AVROSS

Accelerating Transition to Virtual Research Organization in Social Science

Funded by the European Commission,
Information Society and Media
Directorate-General
under Service Contract
No. 30-CE-0029513/00-42
27.10.2006 – 26.10.2007 (extension till 30.11.2007)

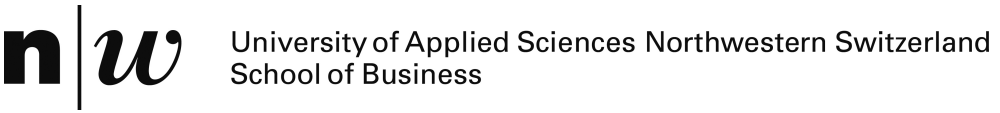
Main AVROSS goals

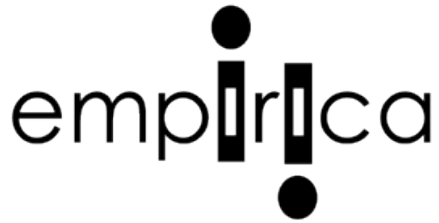
- Understanding the reasons behind the low level of adoption of e-Infrastructures in the social sciences and humanities (SSH)

- Supporting optimisation of Grid and GÉANT developments by
  - Providing an analysis and assessment of the current conditions of use
  - Providing guidance on how e-Infrastructures may be better deployed and exploited (by SSH)

Consortium

- School of Business, University of Applied Sciences Northwestern Switzerland
  (Franz Barjak, coordinator)
- empirica Technology and Communication Research, Bonn/Germany
  (Simon Robinson)
- National Centre for e-Social Science, Manchester/UK
  (Rob Procter)
- National Opinion Research Centre, University of Chicago/USA
  (Julia Lane)



