

Information infrastructure-based (regional) innovation

1

PETER TINDEMANS
PRAGUE, 24 MARCH 2009

My storyline

2

- Data-driven science adds new dimension: examples from biodiversity, systems biology, linguistics,
- Requires infrastructure: European Digital Data Infrastructure
- How to create this infrastructure
- Regions key for economic development and innovation
- Physical research infrastructures can be central for clusters and open innovation
- Building blocks for (generally) distributed Data Infrastructure can equally be stepping stones
- Think innovatively about innovation and regional development!

Biodiversity

‘Traveling through space and geological time’

From molecular data to Tree of Life

Ancient DNA

Computationally very intense
Phylogenies

Biogeography

DNA barcoding

Global Biodiversity Information
Facility (more than 60 countries)

*Life Watch on
ESFRI Road Map*

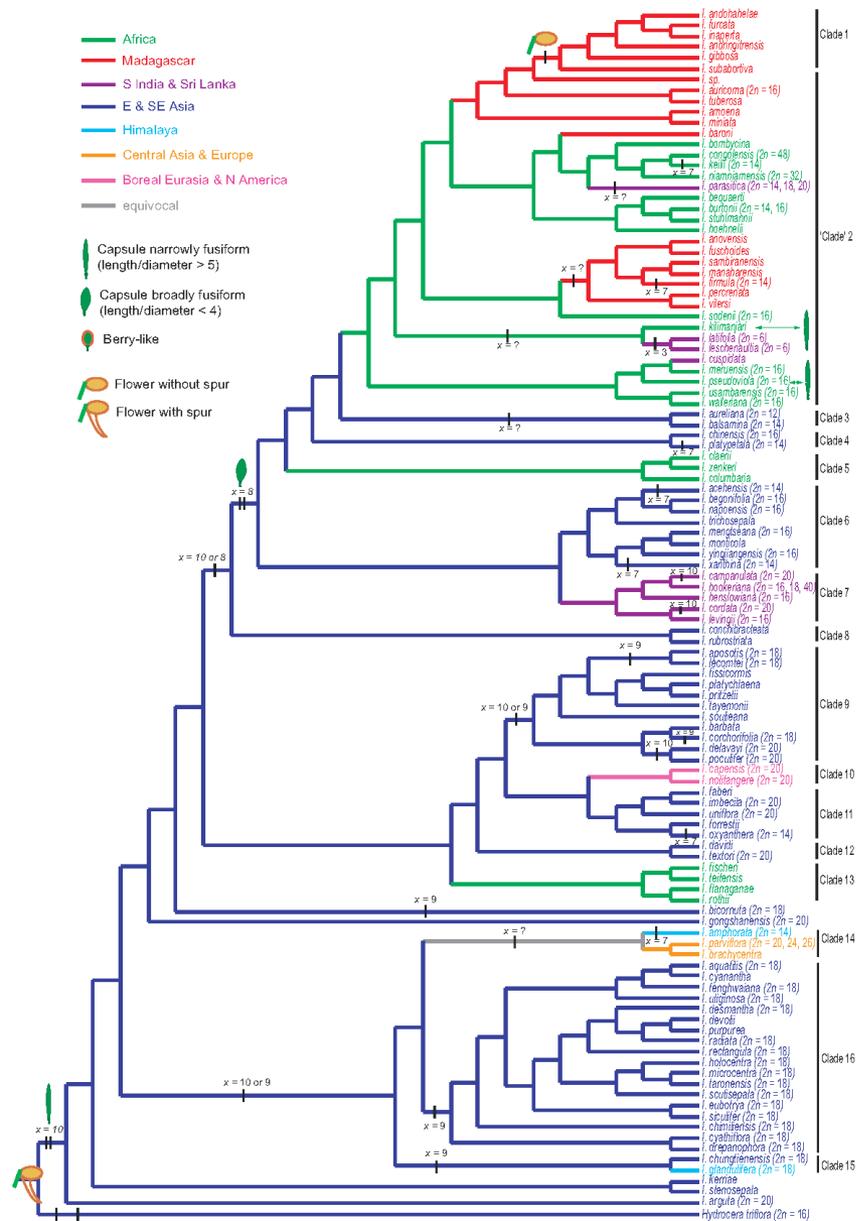
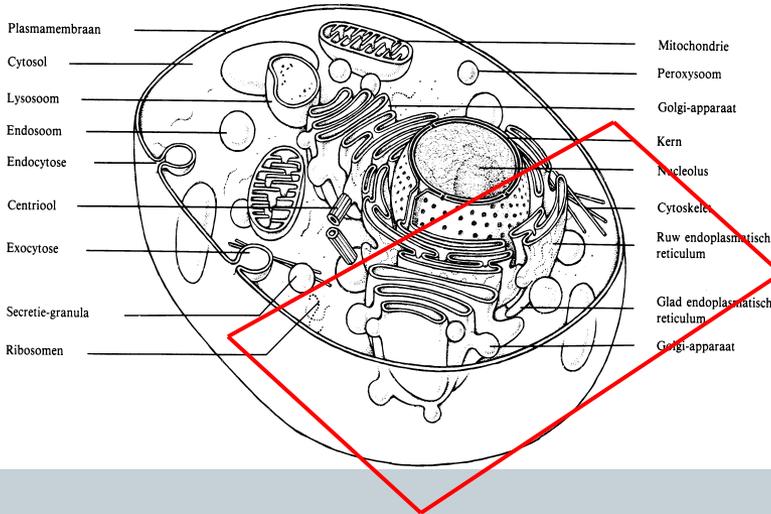


