

# **CLARIN & FLaReNet**

## **New European Infrastructural and Networking Initiatives**

**Nicoletta Calzolari**

Istituto di Linguistica Computazionale del CNR, Pisa, Italy  
[glottolo@ilc.cnr.it](mailto:glottolo@ilc.cnr.it)



# 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

### 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

**Multilingualism**

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

## Which Communities?

**Enabling  
infrastr**

**core**

- Language Resources
- Language Technologies
- Standardisation

**for**

- Humanities
- Social Sciences
- Digital Libraries
- Cultural Heritage
- ...

**on**

- Grid
- Semantic Web
- Ontologists
- ICT
- ...

**Multilinguality**

**Focus on cooperation**

**for**

- Many application domains  
(eculture, egovernment, ehealth, ...)

**CLARIN  
ResInfra**

FLaReNet  
Network

# 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

# What is CLARIN

- Basic idea:
  - European federation of digital archives with language data
  - with uniform access to the data
  - with access to language and speech technology tools as services, to manipulate, enhance, explore and exploit the data
  - target audience SSH researchers, NOT industry
  - all languages are equally important

# Phasing and funding

- Preparatory phase (2008-10) funded by the EU (4.1 M€), 32 consortium partners in 22 countries, > 130 member organisations in 32 countries, with additional funding from national governments (>> 5 M€)
- Construction phase (2011-15)
- Exploitation phase (2016-...)
- Total cost until 2020 (> 200 M €?)
- Construction and exploitation jointly funded by participating countries



# 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





# 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/>

# 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?

core

- Language Resources
- Language Technologies
- Standardisation
- Ontologists
- Content

EU  
Forum

for

- Humanities
- Social Sciences
- Digital Libraries
- Cultural Heritage
- ...

CLARIN  
ResInf

FLaReNet  
Network

Multilinguality

Focus on cooperation

for

- EC
- Funding agencies
- ...

for

- Many **application** domains  
(eculture, egovernment, ehealth, intelligence, domotics, content industry, ...)

# FLaReNet

## Fostering Language Resources Network

<http://www.ilc.cnr.it/flarenet/>

### A European forum

- to facilitate interaction among LR stakeholders

The Network structure considers that LRs present **various dimensions** :

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

must be approached from many perspectives & involve various communities

Addresses also

- multicultural and multilingual aspects**, essential when facing access and use of digital content in today's Europe
- To provide common recommendations (to the EC) for future actions**
- To give priorities**
- Need of 'visions'**

# FLaReNet Objectives

- **Re-Creation of the LR Community**
  - According to a very broad definition of LRs
- **International cooperation**
  - European Forum to facilitate interaction among LR stakeholders (also SMEs & companies)
- **Considers LRs along all the dimensions**
  - technical, but also
  - organisational, economic, legal, political
  - addressing also multicultural and multilingual aspects

# FLaReNet Objectives

## ■ A Roadmap for the LRs

- Consolidation & structuring of the field
- Overall structure and main building blocks
- Looking into the future of the field, proposing new strategies, anticipating the need for new LR types, ...

## ■ Outcome of a directive nature

- Recommendations for initiatives to EC & funding agencies
- Input to policy development both at EU and national level: a EU model for the LRs of the next years

# Objectives & expected results

The **largest Network of LR & HLT players**, with diverse approaches, to **structure the area of LRs of the future** by discussing **new strategies**:

- ⊗ Enable progress toward **community consensus**
  - ⊗ Give an extended picture of LRs & **recast its definition** in the light of recent scientific, methodological, technological, social developments
  - ⊗ **Consolidate** methods & approaches, common practices, frameworks & architectures
  - ⊗ **Integrate** so far partial solutions into broader infrastructures
  - ⊗ Convert existing & experimental LTs related to LRs into useful economic & societal benefits
  - ⊗ Anticipate the needs of **new types of LRs**
  - ⊗ 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**
- ➡ Recommendations in the form of a **plan of coherent actions for the EU and national organizations**

**Ambitious!**

# 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**

To be enlarged

- ◆ with new members &
- ◆ the involvement of the community



# 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**

# 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
- ➡ A **European model for the LRs of the next years**

# CLARIN & FLaReNet, ... 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**



# Title

- Text
  - Text
    - Text
      - Text



# FLaReNet WPs

- **1 Project Management and Dissemination**
- **2 The Chart for the area of LRs in its different dimensions**
- **3 Methods and models for LR building, reuse, interlinking and maintenance**
- **4 Harmonisation of formats and standards**
- **5 Definition of evaluation protocols and evaluation procedures**
- **6 Methods for the automatic construction and processing of LRs**
- **7 Evolving RoadMap**
- **8 Blueprint of actions and infrastructures**