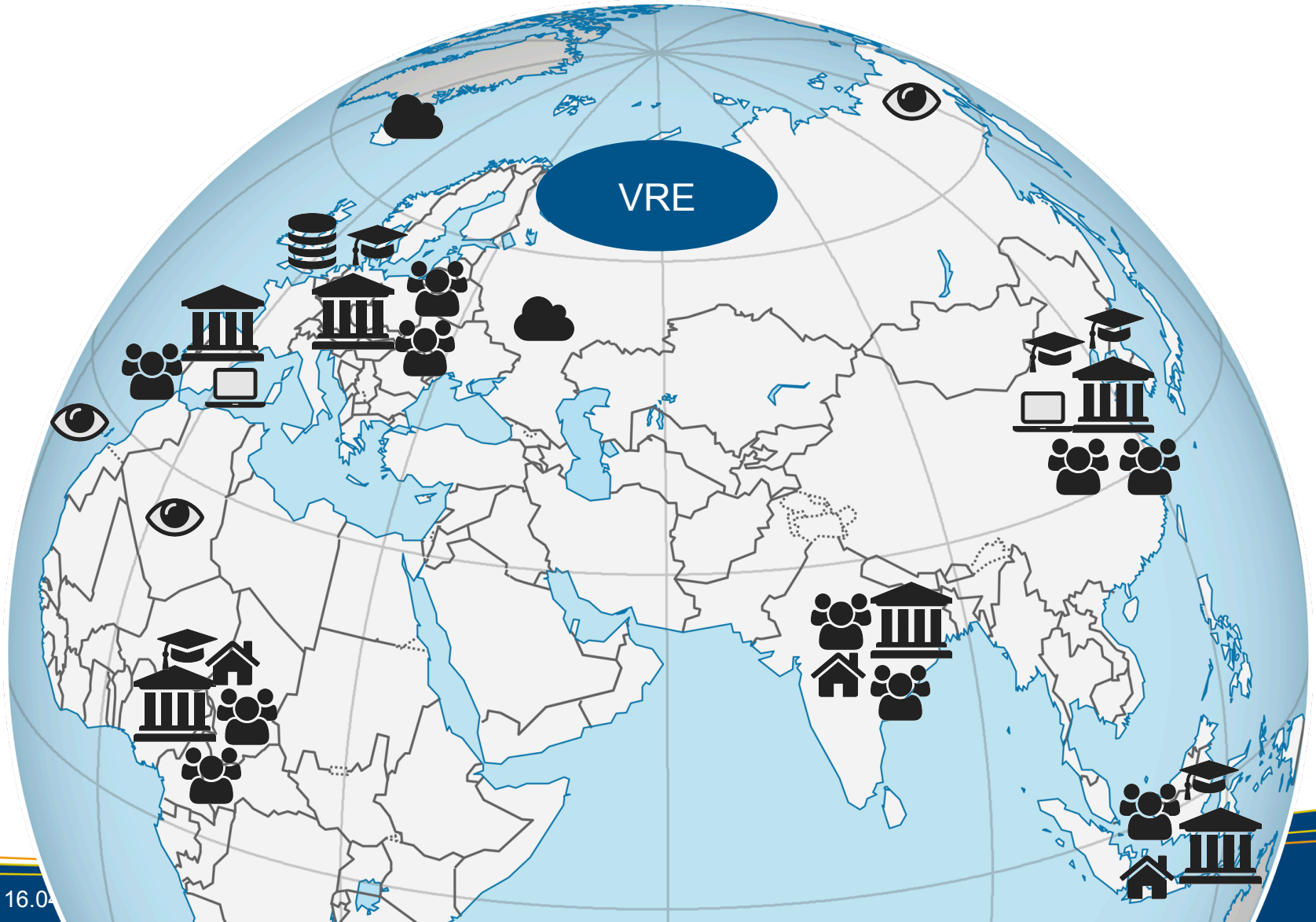
- SURFnet. Collaboration Infrastructure. Report. https://www.surf.nl/en/knowledge-and-innovation/knowledge-base/2009/report-collaboration-infrastructure.html, 2009.
- Amy Carusi and Torsten Reimer. Virtual Research Environment Collaborative Landscape Study. http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx, 2010.
- AG VRE der Allianz-Initiative Digitale Information. Virtuelle Forschungsumgebungen - Ein Leitfaden. http://www.allianzinitiative.de/handlungsfelder/virtuelle-forschungsumgebung/definition.html, 2011.

VRE <=> Open-Science



- Treloar, A., D. Groenewegen, and C. Harboe-Ree (2007), The Data Curation Continuum - Managing Data Objects in Institutional Repositories, D-Lib Magazine, 13(9/10), 13, http://dx.doi.org/10.1045/september2007-treloar
- DFG-Projekt RADIESCHEN (2013): Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Organisation und Struktur. http://dx.doi.org/10.2312/RADIESCHEN_005

So what is a VRE again?