Fig. 4. Parsimonious optimization of the distribution centers of endemism, basic chromosome numbers, and selected morphological characters onto the NJ tree generated from the reduced ITS dataset. Vertical bars indicate positions of character state changes. Chromosome numbers when available are shown after the species names.

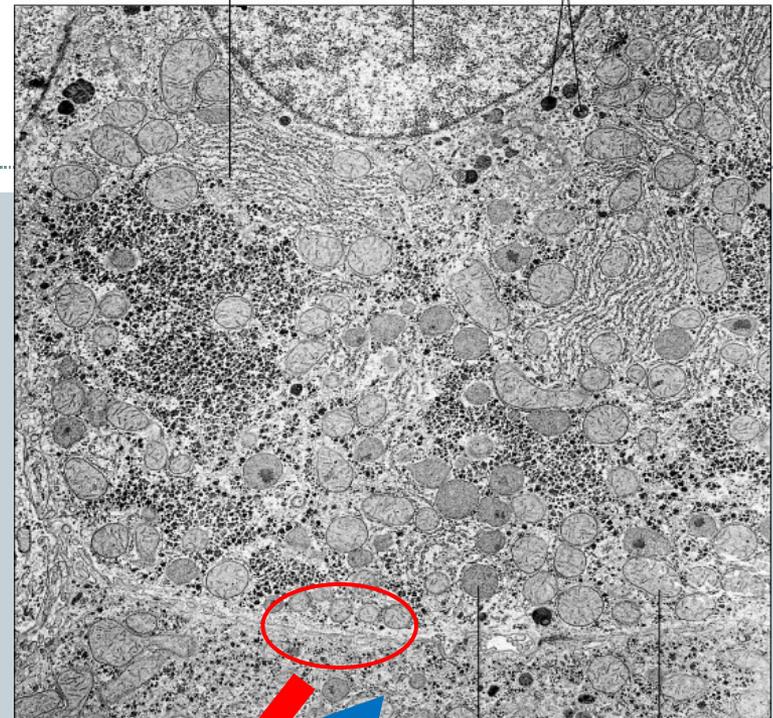
Biology today

Biology tomorrow

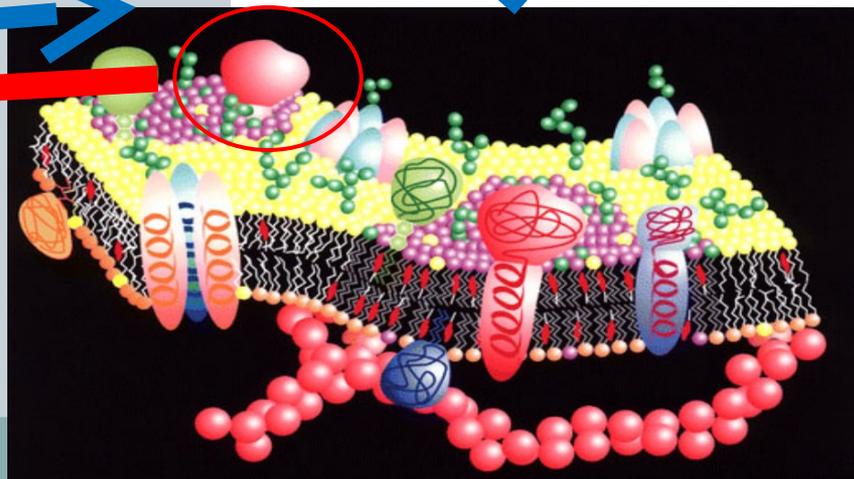
Overzichtstekening van de cel.



