



European research infrastructure work: e-IRG promoting e-Infrastructure and e- Science development

Leif Laaksonen (e-IRG chair)
CSC - IT Center for Science Ltd.
28.5.2009

e-IRG
e-Infrastructure
Reflection Group

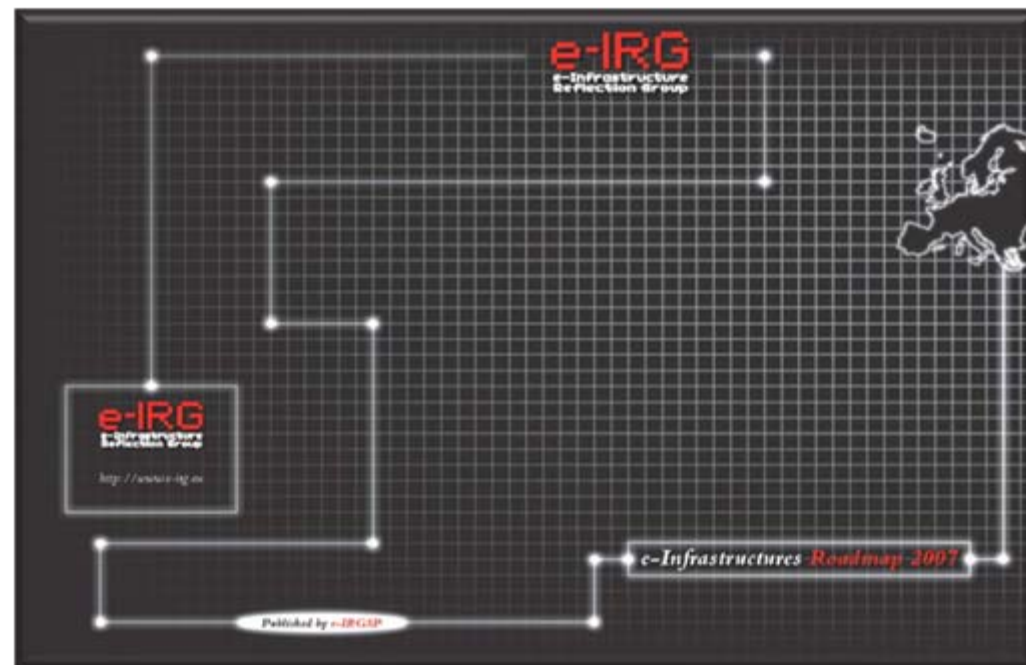
- Some historical perspectives up to the current time.
- The mission and vision of the e-IRG and outcome.
- The European Research Area and ESFRI
- The e-Infrastructure in the European Research Area.
- Research and Innovation in the ERA.
- The e-Infrastructure eco-system (landscape).
- Some observations and summary.

Steps on the path to the European Roadmap for Research Infrastructures

- Competitiveness Council Conclusions, 25-26 November 2004:
 - “In the context of developing research infrastructures of European interest, the Council of the European Union welcomes the development of a strategic roadmap for Europe in the field of research infrastructures and the role of the European Strategy Forum on Research Infrastructures (ESFRI) in this context.”
- The national roadmaps are important as ESFRI activities and its roadmap aim to integrate national resources into a common, pan-European effort.
- Development of legal instruments at EU level for setting up pan-European research infrastructures.
- 44 ESFRI Roadmap 2008 projects with a costs ~ 20 Billion euro in 10 years (~2B€/year)

Relevant policy bodies

ESFRI European Strategy Forum
on Research Infrastructures



ESFRI European Strategy Forum
on Research Infrastructures



- The European Strategy Forum on Research Infrastructures - ESFRI brings together representatives of EU Member States and Associated States, appointed by Ministers in charge of Research, and one representative of the European Commission.
- The role of ESFRI is to support a coherent approach to policy-making on research infrastructures in Europe, and to act as an incubator for international negotiations about concrete initiatives.
- <http://cordis.europa.eu/esfri/>
- The e-Infrastructure Reflection Group (e-IRG) was founded to define and recommend best practices for each of the (pan-) European Grid efforts. Representatives appointed by Ministers in EU Member States, Associated States to the EU Research Framework Programme and the European Commission.
- The main objective of the e-Infrastructure initiative is to support the creation of a framework (political, technological and administrative) for the easy and cost-effective shared use of distributed electronic resources across Europe.
- <http://www.e-irg.eu/>

Purpose of e-IRG

The e-Infrastructure Reflection Group (e-IRG) was founded to define and recommend best practices for the pan-European electronic infrastructure efforts.



Mission and Vision of e-IRG

e-IRG Meeting in Lugano (Switzerland) in 2008:

The e-IRG mission is to pave the way towards a general-purpose European e-Infrastructure.



The vision for the future is an open e-Infrastructure enabling flexible cooperation and optimal use of all electronically available resources.