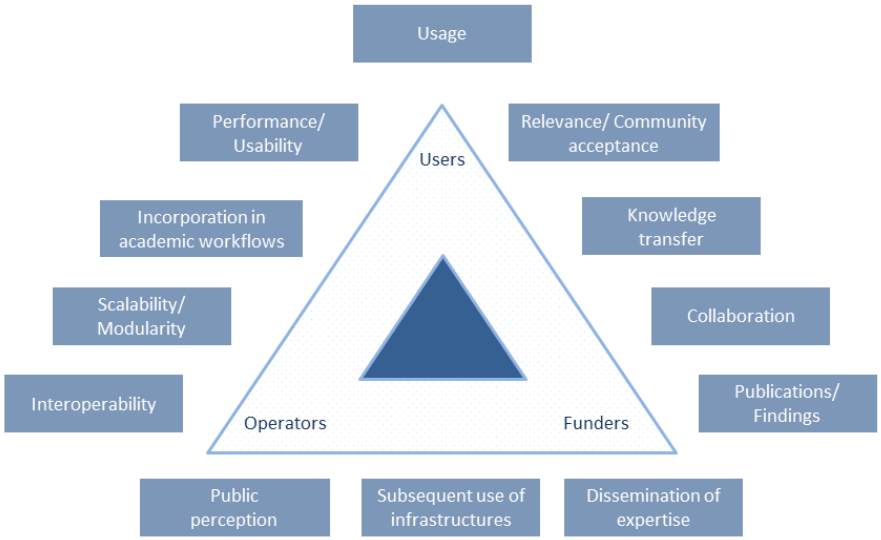
Issues with VRE

- Most VRE do not evolve beyond a prototype
  - Development is technology/software driven
  - Organizational aspects are neglected
- Sustainable long term funding is hard to archive
  - Initial funding is project based and time limited
  - Stakeholders have different agendas regarding the VRE
  - Members of the VRE come from different countries
- VRE are not reaching a satisfactory relevance
  - Scientist cling to their established workflows
  - Coverage of the whole scientific workflow might conflict with other established tools
  - Benefits of the VRE are/appear not big enough
  - Community integration of the VRE might be insufficient

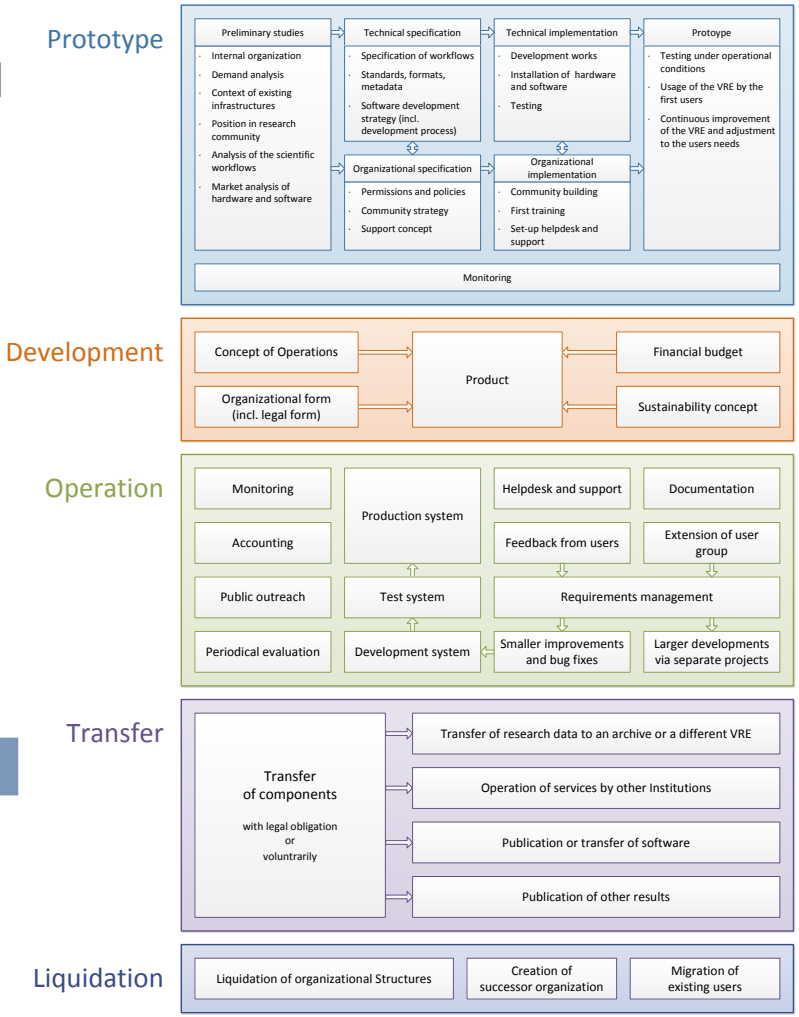
Project DFG-VRE

Success criteria for the development and sustainable operation of virtual research environments

SUB Göttingen, TU Dortmund, AIP (2013-2014)



www.forschungsdaten.org/index.php/DFG-VRE



TextGrid

TextGrid
Virtual Research Environment
for the Humanities

- supports scholarly work in the humanities with digital methods, tools, and services
- started in 2006, funded by the German Federal Ministry of Education and Research (BMBF)
- TextGrid Laboratory
  - individually adaptable and extensible range of applications
  - based on the Eclipse IDE
- TextGrid Repository
  - storage capacity for the VRE
  - permanent, secure and citable long-term archiving

http://www.textgrid.de

TextGrid Repository

```
</sp>
▼<sp>
  <speaker xml:id="tg34.2.22.part1">PAMINA.</speaker>
  <l xml:id="tg34.2.22.part2">Aber liebste Mutter! –</l>
 </sp>
▼<sp>
  <speaker xml:id="tg34.2.23.part1">KÖNIGINN.</speaker>
 ▼<lg>
    <l xml:id="tg34.2.23.part2">Kein Wort!</l>
  </lg>
 ▼<stage rend="zenoPC" xml:id="tg34.2.26">
    <hi rend="italic" xml:id="tg34.2.26.1">Arie.</hi>
  </stage>
  <lb xml:id="tg34.2.27"/>
 ▼<lg>
    <l rend="zenoPLm4n4" xml:id="tg34.2.28">Der Hölle Rache kocht in meinem Herzen,</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.29">Tod und Verzweiflung flammet um mich her!</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.30">Fühlt nicht durch dich Sarastro Todesschmerzen,</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.31">So bist du meine Tochter nimmermehr.</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.32">Verstossen sey auf ewig und verlassen,</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.33">Zertrümmert alle Bande der Natur,</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.34">Wenn nicht durch dich Sarastro wird erblassen!</l>
    <l rend="zenoPLm4n4" xml:id="tg34.2.35">Hört Rache, – Götter! – Hört der Mutter Schwur.</l>
  </lg>
  <lb xml:id="tg34.2.36"/>
 ▼<stage rend="zenoPC" xml:id="tg34.2.37">
    <hi rend="italic" xml:id="tg34.2.37.1">Sie versinkt.</hi>
  </stage>
  <lb xml:id="tg34.2.38"/>
 </sp>
</div>
</div>
</div>
▼<div subtype="work:no" xml:id="tg35" n="/Literatur/M/Schikaneder, Johann Emanuel/Libretto/Die Zauberflöte/2. Akt/9. Auftritt">
 ▼<div>
```