rough endoplasmic reticulum nucleus lysosomes



Courtesy
Roel van Driel



- Grammar, language families
- Language development
- Language learning
- Language and the brain
- Written vs spoken languages
- Language acquisition
- Language and social patterns
-

Linguistics

5

Huge data amounts gathered:
Audio, video, images, 'ordinary' data,



**CLARIN project
on ESFRI Road
Map**

A European Digital Information Infrastructure

6

Identify **core physical digital archives/repositories** in several initial communities and among cross-community organisations. Do this for documents and for data

- These must **OAIS-compliant** to ensure proper archiving, interoperability and long-term preservation
- Framework for **metadata**, Framework for **persistent identifiers**, and number of **registries**, possibly **other standards**
- Cost-effective **preservation methods and services** must be available
- Common framework of **principles and guidelines for management of access and rights** (underlying the technical tools to implement this framework)
- Create **Financial mechanism** for developing and testing implementation tools, techniques and services, and for strengthening collaboration and training
- a. **Certification service providers**, accredited according to
b. Common **European accreditation mechanism**.

Data Resources & Tools

- EMBL-BANK
- UniProt
- ArrayExpress
- Ensembl
- InterPro
- PDB-EBI
- Genomes
- Nucleotide Sequences
- Protein Sequences
- Macromolecular Structures
- Small Molecules
- Gene Expression
- Molecular Interactions
- Reactions & Pathways
- Protein Families
- Enzymes
- Literature
- Taxonomy
- Ontologies
- Sequence Similarity & Analysis
- Pattern & Motif Searches
- Structure Analysis
- Text Mining
- Downloads



European Bioinformatics Institute

About the EBI

- Research
- PhD Studies
- Training
- Industry Support
- Group & Team Leaders
- EBI Funders
- User Support
- EBI Mission
- People
- Events at the EBI
- How to Find us

EBI hosted EU Project Websites

- BioSapiens
- E-MeP
- ELIXIR
- EMBRACE
- EMERALD
- ENFIN
- FELICS
- SYMBIOmatics

Please Note: 7th-9th March 2008 -
Most EBI services will be affected by essential maintenance between 7th-9th March 2008. Work will commence on Friday 7th March at 16:00 GMT. We apologise for any inconvenience.

Research Highlights

Jan 22, 2008
Analysing large-scale proteomics projects with latent semantic indexing
A research paper in the Journal of Proteome Research presents the analysis of HUPO PPP data using latent semantic indexing. Such approaches offer a way of identifying and extracting useful information from large-scale data sets that can impact on proteomics experimental design... [more](#)

Latest News

Jan 22, 2008
International consortium announces the 1000 Genomes Project
Hinxton, 22 January 2008 – The EBI is part of an international research consortium that today announced the 1000 Genomes Project, an ambitious effort to sequence the genomes of at least 1000 people to create the most detailed and medically useful catalogue to date of human genetic variation ... [more](#)

to contribute to realise this infrastructure

8

Key Stakeholders from Science and Science Information committed *at Board level* to develop coherent European solution from which they will benefit themselves

Research organisations

- ESA, CERN, Max Planck Gesellschaft/Max Planck Digital Library, STFC, CNET, CNRS (tbc), EUA (tbc), Helmholtz Gesellschaft,

Funding agencies

- ESF (representing all national funding agencies), JISC (UK)

National libraries and archives, and their consortia

- British Library, Koninklijke Bibliotheek, Deutsche National Bibliothek, Swedish National Archive; Portico (global)

Publishers

- International Association of STM Publishers;

National 'coalitions'

- DPC (UK), NESTOR (DE),(NL)

