views on research data and RDA

Potsdam, 20 November 2014

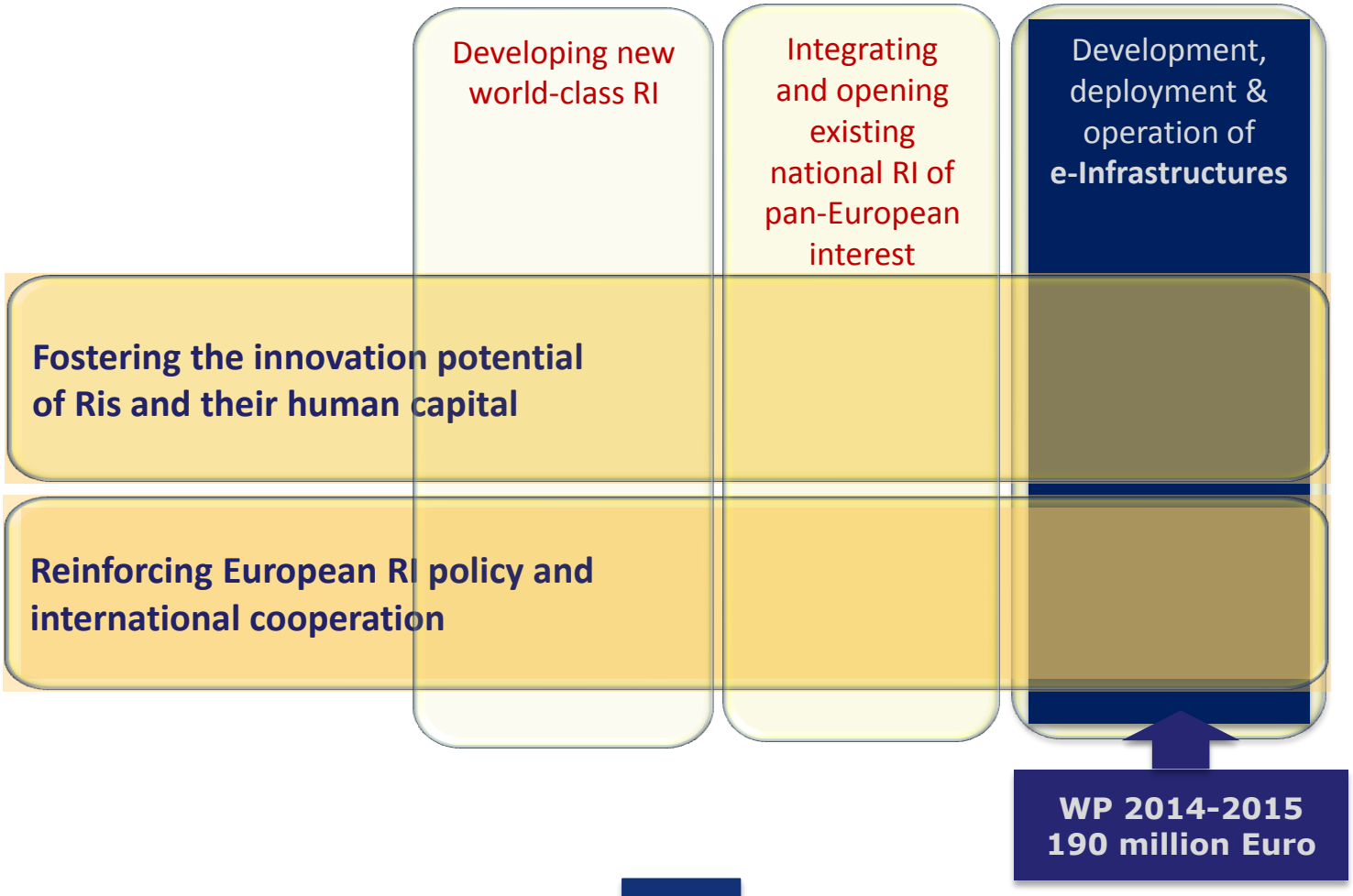
Carlos Morais Pires
carlos.morais-pires(at) ec.europa.eu
European Commission
e-Infrastructures, DG CNECT.C1

