data infrastructures

Barcelona, 28 January 2015

Carlos Morais Pires
European Commission
e-Infrastructures, DG CNECT.C1

selling 3 products

- **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")**

Research Infrastructures

"Men of science […] could formerly work in isolation as writers still can.

Cavendish and Faraday and Mendel depend hardly at all upon institutions and Darwin only in so far as the government enabled him to share the voyage of the Beagle.

But this isolation is a thing of the past.

*Most research requires **expensive apparatus** […]. Without facilities provided by a government or a university, few men can achieve much in modern science."*

from Bertrand Russell in BBC Reith Lectures,1949

data has been and remains key to science

Need for "**expensive apparatus**" 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 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 - 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: <u>Data Management Plan</u>**

Open Science for the 21st century

A declaration of ALL European Academies

presented at a special session
with Mme Neelie Kroes, Vice-President of the European Commission,
and Commissioner in charge of the Digital Agenda

on occasion of the ALLEA General Assembly held at
Accademia Nazionale dei Lincei, Rome, on 11-12 April 2012

EUROPEAN COMMISSION

Brussels, 17.7.2012
COM(2012) 401 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE AND THE COMMITTEE OF THE REGIONS

Towards better access to scientific information:
Boosting the benefits of public investment in research

European Commission

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)

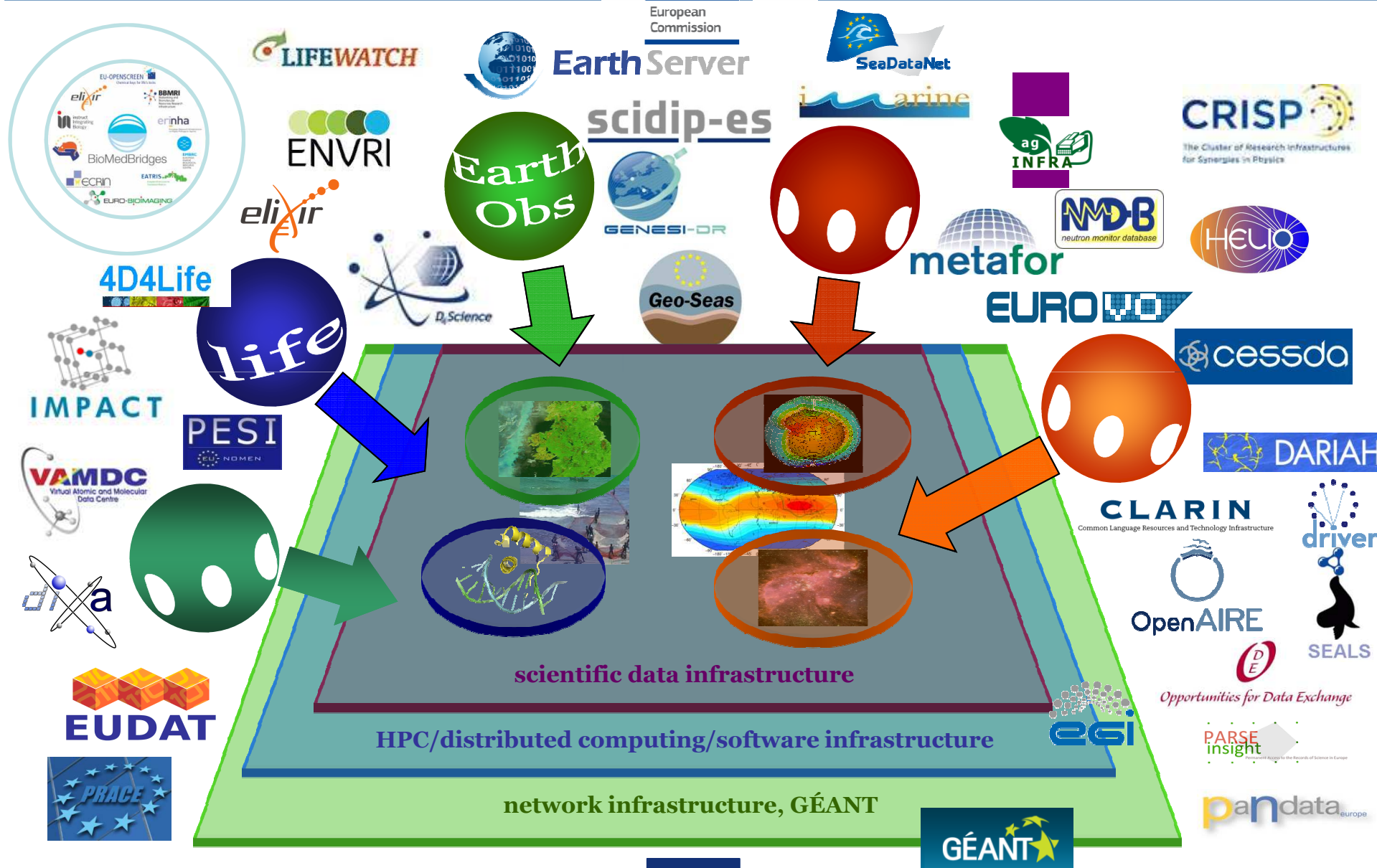*[c.f. White House Memo on "Increasing Access to the Results of Federally Funded Scientific Research"]*