US NSF DataNet

9

DataNet : 5 projects @ 20 M\$ (plus 10 M\$ possible renewal); 2 approved: The Data Conservancy, DataNetONE (earth observation)

- Provide reliable digital preservation, access, integration, and analysis capabilities for science/engineering data over decades-long timeline.
- Achieve long-term preservation and access capability in an environment of rapid technology advances.
- Create systems and services that are economically and technologically sustainable.
- Empower science-driven information integration capability on the foundation of a reliable data preservation network.
- Each project needed to develop a model for shared governance and the standards and protocols to enable interoperability

Joint programming helps matching NSF DataNet

10

Problems in Europe

- EU's FP is for Research and Technological Development, not for operational infrastructure (GEANT exception)
- Cooperation needed between DG INFSO (Content, Technologies, Infrastructures) and DG Research (Research Infrastructures)
- National Funding Agencies only begin to support projects

Two interesting developments might converge

- 'Joint Programming': member states to take initiative
- EUROHORCS (Heads of Research Councils) accepted challenge to common activities with common funding

So why wait?

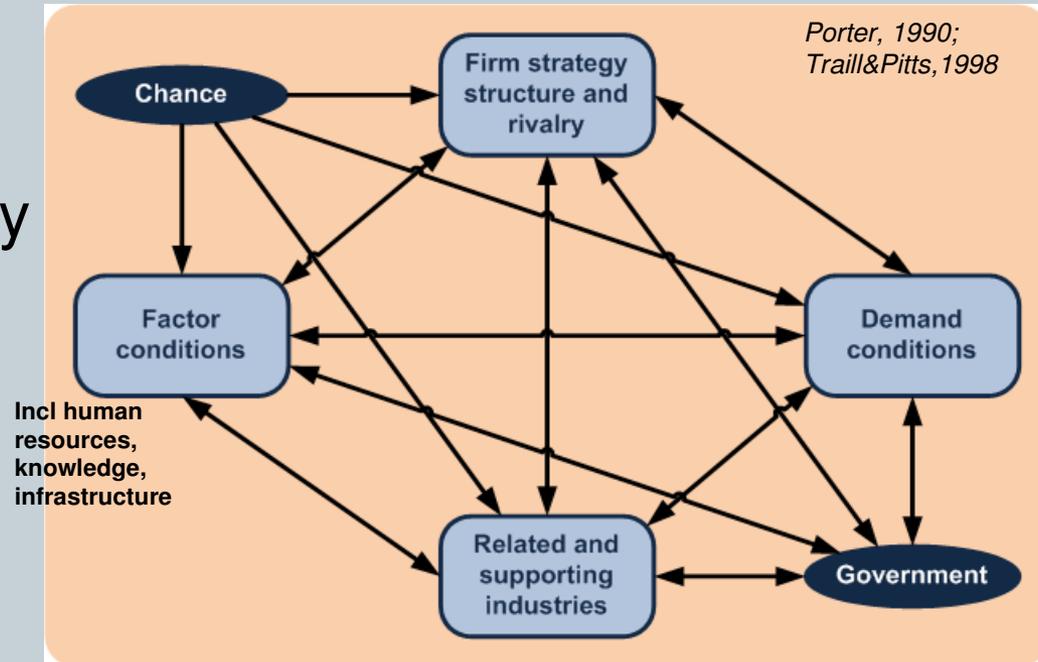
- Let EC and EUROHORCS/ESF define a joint EC-EUROHORCS/ESF programme to create the first large-scale

Regions key for economic development and innovation

11

1. Economics: agglomeration and proximity effects.
2. Open innovation
3. Specialisation creates innovative regions throughout EU

Point 1 illustrated by
M.E. Porter's
'Diamond'