XML poetry …

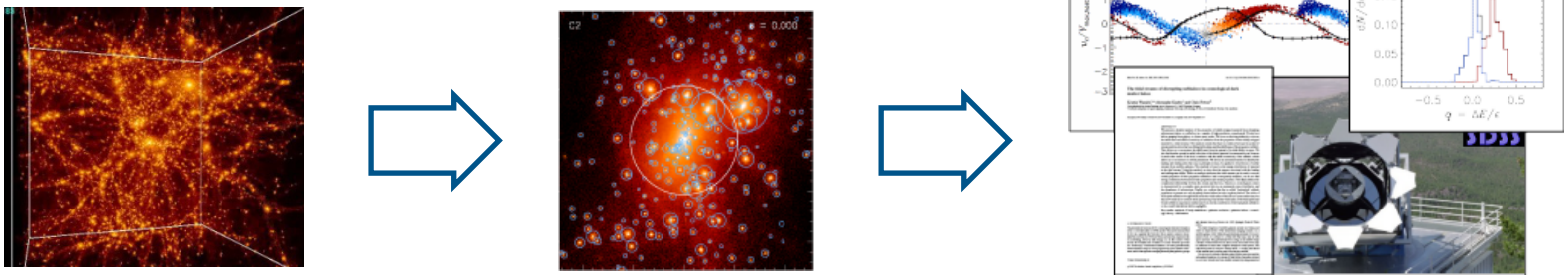
http://www.textgridrep.de

VFU soeb

- VRE for the German Socio-Economic Panel
- Interviews of the same 12000 households once a year since 1984
- No sharing of the actual research data of the study due to privacy restrictions
- Discovery of study descriptions
- Common metadata schema based on DDI
- Collaborative work on metadata, descriptions, source code for analysis (Syntax), documents, …
- Liferay-based CMS + Archive, Forum, Wiki, …

CLUES VRE

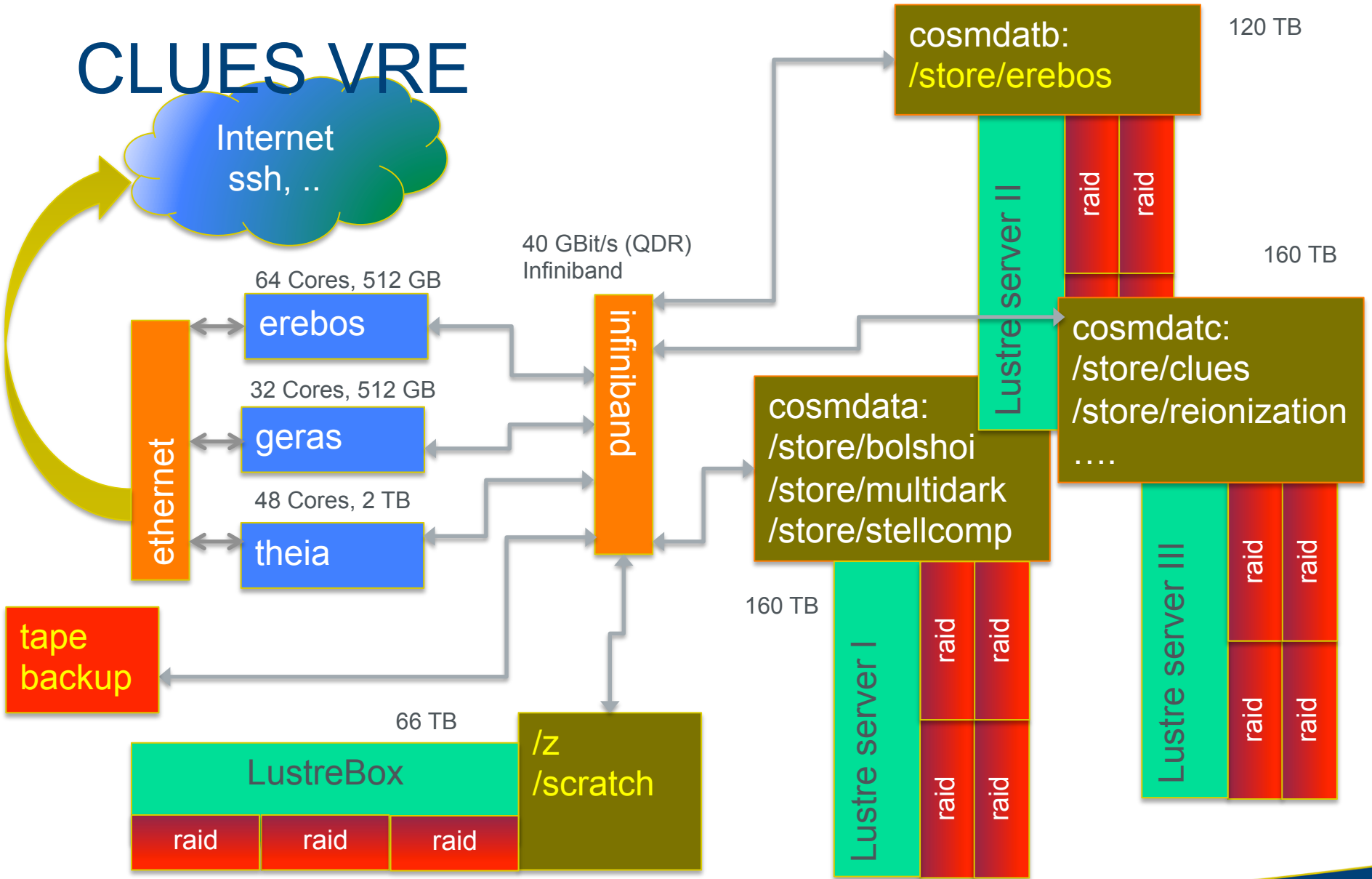
- Constrained Local UniversE Simulations
- > 30 Scientists from Potsdam, Jerusalem, Madrid, Las Cruces, Santa Cruz, Lyon, Hawaii, Bogota, …
- a lot of data from simulations (~ 400 Tb)
  - raw data from the supercomputing centers
  - post processing data products



