

EIRG Workshop on eInfrastructures

Lisbon 11th October 2007

Data Repositories: Where we are and where are we going...

Carlos Morais Pires
European Commission - DG INFSO
Unit "GÉANT and e-Infrastructures"



e-infrastructure

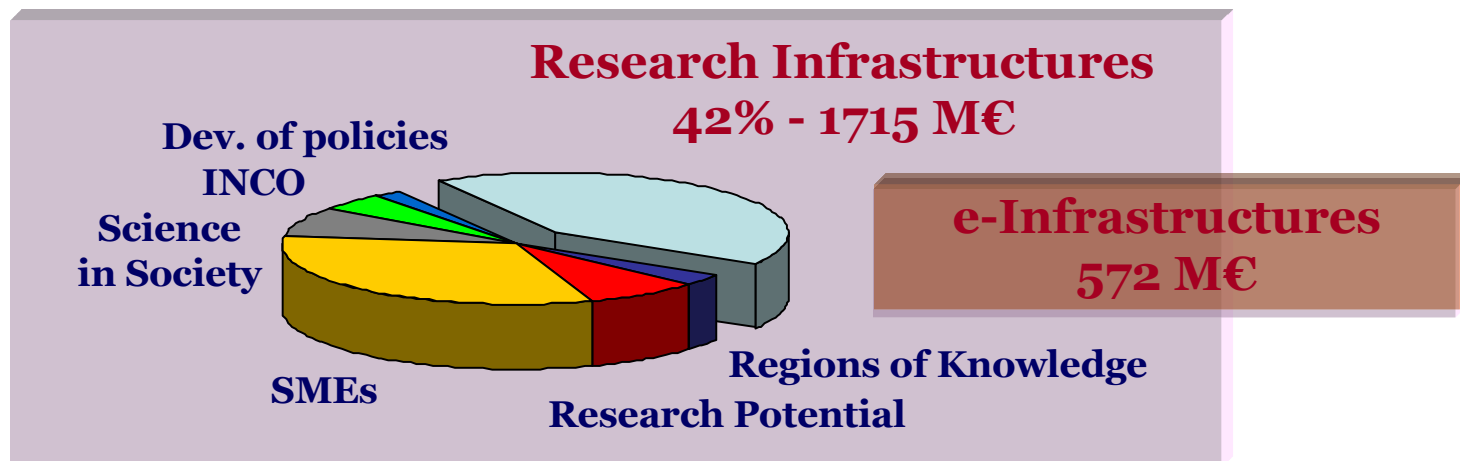
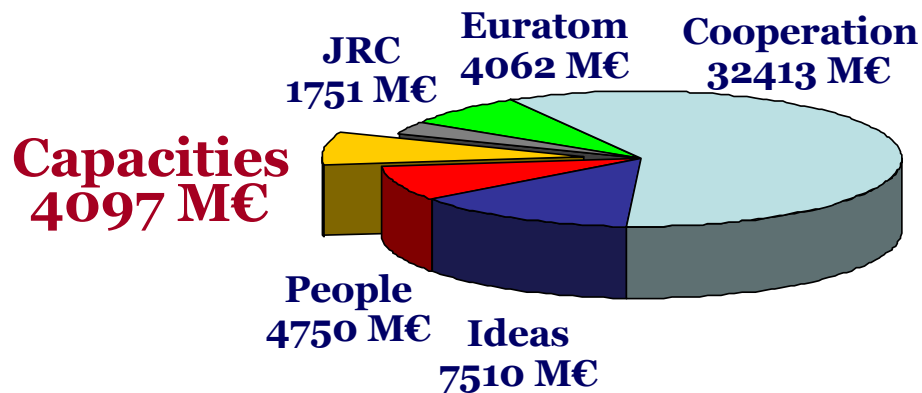


European Commission
Information Society and Media

Where we are and where we are going...