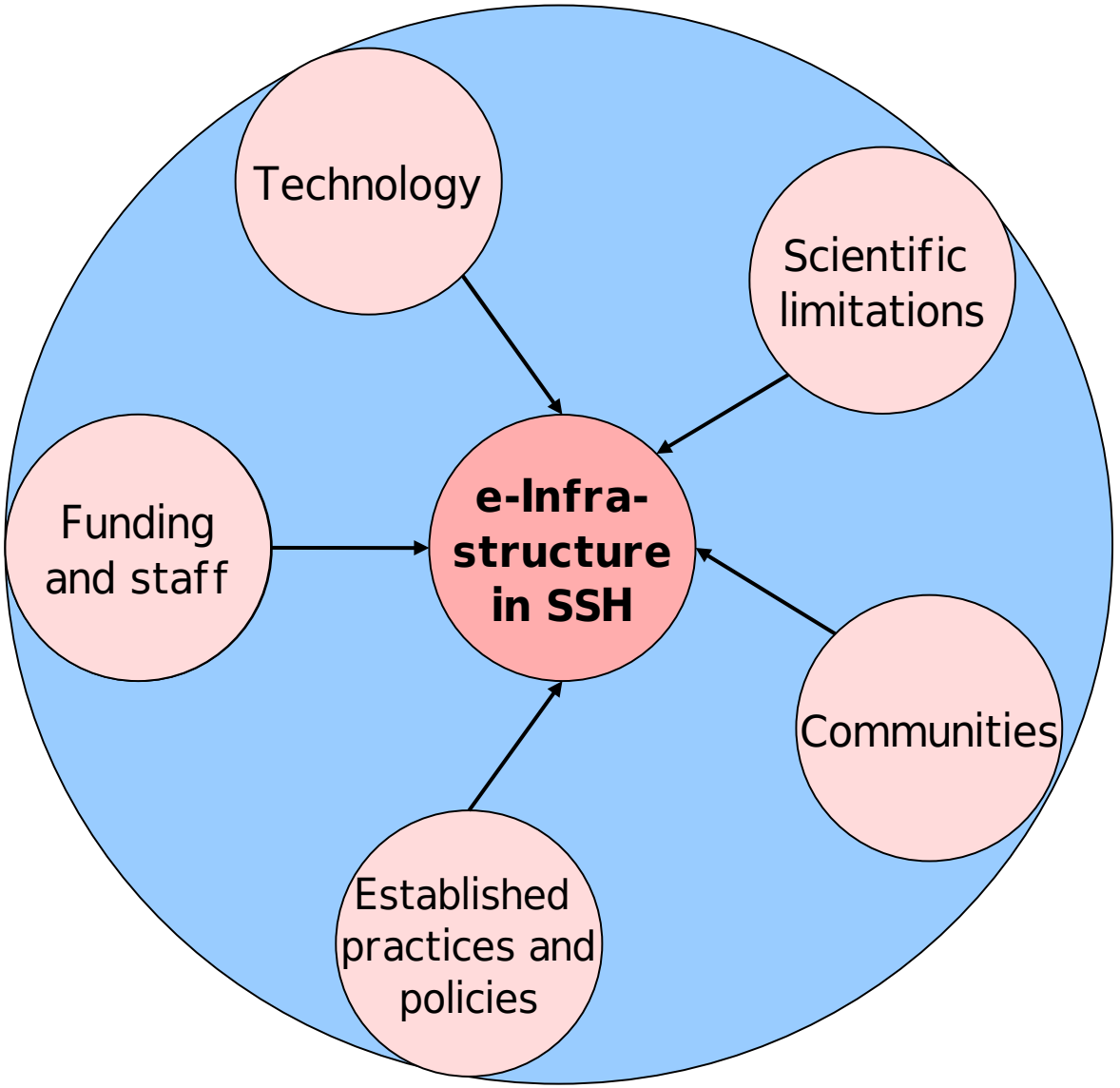


Elements

- WP 1: Theoretical framework:
  - Social shaping of technology
- WP 2: Stock-taking of e-social science initiatives in four fields
  - Survey of early adopters and enthusiasts of eInfrastructures in SSH in Europe, the US and beyond
- WP 3: Analysis of eight approaches to using e-infrastructures
  - Case studies
- WP 4: Policy recommendations

WP 1: Theoretical Framework

Figure 1:
"Social" shaping of technology approach



WP 2: Stock-taking of e-social science initiatives in four fields

Survey of early adopters of e-infrastructures in/for SSH

- Individuals (not institutions) that had appeared in conferences, workshops etc. worldwide → snowballing approach
  - Advantage: broad ranging and informed sample
  - Disadvantage: not representative as we don't know anything on the characteristics of the universe

- Focus on four fields, but responses from outside were included, too
  - Economic and social research
  - Geography and regional science
  - Archaeology
  - Computer linguistics

Survey statistics

- Total “universe”: 1917 mail addresses in 45 countries and 5 institutional TLD (.edu, .com, .org, .net, .gov)
- Total responses: 561
- Gross response rate: 29.2% of the universe
- Usable responses: 448 (23.4% of the universe)

Table 1: Country distribution (most important countries)

Country	„Universe“			Sample		
	TLD	No.	%	No.	%	RR in %
UK	ac.uk, co.uk, gov.uk, nhs.uk	501	26.1%	182	32.4%	35.7%
Educational	edu	587	30.6%	135	24.1%	23.0%
Germany	de	276	14.4%	69	12.3%	25.0%
Commercial	com	100	5.2%	23	4.1%	23.0%
Organisation	org	84	4.4%	19	3.4%	22.6%
Netherlands	nl	39	2.0%	15	2.7%	38.5%
Australia	au	38	2.0%	14	2.5%	36.8%
Canada	ca	20	1.0%	12	2.1%	60.0%
France	fr	26	1.4%	10	1.8%	38.5%
Italia	it	22	1.1%	10	1.8%	45.5%
Switzerland	ch	11	0.6%	8	1.4%	72.7%

