e-Infrastructures

supporting open knowledge circulation

Carlos Morais Pires

Head of Sector for Scientific Data e-Infrastructures
European Commission
Information Society and Media
INFSO/F3

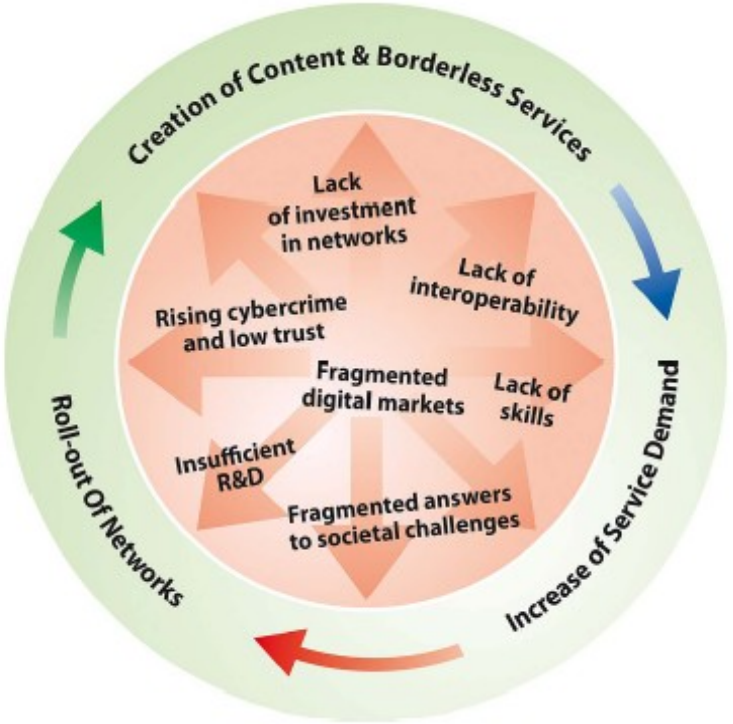
email: carlos.morais-pires **at** ec.europa.eu

European Commission
Information Society and Media

Disclaimer: "The views expressed in this presentation are those of the author and do not necessarily reflect the views of the European Commission"

Policy Context

Digital Agenda
1001100101011101110000100 2010-2020
for Europe



The **DAE** is one of the flagships of "Europe 2020: a strategy for smart, sustainable and inclusive growth".

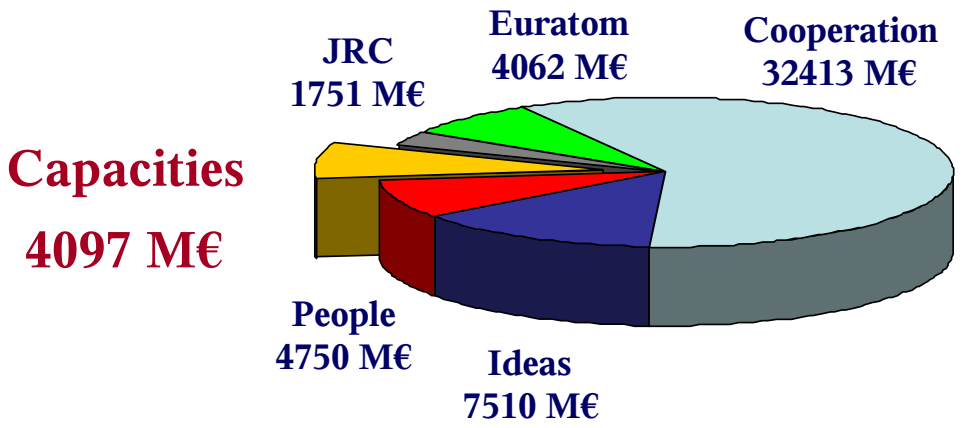
R&I are priorities of DAE

"The Digital Agenda for Europe outlines policies and actions to maximise the benefit of the digital revolution for all. Supporting **research and innovation** is a key priority of the Agenda, essential if we want to establish a flourishing digital economy."