Framework Programme 7 (2007-13)



FP7 - research at the core of EU policies

**World class research infrastructures
including e-Infrastructures
(ICT based infrastructures)**

People

Capacities

Research Infrastructures

e-infrastructure



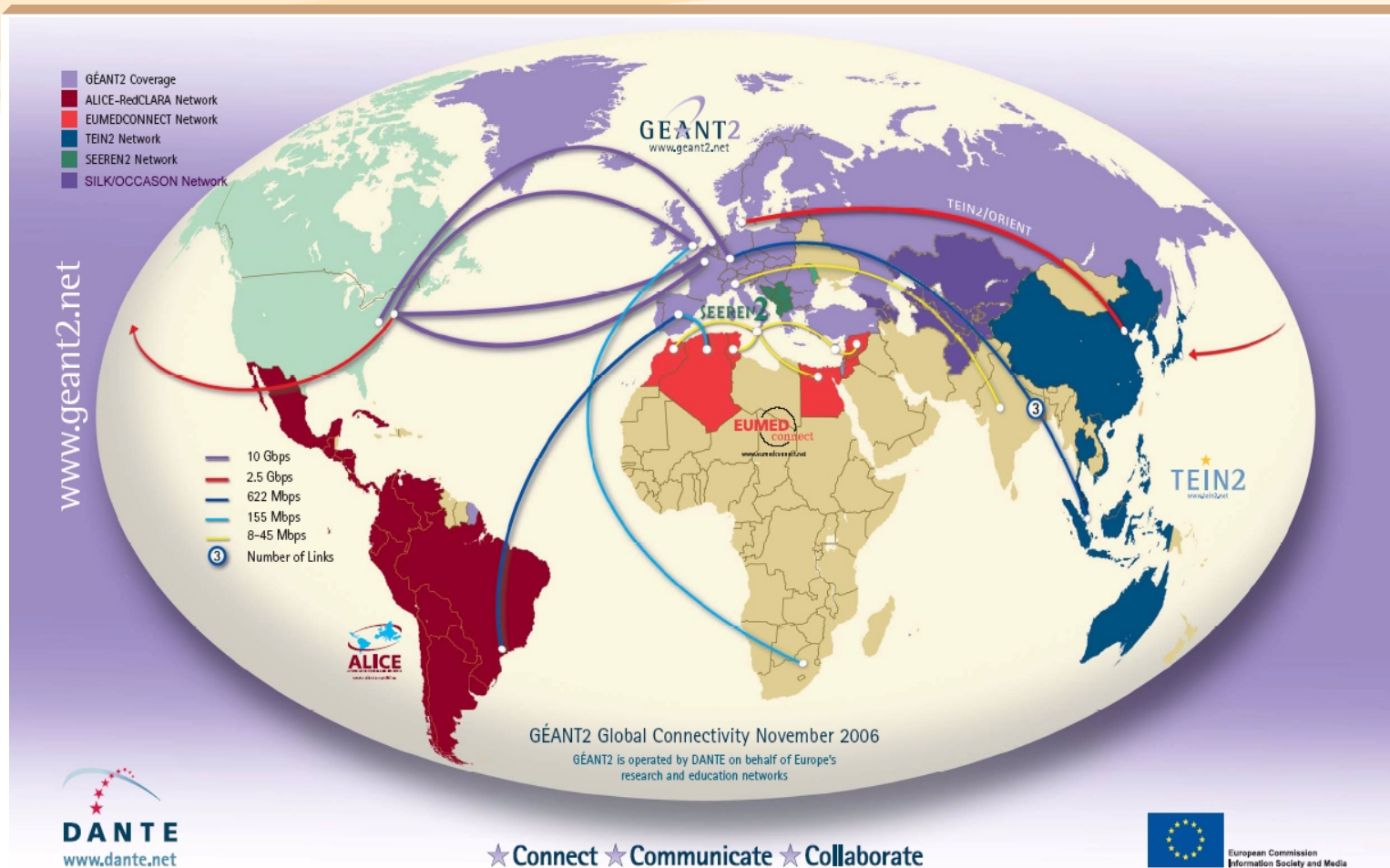
e-infrastructure



European Commission
Information Society and Media



GÉANT: global reach



Science has been a constant dialogue

... scientific dialogue between **Galileo** and **Kepler** (1610)

smaismrmilmepoetaleumibunenugttauiras

salve umbistineum geminatum Martia proles

"Be greeted, double knob, children of Mars."

