

# eInfrastructure@GEOSPATIAL

## Research Data - Infrastructure and Services

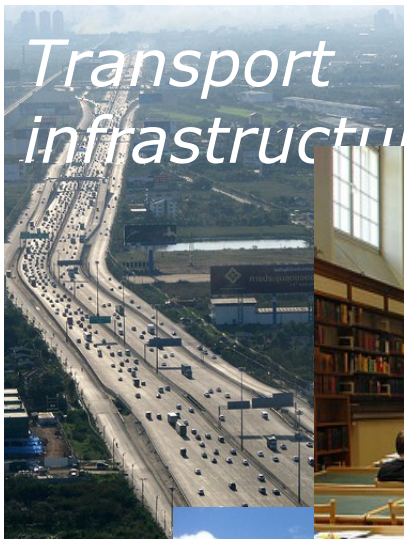


**Wim Jansen**  
European Commission  
DG CONNECT  
eInfrastructure

## **This presentation is about:**

- **Data and Computing e-infrastructures go together**
  - **Research Infrastructures and e-Infrastructures go together**
- **Research Data –European Policy Framework**
- **Research Data Alliance**
  - **Report “Data Harvest” (follow-up of “Riding the Wave”)**

# What are infrastructures?



*Transport  
infrastructure*



*Transport  
infrastructure*



*Education  
infrastructure*



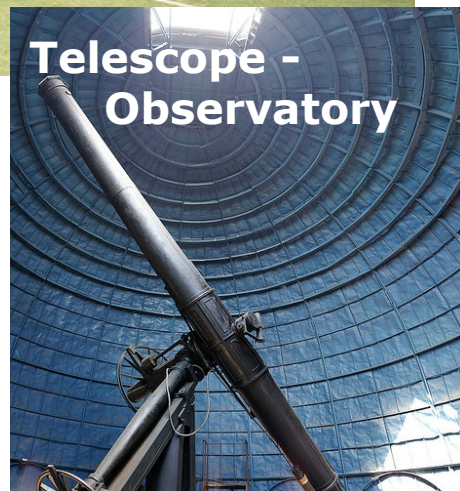
*Energy  
infrastructure*



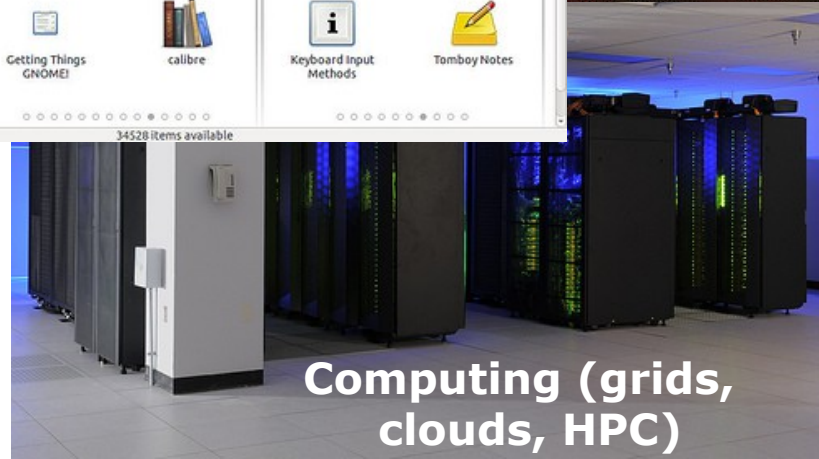
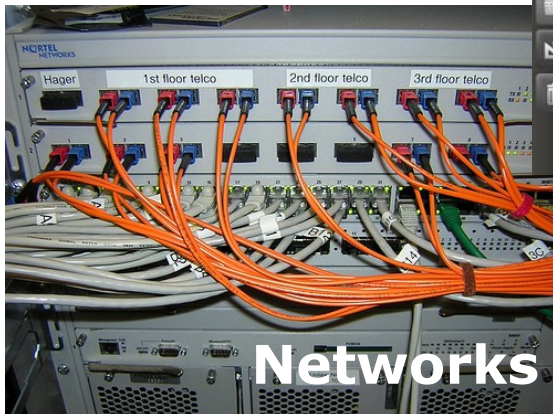
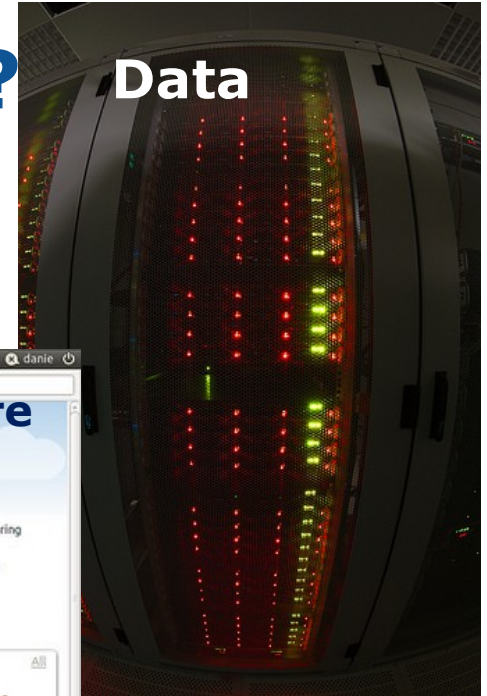
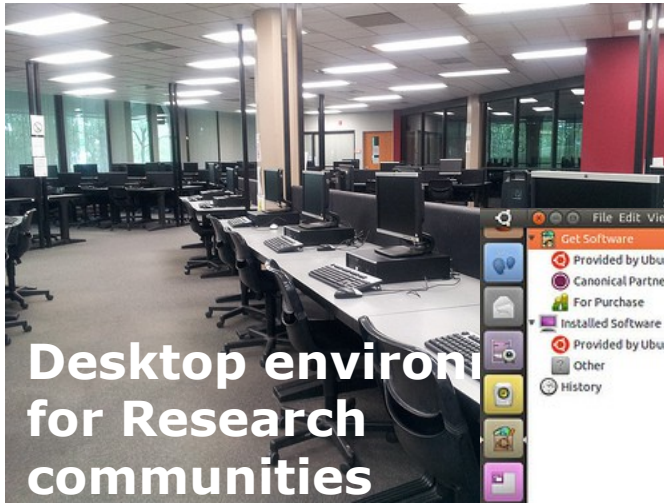
*Energy  
infrastructure*



# What are Research Infrastructures?



# What are e-Infrastructures?



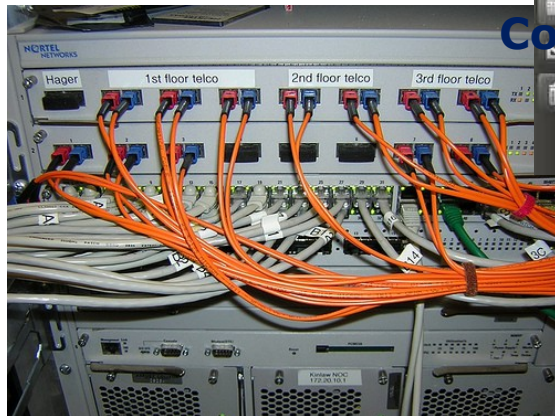
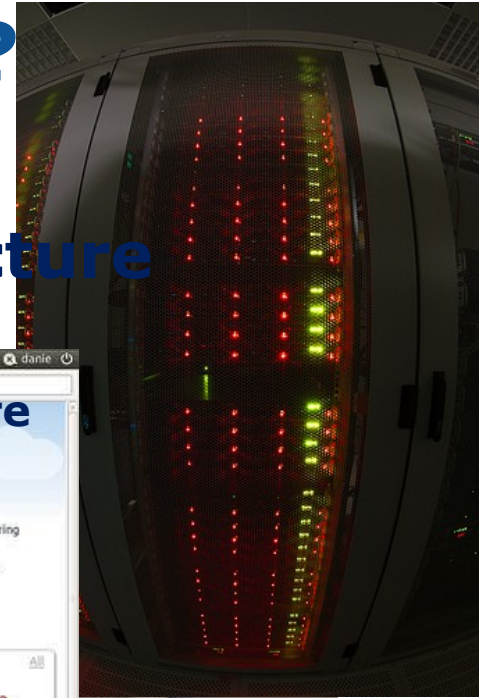




