e-infrastructures enablers of user-centric innovation

Copernicus Climate Data Store Workshop
Brussels-Reading video link, 3 March 2015

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

summary

1 - e-infrastructures

2 - user-centric innovation

3 – making it happen

1 - e-infrastructures

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**

stakeholders

- **Scientists and Research Orgs** want the best research (education)

  access to the best research facilities and instruments, to the best information from peers

- **Industry** wants to be competitive

  leverage on new knowledge and know-how, develop and test new products

- **Infrastructure operators** want to deliver the best services

  respond to needs of scientists, integrate state-of-the-art technology in the mainstream infrastructure, explore innovative tech for the long-term needs

- **Funders** want to support scientific discovery and innovation

  sustainable investments, promoting excellence and inclusiveness

mission and objectives

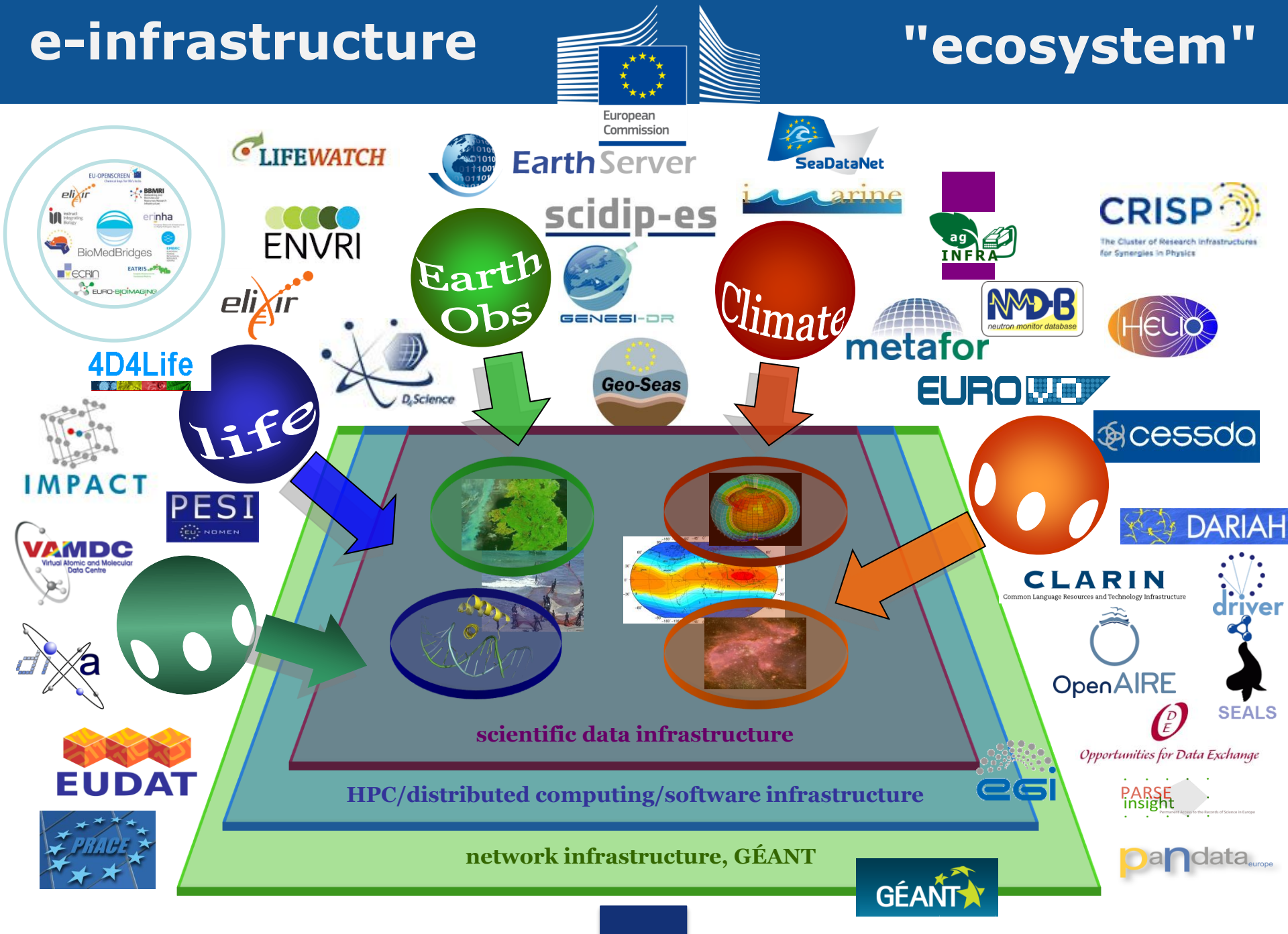
Develop infrastructure capacity to foster European excellence in science