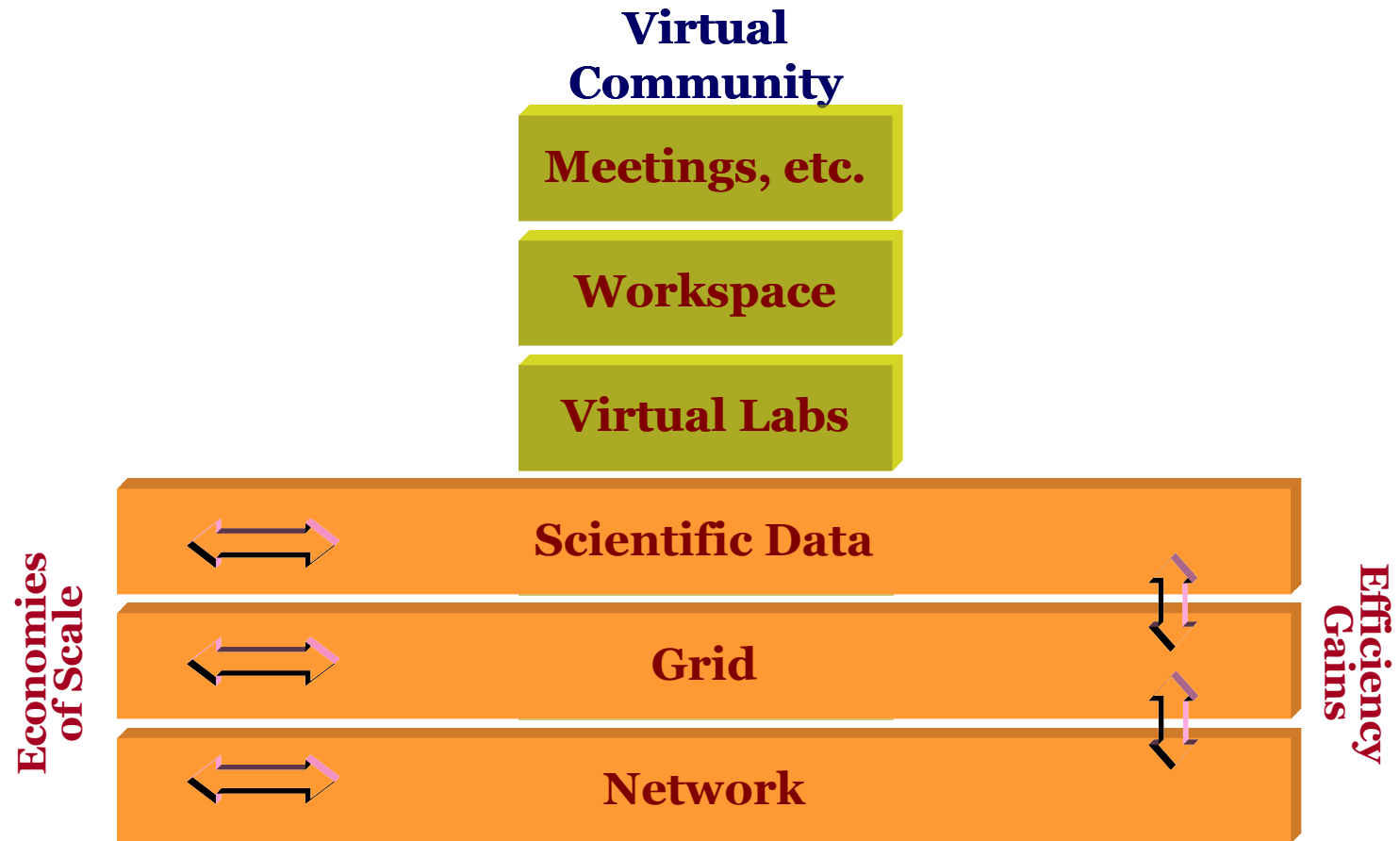
altissimum planetam tergeminum observavi

"I have observed the highest of the planets [Saturn] three-formed".