- scientific programs, compiler, libs, … login via ssh

www.clues-project.org

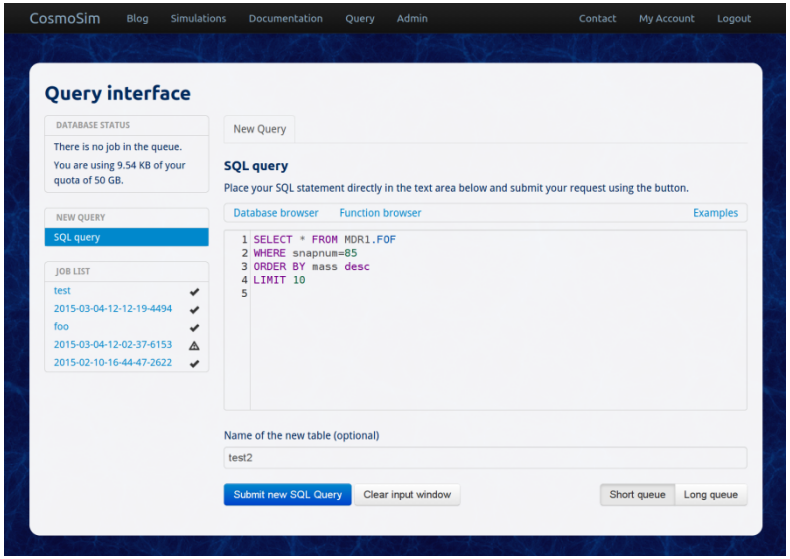
CLUES VRE

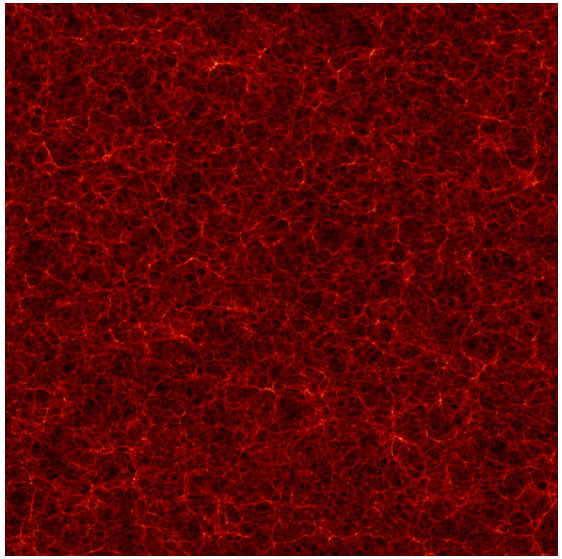

Internet
ssh, ..
ethernet
64 Cores, 512 GB
erebos
32 Cores, 512 GB
geras
48 Cores, 2 TB
theia
40 GBit/s (QDR)
Infiniband
infiniband
cosmdatb:
/store/erebos
120 TB
Lustre server II
raid
raid
160 TB
cosmdatc:
/store/clues
/store/reionization
….
cosmdata:
/store/bolshoi
/store/multidark
/store/stellcomp
160 TB
Lustre server I
raid
raid
raid
raid
Lustre server III
raid
raid
raid
raid
tape
backup
66 TB
LustreBox
raid
raid
raid
/z
/scratch

CosmoSim

- public access to cosmological simulations
- query subsets using SQL
- share the query not the data

```
SELECT * FROM MDR1.FOF
WHERE snapnum=85
ORDER BY mass DESC LIMIT 10
```





www.cosmosim.org

Daiquiri

- Highly customizable framework for database publication
- Many applications – one code base
- Queued SQL queries including SQL query assistance
- Can be used together with PaQu for Parallel queries
- Query result table viewer, quick plotting tool
- User database space, download in different formats
- VO compatible meta data management for databases
- User management, Contact messages, WordPress integration, Meeting organization, …
- Used for CosmoSim, RaveDB, APPLAUSE, Gaia@AIP,…
- Open Source Software escience.aip.de/daiquiri

Leibniz-Institut für
Astrophysik Potsdam

Thank you

jklar@aip.de, @jochenklar, github.com/jochenklar

forschungsdaten.org

Development Model for Research Infrastructures

J. Wächter, M. Hammitzsch, D. Kerschke, and J. Lauterjung

German Research Centre for Geosciences GFZ

GFZ Helmholtz-Zentrum Potsdam

HELMHOLTZ GEMEINSCHAFT

Content

- Framing conditions
- Factors influencing infrastructure development
- From capabilities to maturity
- Summary and perspectives

Research Infrastructure

- Research infrastructures (RIs) are platforms integrating facilities, resources and services.
- RIs include:
  - scientific equipment, e.g., sensor platforms, satellites or other instruments
  - scientific data, sample repositories or archives
  - computing and storage services
- E-infrastructures provide the technological integration platform to interlink distributed RI components.
- The resulting standardised platforms provide the foundation for the design and implementation of a new generation of applications incl. VREs.
- The current development of RIs is strongly driven and enabled by Information and Communication Technology (IT).

