

ECRI 2007[★]


★ *european conference on research infrastructures*

Research Infrastructure Initiatives in the Social Sciences and Humanities

Peter Wittenburg, ECRI 2007, Hamburg, 6.6.2007

Max Planck Insitut für Psycholinguistik



Complex Systems

we are living in the “*century of the complex systems*”

this is what is partly meant by the ESRI Roadmap document:

- *if humans want to take an action they need to predict what will happen*
- *since whatever we do has strong consequences on nature, minds, societies in various dimensions, we need complex models to better understand*

establishing these complex models is the task of science

complexity character is true for natural sciences, life sciences AND *social sciences and humanities*



Challenges in SSH

live in a world with continuous adaptations and changes

(nature, character, health, education etc)

economy requires increasing mobility of humans and

live in a time of gigantic migration streams

(cultural barriers, linguistic barriers, psychological effects, identity forming, group behavior, education etc)

How can we tackle these challenges?

J. Taylor

“eScience is about global collaboration in key areas of science and the **next generation of infrastructures** that will enable it”

Need to build new types of platforms

- to allow researchers to combine existing resources easily to new ones to tackle the big challenges
- to increase the productivity of all interested researchers, since currently too much time is wasted by preparatory work



Research Infrastructures



- Research Infrastructures are not just another set of short-term technology projects
- RIs are long-term investments in service oriented facilities to enable first rate research by sharing and combining resources
- RIs have a technological component and they require new and smart standards
- but even more important is a proper solution of the **organizational/legal/administrative** aspects (trust relations, budgets for years, operation models)

Need to overcome many barriers in SSH



At the specification level:

- data structures need to become well specified
- terminology needs to be defined
- software interfaces need to be documented
- etc etc

At the access level:

- need to digitize and archive data
- need to support localization of data and tools
- data and tools need to become re-usable
- need to support workflows
- etc etc



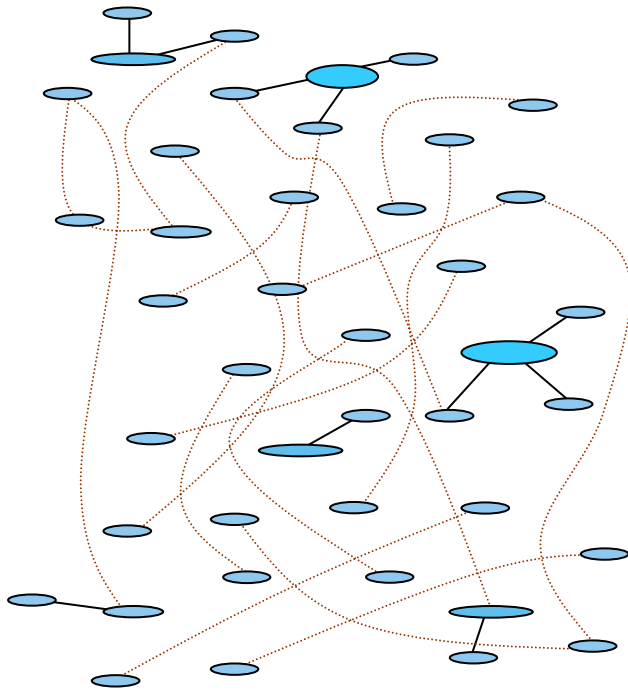
landscape of resources and tools in SSH needs to overcome its fragmentation



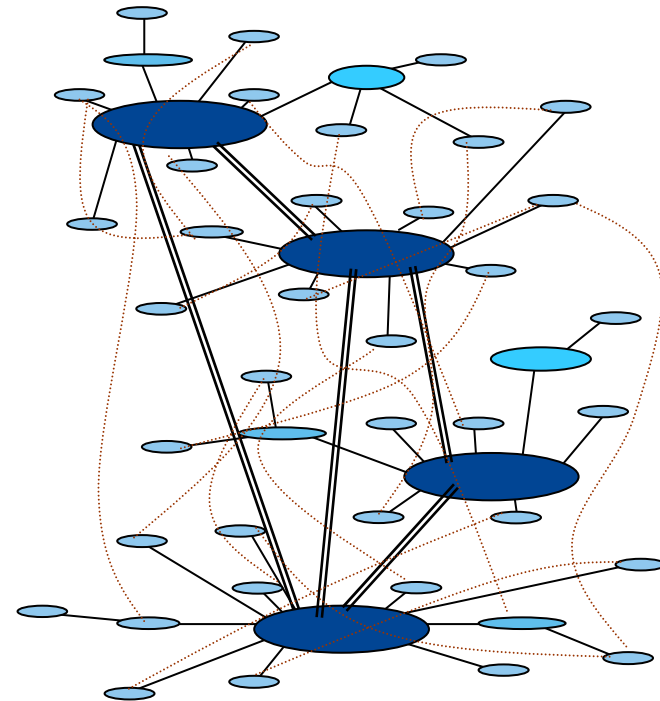
Are RI for SSH different?