Author's views do not commit the European Commission

summary

- **Research Data Alliance**
  - **funders colloquium**
  - **principles (link with G8 – opportunity to repurpose at Rome event)**
- **Report "Data Harvest" (follow-up of "Riding the Wave")**
  - **strategic relevance of research infrastructures and e-infrastructures: moving, processing and managing data**

e-infrastructure/research infrastructure in H2020



data has been and remains key to science

Need for research infrastructures is something that modern science intensified (need for more powerful telescopes, light sources, research boats, geological probes etc)

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

But there is something about research data:

information/data opens new opportunities 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**

context: open science

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 - July2012

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

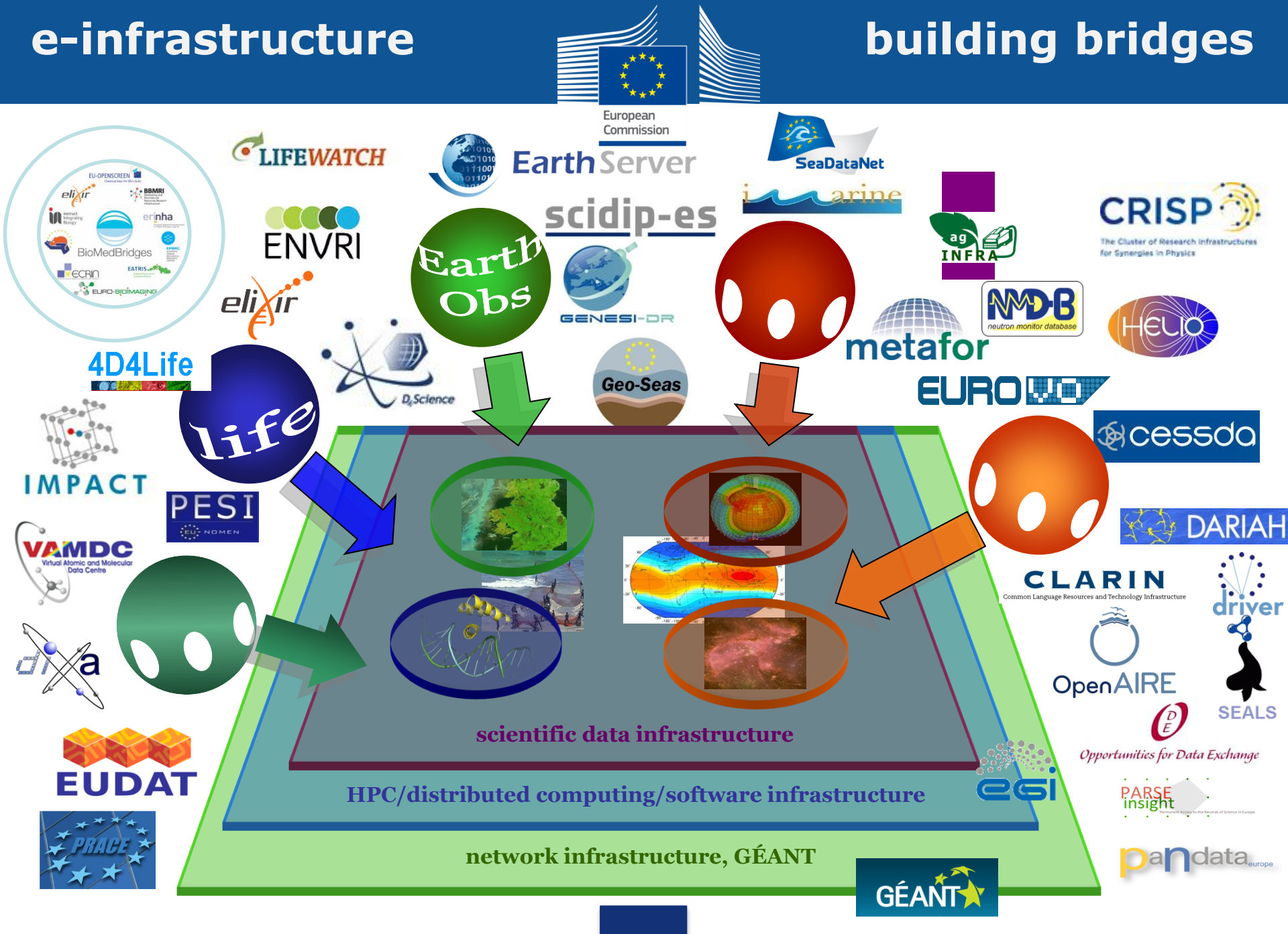
Data: digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings

