



ENVI RI Common Operations of
Environmental Research Infrastructures

Sustaining common operations & Future ambitions

Cees de Laat
University of Amsterdam

Connecting information
and knowledge
from Deep Earth,
land and sea,
the atmosphere

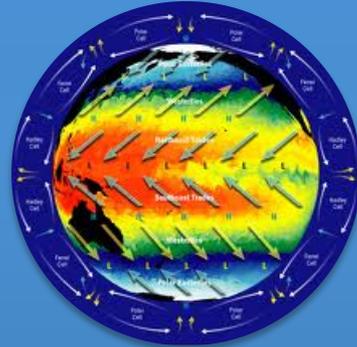
- Both living and dead
environments



Environmental Science



oceanic and
atmospheric
processes



long-term
development of the
climate system



Biological processes
biodiversity



development of the
cryosphere and
lithosphere

Earth as a single complex and coupled system

ESFRI Environmental Research Infrastructures

- Tropospheric research aircraft



COPAL

- Upgrade of incoherent SCATter facility



EISCAT-3D

- Multidisciplinary seafloor observatory



EMSO

- Plate observing system



EPOS

- Global ocean observing infrastructure



EURO-ARGO

- Aircraft for global observing system



IAGOS

- Integrated carbon observation system



ICOS

- Biodiversity and ecosystem research infra

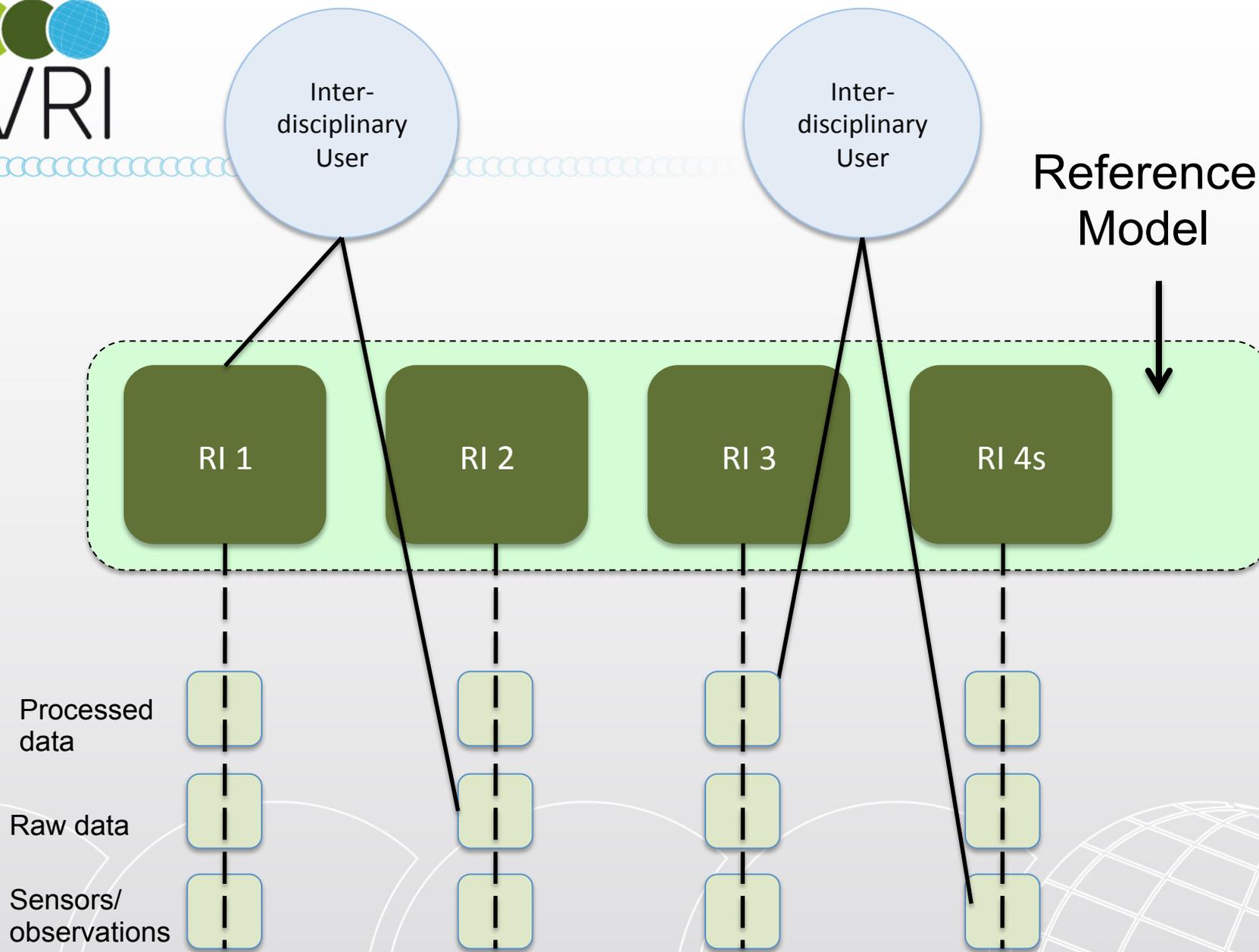


LIFEWATCH

- Svalbard arctic Earth observing system



SIOS



Enable multidisciplinary scientists to **access, study**
and **correlate** data from **multiple domains** for