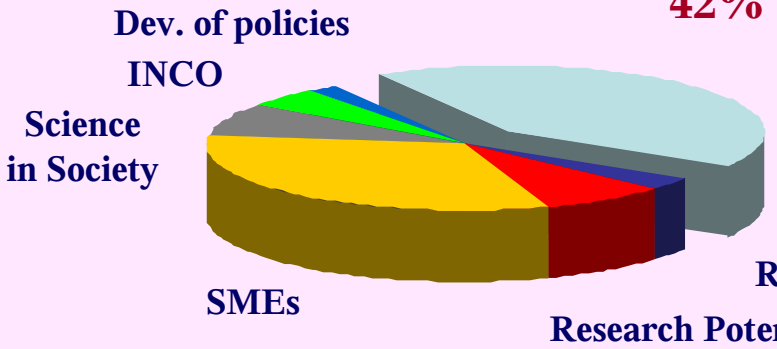
Neelie Kroes,

Vice-President of the EC, responsible for the Digital Agenda

European R&D Framework Programme
FP7 (2007 to 2013)



Research Infrastructures
42% - 1715 M€



e-Infrastructures
(ICT for Science)
572 M€

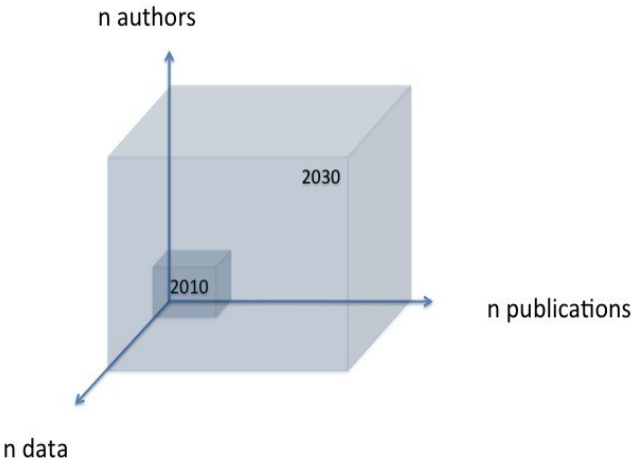
A world in transformation

- One of the factors of change is technology
  - just like in the past
- Internet (instantaneous communication)
  - the power in the edges/nodes of the network
- Miniaturisation (pervasiveness)
- Virtualisation (information/data)

Science in transformation

- More intense collaborations
  - between scientists, between machines
  - across disciplines
- ICT Infrastructures enabling e-Science
  - drivers of social transformations
- **Information becomes an infrastructure**

Science in transformation



- Science 2.0
  - main trends (figure 1)

*"[…] The **data availability landscape transforms** because of interrelated major trends. The cost for accessing data has dramatically lowered: much of the useful statistics and more general data from (often publicly funded) research are now published and freely accessible in raw format on the web. […] much more data collected and archived today than ever before, and the **volume is growing at an exponential rate** […]"*

reference: Science 2.0 (change will happen….) J.C Burgelman, D. Osimo & M. Bogdanowicz

- Scientific Data
  - Information cycle/continuums
  - Costs (associated with quality)
  - Roles and tensions between "today's" stakeholders
  - Scenarios for the future
  - … a New World

Meeting 21st Century Challenges

- strategic to embrace the e-Science paradigm shift and the strategic role of e-Infrastructures as a crucial asset underpinning European research and innovation policies
- e-Science benefiting from pervasive technologies for high-speed communication and information processing
- Science is global: 35% of articles in leading journals result from international collaboration (that was 25% 15 years ago)
- Information infrastructures are key enablers of e-Science

e-Infrastructures supporting Science

e-Infrastructure is “an environment where research resources (hardware, software and content) can be readily shared and accessed wherever this is necessary to promote better and more effective research”.

Such environment integrates networks, grids and middleware, computational resources, experimental workbenches, data repositories, tools and instruments and the operational support that enable global virtual research collaborations.

