

Scientific Data e-Infrastructures: supporting the transition to e-Science

PolSCA Meeting
28 May Brussels

*Carlos Morais Pires
Head of Sector INFSO F3
European Commission*

e-Infrastructures underpinning a creativity machine...

“We humans have built a creativity machine. It’s the sum of three things: a few hundred million of computers, a communication system connecting those computers, and some millions of human beings using those computers and communications.

This creativity machine is **the Internet.**”

Vernor Vinge (*Journal Nature, March 2006*)

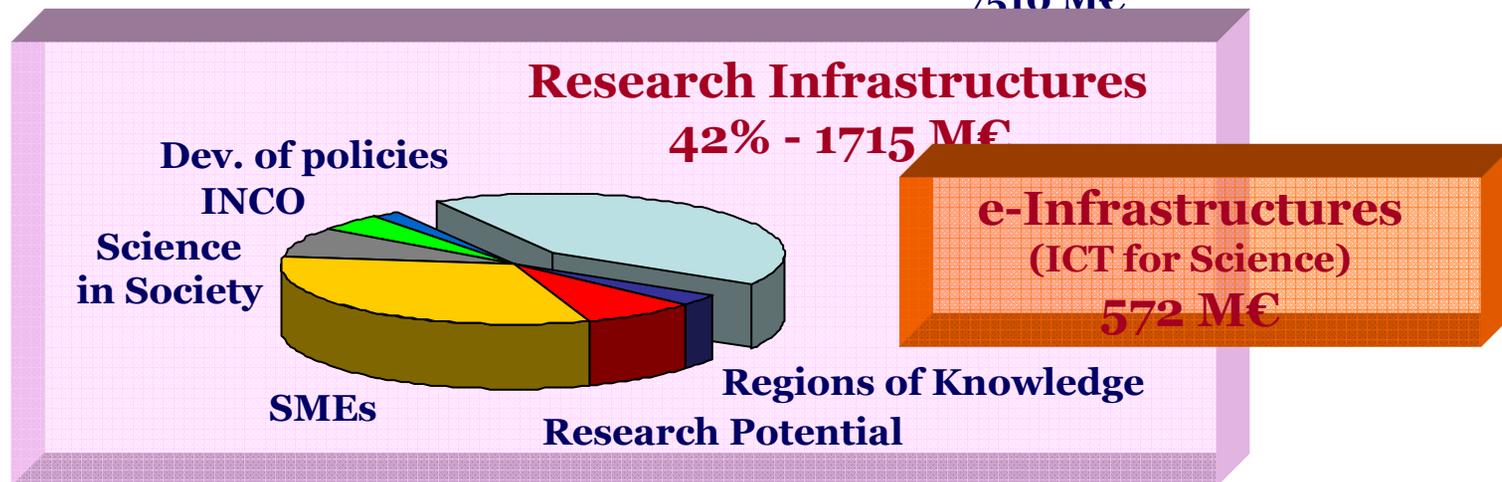
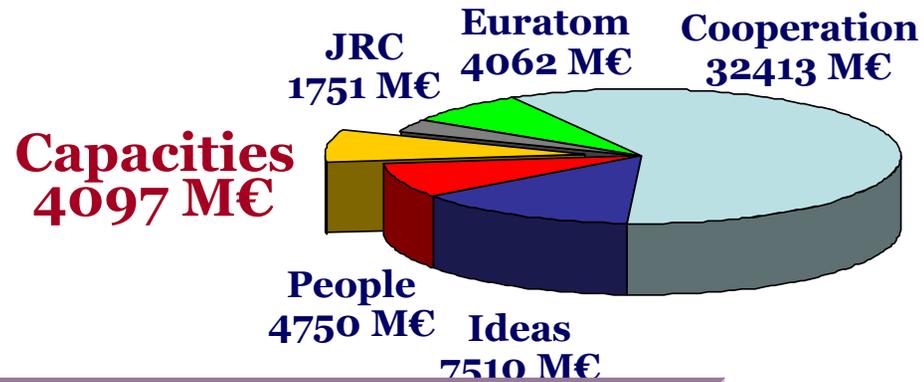