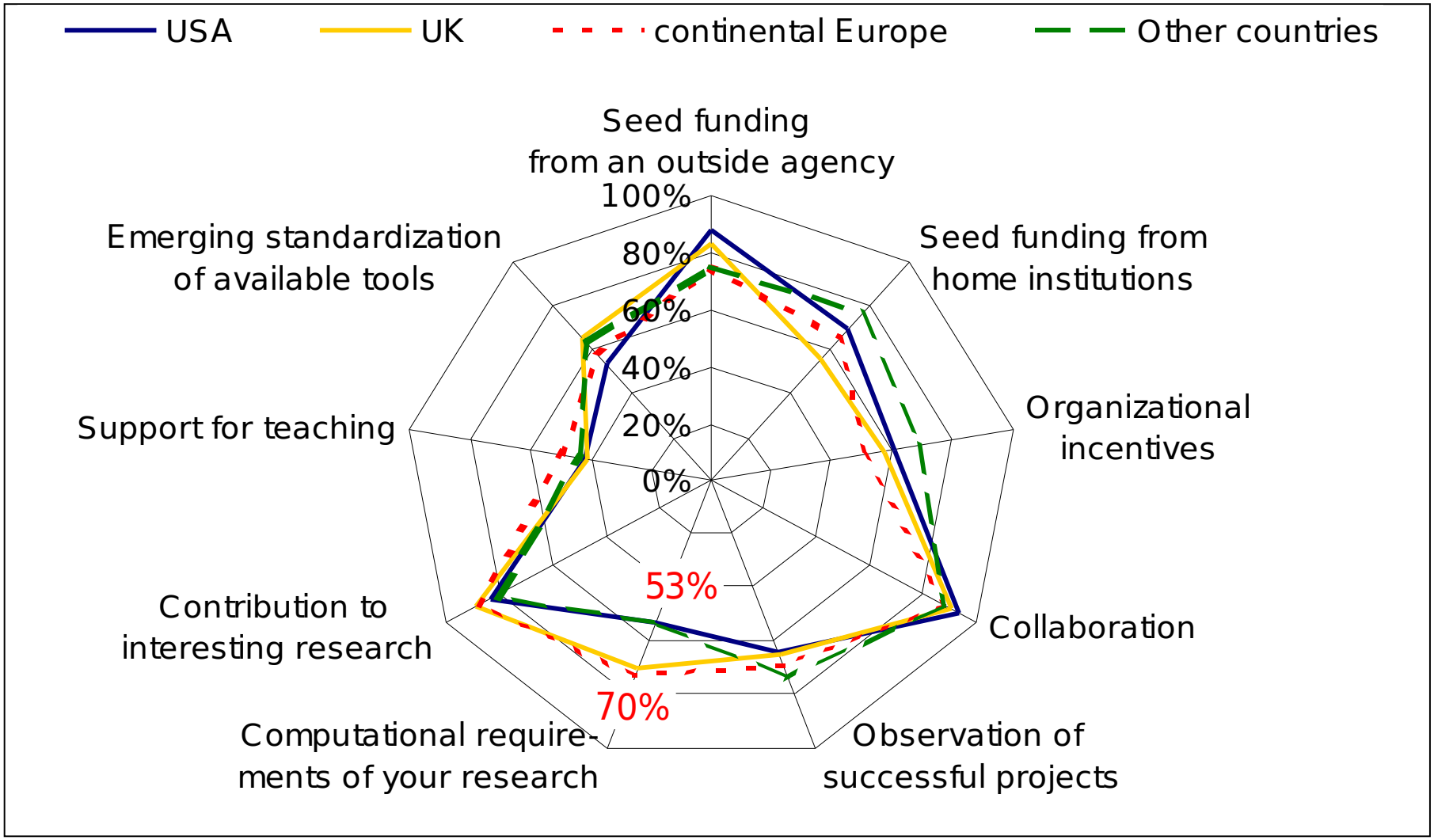
Table 2: Sources of information when beginning to use e-Infrastructure (in % of responses)

Source	Very imp.	Somewhat imp.	Neutral	Somewhat unimp.	Not at all imp.
Meetings or workshops providing information on e-Infrastructure	29.0%	29.0%	20.8%	9.7%	11.6%
Infrastructure or administration people at your own organization	31.6%	28.2%	13.6%	11.2%	15.5%
Infrastructure or administration people from other organizations	32.4%	38.1%	17.1%	4.3%	8.1%
Journal, magazine, or other printed or electronic information source	13.2%	30.4%	26.5%	12.7%	17.2%
Other scientists, colleagues, or collaborators	54.5%	32.9%	9.4%	1.9%	1.4%

Table 3: Catalysts for e-Infrastructure adoption (cases)