Source: ICT infrastructures for e-Science

Scientific Data e-Infrastructures

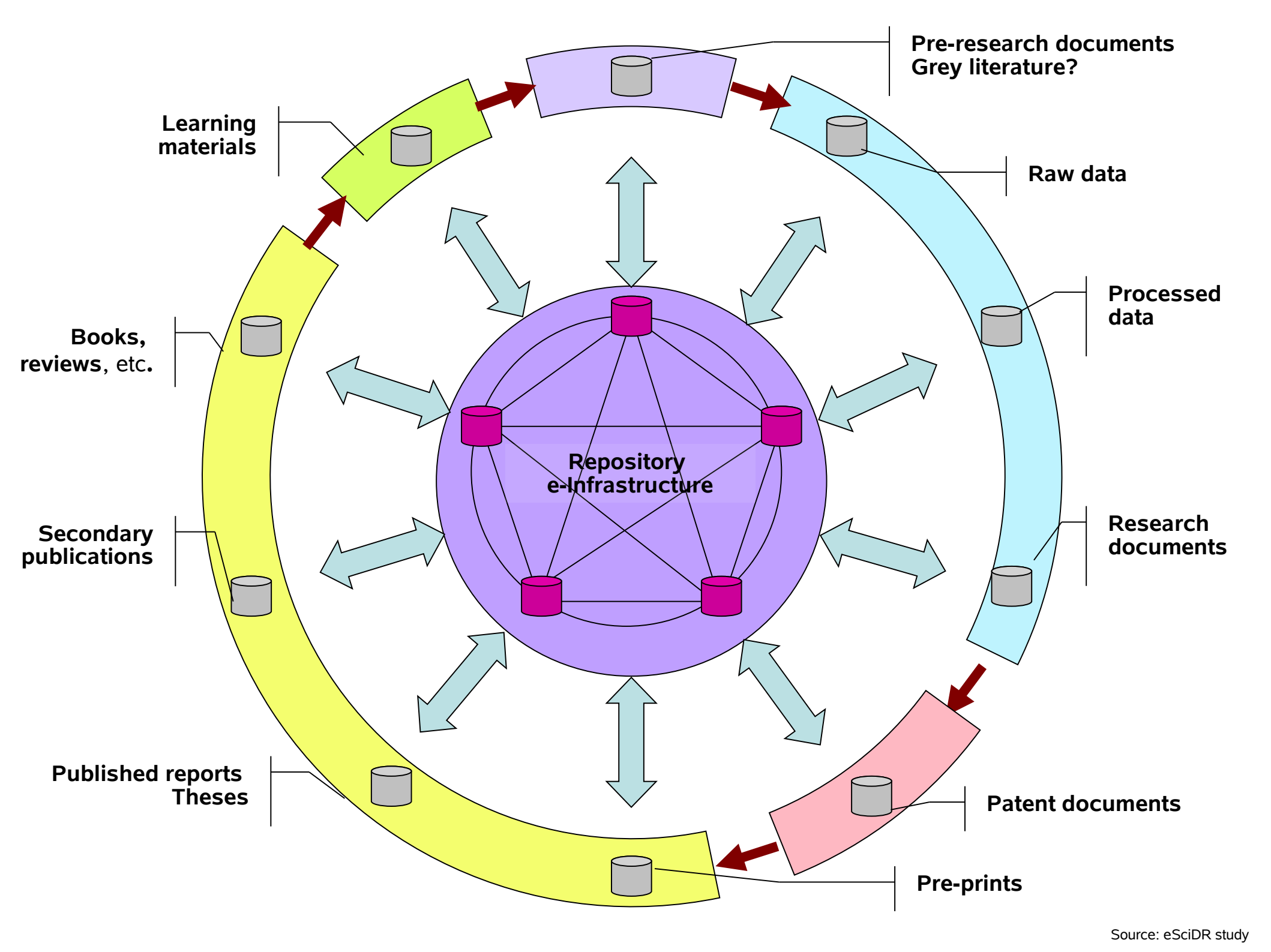
A wider interpretation of e-Infrastructures include technologies of various kinds for creating, collecting, annotating, manipulating, storing, finding and re-using **information** and services such as those to provide user support, training, and preservation.

Further, we include information resources and associated tools such as vocabularies, ontologies, rights management and privacy protection systems, and curation.