European Commission
Information Society and Media

Science and ICT

- Scientific advances more important than ever
 - Global challenges with high societal impact
 - Innovation and economic development
- Adoption of ICT changes the scientific discovery process
 - Computing, simulation and data
 - Tackling the very small, the very big and the very complex
 - Cost efficiency
 - Open, cross-border and cross-discipline collaboration

EU R&D programme: budget split (FP7: 2007 - 2013)



e-Infrastructures for science

ubiquitous research environments
accessing and sharing resources and tools



e-Infrastructures in action today



Innovating the scientific process:
global virtual research communities



Accessing knowledge:
scientific data



Experimenting *in silico*:
simulation and visualisation



Sharing the best computational resources:
e-Science grid, supercomputing



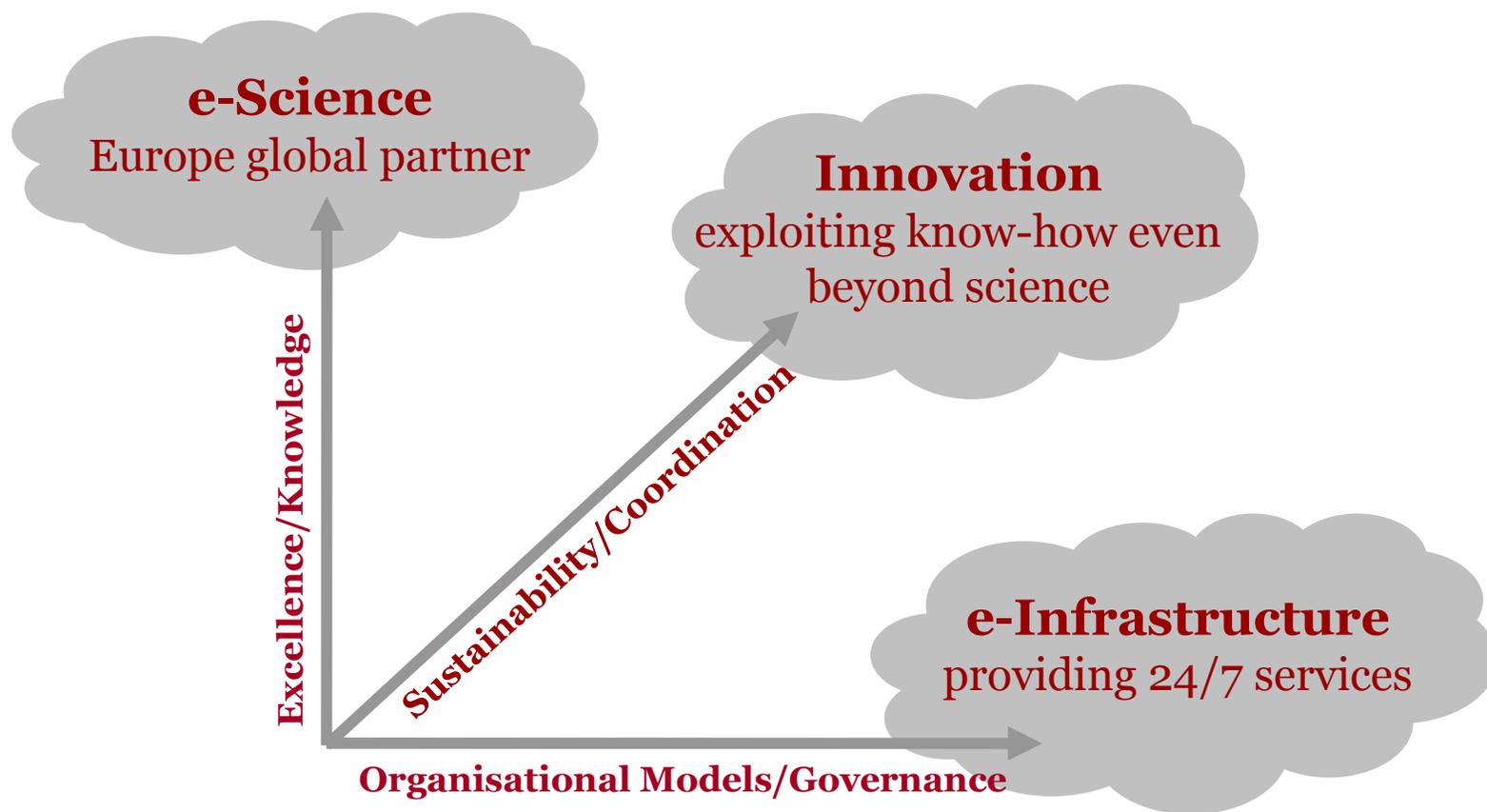
Linking at the speed of the light:
GÉANT