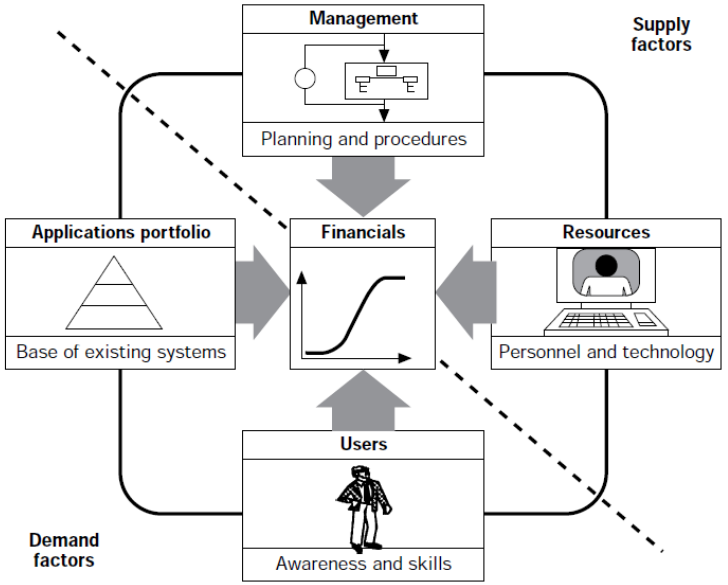
Snapshot of RI development

- Parallel activities on European and national levels with numerous institutes and organisations involved.
- The conceptual and technological maturity of individual scientific domains differs considerably.
- The concrete implementation process consists of independent and parallel development activities.
- Integration of distributed heterogeneous systems and components.
- Individual systems: often complex with a long-term history.
- Different maturity levels: in respect to standardisation of interfaces and data exchange capabilities.
- Most of the funding currently available for RI implementation is provided on a project basis.

Development of IT in Organisations

- **"Productivity Paradoxon"** investments focused on technology failed

  -> Integrated Approach
- Growth factors (according Nolan Stages Theory):
  - Supply: IT resources + IT management
  - Demand: application portfolio + user capabilities
- Key management activity **->** balancing of relevant growth factors
- RI development has to deal with both
  - technology development and
  - organisational development

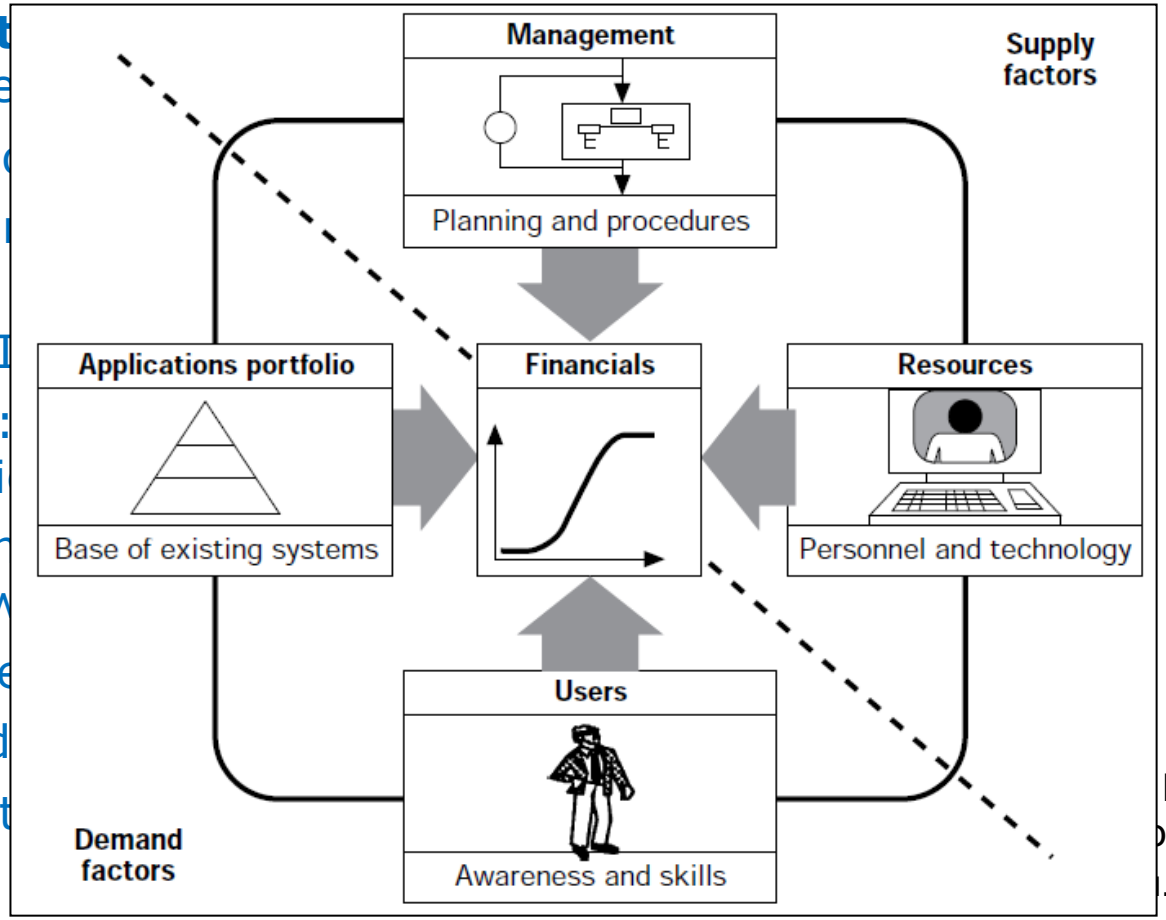


Nolan Stages Theory: Framework for the development of IT in organisations

Mutsaers, E.-J., van der Zee, H., and Giertz, H. (1998)

Development of IT in Organisations

- **"Productivit**
  focused on te
  -> Integrate
- Growth factor
  Theory):
  - Supply: I
  - Demand:
    capabiliti
- Key manager
  relevant grow
- RI developme
  - system d
  - organisat