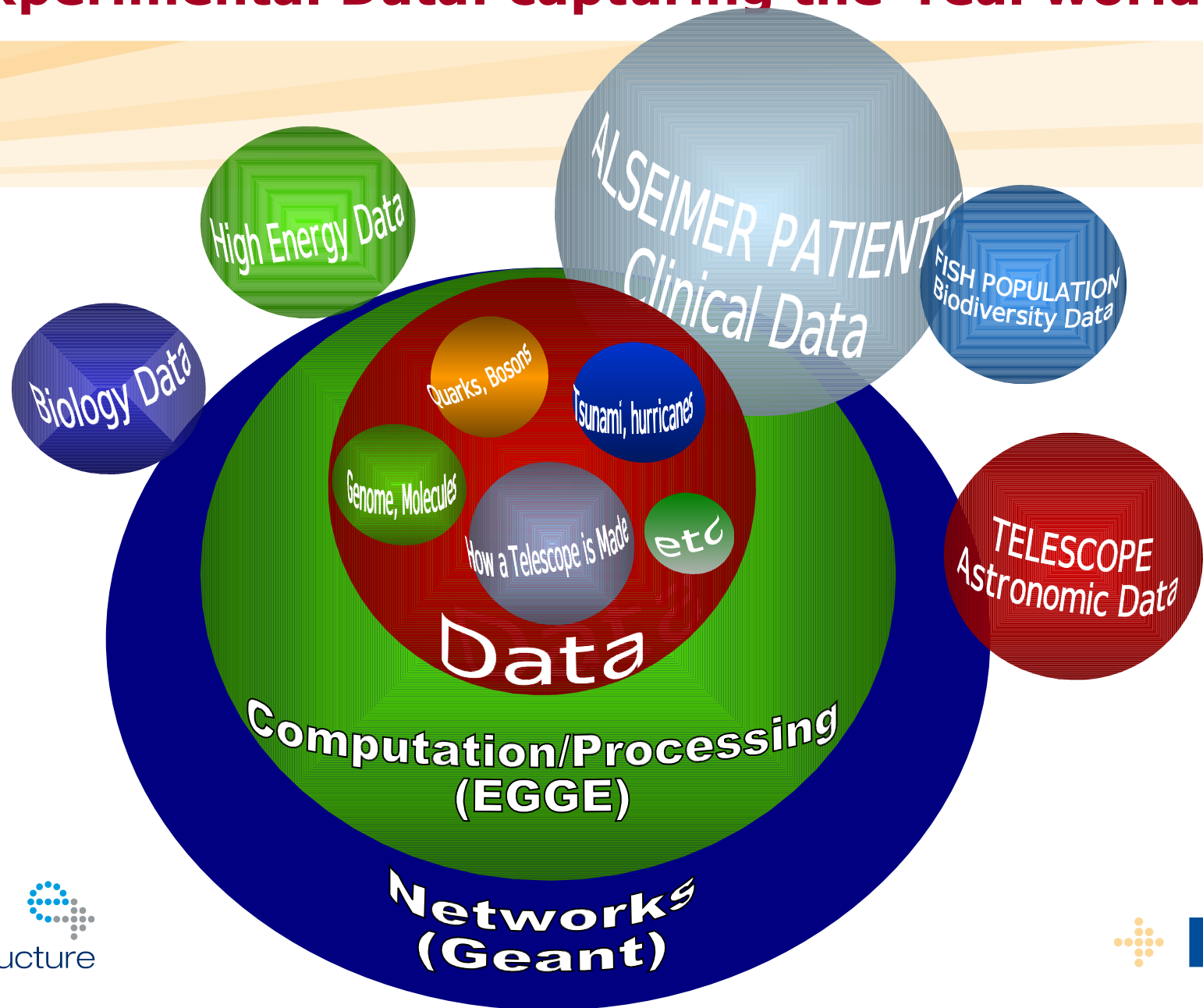
<http://www.mathpages.com/home/index.htm>