ICT infrastructures for e-Science: a Communication to European Institutions COM(2009) 108

- Highlighting the importance of embracing the **e-Science** paradigm shift
- Highlighting the strategic role of **e-Infrastructures** as a crucial asset underpinning European research and innovation policies
- Calling on Member States and the scientific communities, in cooperation with the European Commission, for a reinforced and coordinated effort to further develop world class e-Infrastructures

ICT infrastructures for e-Science

COM(2009) 108



e-Science workspaces



Data/Visualisation

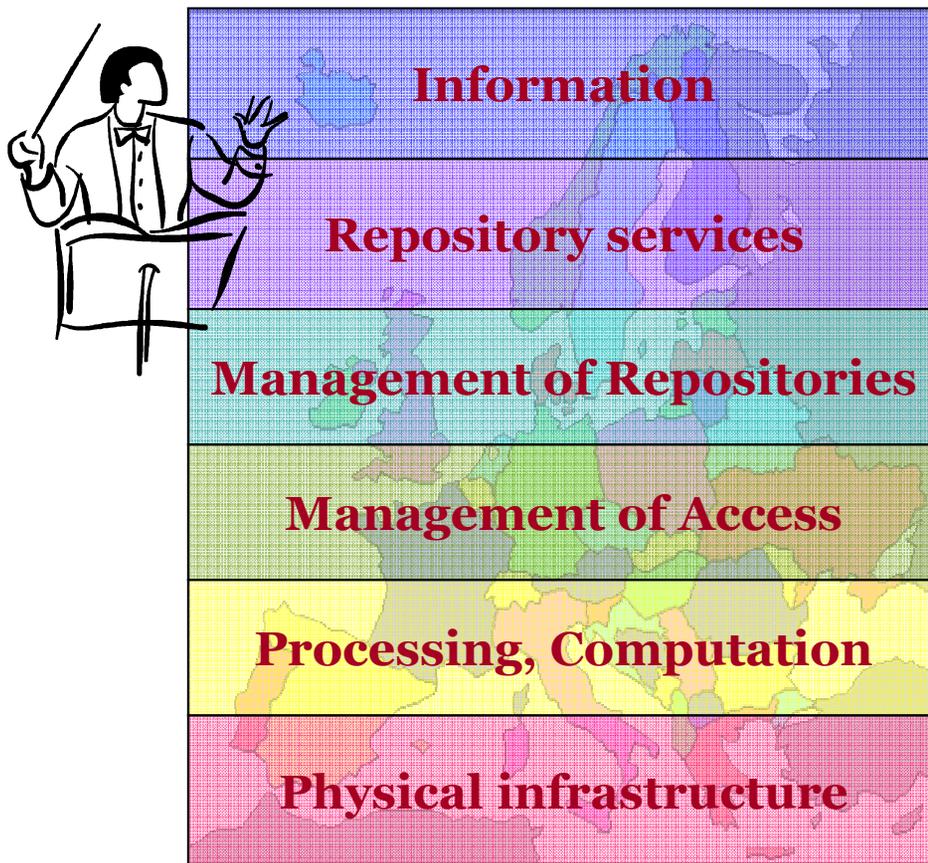
Computer/Simulation

Networking/Connectivity

**Economies
of Scale**

**Efficiency
Gains**

Orchestration within the e-Infrastructure: need for coordination all the elements and layers