- The e-IRG formed by official delegates appointed by the relevant ministries or authorities of their countries
- Structure includes appointed members, accessions, associated states representatives and European Commission officials
- The executive board consists of a chair elected by the members, and three board members representing the rotating EU Presidency (past – current – future presidencies)

The e-IRG Workshops serve as input

- e-IRG maintains the Vision document describing the actions and directions of e-IRG in the field of e-Infrastructure.
- e-IRG also maintains the white paper and roadmaps.
 - The white papers are slightly more technical documents describing the e-Infrastructure possibilities and developments trends from a technological enabling side.
 - The roadmaps define the long term e-infrastructure development trends seen from a European perspective.
- White paper and the roadmap are updated on a 24 month cycle, with white paper and roadmap being revised during alternative years. White papers 2008 and 2010 and Roadmap 2009.

1. Global collaboration
 2. Education and Training in the use of e-Infrastructure
 3. Grid and Cloud Computing
 4. Security: a holistic approach
 5. Virtualisation
 6. Remote instrumentation
 7. Sustainability of the computing-related e-Infrastructure
- Recommendations!**

Sustainability of e-Infrastructures

E-IRG recommendations by the task force:

1. Governments and the Commission should develop policies and mechanisms to encourage increased investment in a more coherent and interoperable way across Europe.
2. The existing e-Infrastructure projects must be superseded by integrated sustainable services at national and European levels.
3. e-Infrastructures must be application-neutral and open to all user communities and resource providers. National funding agencies should be encouraged to fund multi-disciplinary and inclusive infrastructures rather than disciplinary-specific alternatives
4. e-Infrastructures must inter-operate and adopt international standard services and protocols in order to qualify for funding
5. The Commission should, within the seventh Framework Programme, develop a pan-European e-Infrastructure which explicitly encourages the further integration of national e-Infrastructure initiatives.

ESFRI and e-IRG developing the ERA

- The Final Report of one the seven Expert Groups set up by DG Research of the European Commission in the context of the follow-up to the Green Paper “The European Research Area: New Perspectives” adopted by the Commission on 04 April 2007.



2008



Developing World-class Research
Infrastructures for the European
Research Area (ERA)

Report of the ERA
Expert Group



- Europe should reinforce an e-Infrastructure strategy able to boost the creation of virtual collaborative communities of researchers, ensuring the inclusion and participation of students and researchers from all around Europe.
- A coherent and managed layer of scholarly and academic research resources (including research outputs and data) by bringing together Europe's research repositories.
- An adequate and coordinated high-performance computing provision ecosystem and smooth access to its services.
- Generic virtual presence tools able to facilitate virtual research communities.
- Education and training programmes should be put in place to accelerate the exploitation of the e-Infrastructures by younger researchers and to improve the availability to wider user communities.