“system level” research

*by providing solutions and
guidelines for the RIs
common needs*



The ENVRI project is supporting the environmental ESFRI research infrastructures with common solutions.

- The ENVRI Reference MODEL
- Common data discovery and processing tools
- With EUDAT
 - **Semantic annotation**
 - **Real time data and streaming analytics**

These services should be regularly updated and also introduced to other (national and non-European) facilities.

Reference MODEL Viewpoints

Enterprise

Business Aspects

The purpose, scope and policies for the organization that will own the system

What for? why? who? when?

Information

Information System Aspects

Information handled by the system and constraints on the use and interpretation of that information

What is it about?

Computational

Application Design Aspects

Functional decomposition of the system into objects suitable for distribution

How does each bit work?

Technology

Implementation

System hardware & software and actual distribution

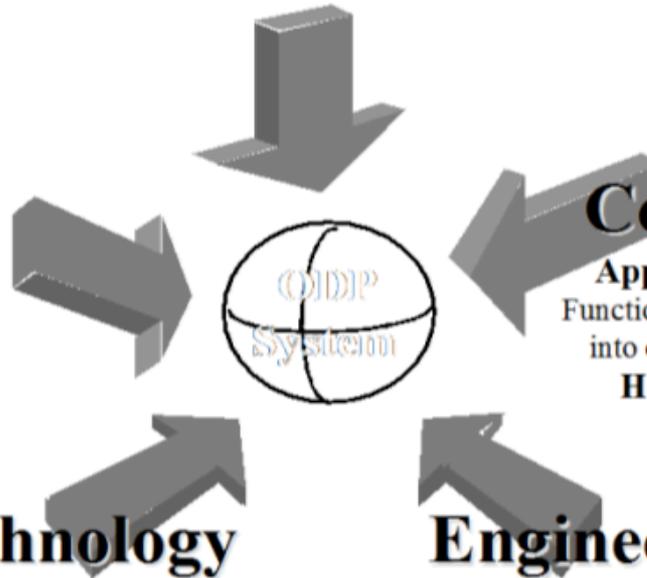
With what?

Engineering

Solution Types & Distribution

Infrastructure required to support distribution

How do the bits work together?