- physics is independent of cultural + linguistic differences – SSH are
- SSH are different to natural sciences where for example the HEP community shares ONE big facility such as CERN
- in SSH resources and tools are created in a distributed manner
 - in Soc Sci many different culture specific polls by different teams
 - in Linguistics as part of the Humanities
 - texts, annotated recordings, etc – language specific
 - lexica, ontologies, etc – language and project specific
 - parsers, aligners, etc – partly language and project specific
 - all created by a large number of teams – only little is explicit in all dimensions
- to overcome fragmentation, we certainly need to establish a network of centers to maintain the rich landscape of resources and tools and to make them accessible

Are SSH RI different?



from a basically unorganized domain
of small occasional networks and
temporary collaborations



to an added structured domain with
respected centers that offer stable
access in a quickly changing world

Research Infrastructures in Social Sciences

CESSDA (Council of European Social Science Data Archives)

- distributed RI providing and facilitating access of researchers to high quality data and supporting their use.
- is network of institutes from 21 countries in Europe holds **15,000** data collections and provides access to 20,000 researchers
- **proposed is a major upgrade**



ESS (European Social Survey)

- operational from 2001 to monitor long term changes in social values throughout Europe covering 27 EU countries
- produce data relevant to academic debate, policy analysis and better governance
- **proposed is an upgrade and long-term support**



SHARE (Survey of Health, Ageing and Retirement in Europe)

- provides data infrastructure for economic and social science analyses of the on-going changes due to population ageing
- original 8-country survey has already being expanded to cover two new Member States;
- **proposed is an extension to all 25 Member States**



Research Infrastructures in Humanities

CLARIN (Common Language Resources and Technology Infra)

- want to make language resources and technology available and useful to scholars of all disciplines working with language material
- overcome the present fragmented situation by harmonising structural and terminological differences,
- basis is a Grid-type of infrastructure and Semantic Web technology.
- proposed is a EU wide collaboration (31+ countries)

DARIAH (Digital RI for the Arts and Humanities)

- want to be build a framework for unification and collaboration in the humanities
- will be based upon an existing network of Data Centres and Services (currently 12 centres)
- model is open also to new disciplines
- proposed is an organization with a glue function





How to organize such a community – the CLARIN example ?

- Language Resource and Technology field has long history of collaboration
- always project oriented where each partner had individual interests
- also many conferences over the years where most of the community meets
- but no broad coverage until now when looking for example to standards
- needed 3 special meetings to make the field ready for a RI proposal
 - every relevant institute in Europe (now about 90 from 31 states)
 - in each EU country and beyond we localized respected experts
 - national group formation by bottom-up process – different per country
 - needed to transmit the main message: **minds need to change!**
 - LRT field is becoming a service field to a certain extent
 - LRT needs to be accessible and shared
 - LRT people need to adhere to standards and agreements
 - LRT landscape will change extremely
- needed some committed persons without own hidden agenda

It takes a long time to bring excellence to maturity.

Publilius Syrus (~100 BC)

- building RI in SSH costs time
 - due to complex community processes
 - stakeholders need to be convinced
 - researchers need to be convinced to rely on RI in competitive times
- all these aspects are more difficult to achieve in SSH
- building RI is costly
- building RI is nothing for impatient people



Let's thank those bright minds and the EC who organized the ESFRI process and the plan to fund Research Infrastructures. This is all just in time.



Thanks for your attention!