Source: eSCIDR study

The Information Watermill

- In the digital connected world information
  - is produced in large volumes
  - becomes more complex expressions of knowledge
  - is not only human readable
  - for machine-to-machine communication
  - struggle between volatility and permanence
  - puts pressure on traditional organisations
  - is the basis for e-Science



Source: eSciDR study

nature

www.nature.com/nature

Vol 461 | Issue no. 7261 | 10 September 2009

Data's shameful neglect

More and more often these days, a research project's success is measured not just by the publications it produces, but also by the data it makes available to the wider community. Pioneering archives such as GenBank have demonstrated just how powerful such legacy data sets can be for generating new discoveries — espe-

All but a handful of disciplines still lack the technical, institutional and cultural frameworks required to support such open data access (see pages 168 and 171) — leading to a scandalous shortfall in the sharing of data by researchers (see page 160). This deficiency urgently needs to be addressed by funders, universities and the researchers themselves.

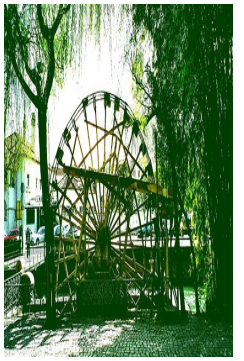
Research funding agencies need to recognize that preservation of and access to digital data are central to their mission, and need to be supported accordingly. Organizations in the United Kingdom, for instance, have made a good start. The Joint Information Systems

combined in unanticipated ways, is software that can keep track of which pieces of data came from whom. Such systems are essential if tenure and promotion committees are ever to give credit — as they should — to candidates' track-record of data contribution.

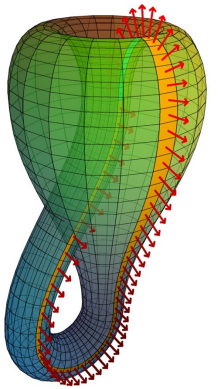
"Data management should be woven into every course in science."

Who should host these data? Agencies and the research community together need to create the digital equivalent of libraries: institutions that can take responsibility for preserving digital data and making them accessible

scientific Information con·tin·u·ums



- between experimental data and publications (new paradigm)
- between different scientific disciplines (multidisciplinary)
- between past, present and future (preservation)
- between different institutions (organisation)
- between humans and computers (e-Infrastructure)
- between research and education (public mission)



Reference: Klein Bottle with Moebius Band. Reference to article "Imaging maths - Inside the Klein bottle" at http://plus.maths.org/issue26/index.html. The Klein bottle is a non-orientable surface found by Felix Klein in 1882 while working on a topological classification of surfaces.

Riding the wave

How Europe can gain from the rising tide of scientific data

Final report of the High Level Expert Group on Scientific Data
A submission to the European Commission

