

# META-NET and META-SHARE: An Overview

**Georg Rehm** 

**DFKI, Germany** georg.rehm@dfki.de

Human Language Technologies – The Baltic Perspective Riga, Latvia, October 8, 2010



### **Outline**



- Introduction
- The three areas of META-NET:
  - (1) Community Building and Development of a SRA
  - (2) META-SHARE
  - (3) Innovative MT Research
- The META-NET Website
- Extension of META-NET
- Collaborations



## Introduction

## A Socioeconomic Experiment META NET



- Core component of the European Union: a common market with a single information space.
- Two dozen national and many regional languages.
- Europe is an ambitious endeavour and an unprecedented socioeconomic experiment.



### **Two Contradictory Goals?**



- **Goal 1:** to preserve the diversity of Europe's cultural heritage (especially its 23 official languages and numerous local variants).
- **Goal 2:** to provide all EU-inhabitants with equal access to information and to enable them to participate.
- ♦ Solution: High-performance, robust, crosslingual and multilingual Language Technology!
  - enables communication across language boundaries,
  - secures equal **access to information**, and
  - supports each language in the advancement of networked LT ...

#### ... at least in theory.

### LT: The Current Situation



- **Problem:** Pace of progress is too slow to bring LT to a level that can fulfil these goals within the next 10 to 20 years.
- We have to accelerate progress!
  - But: Language Technology landscape in Europe is highly fragmented.
  - All stakeholders such as research communities, LT industry, user industries, language communities, funding programmes and policy makers have to team up for a dedicated major push.
- We need an alliance of stakeholders dedicated to a large coordinated effort:
  - META The Multilingual Europe Technology Alliance
  - Goals: to assemble and to represent the European LT community, to speak with one voice, to pursue a convincing research agenda.

### A Simple Mechanism



### a strong community

with ties into other R&D communities
with a network into industry and public administration
with visions and plans

visibility and credibility

powerful support from many sectors

massive additional resources

### **Examples from other Fields**



- ARTEMIS the European Technology Platform for Advanced Research and Technology for Embedded Intelligence and Systems
- AAL The Ambient Assisted Living Joint Programme
- **EPoSS** The European Technology Platform on Smart Systems Integration
- NEM The European Technology Platform on Networked and Electronic Media
- NESSI The European Technology Platform Networked European Software and Services Initiative

### **META-NET:** Initial Funding



- T4ME Net: "Technologies for the Multilingual European Information Society"
  - Funded under FP 7
  - 13 partners in 10 countries.
  - Coordinated by DFKI.
  - Area: Language-based interaction (ICT-2009.2.2)
  - Grant agreement no.: 249119
  - Total cost: €7.62 million (EU contribution: €5.99 million)
  - Execution: From Feb 1, 2010 to Jan 31, 2013 (36 months)
- To avoid multiple brand names, META-NET is used in external communication exclusively.

### **Founding Members**



- Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Germany
- Barcelona Media Centre d'Innovació, Spain
- Consiglio Nazionale Ricerche Instituto di Linguistica Computazionale "Antonio Zampolli", Italy
- Institute for Language and Speech Processing, R.C. "Athena", Greece
- Charles University in Prague, Czech Republic
- Centre National de la Recherche Scientifique
   Laboratoire d'Informatique pour la Mécanique et les Sci.s de l'Ingénieur, France
- Universiteit Utrecht, The Netherlands
- Aalto University, Finland
- Fondazione Bruno Kessler, Italy
- Dublin City University, Ireland
- Rheinisch Westfälische Technische Hochschule Aachen, Germany
- Jožef Stefan Institute, Slovenia
- Evaluations and Language ResourcesDistribution Agency, France



### **META-NET:** Objectives



- META-NET is a NoE dedicated to fostering the technological foundations of the European multilingual information society:
  - Build META, a strategic alliance that includes multiple stakeholders to prepare the ground for a large-scale concerted effort.
  - Strengthen the European research community.
  - Approach open problems in MT in collaboration with other fields.