Roadmap to an ESFRI e-Infrastructure eco-system

*ESFRI addressing
fields of Research and
major research
challenges*

**SOCIAL SCIENCES &
HUMANITIES**

ENVIRONMENTAL SCIENCES

ENERGY

**BIOMEDICAL AND LIFE
SCIENCES**

MATERIAL SCIENCES

**ASTRONOMY, ASTROPHYSICS
AND NUCLEAR PHYSICS**

**COMPUTER DATA TREATMENT,
PARTICLES & SPACE PHYSICS**

*e-IRG building the
e-Science e-Infrastructure*

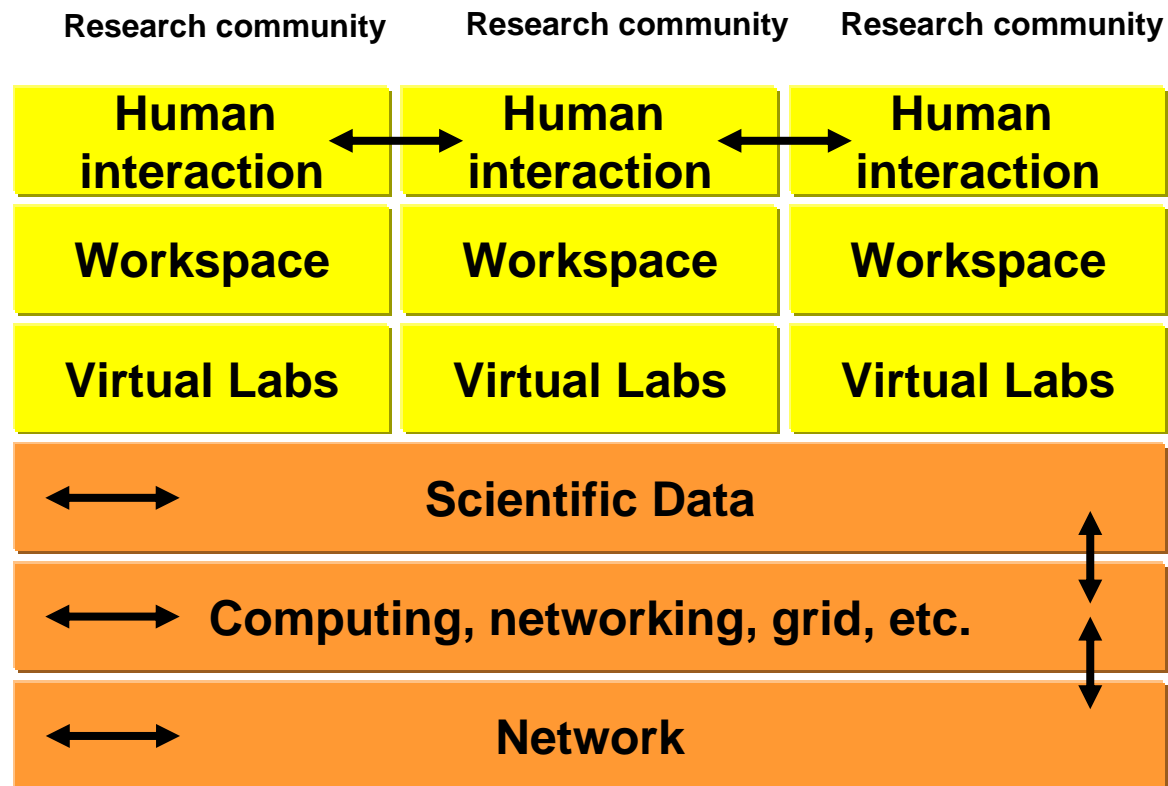
Crossing the boundaries of science

Resources/services (supercomputers, sensors, data)

Middleware and organisation

Networking Infrastructure

e-Infrastructure ecosystem



e-Infrastructure domains providing functions and services

- GÉANT is the world's largest multigigabit communication network dedicated to research and education.
- e-Science Grids respond to the requirements of the most demanding scientific disciplines (e.g. high-energy physics, bioinformatics) to share and combine the power of computers and sophisticated, often unique scientific instruments.
- The scientific data domain tackles the accelerated and uncontrolled proliferation of data elementary of the scientific discovery process.
- Supercomputing e-Infrastructures address the data-intensive and complex challenges of providing modern science with the new computing and simulation capabilities.
- Global Virtual Research Communities, anticipating the advent of research 2.0 paradigms, opens new perspectives for cross-border multi-disciplinary collaboration among research communities.

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF>

Science and society builds on data

Data ↔ Information ↔ Knowledge ↔ Wisdom

Data: symbols

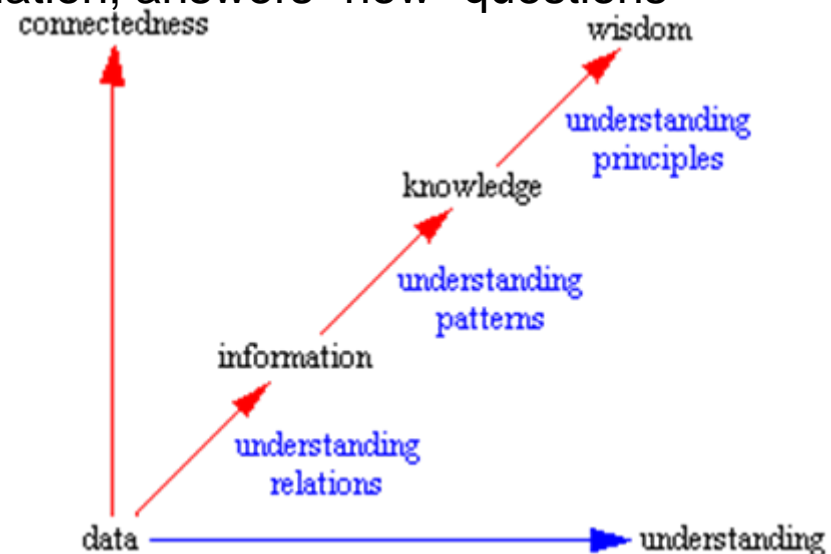
Information: data that are processed to be useful; provides answers to "who", "what", "where", and "when"

Knowledge: application of data and information; answers "how" questions

Understanding: appreciation of "why"

Wisdom: evaluated understanding.

- High-volume Data
- Long Term Storage
- Curation
- Cumulative
- Multidimensional
- Constant attention



The importance and value of data in research

(P. Arzberger et al. DSJ 2004)

- Publicly funded research data should be openly available to the maximum extent possible.
 - Data from publicly funded research are a public good produced in the public interest.
 - Both the data from publicly funded research and research itself have strong public good characteristics that support their open availability to the public, and especially to other researchers.
 - Factual data are central to the scientific research process.
 - The production and open dissemination use of factual data are essential attributes of, and inputs to, modern systems of scientific research and technological innovation. Recognizing the role of digital data as fundamental to the value chain of science, technology, and innovation will enable an optimum return on public investments.
 - Data access and sharing issues are international in scope.
 - To more fully exploit the possibilities of global digital networks, and to capture their benefits for the global community, policy issues concerning access to and sharing of publicly funded scientific research data must be addressed, not only at the institutional and national levels, but also at the international level.