October 2010

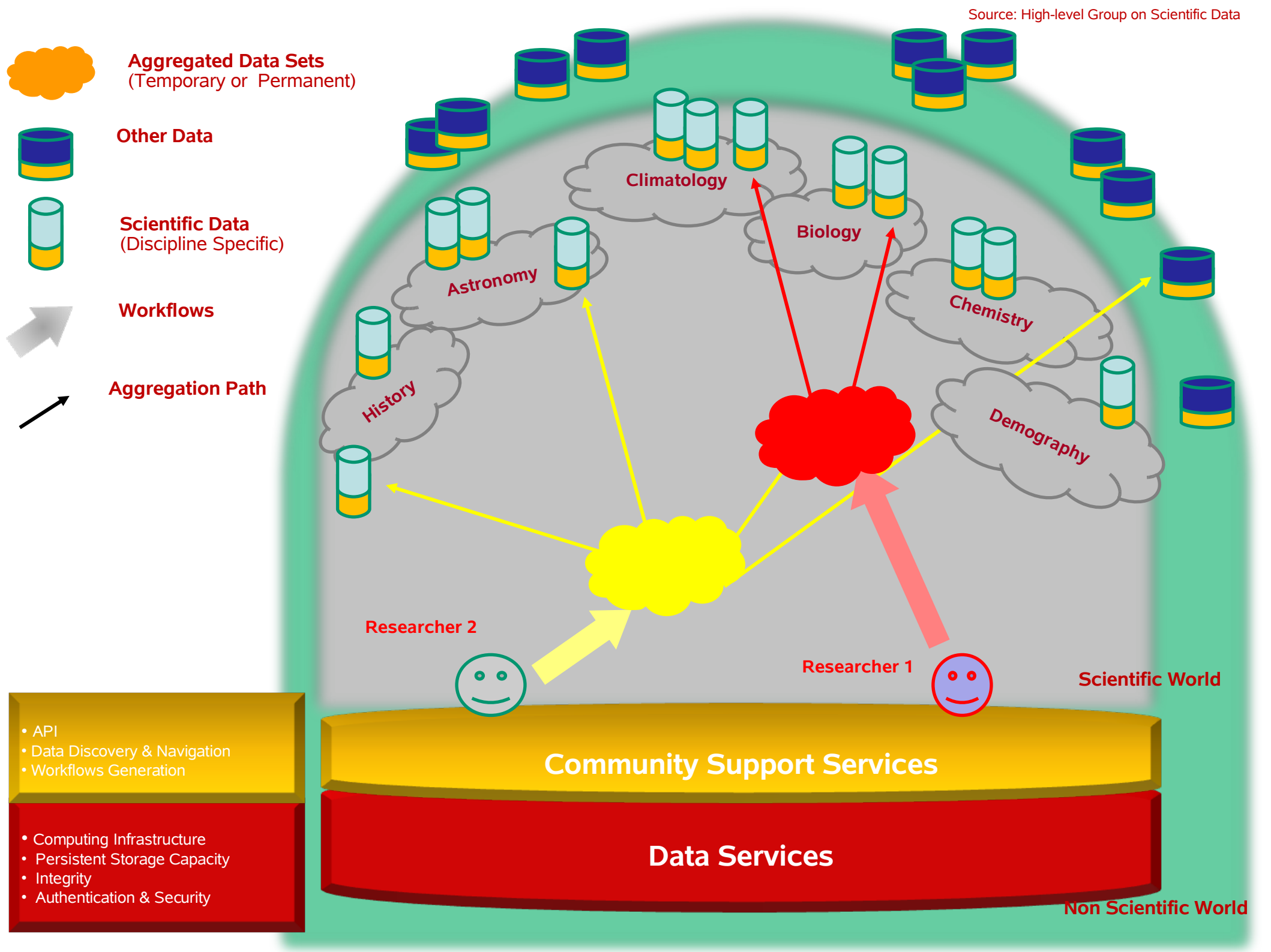
Rising tide of data

- “A fundamental characteristic of our age is the rising tide of data – global, diverse, valuable and complex. In the realm of science, this is both an opportunity and a challenge”

  Report of the High-Level Group on Scientific Data, Oct 2010

  “Riding the wave: how can Europe gain from the rising tide of scientific data”

Source: High-level Group on Scientific Data



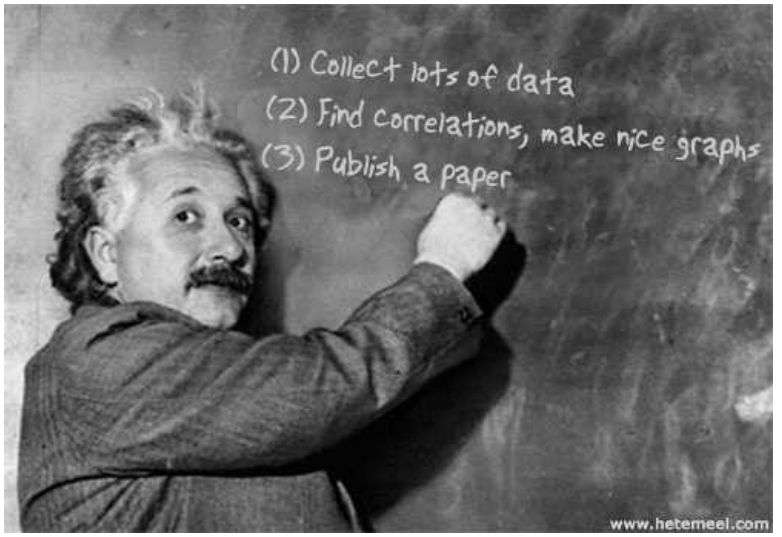
Open Access to Scientific Information

- Open Access Pilot in FP7
- OpenAIRE e-Infrastructure
- Policy issues at stake
  - Digital Agenda and Innovation Union
  - Coordination with EU Member States
  - COMs being prepared for 3Q of 2011 and 2012

Access to Information is key to Open Science

- Science has been always based on open communication and collaboration across geographic boundaries.
- What about organisational and disciplinary boundaries?
- Unlocking access for wide use and re-use of scientific information accelerates the research process and improves research efficiency

Future steps...