Building a community with a shared vision and strategic research agenda

Building META-SHARE, an open resource exchange

Building bridges to neighbouring technology fields



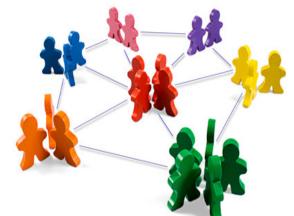
Community Building

## **Shared Vision and SRA**

## Building a community with a shared vision and strategic research agenda



- Mobilize researchers, decision makers, users and providers of LT,
   R&D programmes for cooperation and collaboration in META.
- Large-scale joint action to
  - Surveying and charting the field
  - Creating a shared vision
  - Preparing a Strategic Research Agenda
  - Building a multiplayer community via external liaisons, promotion, etc.
  - Identifying stakeholders, informing them about our initiative, dangers and challenges of the situation, LT support for their language etc.



### From Visions to the SRA



- **Vision Groups** bring together researchers, developers, integrators and (actual or potential, corporate or professional) users of LT-based products, services and applications (ca. 20 members each).
- A Technology Council will prepare a vision paper "European Multilingual Information Society 2020", a comprehensive roadmap including individual roadmaps for key socio-economic sectors, light-tower application visions.
- □ Finally, a **Strategic Research Agenda** will be presented by the Council (has to reflect the opinions and needs of all stakeholders).

### **Three Vision Groups**



- Translation and Localisation (technical documentation, official bulletins, GUI localisation, games, services etc.)
  - Target stakeholders: large users of translation services, (machine)
     translation, software companies, game companies, localisation industry
- Media and Information Services (audiovisual sector, news, digital libraries, portals, search engines etc.)
  - Target stakeholders: media industries, search engine providers, archives
- Interactive Systems (mobile assistance, dialogue translation, call centres, etc.)
  - Target stakeholders: mobile software and service providers, telecom industry, call centres

### **Vision Group Meetings**



- Vision Group Translation and Localisation
  - 23/07/2010 Berlin, Germany
  - 28/09/2010 Brussels, Belgium
- Vision Group Media and Information Services
  - 10/09/2010 Paris, France
  - 15/10/2010 Barcelona, Spain
- Vision Group Interactive Systems
  - 10/09/2010 Paris, France
  - 05/10/2010 Prague, Czech Republic

### The Process



**Visions** 

Roadmap

#### Strategic Research Agenda

communication to policy makers
communication in the funding bodies, public
wider LT community

and among other stakeholders

within META-NET (META-VISION) and among other stakeholders

2010 2011 2012

### **Vision Groups: Initial Results**

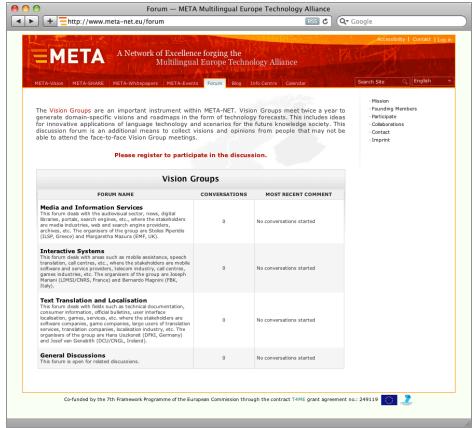


- Initial results (Translation and Localisation Vision Group).
- Five topics from the first meeting have strong visionary potential:
  - 1. New modes of research studying and involving human translators.
  - 2. A new role for the human in the machine translation process.
  - 3. New computing and deployment models (e.g., specialized services, trust centres, standard data formats, standard platforms such as UIMA).
  - 4. New ways of sharing resources and knowledge in R&D.
  - 5. New, transparent business models.

## http://www.meta-net.eu/forum META NET



- Please provide input in the discussion forum on our website.
- Tell us your ideas and priorities!





**META-NET** 

## **META-SHARE**

### Building META-SHARE, an open resource exchange



21

- Data has become a key factor in LT (see LREC, corpora mailing list).
- Expensive data become valuable through sharing.
- However, the long demanded and well-contemplated instruments for managing and sharing this data are still missing.

### **META-SHARE:** Key Features



- META-SHARE is an open, integrated, secure, and interoperable exchange infrastructure for language data and tools for the Human Language Technologies domain.
- A marketplace where language data and tools are documented, uploaded and stored in repositories, catalogued and announced, downloaded, exchanged, aiming to support a data economy (includes free and for-a-fee LRs/LTs and also services).
- Standards-compliant, overcoming format, terminological and semantic differences.