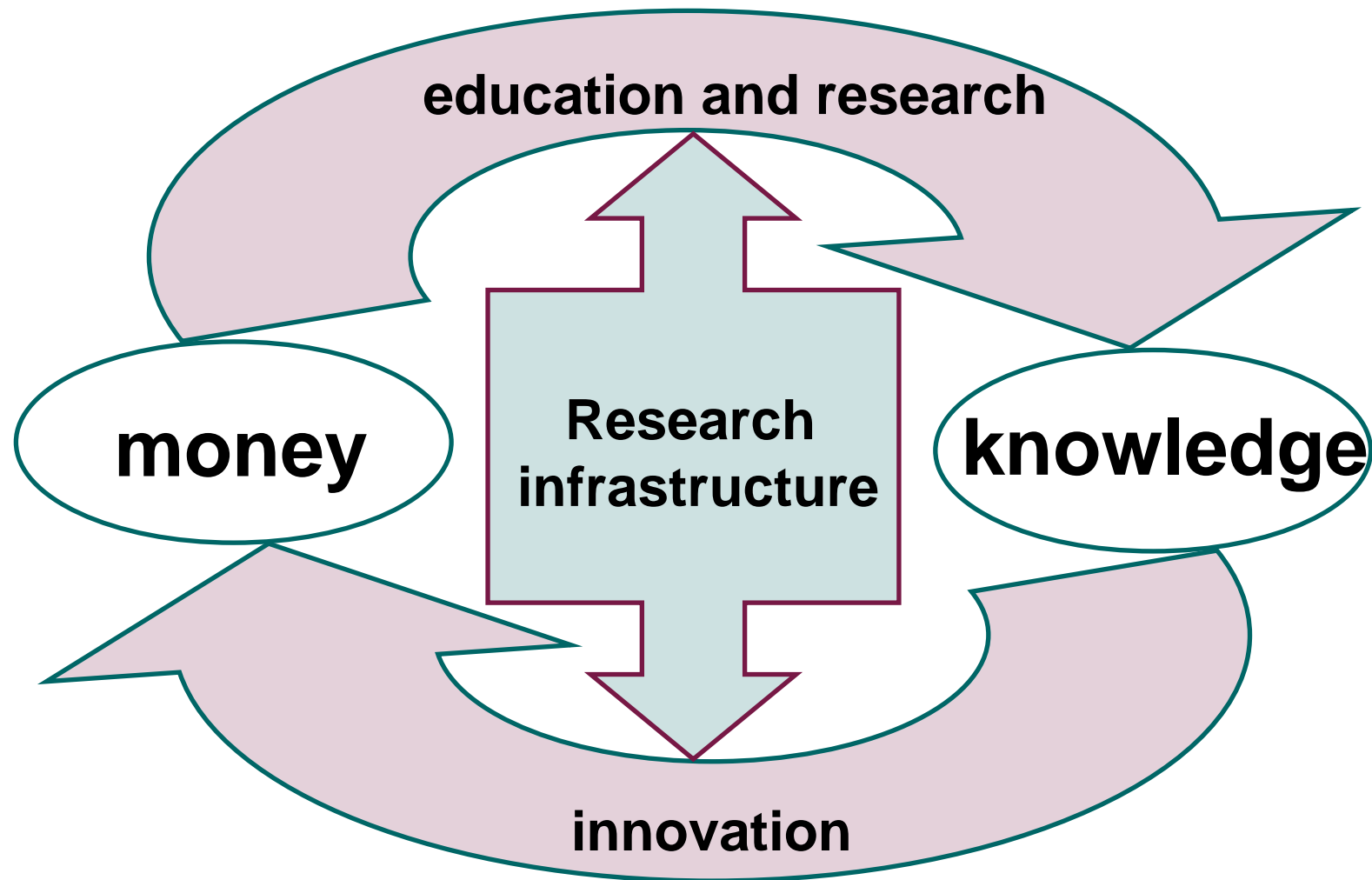
The Process of Innovation

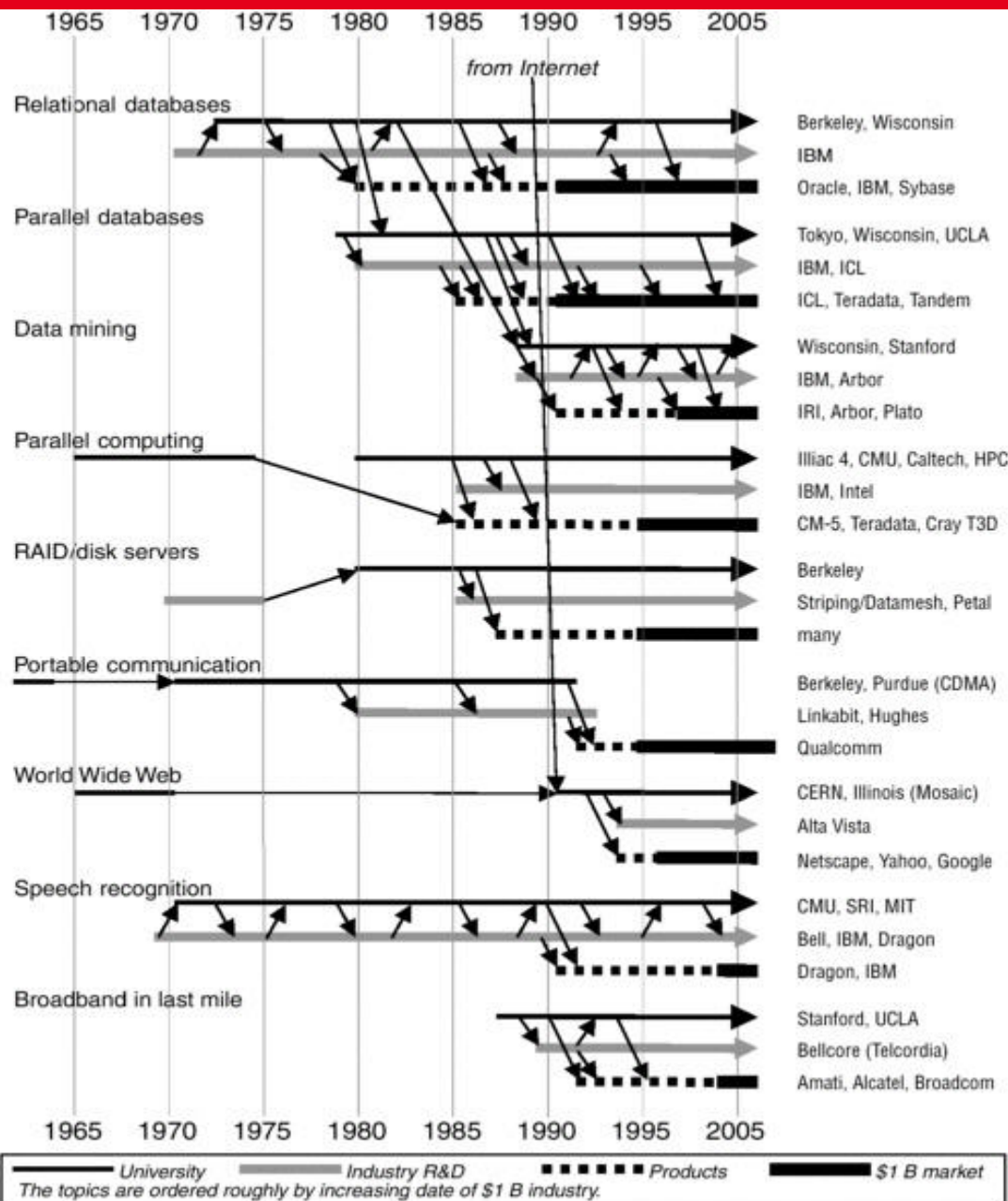
- **Research** = translating money into knowledge
- **Innovation** = translating knowledge into money
- Innovation can be **non-technological**: new or improved processes, organisation, design, marketing, etc.
- Innovation can emerge **without new knowledge / research**: the new combination of existing knowledge can be innovative
- Innovation = **interactive process**:
 - it depends not only on the **performance** of enterprises and research institutes, but
 - also the **manner and frequency in which they interact** to share knowledge, know-how, market flair, production / marketing capacities etc. is favourable for innovation
 - **Proximity** between different actors facilitates the acquisition, accumulation and use of knowledge (clusters, poles, etc.)

See: Oslo Manual on definition of innovation (OECD):