- We don't know how scholar communication will adapt to new communication paradigms
- “Scientific data has the power to transform our lives for the better – it is too valuable to be locked away.”

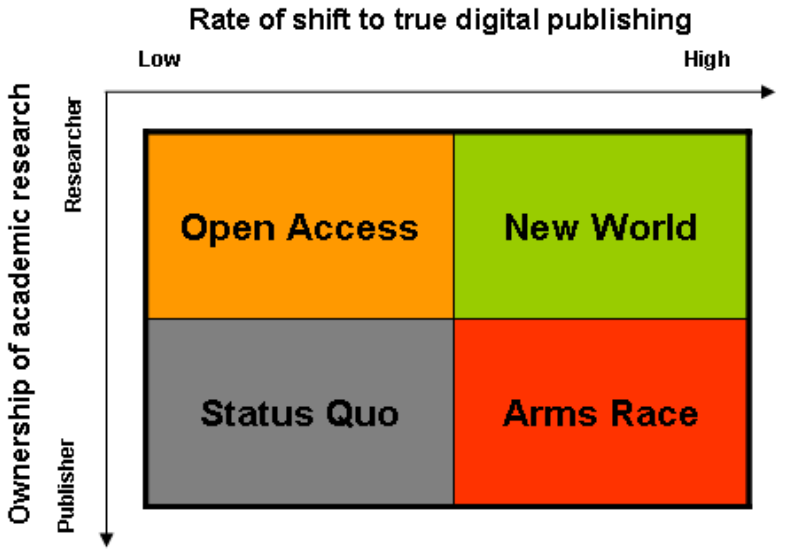
publishers, libraries, research organisations, researchers, funding agencies,…

- Growing budget pressures (libraries, publishers,…). That long black cloud is coming down?

  "The recession may trigger a fundamental shift in how scientific, technical and medical research is disseminated. Traditional for-profit journal publishers will be threatened by near-term academic library budget cuts and by technology and the Open Access movement in the long term." [from Bernstein Research July 2009]

- How to face the digital world?

- Scenarios:



	do nothing	focus on OA	new world (with digital OA)
Optimisation of public resources invested in R&D	Governments and public funding agencies maintain current models to support R&D and access to its results (grants to researchers and funding access costs) Unjustified barriers to access information may continue with negative impact on education, research and innovation	Increase role and power of funding agencies who incorporate costs of Access into the R&D funding Research communities assume great part of responsibilities of quality-assessment process (publishers may have a role depending on copyright agreements)	Governments and public funding agencies invest on infrastructure to facilitate pure-digital publishing (article + data) Research communities – possibly in partnership with publishers - assume responsibilities in peer-review process
Pro-Science, Pro-Innovation	Research organisations and Higher-Education institutions lose opportunity to re-structure. Webs of science do not develop or develop in "silos"…	Flexible licensing schemes stimulate access and reuse Publishers insist on "authors –pay" OA and embargo periods compatible with "established" business models	Scientific (experimental) data and scientific paper become an aggregated "Scientific Result". The "user" is at the centre New services and competences for hosting information and bulky data (for Open Access)
Stakeholders transformation	Libraries suffer from "public budget cuts" and are unable to shift to the digital world (except in very few cases) Publishers engage in "arms race" struggles to gain market shares	Responsibilities of long-term preservation in the digital world are left undefined Scientific (experimental) data is not dealt with or it is in a fragmented way	Libraries (readers) engage in exercises of demand aggregation to make the best use of underlying infrastructures European coordination would bring great benefits for negotiations of costs and cope with preservation issues (aggregation of demand)
Deployment	Major publishers (non-traditional or without legacy) invest significantly in new platforms and tools increasing their control in the process. Smaller players exit the market Wait and see…	Discussion focused on changing business models Intense policy discussion (with a lot of lobbying) focused on copyright legal frameworks	Combination of R&D incentives, education policies, investment in infrastructure and… copyright framework

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)