### **META-SHARE**



Data Centres ELRA, LDC, NICT

LT industry, SMEs

Acquisition projects
PANACEA, TTC,
ACCURAT, LET's MT

Regional & national LR projects & initiatives

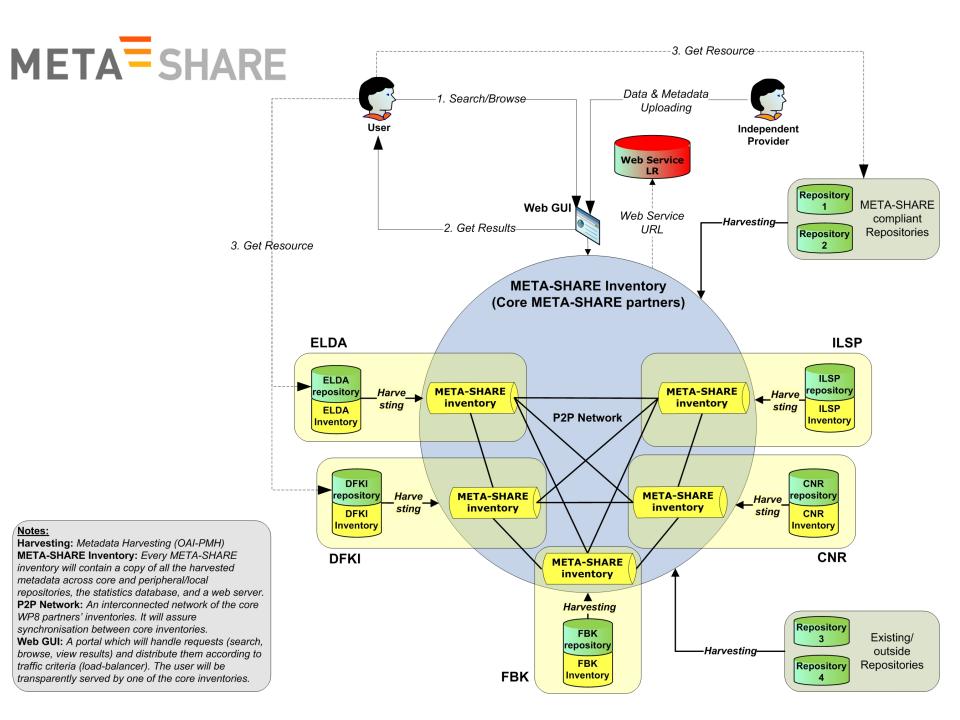


Academic catalogues & repositories

**CLARIN** 

Harvesting initiatives
LRE Map,
Harvesting Day

National data centres



## What we're offering



- A channel to share and distribute language data and tools.
- Technical solutions for building your own repositories.
- Protocols and mechanisms for making the descriptions of your resources (and the actual resources) harvestable.
- Guidelines and recommendations on standards used in the LR production and documentation processes.
- Recommendations on data and tools licensing issues.
- Access to large catalogues of documented, high-quality resources, as well as the actual data and tools.

### **Sneak Peak**





Sneak Peak (Version o)





META-SHARE is developed within the META-NET Network of Excellence

#### » browse metadata catalogue

#### What is it? About the project.

META-NET aims at creating META-SHARE, a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. Data and tools can be both open and with restricted access rights, free and for-a-fee. META-SHARE targets existing but also new and emerging language data, tools and systems required for building and evaluating new technologies, products and services. In this respect, reuse, combination, repurposing and re-engineering of language data and tools play a crucial role.

META-SHARE will eventually be an important component of a language technology marketplace for HLT researchers and developers, language professionals (translators, interpreters, content and software localisation experts, etc.), as well as for industrial players, especially SMEs, catering for the full development cycle of HLT, from research through to innovative products and services.

#### How can I use it? First steps.

More specifically, META-SHARE will be a freely available facility, supported by a large user and developer community, based on distributed networked repositories accessible through common interfaces. Users (consumers, providers or aggregators) will have single sign-on accounts and will be able to access everything within the repository. Each language resource will be given a permanent locator (PID). One of the key features of META-SHARE will be metadata harvesting, allowing for discovering and sharing resources across many repositories.

