

New European Infrastructural and Networking Initiatives

Nicoletta Calzolari

Istituto di Linguistica Computazionale del CNR, Pisa, Italy
glottolo@ilc.cnr.it



LangTech, Roma, February 2008

1

Today ...

The wealth of data & of basic technologies is such that:

■ We should reflect again at the field as a whole & ask if

■ Standards

← Content interoperability

■ Creation of LRs

← Collaborative creation & Manag.

■ Automatic acquisition

← Dynamic LRs

■ Distribution

← Sharing

are still “the” important components,
or how they have changed/must change

could be at the basis of a
new Paradigm for LRs & LT
& of a new Infrastructure ??

... Which new challenges towards a
new & more mature infrastructure of LRs & LTs??

Why an Infrastructure & a Network of LRs & LTs?

Many dimensions around the notion of language

Political issues

e.g. a commonly agreed list of minimal requirements for
“national” LRs: **BLARK**

Putting together **technical, organisational, strategic, economic, political issues** of LRs

Multilingualism

Cultural issues

→ Language
→ Cultural identity
→ Humanities

Need of bodies for a broad research agenda & strategic actions for LT&LRs (W/S /MM)

Economic, social

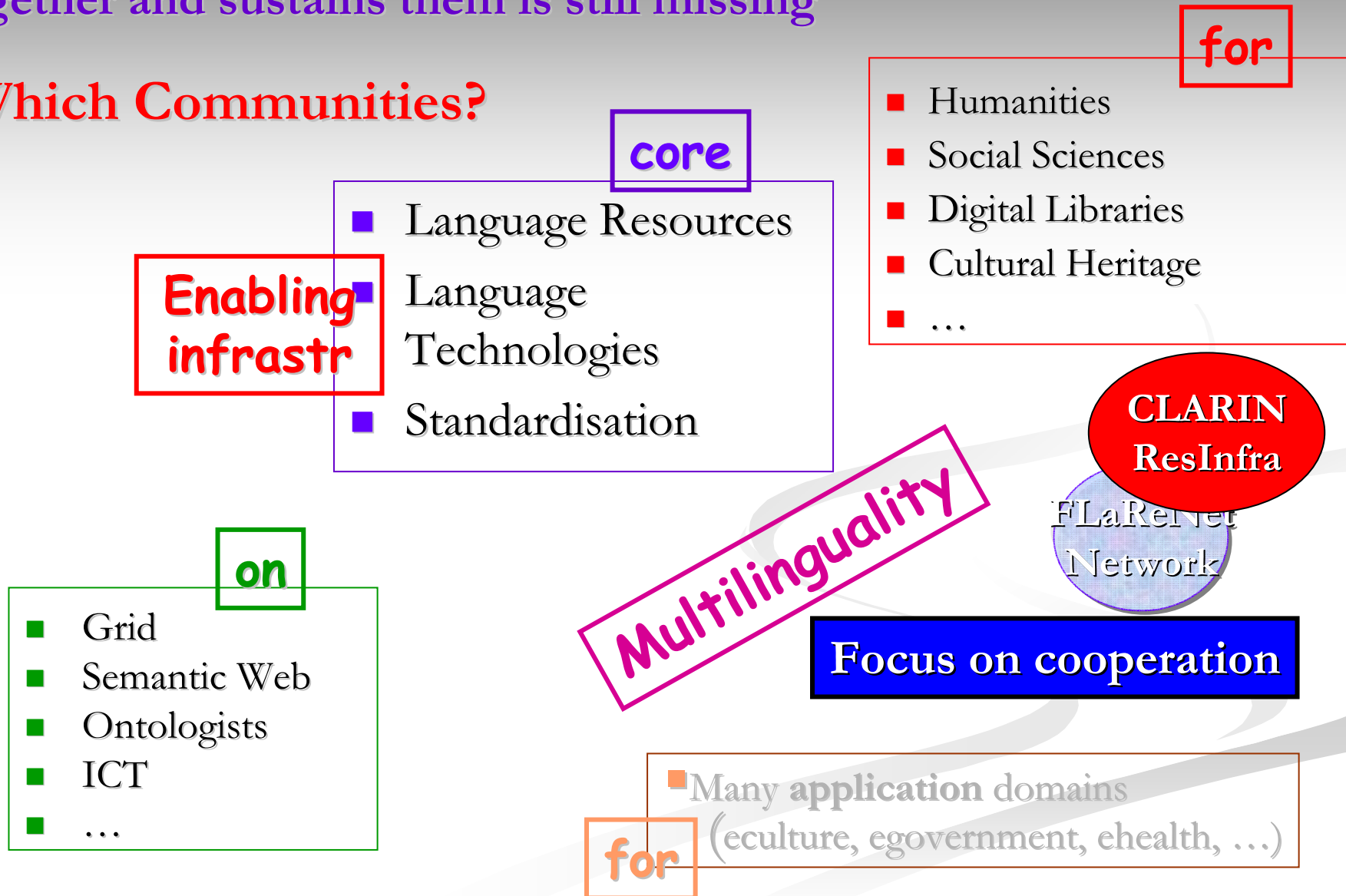
- Applications
- Services

Interdisciplinarity & Multidisciplinarity

Technical issues

Technologies exist, but the infrastructure that puts them together and sustains them is still missing

Which Communities?