Open innovation

12

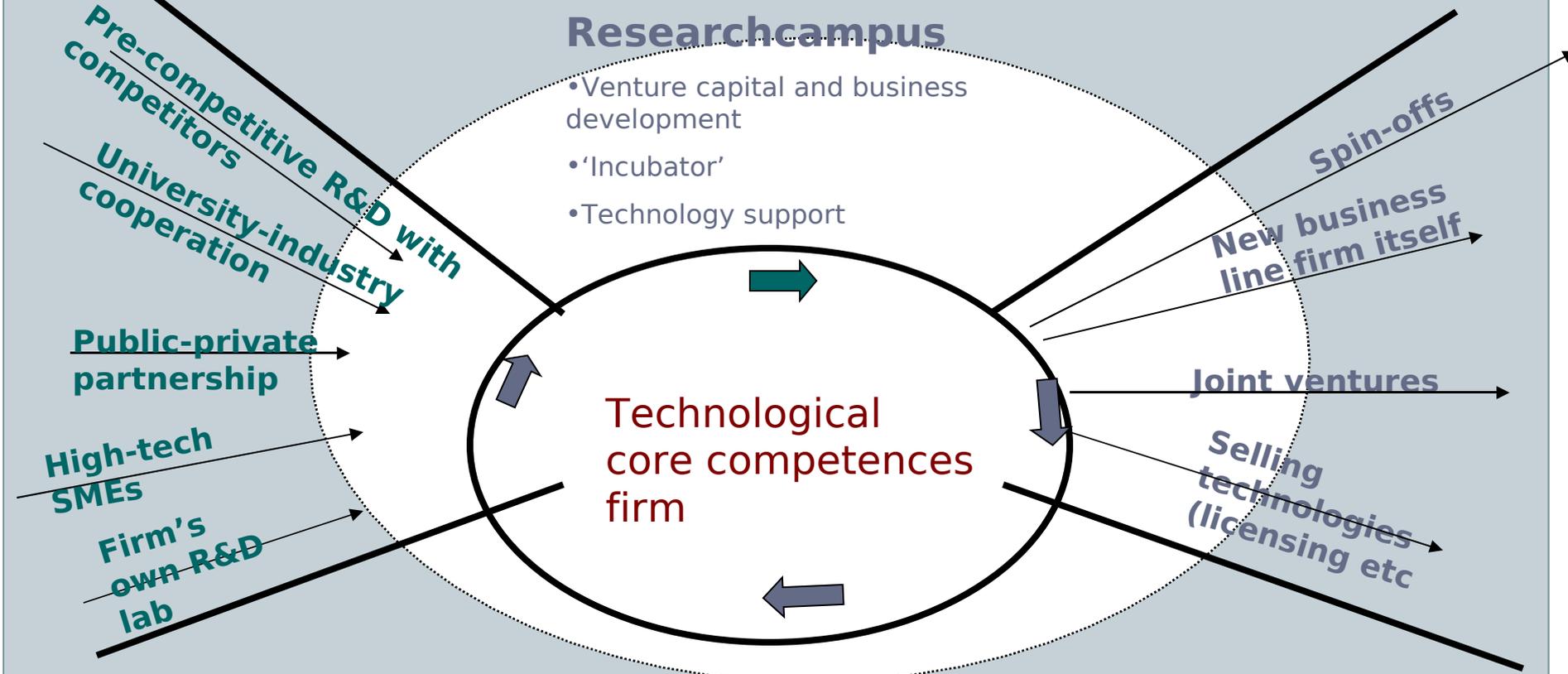
Worldwide search for knowledge and technologies

Value creation

Researchcampus

- Venture capital and business development
- 'Incubator'
- Technology support

Technological core competences firm



Pre-competitive R&D with competitors
University-industry cooperation

Public-private partnership

High-tech SMEs

Firm's own R&D lab

Spin-offs
New business line firm itself

Joint ventures

Selling technologies (licensing etc)

