Many LRs & LTs exist, but a global vision, policy & strategy is needed

Which Communities?
- Language Resources
- Language Technologies
- Standardisation
- Content/Ontologies
- System developers
- Integrators
- SSH

Many dimensions

Many applications/domains
- MT
- CLIR
- ...
- e-government
- content industry
- intelligence
- e-culture
- e-health
- domotics...

Need to consider together
- technical
- organisational
- strategic
- economic, social
- cultural
- legal
- political issues wrt LRs & LTs

Focus on cooperation

EU Forum

CLARIN for SSH

FLaReNet Network

META-NET NoE

Multilinguality

for

with

EC
- National funding agencies
- Industry

N. Calzolari

Leipzig, July 2010
An international Forum to facilitate interaction, to

- **Overcome the fragmentation** in LR & LT & recreate a community
- **Anticipate** the needs of **new types of LR & LT & Language Infrastructures**
- Create a **shared policy** for the next years
- Foster a **European strategy** for consolidating the sector

- A “roadmap”: a **plan of actions as input to policy development**
- A (Eu) model for the LRs/LTs area of the next years

Community mobilisation:

- 102 National contacts from 76 countries
- **99 Institutional Members** From 33 countries
- **383 Individual Subscribers**

N. Calzolari
Multilingual Web, Madrid, October 2010
Since few years

A new paradigm of R&D in LRs & LT

Open & distributed linguistic infrastructures for LRs & LT

Adopting the paradigm of accumulation of knowledge & results, so successful in more mature disciplines, based on sharing LRs, tools & results

Ability to build on each other achievements, allowing controlled & effective cooperation of many groups on common tasks (see HumanGenomeProject)

Emerging concept of collective intelligence

Emphasize interoperability among LRs & LT

e. g. initiatives to achieve international consensus on annotation guidelines
META-SHARE is a big step ... that needs a real Paradigm shift

Assemble a comprehensive "map of language data and mechanisms" for the planet’s languages (LRE Map)

LR building as collaborative “common shared task”

New methodology of work

Interoperability acquires even more value

Needs consensual planning of common strategies towards shared objectives

Not just the sum of many individual efforts

But an organised, well-structured, collective enterprise

We wanted more & more data

... Have we been too successful ?!?
We experience today a sort of statistical “intoxication”!

- It started as a new strategy, a revolution maybe? But it has turned to tactics. Stuck with it? In a narrow loop of small advances, not linked to each other.
- Can we add also a new strategy? and hopefully a vision?

Main Statement

- We tend to forget about “language” & the need to understand its properties & complexities.

- Where do we (try to) encode what we know about language properties?

In annotations/encodings

- Is there any serious methodology of studying and exploiting the interactions among the various annotation layers?
To a concrete plan ...

**Vision**

*Like the Genome project, ... a huge Language initiative*

**Strategy**

A “Multilingual Annotation Library”
As a Very Large International Initiative

**MANY** (parallel?) texts for **MANY** languages
With **ALL** possible annotation layers

Similar to **more mature sciences**, e.g. physics, ...
with **thousands of people working together** on the same big experiment
Create **a sustainable infrastructure** for a **large Language repository plan**

Where we **accumulate all the knowledge we have about language**

**Encourage analysis of linguistic interrelations**

---

**As a Collaborative Resource:**
A new paradigm for a big language map

---

The challenge of interoperability among layers with different annotations

**The major challenges:**

- At the **organisational/design** level?
- At the **community involvement** level?
Means a change of mentality: going beyond “individual” research interests.

From “my approach” to some “compromise” allowing to go for big amounts/integration/building on each other/…

Could be a great framework for practicing & experimenting interoperability.

Also multilingually.

Interoperability issues.
Sinergie fra ricerca di base & ricerca applicata

Promotori di nuovi “paradigmi”
- Estrazione di conoscenza da dizionari/testi: precursori
- Settore delle “Risorse Linguistiche”: lanciato da noi
- ...

→ indirizzare la ricerca

Competitività sul mercato della ricerca
- Contratti Europei (ILC-CNR: 51 & 15: Coordinatori)
- Progetti Nazionali

Riconosciuti centri di eccellenza
- Collaborazioni internazionali, nazionali, industrie, ministeri, regione, PPAA, ...
- Principali Comitati/Board/Associazioni (ELRA, ELSNET, ICCL, ICCWLRE, SIGLEX, SENSEVAL, etc.)
Attività Strategiche di Coordinamento:
.... ruolo dell’Italia

- **Coordinatori Rete Europea dei Progetti Nazionali** per le TL
- **Coordinatori** iniziative internazionali (EU-US + Asia) per la definizione di **Standard** per le TL: ora in **ISO**
- **Coordinatori** iniziative internazionali infrastrutturali per **creazione** e acquisizione di **Risorse Linguistiche**
- **Coordinatori** progetti **Traduzione Automatica del Parlato**
- ....
- LREC, ELRA, ICCWLR, ELSNET, ...
- ....
- Promotore di proposta nello schema **ERA-Net 6PQ**: coordinamento **Progetti Nazionali di TL in Europa**
- ....
Roma, Conferenza stampa, 28 Settembre 2004

Ricadute Economiche e Occupazionali...

- **Autofinanziamento** consistente
  (“anomali” nel settore delle Scienze Umanistiche)

- Contratti a tempo determinato
- Trasferimento tecnologico verso PMI

→ formazione

... ricadute Culturali

Sinergie fra

- trattamento e promozione dell’Italiano (anche in rete)
- trattamento del ‘contenuto digitale’
- gestione patrimonio culturale, letterario, librario, ...
- ...

Roma, Conferenza stampa, 28 Settembre 2004
.... Necessità di investimento pubblico in Italia

- Per evitare una Europa a due velocità
  (v. rapporto Euromap)

- con rafforzamento delle tecnologie linguistiche solo per le lingue più interessanti dal punto di vista
  - economico, commerciale
  - politico, strategico…

- senza considerare gli aspetti culturali, …
Technical scientific issues are important **BUT** …  
**Organisational, coordination, political** issues play a major role…

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

The implementation of the notion of **open distributed infrastructures** for LRs & LT:

- could act as a major technological & organisational challenge **around which synergies** must be sought (also with other communities)
- organisational, coordination, strategic, political issues acquire a decisive relevance & can naturally **lead to the creation of an International Forum**