(*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



Implementation cluster for Environment

ENVRI

EU funding : 3.7 M€, started in 2011

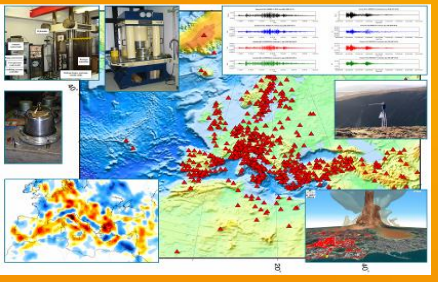
- Development of common reference model , standards, and common components for data pre-processing and post-processing
- Contribution to GEOSS (Global Earth Observation System of Systems) and compliance with INSPIRE EC Directive
- Large participation of ICT and e-infrastructures actors (key partners from D4SCIENCE, GENESI, EGI, EUDAT, PRACE…)

EURO-ARGO

SIOS

LIFE-WATCH

EPOS



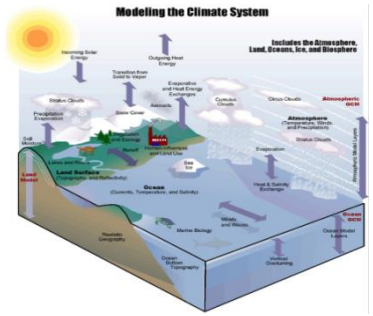
ICOS

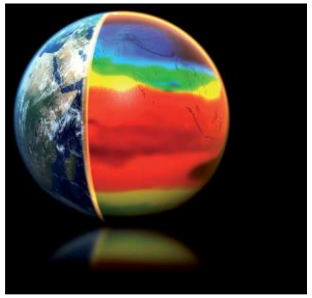
EISCAT

EMSO

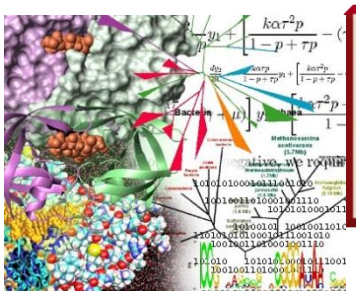
HPC

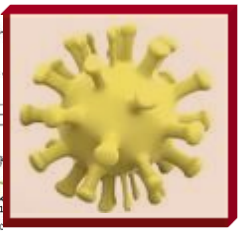
European Commission

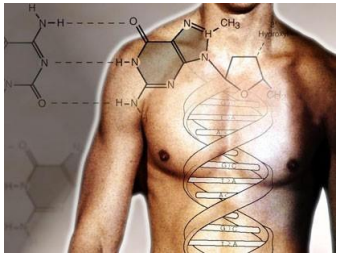




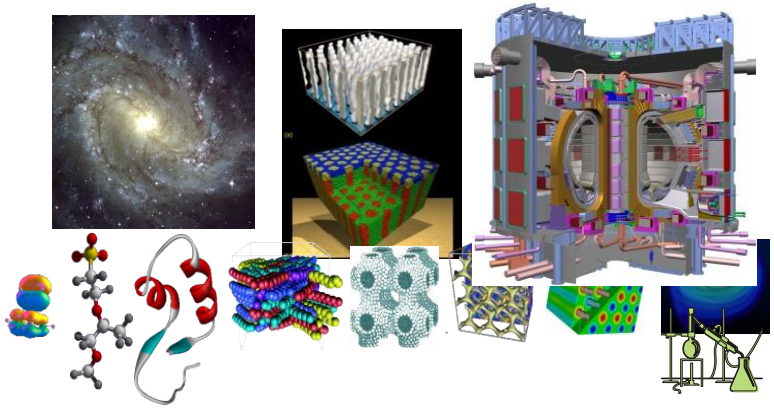