http://www.oecd.org/topic/0,3373,en_2649_37417_1_1_1_1_37417,00.html

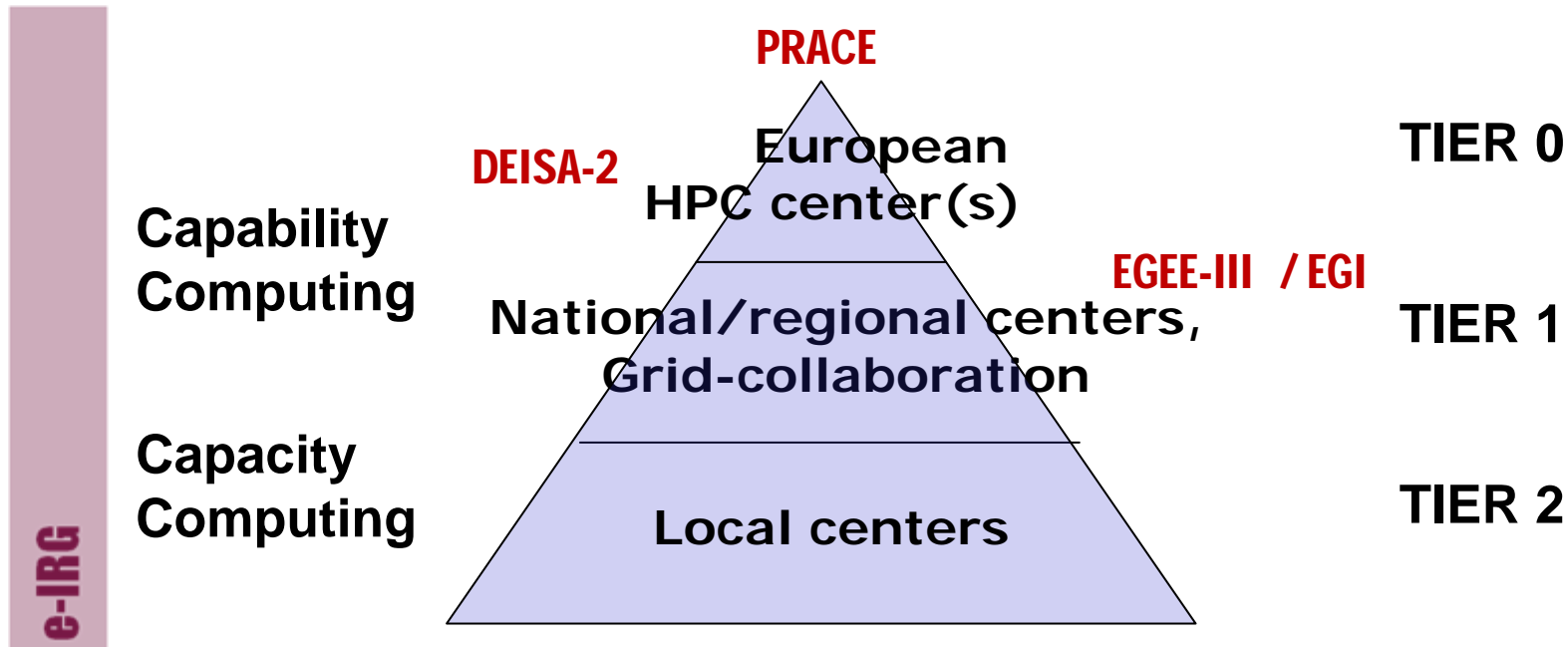
Research, education and innovation process





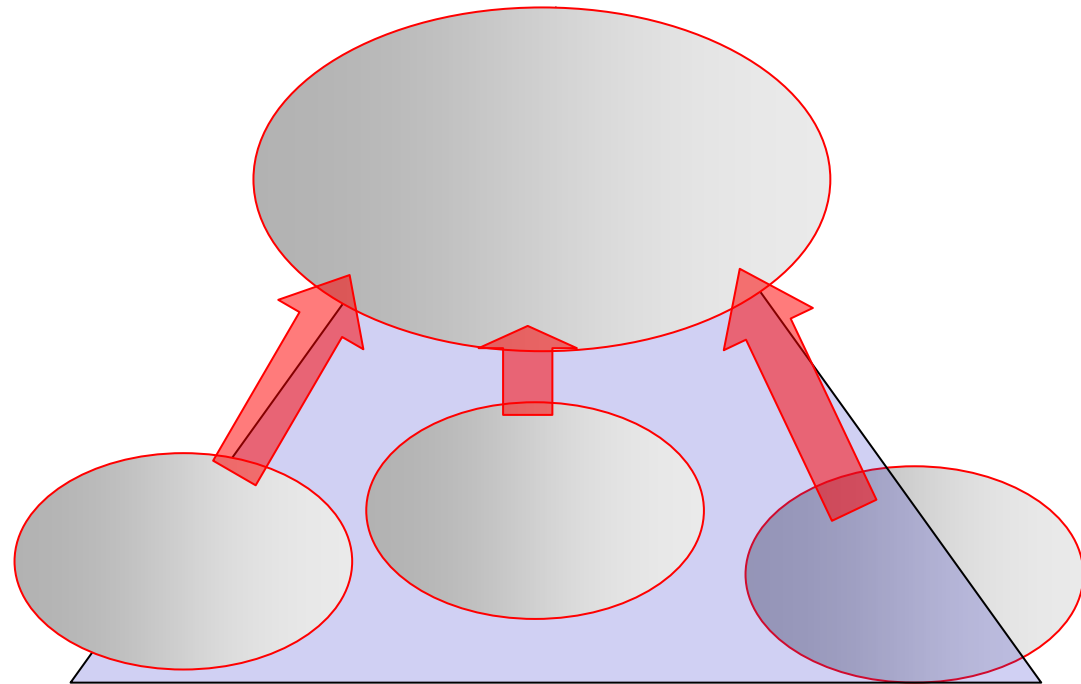
Innovation in Information Technology (2003) Computer Science and Telecommunications Board (CSTB)