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

European Commission

LIFEWATCH   EarthServer   SeaDataNet   CRISP The Cluster of Research infrastructures for Synergies in Physics

EU-OPENSCREEN   elixir   BBMRI   scidip-es   i marine   ag INFRA

instruct Integrating Biology   erinha   ENVRI   Earth Obs   GENESI-DR   NMDB neutron monitor database   HELIO

BioMedBridges   EATRIS   elixir   D4Science   Geo-Seas   metafor   EURO VO

ECRIN   EURO-BIOIMAGING

4D4Life   life   cessda

IMPACT   PESI EU-NOMEN   DARIAH

VAMDC Virtual Atomic and Molecular Data Centre   CLARIN Common Language Resources and Technology Infrastructure   driver

dixa   OpenAIRE   SEALS

scientific data infrastructure   Opportunities for Data Exchange

EUDAT   HPC/distributed computing/software infrastructure   egi   PARSE insight Permanent Access to the Records of Science in Europe

PRACE   network infrastructure, GÉANT   pandata europe

GÉANT

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 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"

Research Data Alliance
Research Data Sharing

- 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**
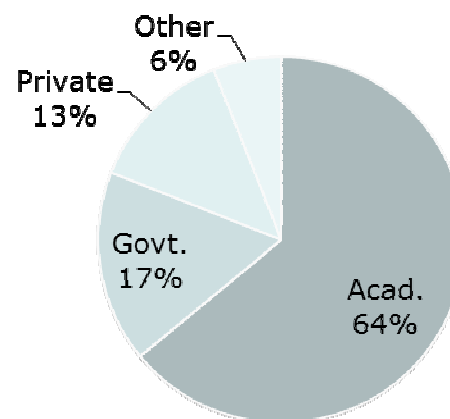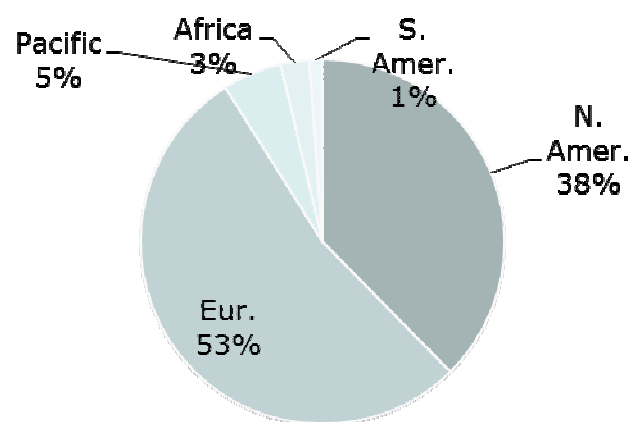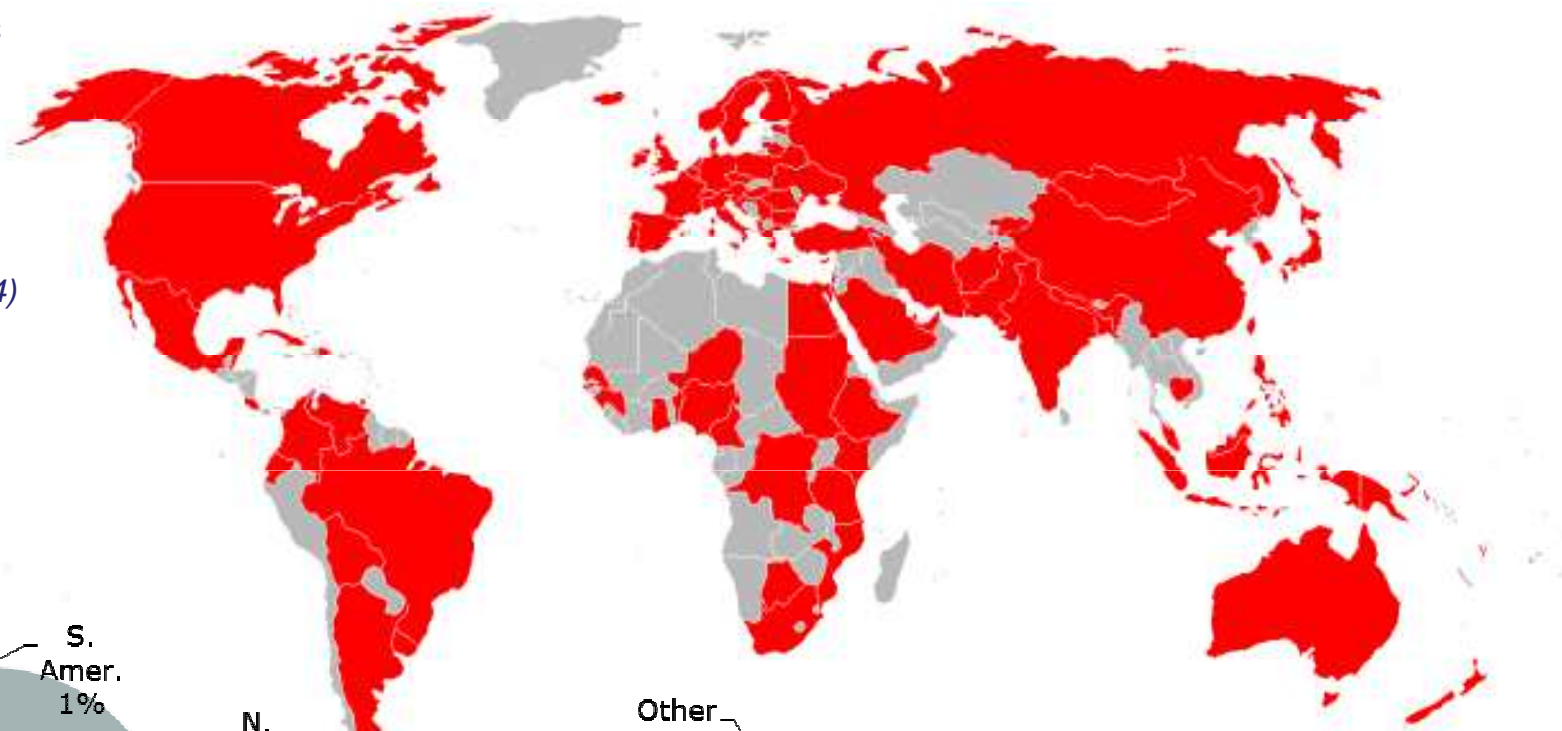

Plenary 2
Washington, DC

CREATE → ADOPT → 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

- Distribution of
  2,538 RDA
  Members
  in
  92 Countries
  (3 December 2014)

Pacific 5% · Africa 3% · S. Amer. 1% · N. Amer. 38% · Eur. 53%

Other 6% · Private 13% · Govt. 17% · Acad. 64%

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

Carlos Morais Pires
carlos.morais-pires(at) ec.europa.eu

Thank you!