Weather, Climate & Earth Sciences



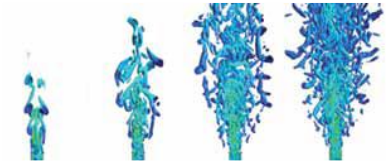


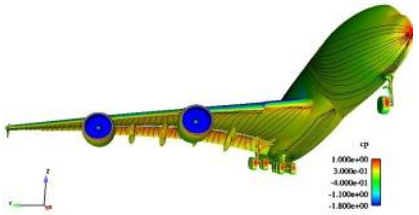


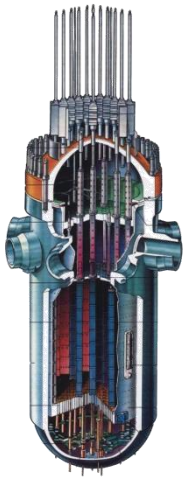
Bio/Life Sciences

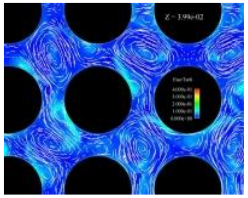


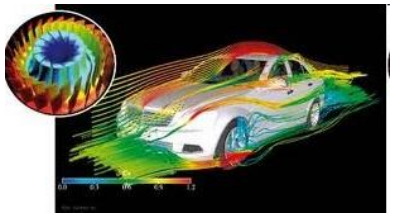
Fundamental sciences: Physics, Chemistry, Material Sciences, Astrophysics











Industrial & Engineering

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**(*):

- How data can be **linked/networked**

  (discoverable, accessible, processed…)

- How to envision and set up data **governance** on a global scale

  (shared, re-usable, preserved,…)

- How the EU can **play a leading role and steer**

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

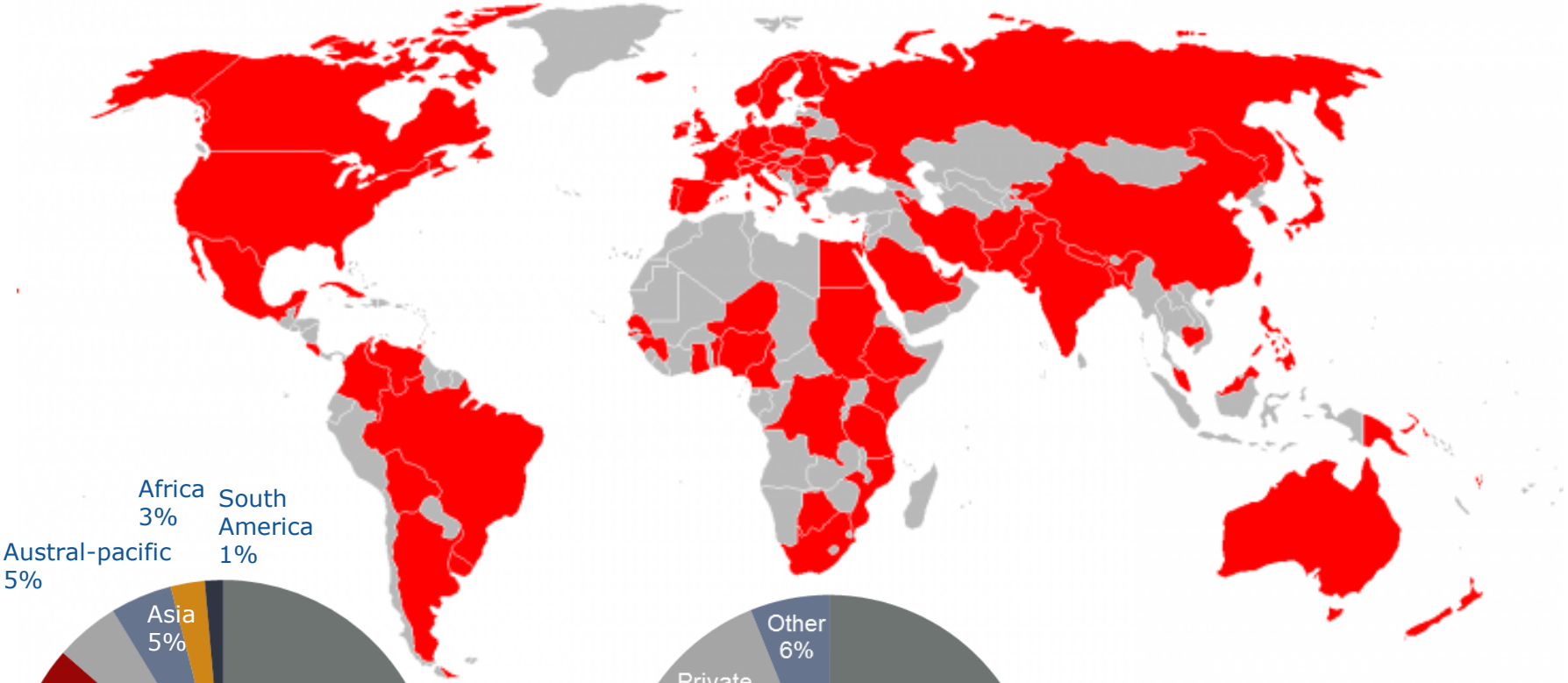
Research Data Alliance: Europe as a Global Partner