HPC Performance Pyramid



HPC Ecosystem to support the top

- The upper layers of the pyramid
 - HPC centers / services
 - European projects (HPC/Grid, networking, ...)
- Activities which enable efficient usage of upper layers
 - Inclusion of national HPC infrastructures
 - Software development and scalability issues
 - Competence development
- Interoperability between the layers



Infrastructure @ CSC is evolving

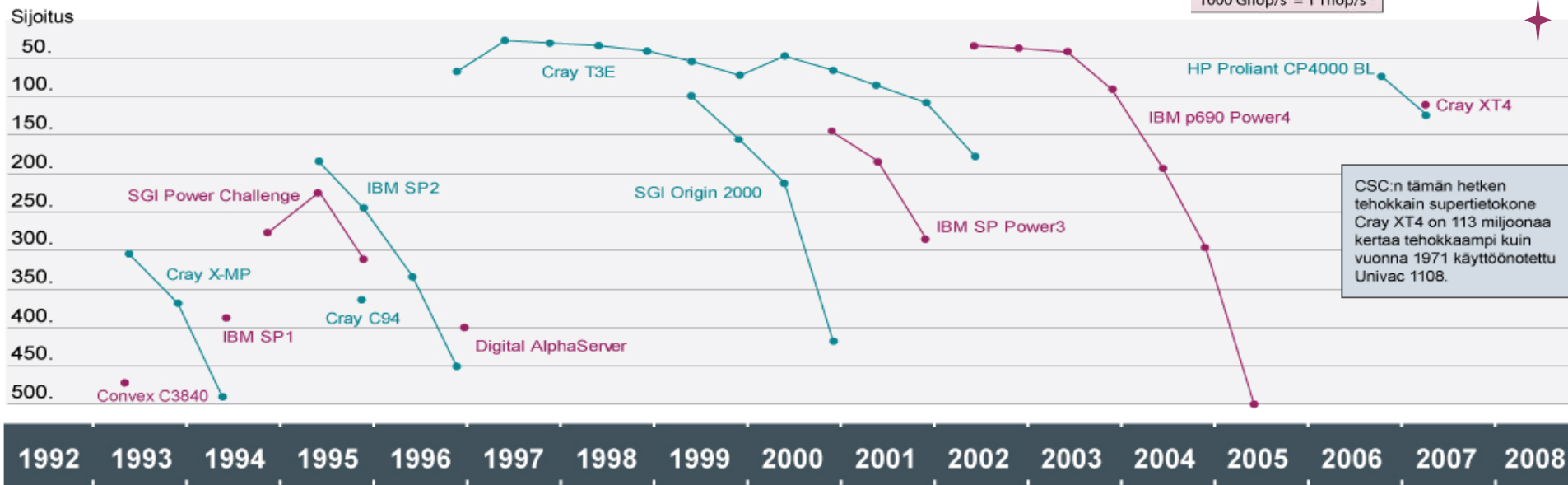


Univac 1108
-käyttöön 1971
-1 prosessori
-1 MB muistia
-huipputeho 0,093 Mflop/s
-oli aikanaan tehokas kone, mutta sitä ei kutsuttu supertietokoneeksi

CSC supercomputers on the TOP500 list
<http://www.top500.org/>

a

**Cray XT5
2008**



Cray X-MP
-käyttöön 1989
-4 vektoriprosessoria
-0,5 GB muistia
-huipputeho 935 Mflop/s



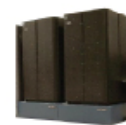
Cray C94
-käyttöön 1995
-4 vektoriprosessoria
-2 GB muistia
-huipputeho 4 Gflop/s



Cray T3E
-käyttöön 1997
-544 prosessoria
-64 GB muistia
-huipputeho 384 Gflop/s



SGI Origin 2000
-käyttöön 1998
-128 prosessoria
-160 GB muistia
-huipputeho 76,8 Gflop/s



IBM SP Power3
-käyttöön 2000
-128 prosessoria
-64 GB muistia
-huipputeho 192 Gflop/s



IBM p690 Power 4
-käyttöön 2002
-512 prosessoria
-512 GB muistia
-huipputeho 2,2 Tflop/s