Framework for the
organisations
., and Giertz, H. (1998)

Growth Factors mapped on RIs

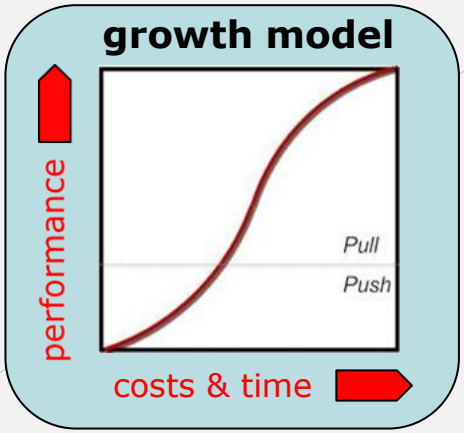
Application Portfolio

- Portals + collaboration
- Data management tools
- Integration and modeling
- Research Software

Demand Factors

Scientists

- Competencies and skills
- Integrated modeling
- Software development
- Data management
- Visions for integration



Resources

- Data archives
- Sensor platforms
- Monitoring systems
- Storage and computing
- Service platforms
- Registries

Supply Factors

IT Organisation

- Software engineering
- Data management
- Framework development
- Service Management

Lessons Learned: Development of SDIs

- The development of Spatial Data Infrastructures (SDIs) was initiated in 1994 by the "Clinton Adminstrative Order"
- Trigger for initiatives on the national and international level:
  - Standardisation processes: interfaces and encoding
  - Community processes harmonising the common understanding of data and semantics
- Until 2007: European activities mainly organised on the national level
- INSPIRE initiative of the European Commission:
  - Policy framework including directives and regulations
  - Strict time schedule
  - Reference models for the implementation process

Growth Factors cont.

Policies

- IT Governance
- Research assessment
- Intellectual property
- Data and software publication
- Quality models

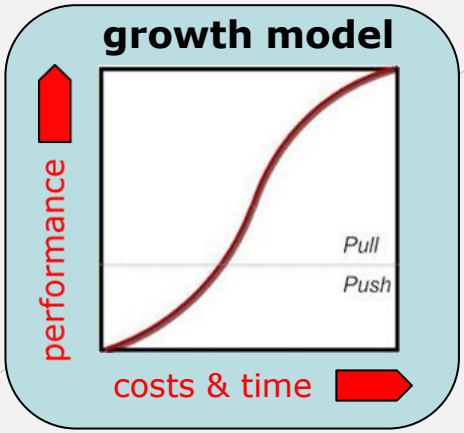
Application Portfolio

- Portals + collaboration
- Data management tools
- Integration and modeling
- Research Software

Demand Factors

Scientists

- Competencies and skills
- Integrated modeling
- Software development
- Data management
- Visions for integration



Resources

- Data archives
- Sensor platforms
- Monitoring systems
- Storage and computing
- Service platforms
- Registries

IT Organisation

- Software engineering
- Data management
- Framework development
- Service Management

Supply Factors

Reference Architecture

- Interfaces
- Encoding
- Conceptual metadata
- Design patterns
- Framework

Levels of Interoperability Model

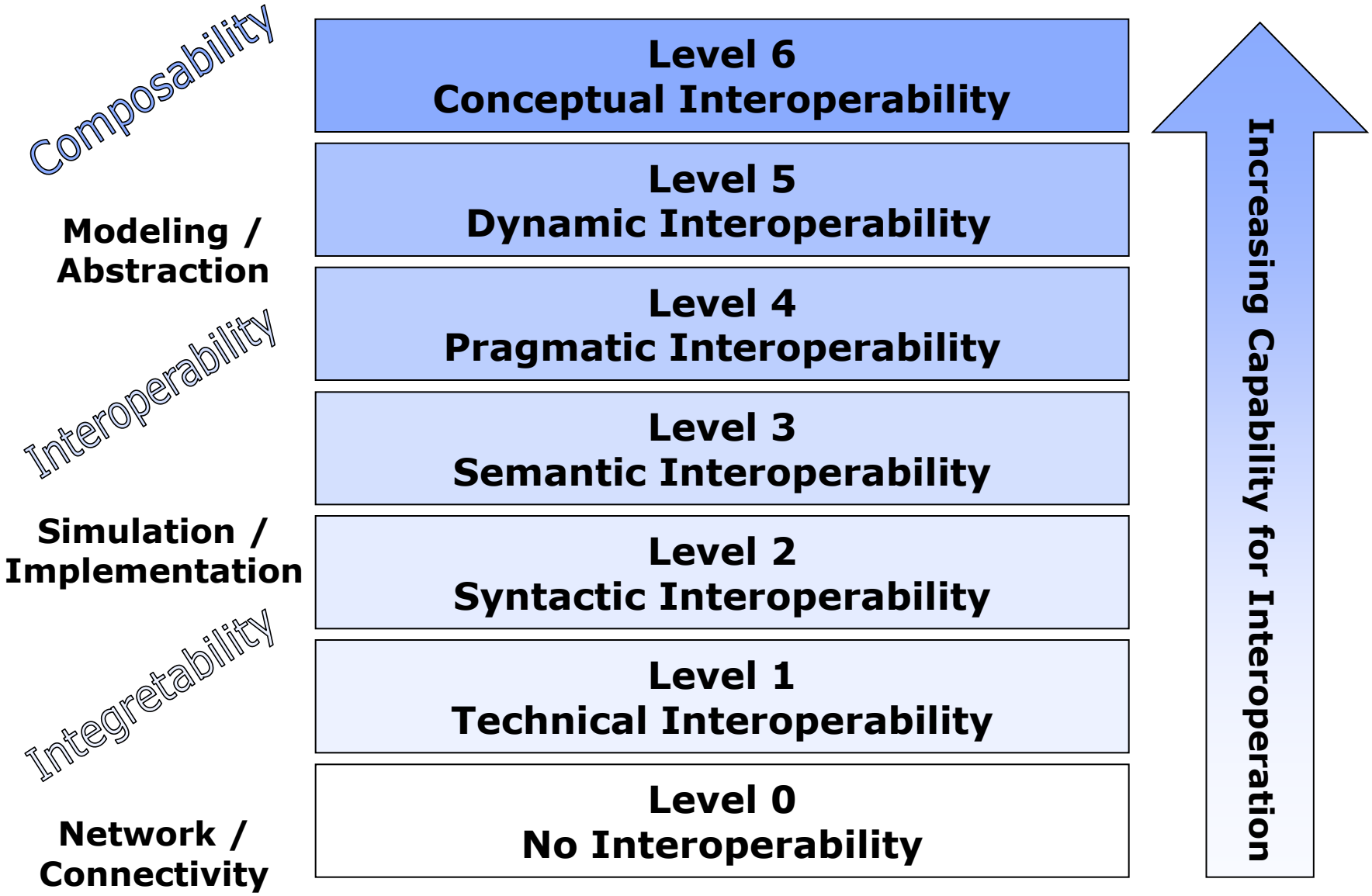
"While a picture says more than 1,000 words …, an executable Modeling & Simulation application says more than 1,000 pictures!" (Tolk et. al. 2007)