- In the Communication to the Council and EP, the Commission asks the involvement of Member States and key stakeholders to build robust, dynamic and innovative e-Infrastructures for scientific data
- This cooperation started already by launching 15 projects, including the FP7 Open Access Pilot (40 Mio Euro)
- There is still a long way to go...

On Scientific Information in the Digital Age: Access, Dissemination and Preservation

COM(2007) 56

- Already in the Communication "On Scientific Information in the Digital Age: Access, Dissemination and Preservation"

the European Commission has pointed out that building in Europe a dynamic information society requires providing wide access and ensuring long term preservation of scientific information

the communication stresses that **"The Internet [...] opened new ways to use masses of data resulting from experiments and observations in the scientific process and to extract meaning from this data stored in repositories in combination with other scientific information resources. This leads to a 'continuum' of scientific information space from raw data to publications across different communities and countries"**.

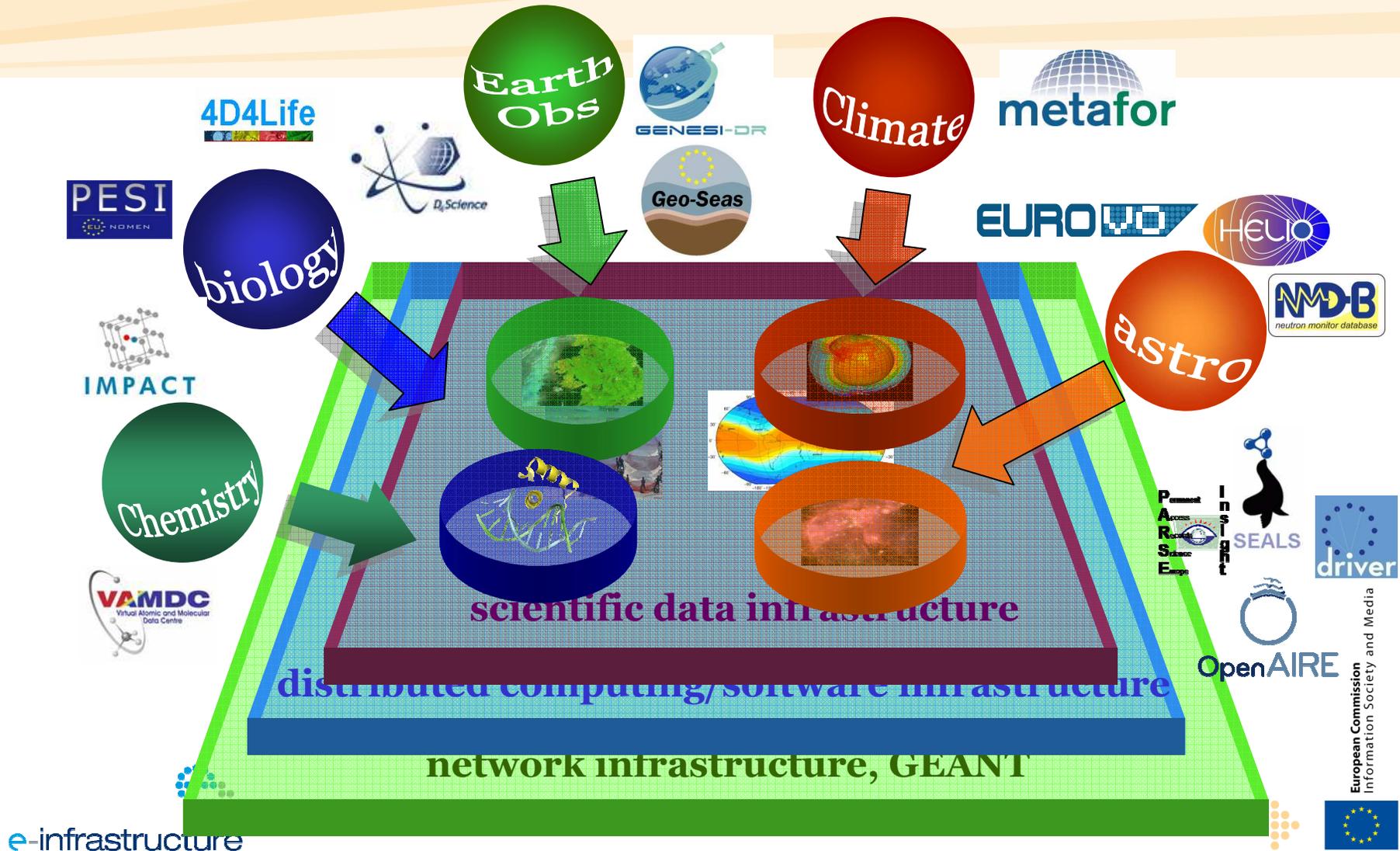
Information deluge...