PROVIDE SOFTWARE TOOLS TO

Promote Accessibility

discover data
which are
heterogeneous in
format, content,
and metadata
description

**harmonise,
integrate and
analyse data**
across domains
and RIs

Preserve Specificity

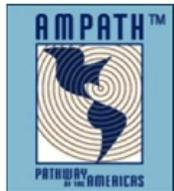
- RIs are contacted individually to develop a “train-the-trainers” scheme for adapting the Reference Model and common ENVRI tools. The scheme includes online trainings (WebEx) and face-to-face meetings
- EGI Community Forum, May 2014 in Helsinki
- PIRE workshop, June 2014 in Amsterdam
- RDA conference, September 2014 in Amsterdam

Announcement June 2014

PIRE Workshop Amsterdam



- OpenScienceDataCloud.org
- PIRE Fellowship Application (+/- 15)
- The OSDC PIRE Program is six to eight week fully funded fellowship for US graduate student researchers with an information technology background.
- Format:
 - 1 week tutorials and hands on training
 - 2 months research at a participating institute
 - Results in science/tools and papers/posters/



THE UNIVERSITY OF
CHICAGO

COMMIT/



UNIVERSITY OF AMSTERDAM

PIRE - OpenScienceDataCloud.org

1000 Genomes Project

Human sequence data from populations around the world with the goal of cataloging human genetic variation.

Total Size: 383.5TB Categories: [genomics](#), [biology](#)

ASTER

ASTER Level-1B Registered Radiance at the Sensor

Total Size: 12.7TB Categories: [earth science](#)

Complete Genomics Public Data

Whole human genome sequence data sets provided by Complete Genomics, containing 69 standard, non-diseased samples as well as two matched tumor and normal sample pairs.

Total Size: 47.1TB Categories: [genomics](#), [biology](#)

Earth Observing-1 Mission

Data gathered by the Advanced Land Imager (ALI) Hyperspectral Imager (Hyperion) instruments on NASA's Earth Observing-1 Mission (EO-1) satellite.

Total Size: 45.2TB Categories: [earth science](#), [satellite imagery](#)

City of Chicago Public Datasets

Data set from the City of Chicago Data Portal in JSON format for tabular data and the raw files for "blob" data.

Total Size: 9.7GB Categories: [social science](#)

EMDataBank

Unified Data Resource for 3-Dimensional Electron Microscopy

Total Size: 91.3GB Categories: [biology](#)

Enron Emails

Data sets based on the original Enron emails released to the public by the Federal Energy Regulatory Commission as part of their investigation.

Total Size: 155.9GB Categories: [social science](#)

FlyBase

FlyBase is the leading database and web portal for genetic and genomic information on the fruit fly *Drosophila melanogaster* and related fly species.

Total Size: 614.8GB Categories: [biology](#), [genomics](#)

.....



Research Data Alliance

- <https://rd-alliance.org>
- The Research Data Alliance implements the technology, practice, and connections that make Data Work across barriers.
- The Research Data Alliance aims to accelerate and facilitate research data sharing and exchange.
 - Working groups and interest groups
 - Joining groups and attendance at the twice-yearly plenary meetings is open.
- P4 Sep 22-24 2014 hosted by the Netherlands - Amsterdam
 - Conference Management Team (CMT) Chair: Peter Doorn (DANS)
 - Program Committee (PC): chair Cees de Laat (UvA)
 - Satellite Events Committee (SEC): Jeroen Rombouts (TUD)



ENVRI Common Operations of
Environmental Research Infrastructures

Ambitions for the upcoming years

And the implications for

- Sustainability of new services
- The role of e-Infrastructures

- **Innovation in novel physical instrumentation**
- **Optimize the data pipeline from (distributed) data generators to storage, to use**
- **Virtualization: enhanced networked/interrelated instruments, other facilities, web services, data, operators, diverse users (incl. citizen scientists)**
- **Data quality**
- **Support common operations and roadmapping**

- e-Infrastructure providers supported the development of various tools or are interested to be involved in other developments.
- These are partly the EU supported providers such as GEANT, PRACE, EGI and EUDAT.
- But also increasingly others, such as ESA, CNR(ISTI), CERN or computing centers in Europe.

Questions:

- Which business model is offered by e-Infrastructures to secure their sustained services?
- What are the implications for various ESFRI infrastructures depending on offered e-Infrastructure services?
- What is the long-term perspective?

THANKS