Global Virtual Research Community



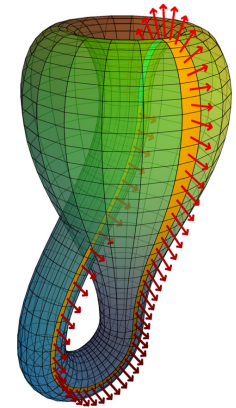
Experimental Data: capturing the 'real world'



con·tin·u·ums

- Scientific Data and the continuums:
 - between past, present and future
 - between raw experimental data and publications
 - between different scientific disciplines
 - between different institutions
 - between research and education

Figure: 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.



What are the issues we are addressing?

- Modern Science makes increasing use of information and communication technologies
- Volumes of Scientific Data produced in 'European Infrastructures' is growing very fast
- Data 'is' infrastructure for Science and Research
- Each 'Science' is 'a' 'Science'
- European Member States would like to see clear...



What would you do in a Deluge?



The animals boarding Noah's Ark two by two

painting by the American Edwards Hicks (1780-1849)
(from Wikipedia)

What would you do in a Deluge?



a more recent version
(couldn't identify the author)

(picked from google image search)



e-infrastructure



European Commission
Information Society and Media

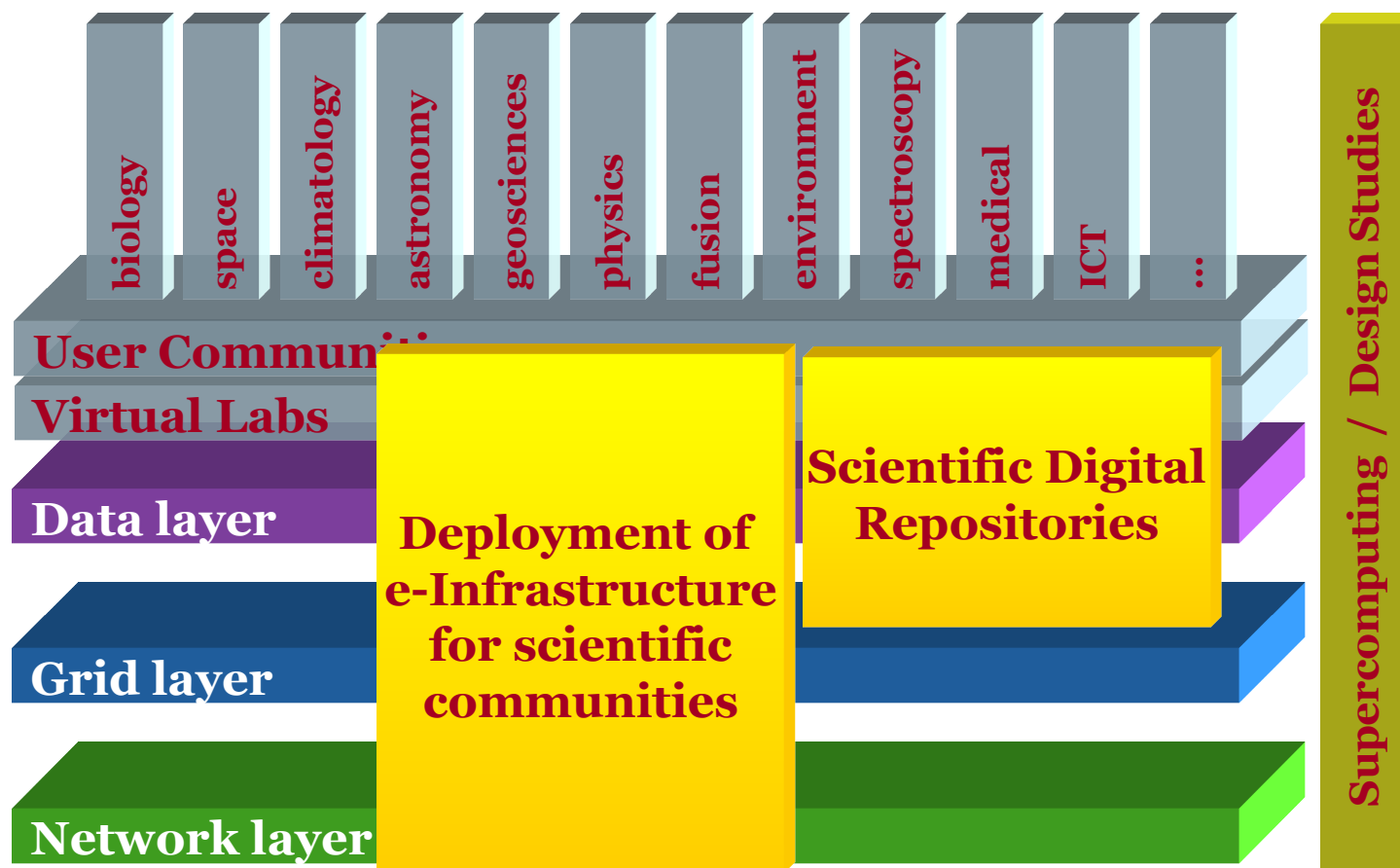


What are we doing...

- Planning two calls under FP7/Capacities
 - call 1 with 15 M€ (done)
 - call 2 with 20 M€ (2008)
- Launched a study (eSciDR)
 - results almost ready to be made widely and publicly available (Oct/Nov 2007)
- Evaluated already call 1 (July 2007)
 - 7 projects being negotiated (start early 2008)



e-Infrastructure Call 1 main objectives



e-Infrastructure: data repositories*