The animals boarding Noah's Ark two by two by two painting by the American Edwards Hicks (1780-1849)



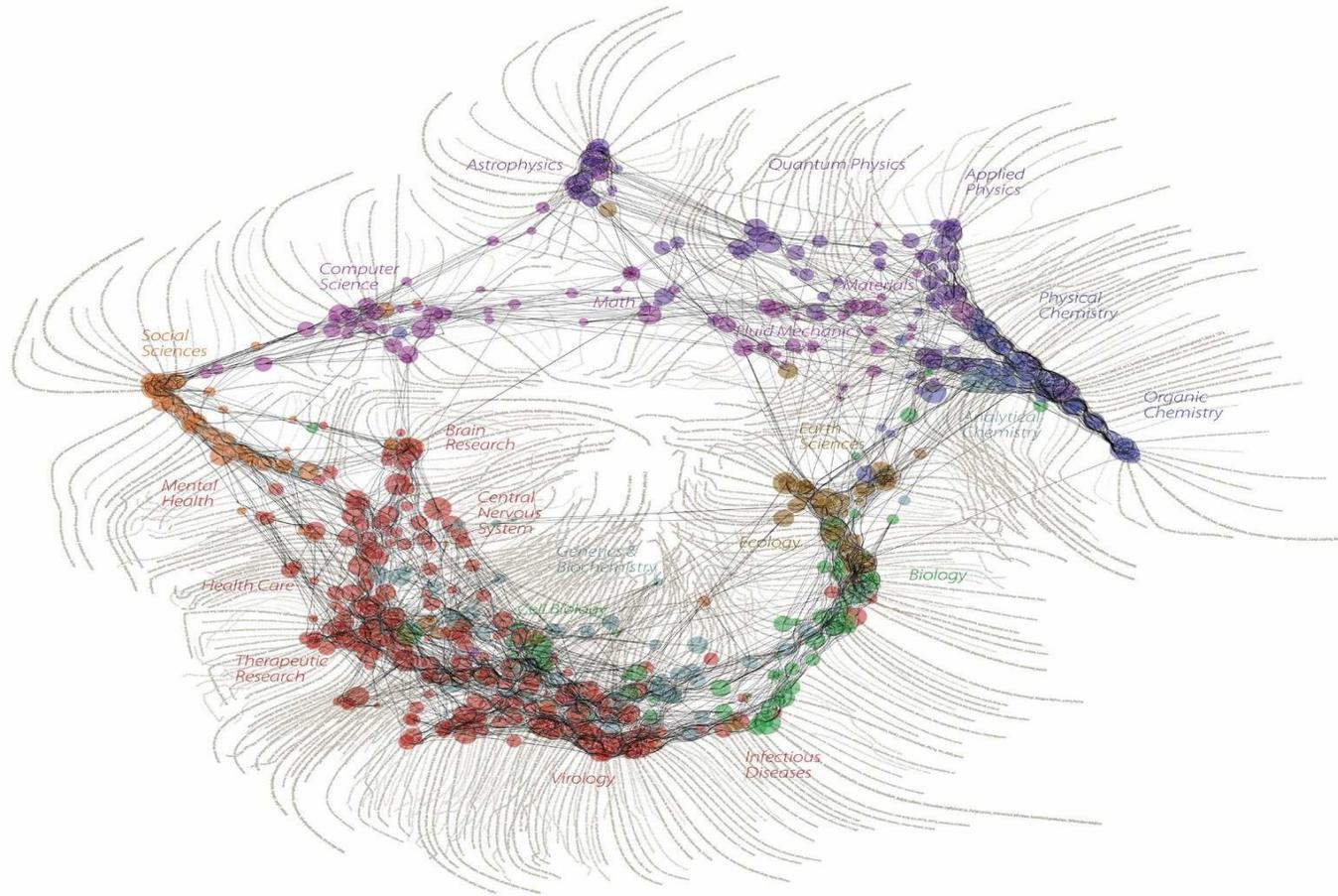
- European Universities and Labs producing more and more Data (databanks at EBI, 'LHC at CERN, ESA, ...)
- Biodiversity data is a very good example
- Availability of Information is at the basis of Knowledge creation
- Need for e-infrastructure addressing Scientific Information however, each 'Science' has different requirements:
 - Access modalities to papers and data vary across communities
 - Identification of researchers as "authors"
 - Identification of Digital Objects to allow location, discoverability, sharing and use
 - Allocation of the responsibility of long-term preservation