Catalyst	Very imp.	Some-what imp.	Neutral	Somewhat unimp.	Not at all imp.
Seed funding from an outside agency	100	40	16	5	12
Seed funding from home institutions	59	52	26	16	18
Organizational incentives within your institution	44	53	38	11	22
Collaboration	117	45	14	3	0
Observation of successful projects in other areas	42	70	37	11	7
The computational requirements of your research	54	54	38	11	13
Contribution to interesting research expected	94	54	22	0	3
Support for teaching activities	26	49	42	29	25
Emerging stand. of available tools	39	60	33	22	14

Figure 2: Catalysts for e-Infrastructure adoption by country (% of respondents considering a catalyst at least important)



Barriers to e-Infrastructure adoption

- Main barriers: funds & staff
  - Lack of initial funding
  - Costs associated with e-infrastructure development
  - Lack of staff available to help with development
- For later adopters (after 2000):
  - higher importance of problems related to protecting confidentiality of data
- Success depends very much on involvement of and interaction with users according to open-ended questions

Some quotes on user involvement

"Keep users involved in all stages and **find 'champions' among domain scientists"**

"**Leadership must come from members of the domain community** (e.g., a humanities or social science faculty member) - and not from a computer or computational scientist. Relying on CI centers (e.g., NCSA or SDSC) only engenders 'learned helplessness.' It is better to adopt less ambitious technology that can be controlled/customized by humanities/social scientist users then to depend on the latest thing from the centers (which produces a state of dependency)."

"Keep it practical and applied. Developing a tool is applied work for the community, it is NOT your ticket to a long ride on the academic granting gravy train. People who use these programs to advance their academic career rather than produce robust tools in a timely manner are destroying some schemes. **Equally, technologists who have little idea about what researchers need are responsible for many expensive projects which are never used**."

"Don't wait for the tool to be 'perfect' - get using it for research as soon as possible because the **development of the tool should obviously be in the context of particular research projects**. The tool is useless if it isn't being used to generate research outcomes that are being published in respected social science journals."

Some facts on e-Infrastructure projects in SSH

- Budget:
  - Median budget: 335,000 Euros
  - Median annual budget: 108,000 Euros
- Average duration: 36 months
- Average staff:
  - In total 14 individuals of which are
  - 5 (36%) scientists
  - 3 (21%) graduate students
  - 6 (43%) technicians and others