IMPACT	bio-informatics
NMDB	space physics
METAFOR	climatology
EuroVO-AIDA	astronomy
GENESI-DR	geosciences
DRIVER II	federated digital repositories

* Proposals under negotiation at this stage



Top ranked proposals

IMPACT unifies data from 10 major databases related to protein families.

Recommended funding: 3 M€

Consortium: 8 partners coordinated by EMBL

Scientific area: **bio-informatics**

NMDB establishes a digital repository for cosmic-ray data, and develop a real-time database from many neutron monitoring stations.

Recommended funding: 0,5 M€

Consortium: 12 partners coordinated by Universität zu Kiel

Scientific area: **space physics**



Top ranked proposals

DRIVER II federates scientific repositories based on open standards, supporting complex information objects, cross-discipline.

Recommended funding: 2,7 M€

Consortium: 13 partners coordinated by Nat. Kapodistrian Univ. Athens

Scientific area: **federated digital repositories**

METAFOR defines a Common Information Model (CIM) for climate data handling heterogeneous metadata standards.

Recommended funding: 2,2 M€

Consortium: 11 partners coordinated by Univ. Reading

Scientific area: **climatology**



Top ranked proposals

EuroVO-AIDA unifies digital data collections of astronomy, integrating European data centres into a global Virtual Observatory.

Recommended funding: 2,7 M€

Consortium: 8 partners coordinated by CNRS

Scientific area: **astronomy**

GENESI-DR provides access to Earth Science Digital Repository (data from space, airborne, in-situ sensors) for scientists.