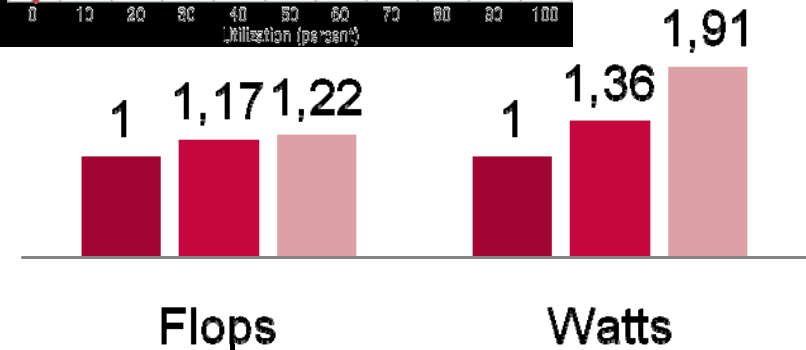
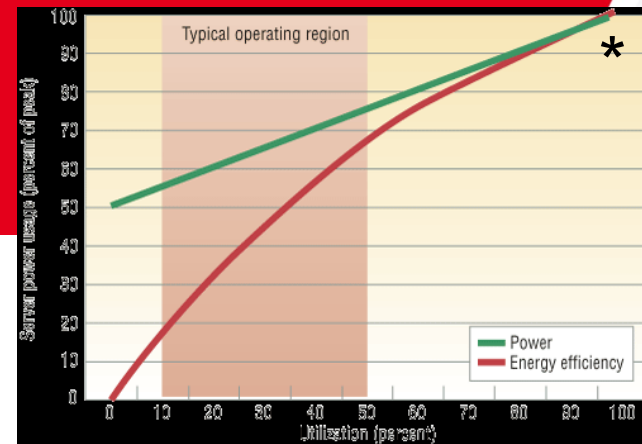
HP Proliant CP4000 BL
-käyttöön 2007
-1024 prosessoria
-4096 GB muistia
-huipputeho 10,6 Tflop/s



Cray XT4
-käyttöön 2007
-1012 prosessoria
-2024 GB muistia
-huipputeho 10,5 Tflop/s

Energy is a major issue

- 50 – 150W on a postage stamp
- Watts/socket ~constant
- Sockets/rack increasing
- Multicores demand memory
 - Each memory DIMM takes 5- 15W



“Virtualization may offer significant energy savings for volume servers because these servers typically operate at an average processor utilization level of only 5 to 15 percent” (Dietrich 2007, US EPA 2007).¹¹

“The typical U.S. volume server will consume anywhere from 60 to 90 percent of its maximum system power at such low utilization levels” (AMD 2006, Bodik et al. 2006, Dietrich 2007).”

*“The Case for Energy-Proportional Computing” (Luiz André Barroso and Urs Hölzle, Google 2007)

Power consumption

- Top500* results about the general power consumption reported of supercomputers include:
 - Average Power consumption of a TOP10 system is 1.08 MW and average power efficiency is **228 Mflops/W**.
 - Only 14 systems on the list are consuming less than 1 MW of power. Average Power consumption of a Green 500 system is 1.08 MW and average power efficiency is **228 Mflops/W**.
 - Average Power consumption of a Green 500 system is 358 kW and average power efficiency is **228 Mflops/W**.
- Top place in Green 500 list:
 - 44th place in Green 500 list, number 7 in Green 500 list:
NNSA/LANL BladeCenter QS22/LS21
Processor: Intel Xeon 8i 3.2 Ghz / Opteron DC 1.8 GHz , Voltaire
Total power **2483.47 KW**.

Unfortunately Flops are usually theoretical while Wattage is for real!

*<http://www.top500.org/lists/2008/11/highlights>
<http://www.green500.org/lists/2008/11/list.php>

Global virtual research communities

- No geographic region has a monopoly on intelligence or creativity.
- Virtual presence tools using the network are supporting the creation of collaborative research communities, boosting the research process and underpinning new ambitious goals and visions to be realized. Science 2.0 refer to new approaches to research enabled by Internet for discovering and accessing these services.
- Virtual classrooms can remove the educational quality gap between developing and rich parts of the world.
- Virtual laboratories and remote access to rare and expensive scientific instruments mean small, low-budget universities can enjoy access to infrastructure of the same quality as large, well-endowed ones.
- A digitally literate generation of young researchers is arriving having a different perspective on, and requirements of, the research process due to increased familiarity and use of ICT, and the Internet in particular.
 - Blogs, wikis, social networking, mashups enable gifted individuals to interact with and contribute to the world's leading research communities;
 - Enabling the inclusion of new groups into the research process.

Based on input from Martin Antony Walker

Challenges in e-Infrastructure deployment

- Petascale computing is driving a new trend - how will the scientific software evolution keep up to program hundreds of thousands of processors?
- The development and maintenance of scientific software is an essential, integral component of e-Infrastructure to support research. Software development and life cycle management needs improving and supporting.
- Education and training is urgently required to be able to utilize the benefits from the investments in e-infrastructure. A thorough dissemination of already available e-infrastructure knowledge to a much wider workforce and potential user community is required.
- e-Infrastructure provision must be directed by the needs of the research community and its requirements to carry major global research efforts.

e-IRG secretariat

**c/o CSC - IT Center for Science Ltd.
(Keilaranta 14)**

P.O. Box 405, FI-02101 Espoo, Finland

Phone: +358(0)9 457 2281

www.e-irg.eu

www.e-irg.eu/support

secretariat@e-irg.eu