Figure 3: Median project budget by country

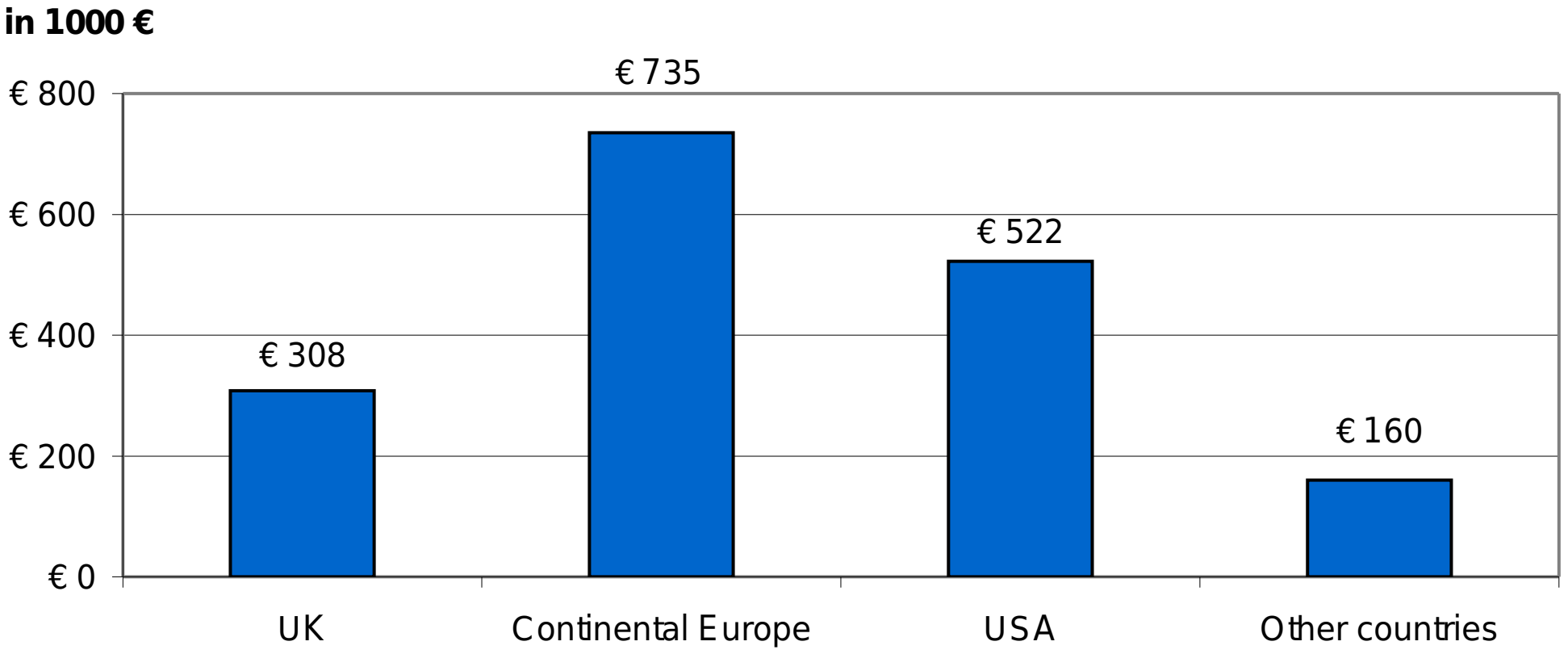


Figure 3: Median project budget by participating field

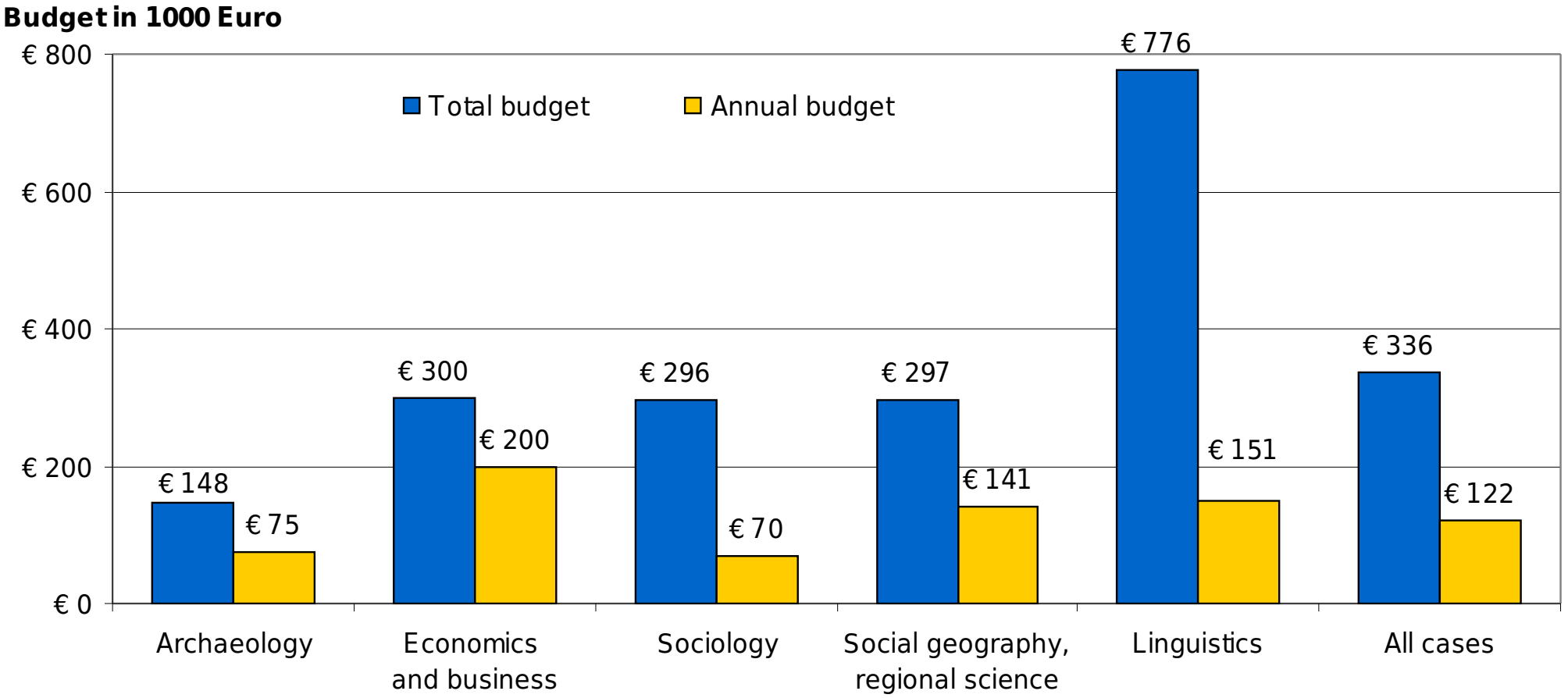


Table 4: Technological items of the projects by country (% of responses)

	USA	UK	Continental Europe	Other countries
High performance computing	45%	39%	38%	38%
High performance communication	62%	46%	45%	71%
High bandwidth	76%	53%	77%	65%
Distributed data, data repository	75%	82%	93%	82%
Collaboration tools/systems	83%	77%	84%	89%
Learning environments	53%	36%	45%	31%
Grid-enabled videoconferencing	37%	44%	21%	44%
Virtual/3D environments	24%	18%	18%	13%
Innovative data collection methods	45%	39%	43%	53%

WP3 Analysis of eight approaches to using e-infrastructures

Approach

- Selection of cases based on project characteristics
  - Technology
  - Success
  - Size
  - Accessibility

  → 3 continental European
  3 UK
  2 US cases

- Data collection through
  - Semi-structured interviews
  - Archival research

Some stylised findings

1. Cross-disciplinary communication and collaboration are problematic
   - SSH: limited IT skills, support and willingness to invest time on this
   - Developers: technical language and focus, often not capable of dealing with field-specific practices, conventions and standards

Possible solutions:

- Micro-teams of domain and computer scientists
- Institutionalised user-developer collaboration
- Engaging “translators”, individuals trained in both fields

1. Establishing a user community in the field is important but difficult

- Possible solution: Involve leading domain scientist

3. Service-oriented business models would be appropriate
   - SSH demand rather small-scale advanced computing and support services (that are the opposite to the HEP grid approach)
   - Revenue models are still missing

Possible solution: publicly funded services/e-infrastructure ...?