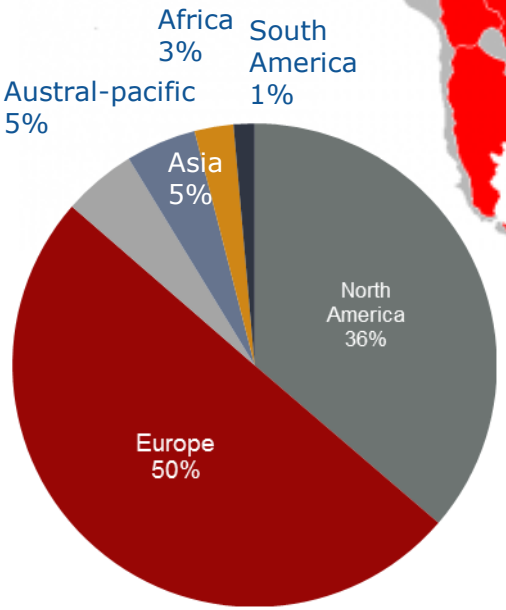
Societal challenges of our time **transcend borders**

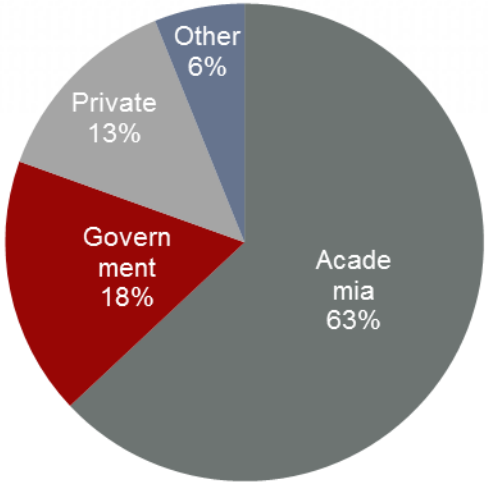
Data and computing intensive **science** is made of **global collaborations**

Research data are **global**

Research Data Alliance: enable data exchange at global scale





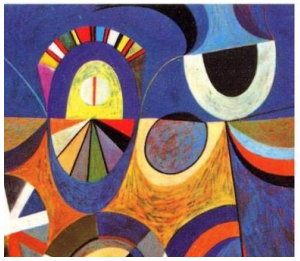


Map courtesy traveltip.org

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, ...

final remarks

RDA was thought to solve a well identified difficulty

RDA came from a real need ("*desperate need*")

RDA got almost immediate support from the community

RDA is a very good idea

RDA comes at the right time

RDA was able to get a first shape thanks to many of you here…

… but RDA **is not there yet**!

I would like to **thank you** for the support and ask you to please **continue supporting RDA**!

Thank you!