CLARIN

Common Language Resources and Technologies Infrastructure for the Humanities & Social Sciences

Large-scale pan-European collaborative effort (**31+ countries**)

- **Make LRs & LTs available** & readily usable **to scholars of humanities & social sciences** (& all disciplines)
- Need to **overcome the present fragmented situation by harmonising** structural and terminological differences
- Basis is a **Grid-type** infrastructure and Semantic Web technology
- The benefits of computer enhanced language processing become available only when a **critical mass of coordinated effort is invested in building an enabling infrastructure**, which can **provide services in the form of provision of tools & resources** as well as **training & counseling** across a wide span of domains
- The infrastructure will be based on a number of resource, service and expertise **centres**

CLARIN Mission

- Create a **comprehensive and free to use distributed archive of LRs & LTs** covering not only the languages of all member states, but also other languages studied and used in Europe
- Through the fact that the **tools & resources** will be **interoperable across languages & domains**, contribute to preserving and supporting **multilingual & multicultural European heritage**
- An operational **open infrastructure of web services** will introduce a **new paradigm of distributed collaborative development**
- Allow **many contributors to add all kinds of new services** based on existing ones, thus ensuring reusability and allowing scaling up to suit individual needs



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



from P. Wittenburg

eScience Vision

CLARIN establishes such a **new generation** of extended infrastructure

Thus CLARIN is **not** about creating and building new language resources and technology, but

- making them available and accessible
- as **services**
- in a stable and persistent infrastructure

to allow tackling the great challenges

CLARIN: <http://www.clarin.eu>

Grid Project: <http://www.mpi.nl/dam-lr>

ISO TC37/SC4: <http://www.tc37sc4.org>

Standards Project: <http://lirics.loria.fr/>

from P. Wittenburg

We have still a long path ...

& also a “new project”

in an **e-Contentplus** Call for a:

- **“Thematic Network on Language Resources”:**

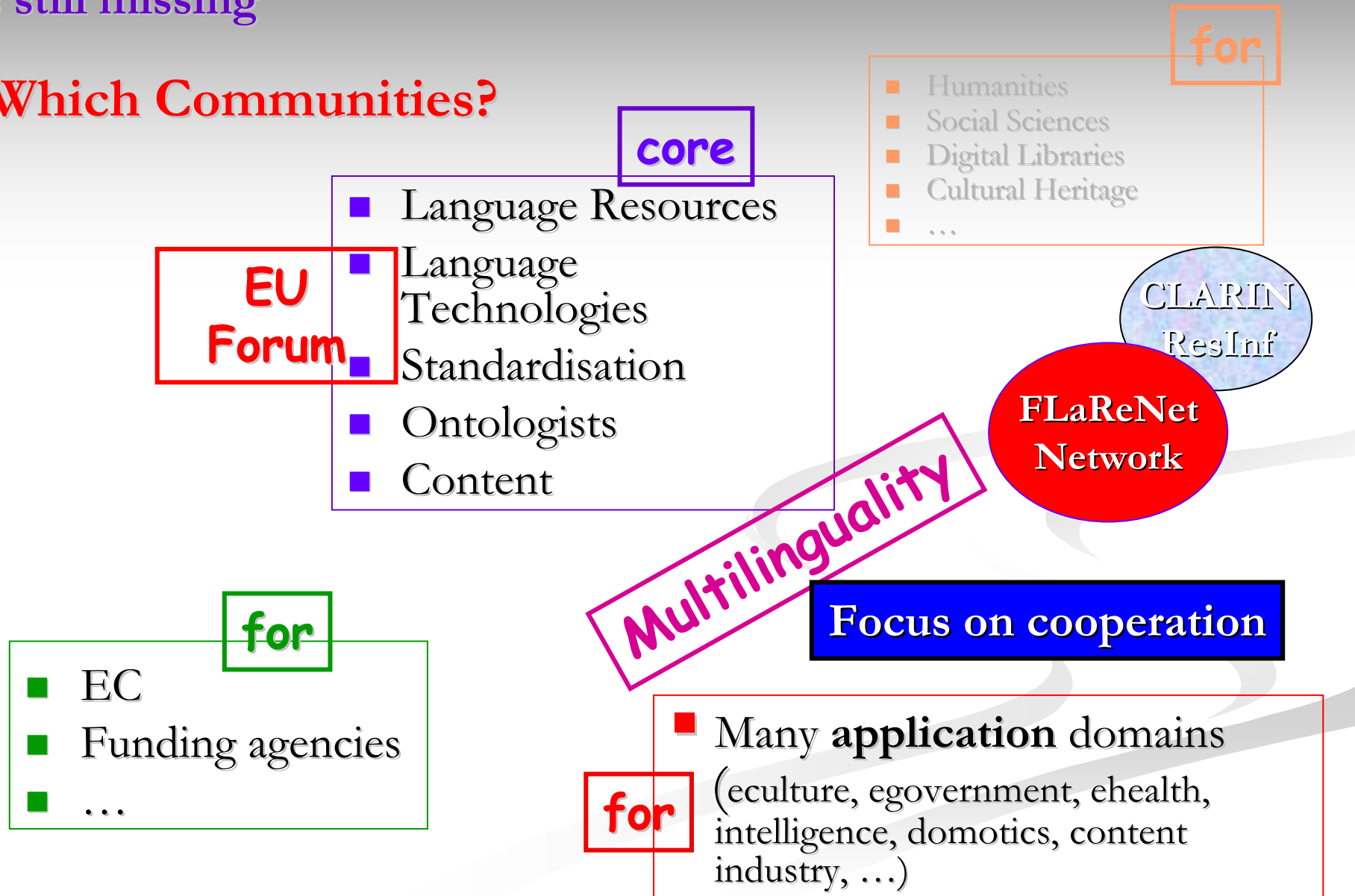
FLaReNet

- To provide common recommendations (to the EC) for future actions
- To give priorities
- Need of ‘visions’

In a global context, in cooperation with **CLARIN**
& also with *non-EU members*

LRs & LTs exist, but a global vision, policy and strategy is still missing

Which Communities?



FLaReNet

Fostering Language Resources Network

A European forum

- to facilitate interaction among LR stakeholders

The Network structure considers that LRs present **various dimensions** and must be approached from many perspectives:

- technical, but also
- organisational
- economic
- legal
- political

Addresses also

- **multicultural and multilingual aspects**, essential when facing access and use of digital content in today's Europe

Organised in Thematic Working Groups

A **layered structure**, with **leading experts & groups** (national and European institutions, SMEs, large companies) for all relevant LR areas (**about 40 partners**)

- ◆ in collaboration with CLARIN
- ◆ to **ensure coherence of LR-related efforts in Europe**

FLaReNet will

- ⊙ **consolidate** existing knowledge, presenting it analytically and visibly
- ⊙ contribute to **structuring the area of LRs of the future** by discussing **new strategies** to:
 - ◆ convert existing and experimental technologies related to LRs into useful economic and societal benefits
 - ◆ integrate so far partial solutions into broader infrastructures
 - ◆ consolidate areas mature enough for recommendation of best practices
 - ◆ **anticipate the needs of new types of LRs**

Thematic Areas

- The **Chart for the area of LRs** in its different dimensions
- Methods and models for **LR building, reuse**, interlinking and maintenance
- **Harmonisation** of formats and standards
- Definition of **evaluation protocols** and evaluation procedures
- Methods for the **automatic construction and processing** of LRs

To build together:

- **Evolving RoadMap**
- **Blueprint of actions and infrastructures**

Objectives & expected results

The **largest Network of LR and HLT players**, with diverse approaches, efforts and technologies

- ④ Enable progress toward **community consensus**
- ④ Give an extended picture of LRs & **recast its definition** in the light of recent scientific, methodological, technological, social developments
- ④ **Consolidate** methods & approaches, common practices, frameworks and architectures
- ④ A **“roadmap”** identifying areas where consensus has been achieved or is emerging vs. areas where additional discussion and testing is required, together with an indication of **priorities**
- ④ Recommendations in the form of a **plan of coherent actions for the EU and national organizations**
- ④ A **European model for the LRs of the next years**



Ambitious!

Outcomes of FLaReNet

The outcomes will be **of a directive nature**

- to help the EC, and national funding agencies, **identifying priority areas** of LRs of major interest for the public that need public funding to develop or improve

A **blueprint of actions** will constitute **input to policy development both at EU and national level**

- for identifying new language policies that support linguistic diversity in Europe
- in combination with **strengthening the language product market**, e.g. for **new products & innovative services**, especially for less technologically advanced languages

These Initiatives, ... together

- Call for **international cooperation also outside Europe**

and will be relevant for

- setting up a global **worldwide Forum of Language Resources and Language Technologies**