WP4 Policy recommendations

Recommendations are still in the making …

… one issue related to today's topic:

→ **Storage and reuse of SSH data should be supported, because content is as important as technology**

- Increase requirements for projects to tag and share data generated with public funds,
- Make use of data archives mandatory for primary data (e.g. CESSDA),
- Make sure that data archives are capable of guaranteeing different degrees of anonymisation and data protection,
- Nudge SSH fields to design solutions for scientific credit and ownership rights to data.

Detailed presentation and discussion of findings:
Workshop with the European Commission
Brussels, November 27th, 2007

For further information:
http://www.fhnw.ch/plattformen/avross

Thank you for your attention and don't hesitate to talk to us!
franz.barjak@fhnw.ch

Extra slides

Definition of e-infrastructures

"For the purposes of this study, eInfrastructures are defined as integrated ICT-based research infrastructures. Key elements include networking infrastructures, middleware and organisation and various types of resources (such as supercomputers, sensors, data and storage facilities). The definition includes "old" components like supercomputers, the World Wide Web, or e-mail, but requires them to be part of an integrated system. The only requirement for any component is that it should be able to exchange information at some point through a standardized interface like a grid protocol." (based on the e-IRG definition)

Country	Universe			Sample		
	TLD	No.	%	No.	%	RR in %
New Zealand	nz	13	0.64%	7	1.25%	54%
Austria	at	16	0.79%	6	1.07%	38%
Governmental	gov	31	1.53%	6	1.07%	19%
Greece	gr	7	0.35%	5	0.89%	71%
Sweden	se	21	1.04%	5	0.89%	24%
Estonia	ee	4	0.20%	4	0.71%	100%
Norway	no	8	0.39%	4	0.71%	50%
Slovenia	si	5	0.25%	4	0.71%	80%
Croatia	hr	3	0.15%	3	0.53%	100%
Ireland	ie	8	0.39%	3	0.53%	38%
Portugal	pt	11	0.54%	3	0.53%	27%