# What are e-Infrastructures?



**Integration : e-Infrastructure for all...**



**Software**

**Networks, connectivity**

**Data**

**Computing (grids, clouds, HPC)**

**Virtual Research Environments**

A screenshot of the Ubuntu Software Center window. The window title is "Get Software". It shows a sidebar with categories like "Provided by Ubuntu", "Canonical Partners", "For Purchase", "Installed Software", "Provided by Ubuntu", "Other", and "History". The main area displays "Departments" such as Accessories, Developer Tools, Education, Fonts, Internet, Office, Sound & Video, and Universal Access. There are also sections for "Featured" and "What's New".

## e-Infrastructure and Einstein

**E**-Infrastructure = **M**emory x **C**onnectivity x **C**omputing

$$E = M \times C \times C$$

$$E = M C^2$$

## Theme of this Conference:

### CONVERGENCE:

1. the act of converging and especially moving toward union or uniformity
2. the merging of distinct technologies, industries, or devices into a unified whole

**e-Infrastructures**



# Vision

- ACHIEVING DIGITAL ERA
- BRIDGE DIGITAL DIVIDES
- EVERY RESEARCHER DIGITAL





# Approach

## Transversal

Cutting across disciplines and sectors

Support tomorrow's science

Open science, open access, best solutions

Enabling innovation

Developing and testing innovative solutions

Servicing industry and SMEs

Spinning out technologies

# DRIVERS for change

- **BIG DATA**
- **MORE COMPUTING POWER**
- **GLOBAL CONNECTIONS**
- **GLOBAL PARTICIPATION**
- **OPEN IS BETTER**
  - **WITHIN AND BETWEEN  
SCIENTIFIC COMMUNITIES**
  - **BETWEEN SCIENCE AND SOCIETY**





# GLOBAL CONNECTIONS ...



## data was, is and remains key to science

Need for "**expensive apparatus**" is something that modern science intensified (need for more powerful telescopes Satellites for Earth Observation, light sources, research boats, wind tunnels, geological probes etc)

Intrinsic to the ambition that European researchers remain at the vanguard of scientific discovery

But there is something about research data:

**information opens new worlds for science**

## research logic machines

**Research Data** collected at observation or experimentation phase were registered in the **scientists notebooks**, which used to be paper books

Now research data is stored in digital form. Easier to be processed by "**logic machines**" programmed with complex models able to dig into the data

Logic machines are made of **human scientific knowledge and creativity, software** and the underlying **hardware**

Scientist notebooks can now be **linked** to a huge amount of other **data resources** (including scientific papers), **computers** with unprecedented capacity, eventually connected to **global networks**





## A Reinforced European Research Area Partnership for Excellence and Growth, COM(2012) 392 – July 2012

## Towards better access to scientific information: boosting the benefits of public investments in research, COM(2012) 401 final - July 2012

## Commission, Recommendation on access and preservation of scientific information, C(2012) 4890 final – July 2012

## Horizon 2020

- Open Access to Scientific Publications
- Pilot on research data: Data Management Plan



## useful definitions

**Data:** digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings

(not include lab notebooks, preliminary analysis, drafts of scientific papers, plans for future research, peer review reports, communication with peers, physical objects, lab specimens)