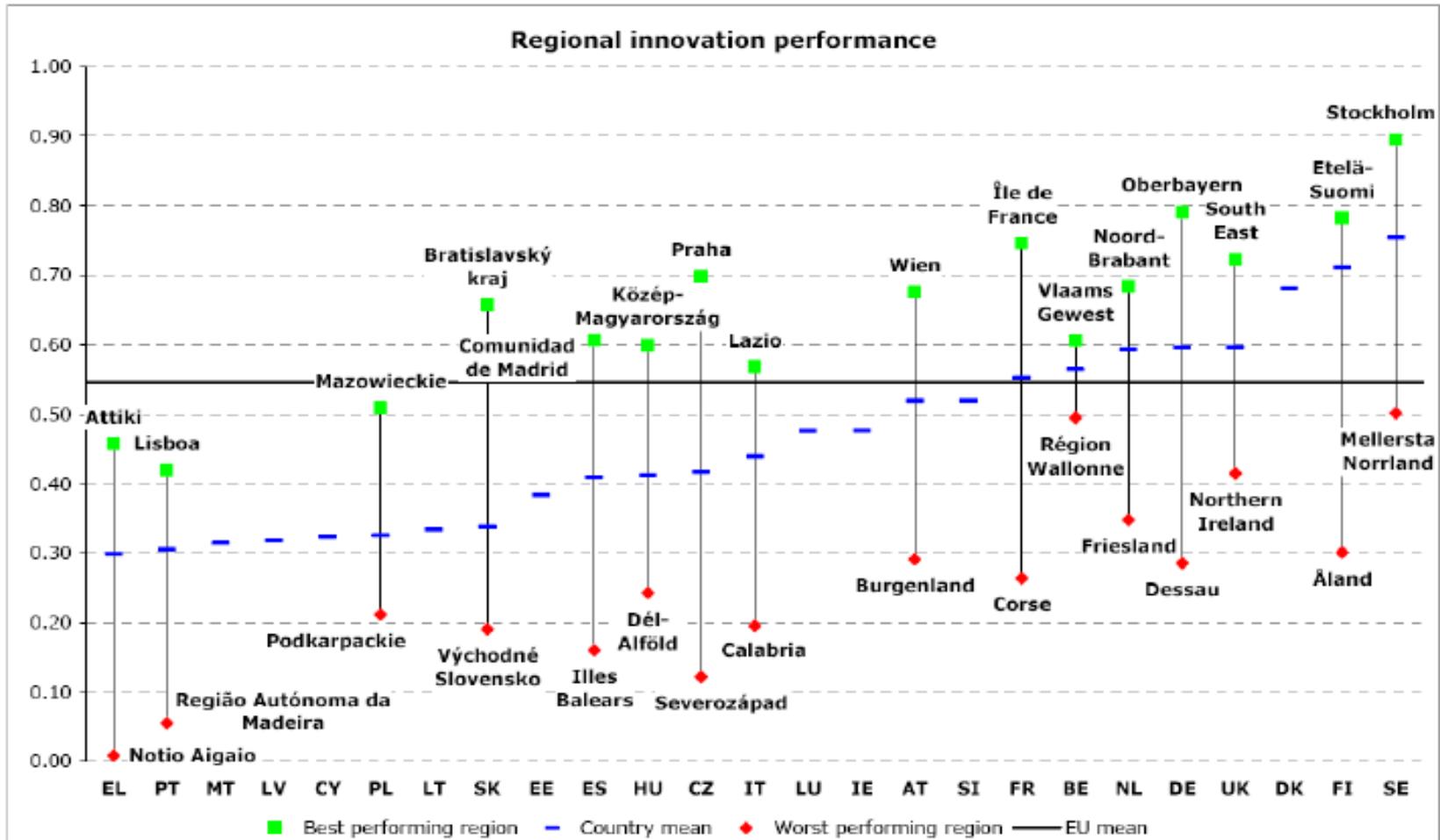
wider scan of technology areas

more focus and resources for core competences

Faster and more value creation

Innovative regions throughout EU

13



Source: UNU-MERIT

Research Infrastructures and regions

14

- Physical infrastructures will play increasing role in strengthening regional 'clusters':
example Grenoble: CEA-Leti, CNRS, ILL, ESRF, STMicroelectronics, NXP (former Philips), Minatec
- For first time regional development key argument in case of all 3 site contenders for European Spallation Source (Bilbao, Debrecen, Lund)
 - Not: local expenses during construction and operations
 - Not: medium- and long-term impact through technologies based on scientific output
 - ***But: cluster/network/agglomeration/spill-over***

What about distributed RI, e.g. Data Infrastructure

15

Key: look for specialisation and combinations

Examples

- Amsterdam: major computer science centre (CWI); strong computer science departments at two universities; Amsterdam Internet Exchange (world's largest); Supercomputer Centre; very modern CAVE (Virtual Reality Environment); one of LHC key data centres; strong focus (also in research) on media, networking, creative industry, ICT companies (hardware, software, content)
- Why not Frascati? With Earth Observation data as starting point
- Research and Business Campus Geleen (Netherlands) built around DSM's performance materials (incl biomaterials). Offer SMEs on-site access to and support for very wide gamut of state-of-art, expensive, remote imaging and analytical equipment