- Origin of the Levels of Interoperability Model (LCIM):
  - The military context
  - Interoperability in network-centric environments
  - Command and Control Systems (CCS)

- System-of-Systems environment
- Independent system collaborate to fulfil a common task
- Modelling and simulation are important components of CCS.
- Systems have to understand each other.

http://www.cotsjournalonline.com/files/images/2330/COTS1207_Radisys_Fig01_large.jpg

Levels of Conceptual Interoperability Model LCIM



LCIM Levels

L6 Conceptual: Systems are completely aware of each others information, processes, contexts, and modeling assumptions.

L5 Dynamic: Systems are able to reorient information production and consumption based on understood changes to meaning, due to changing context as time increases.

L4 Pragmatic: Interoperating systems will be aware of the context (system states and processes) and meaning of information being exchanged.

L3 Semantic: Interoperating systems are exchanging a set of terms that they can semantically parse.

L2 Syntactic: Agreed protocols to exchange the right forms of data in the right order, but the meaning of data elements is not established.

L1 Technical: Have technical connection(s) and can exchange Data between systems

Wang W.G., Tolk A., Wang W.P., 2009: The levels of conceptual interoperability model: Applying systems engineering principles to M&S. Spring Simulation Multiconference (SpringSim'09). San Diego, CA, USA.

From Capabilities to Maturity

Policies

- IT Governance
- Research assessment
- Intellectual property
- Data and software publication
- Quality models

Application Portfolio

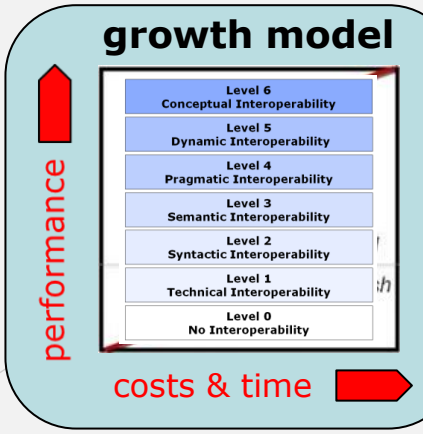
- Portals + collaboration
- Data management tools
- Integration and modeling
- Research Software

Demand Factors

Scientists

- Competencies and skills
- Integrated modeling
- Software development
- Data management
- Visions for integration

growth model



Resources

- Data archives
- Sensor platforms
- Monitoring systems
- Storage and computing
- Service platforms
- Registries

IT Organisation

- Software engineering
- Data management
- Framework development
- Service Management

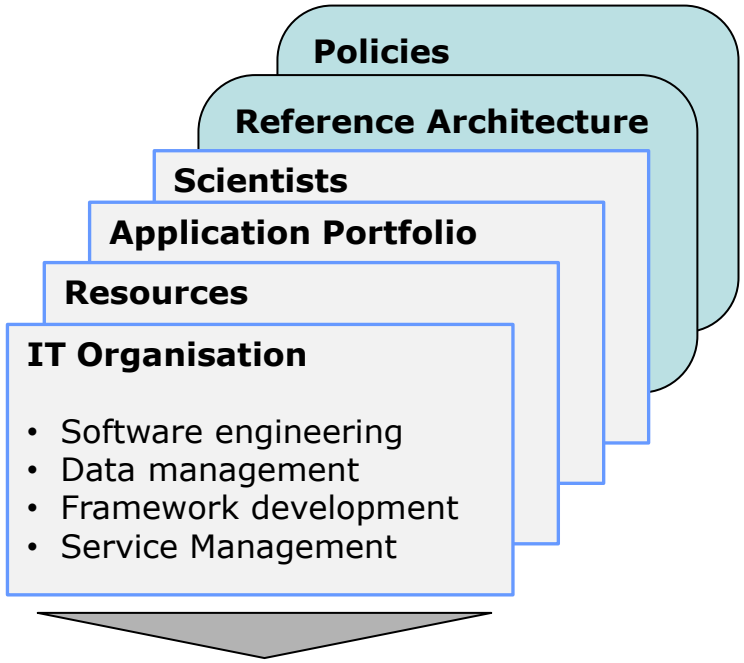
Supply Factors

Reference Architecture

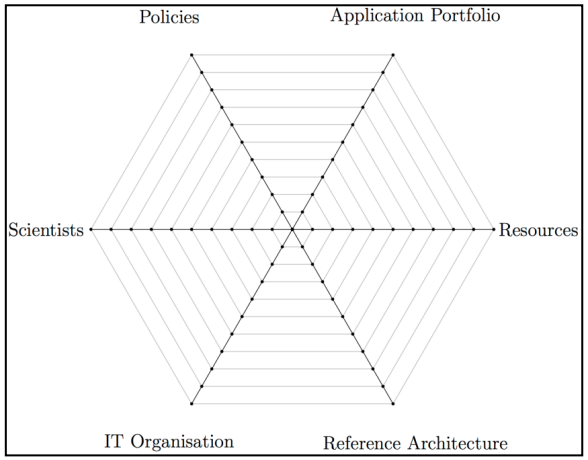
- Interfaces
- Encoding
- Conceptual metadata
- Design patterns
- Framework

Research Infrastructure Maturity Model – RIMM

- Based on six Growth Factors incl. technology and organisation
- Growth Factors deliver a catalogue of relevant capabilites for RI development.
- Application: (Self-) evaluation of organisations active in or responsible for RI development.
- Quality criteria for the determination of maturity are derived from the LCIM interoperability levels.
- The resulting report can be used to identify critical situations and improve relevant capabilities.