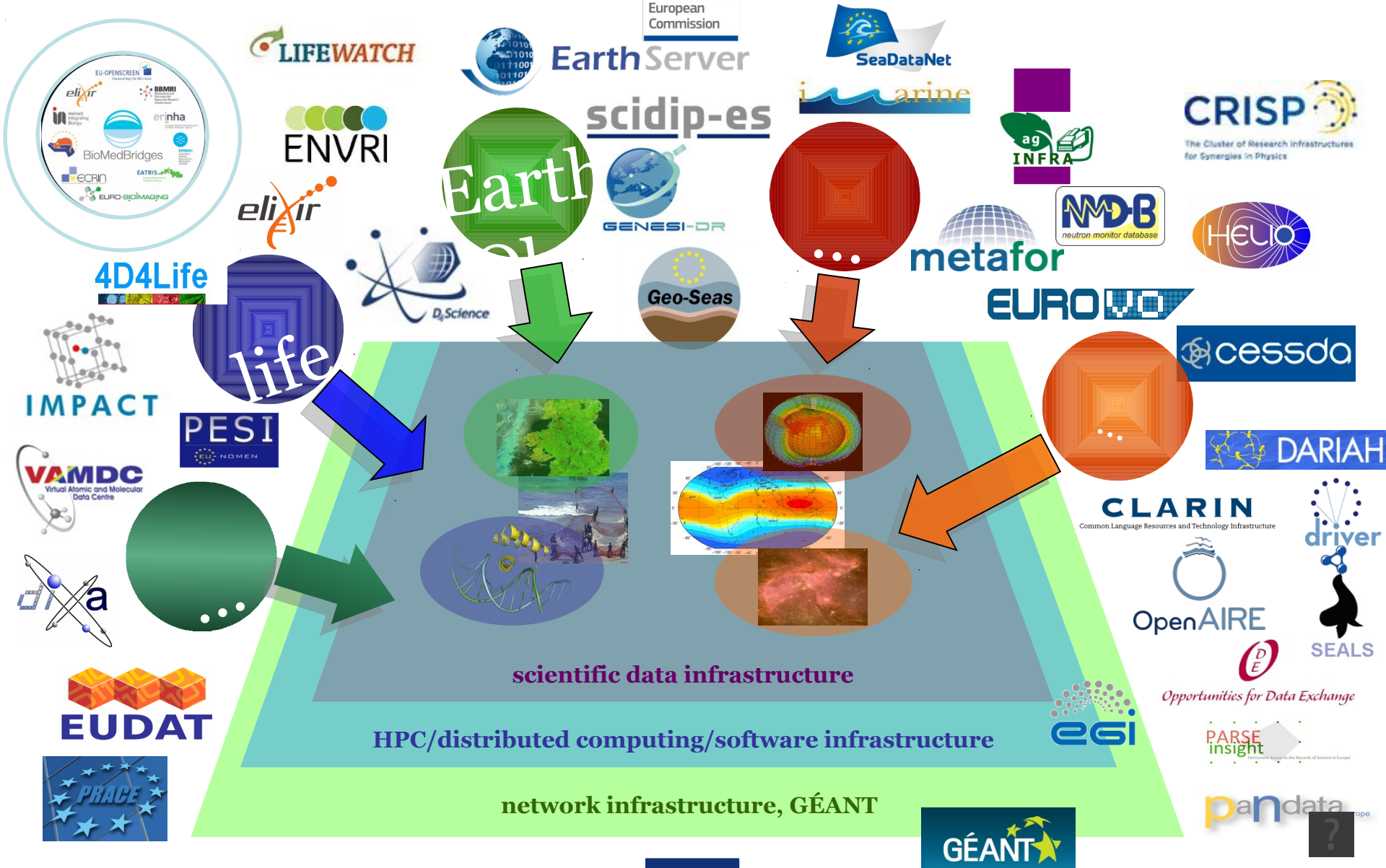
*[c.f. White House Memo on "Increasing Access to the Results of Federally Funded Scientific Research"]*



**Data infrastructures:** services, applications, tools, knowledge and policies for research data to be discoverable, understandable, accessible, preserved and curated... and available 24/7

# e-infrastructure

# building bridges





## issues to be addressed (e-infrastructure)

The EC in coordination with EU Member States is looking after research data as an infrastructure