In particular for e-Infrastructures

- **Bring ICTs to all researchers, educators, students**
- **Enable e-Science (data and computing-intensive science)**
- **Support Open Science**

2 - user-centric innovation

e-infrastructure "ecosystem"



user-centric e-infrastructure innovation

- User-driven design and prototyping of innovative e-infrastructure services and applications
- Meet the scientific needs of those communities that push the envelope in scientific and technological domains requiring top-of-the range capacity in the long term
- Open e-infrastructure resources to pan-European initiatives
- E-infrastructures should enable fast prototyping and development of innovative networking, data and computing intensive application and services promoting adaptation, extension and repurposing of basic services

user-centric e-infrastructure innovation

- Accelerate the development of innovative data and computing intensive services in areas of policy relevance for Europe such as Health, Environment monitoring and management
- Foster the use of open e-infrastructures eco-systems to innovative use promoting smooth collaboration among and between the large European Policy data intensive initiatives
- Bridge the gap between adjacent but not connected scientific communities and promote wide dissemination of data including to the citizens engaged in science
- Support collaboration in data provision and exchange across regional and national related infrastructures allowing the integration of data from a myriad of resources and research communities

3 – making it happen

take 5

5 principles describing benefits of global research data infrastructure

Publicly funded research data is:

- **Discoverable** – IDs, Descriptive Metadata...
- **Accessible** – Citation, License/Terms of Use, Intell. Property, Legal...
- **Understandable (Assessable)** – Semantics, Analysis, Quality, Language...
- **Manageable** – Responsibility, Costs, Preservation...
- **People (Usable)** - Workforce, Cultural, Training...

responding to the 5 Principles

Discoverable: are the data and associated software discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)

Accessible: are the data and associated software accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)

Assessable, and intelligible: are the data and associated software assessable and intelligible by third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review?)

Useable beyond the original purpose for which it was collected: are the data and associated software useable by third parties even long time after the collection of the data (e.g. are the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation?)

Interoperable to specific quality standards: are the data and associated software interoperable allowing data exchange between researchers, institutions, organisations, counties, (e.g. adhering to standards for data annotation, data exchange, etc.)

Research Data Alliance

Create - Adopt - Use

RDA focuses on building social, organizational and technical infrastructure to

reduce barriers to data sharing & exchange

accelerate the development of global data infrastructure

http://rd-alliance.org/

Plenary 2 Washington, DC

Focused pieces of adopted code, policy, infrastructure, standards, or best practices enabling data to be shared and exchanged

"Harvestable" efforts for which 12-18 months of work can eliminate a roadblock for a substantial community

Efforts with substantive applicability to "chunks" of the data community, but may not apply to everyone

Scientists and researchers can adopt today while more long-term or far-reaching solutions are discussed in other venues

slide adapted from RDA Council

final remarks

The European Commission is interested in the sustainable operation and development of innovative e-infrastructures services and in supporting open science

The EC tools for stakeholders to be involved are:

- Open Science Framework
- Horizon 2020 workprogramme

The EC supports European research and e-infrastructures communities to partner globally to reach interoperability and foster collaboration

Thank you!

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