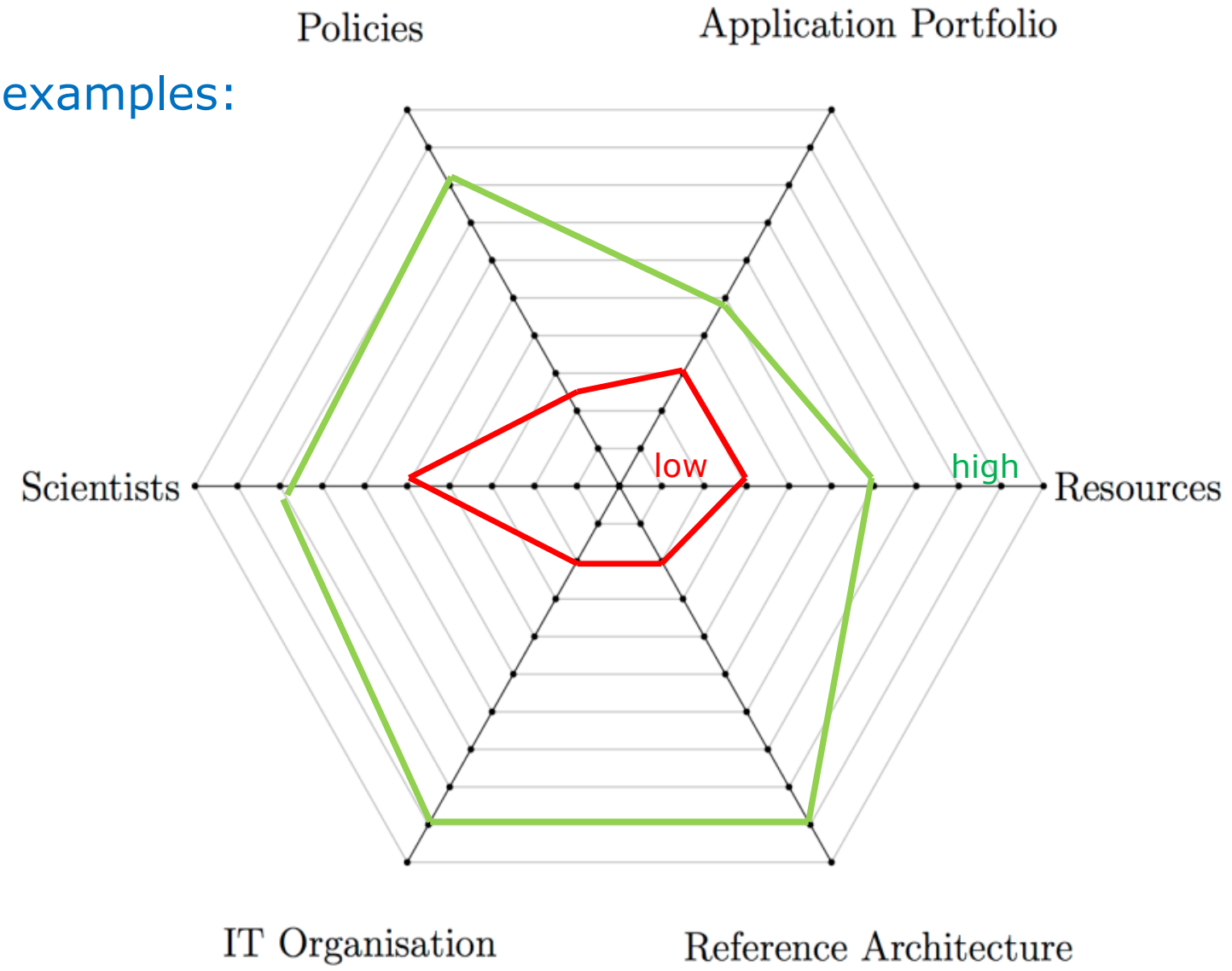
RIMM



RIMM Chart – Example

Maturity plots examples:

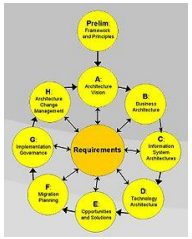
- **positive**
- **negative**



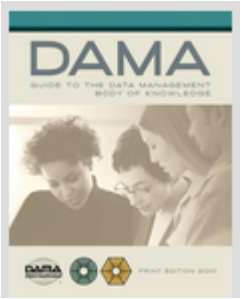
Summary and Perspectives

- The development of next generation RIs is a challenging development process in a complex organisational environment.
- RIMM offers a management and engineering framework to reduce the risk of failure.
- RIMM is based on the identification of technological and organisational development factors and capabilities.
- Maturity determination based on quality criteria derived from the LCIM interoperability levels.
- RIMM should be applied for the (self-) evaluation of organisations.
- Science should take advantage of the rich experience of the IT domain.

THE Open GROUP



TOGAF®



DAMA-DMBOK

OGC
Office of Government Commerce

ITIL®

http://www.opengroup.org/subjectareas/enterprise/togaf/
http://www.dama.org/content/body-knowledge