Recommended funding: 4,4 M€

Consortium: 13 partners coordinated by ESA

Scientific area: **geosciences**



Coverage by the 'top-ranked' proposals

- Proposals the challenge of managing growing (massive) amounts of experimental data
 - with quality-verification concerns
- Repositories are used to manage, harvest data and metadata,
 - simulation models/software, experimental output, published papers by peers to strengthen existing and formulating new hypothesis
- Scientific Repositories are used to make a wealth of information useful and re-usable for scientists and researchers in various communities



Sharing Scientific Data

- Knowledge Based Economy
 - knowledge has value
- Knowledge is built from information
 - information has value
- Data is information
 - data has value



The Policy and the Attitude

- As emphasised by Commissioner Reading, the European policy for the Information Society (i2010) aims at
 - (1) the promotion of a inclusive Information Society and sustainable economy,
 - (2) a common information space for science and research,
 - (3) strengthening of ICT R&D and its deployment in Europe
- So, the attitude is:
Pro-Science and Non-Discriminatory



A concrete example

- EU FP7 Grant Agreement Model Contract contains a special clause specific for environment research
 - Clause 29 – Access Rights to Foreground For Policy Reasons and Transfer of Ownership of Foreground (specific to environment research)



Some other examples

- The ‘policy advisory groups’ eIRG and ESFRI launched working groups to address the issues of Scientific Digital Repositories
 - policy requirements are now being debated concerning aspects of
 - availability
 - permanency
 - quality
 - rights of use
 - interoperability



Points for Reflection

- Scientific Infrastructures generating experimental data and information face a number of challenges ranging from the validation, quality assurance to the long term preservation.
- Each community or institution will have probably the tendency to consider its own requirements and start shaping its 'own' infrastructure.
- What has to be brought to discussion is the need to see beyond these frontiers and use e-Infrastructures to really put in practice economies of scale at infrastructural level.
- One cannot underestimate the differences across Europe in the level of use and penetration of repository technologies.



More Points for Reflection

Other issues of extreme relevance for the policy discussion are:

- what funding models applying to the maintenance of repositories, for their own efficiency, sustainability the preservation of content?
- what can be done to harmonise and simplify authentication and authorisation mechanisms across Europe to gain access to e-Science resources
- which mechanisms of incentives are needed to encourage data generators to deposit (share) their data, and provide good-quality metadata.

These are just a few examples of the debate that has to be done in the coming years in parallel with the launch of the e-Infrastructures projects under the CAPACITIES programme.



Sharing Scientific Data

- More specifically, improved access to, and sharing of, data:
 - Reinforces open scientific inquiry and promotes new research;
 - Encourages diversity of analysis and opinion;
 - Makes possible the testing of new or alternative hypotheses and methods of analysis;
 - Facilitates the education of new researchers;
 - Enables the exploration of topics not envisioned by the initial investigators;
 - Permits the creation of new data sets when data from multiple sources are combined.



Sharing Scientific Data

- Accessibility to research data has become an important condition in:

The good management of the public investment

The creation of strong value chains of innovation

The enhancement of value from international co-operation



Barriers to overcome

Lack of sharing culture? Are scientific communities ‘closed’ ? - after all, we need to keep the incentives for high-quality contributions...

Are there concerns about ownership and IPR ? – Are scientific communities ready to engage on new IPR models based on ‘accessibility’?

Costs – after all, new paradigms are possible due to (IC) technology. This may imply costly investments to start and maintain...



Impact Assessment

Studies launched 1 year ago to help getting a birds-eye view on e-Infrastructures in Europe

eSciDR

ERINA

AVROSS

We are all looking forward to get and read the final reports...



Thank You



e-infrastructure



European Commission
Information Society and Media



Further information



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

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



e-infrastructure



European Commission
Information Society and Media



Reserve Slides



Further Information & Contact

- Contact and Further Information:

Carlos Morais Pires

carlos.morais-pires@ec.europa.eu

Further information

<http://cordis.europa.eu/ist/rn/>