Scientific Data Infrastructure



The 'map of science'

Journal Nature (Dec 2006): This map was constructed by sorting roughly 800,000 published papers into 776 different scientific paradigms (shown as pale circular nodes) based on how often the papers were cited together by authors of other papers. [...]



e-infrastructure for OA pilot

- FP7 Open Access pilot follows the Conclusions of the EU **Competitiveness Council** of 22 and 23 November 2007
- Its key objective is to enable fast and reliable access to EU-funded research results, in particular **peer reviewed research articles**
- Projects from **7 areas of FP7** are required to deposit articles and make their best effort to ensure OA after an embargo period
- **OpenAIRE project** was selected in an open call to provide an e-infrastructure supporting mechanisms for the identification, deposit, access, and monitoring of articles

Scientific Data - Looking ahead



We need to exploit the growing sensor/effector layer to make the world itself a real-time database.

(from the creativity machine, V. Vinge)

- “Big, complex data-intensive science” of global dimension is here to stay; hence the increasing value of observational and experimental data in virtually all fields of science
- Europe pays particular attention to the aspects of accessibility to scientific information, its quality assurance and preservation
- Multi-disciplinary approaches, new participative paradigms and global research communities are an essential part and driver of the strategy
 - ...but organisational, governance and financing models need attention, informed by sociological, political and cultural considerations

e-SciDR study draft recommendations

Build an e-Infrastructure which ensures “research continuity”

Funding, Governance and management, leverage on other e-Infrastructure layers

Engage users and service providers

Support for data producers, Trust and recognition, Training and awareness

Provide access to researchers, educators and students

Discovery and navigation, OA to publicly funded data, International collaborations

Maintain and preserve information

Collections management, selection and appraisal for sustainability

source: eSciDR study (adapted)

e-Infrastructures – future directions

- “Data’s shameful neglect” (Nature, 10 September 2009)
- Developing an ecosystem of European Digital Repositories, federating and adding value to national or discipline-based repositories will be necessary
- A new call to be open 3Q 2010...
 - workprogramme and priorities of the call still under discussion with the Member States

For further information



www.cordis.europa.eu/fp7/ict/e-infrastructure/