As a valuable and a strategic resource, research data opens at least three key issues to be addressed<sup>(\*)</sup>:

- **How data can be networked**
- **How to envision and set up data governance on a global scale**
- **How the EU can play a leading role in helping start and steer this global trend**

*(\*) Fred Friend, Jean-Claude Guédon Herbert van Sompel  
"Beyond Sharing and Re-using: Toward Global Data Networking"*



- RDA community focuses on building **social, organizational and technical infrastructure** to
  - **reduce barriers to data sharing and exchange**
  - **accelerate the development of coordinated global data infrastructure**



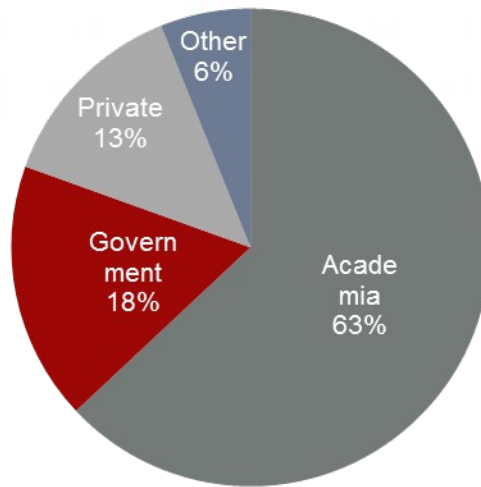
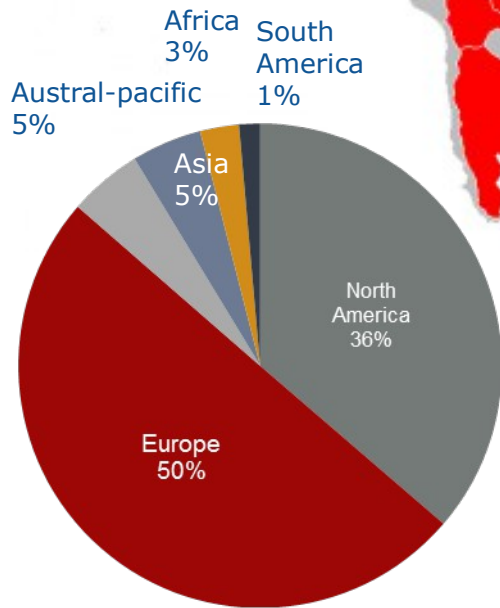
## CREATE $\rightleftharpoons$ ADOPT $\rightleftharpoons$ USE

### RDA Working Group Infrastructure Deliverables are:

- **Focused pieces of adopted code, policy, infrastructure, standards, or best practices** that enable data to be shared and exchanged
- **“Harvestable” efforts** for which 12-18 months of work can eliminate a roadblock for a substantial community
- **Efforts that have substantive applicability** to “chunks” of the data community, but may not apply to everyone
- **Efforts for which working scientists and researchers can start today** while more long-term or far-reaching solutions are appropriately discussed in other venues



Map courtesy [traveltip.org](http://traveltip.org)



## Distribution of 2,353 Individual RDA Members in 96 Countries

12 September 2014





## take five



**5 principles** describing the benefits of a global research data infrastructure (G8+O6)

Publicly funded research data is:

**Discoverable** – IDs, Descriptive Metadata, ...

**Accessible** – Acknowledgment, License, Terms of Use, Intellectual Property, Legal ...

**Understandable** – Semantics, Analysis, Quality, Language translation ....

**Manageable** – Responsibility, Costs, Preservation ...

**People (Usable)** - Workforce, Cultural, Training, ...



# **The Data Harvest Report**

## **How sharing research data can yield knowledge, jobs and grow**

### **A RDA Europe Report**

The Data Harvest, December  
2014 © RDA Europe





**Wim Jansen**  
**[Wim.Jansen@ec.europa.eu](mailto:Wim.Jansen@ec.europa.eu)**

**Thank you!**