META-SHARE will also cater for advanced metadata schemata regarding description, harvesting and discovery of resources. It will be accompanied by a search function, so that users can search and navigate through its resources in the most flexible way possible. META-SHARE will allow the easy integration of new functionalities and services. In order to ensure modularity and robustness, it will follow a service-oriented architecture. Moreover, it will handle equally effective diverse file types. Finally, META-SHARE will provide the ability to compile and produce statistical reports, according to the different user types. The repository administrator will be able to supervise all of the above.

#### Connect and share! Our vision.

META-SHARE will start by integrating nodes and centres represented by the partners of the META-NET consortium. It will gradually be extended to encompass additional nodes/centres and provide more functionality with the goal of turning into an as largely distributed infrastructure as possible.

This is the first prototype version of META-SHARE. © META-NET 2010, all rights reserved. Co-funded by the 7th Framework Programme of the European Commission through the contract T4ME grant agreement no.: 249119.





#### Browse Metadata Catalogue

Contains all known metadata information

#### » back to META-SHARE

Title	Date	Provider
Greek Dependency Treebank	09/24/2010 at 14:51	ILSP Fedora (OAI DublinCore Provider)
Λεξικός πόρος γεγονότων της NE Event type GR lexical resource	09/24/2010 at 14:51	ILSP Fedora (OAI DublinCore Provider)
Εθνικός Θησαυφός Ελληνικής Γλώσσας Hellenic National Corpus	09/24/2010 at 14:51	ILSP Fedora (OAI DublinCore Provider)
INTERA corpus	09/24/2010 at 14:51	ILSP Fedora (OAI DublinCore Provider)
MUSA Multilingual Multimodal Corpus	09/24/2010 at 14:51	ILSP Fedora (OAI DublinCore Provider)

This is the first prototype version of META-SHARE. © META-NET 2010, all rights reserved.

Co-funded by the 7th Framework Programme of the European Commission through the contract T4ME grant agreement no.: 249119.





#### Browse Metadata Catalog

Contains all known metadata information available from META-SHARE.

#### » back to catalogue

Header Fields

Title Λεξικός πόφος γεγονότων της ΝΕ Event type GR lexical resource

Date 09/24/2010 at 14:51

Provider ILSP Fedora (OAI DublinCore Provider)

DublinCore Fields

Title Λεξικός πόφος γεγονότων της NE

Event type GR lexical resource

Description ~ 400 μονολεκτικά φηματικά κατηγοφήματα

Ινστιτούτο Επεξεργασίας του Λόγου / Ερευνητικό Κέντρο "Αθηνά"

Publisher Στέλιος Πιπερίδης

Ελίνα Δεσύποη

Type dataset

Format MS Access

Identifier ilsp.langres:3

Language el

This is the first prototype version of META-SHARE. © META-NET 2010, all rights reserved.

Co-funded by the 7th Framework Programme of the European Commission through the contract T4ME grant agreement no.: 249119.

### **Features**





- Open Source
- Service-Oriented
- Distributed
- Replication/Backup
- Reporting & Statistics

- Single Sign-On
- Easy Administration
- Metadata Harvesting
- Persistent Identifiers (PIDs)
- Intuitive Search

### **META-SHARE: Next Steps**



#### META-SHARE Version 0: November 2010

• First prototype to be demo'ed at META-FORUM 2010 in Brussels (November 17/18).

#### META-SHARE Version 1: July 2011

 Stable, working version of META-SHARE to be rolled out within the META-NET consortium.

#### META-SHARE Version 2: February 2012

Stable version, ready for production use.



#### **META-NET**

## **Innovative MT Research**

### Building bridges to neighbouring technology fields



- Cooperation and building bridges
  - within the project (machine translation community)
  - outside (neighbouring research communities such as, for example, machine learning, knowledge management, etc.)
- Four scientific goals that have not been addressed yet in a systematic way in the MT community are pursued:
  - Bringing more semantics into translation
  - Optimising division of labour in hybrid Machine Translation (consolidation of paradigms)
  - Exploiting the "context" for translation (e.g., multimedia)
  - Extending the empirical base for machine translation (using advanced machine learning)



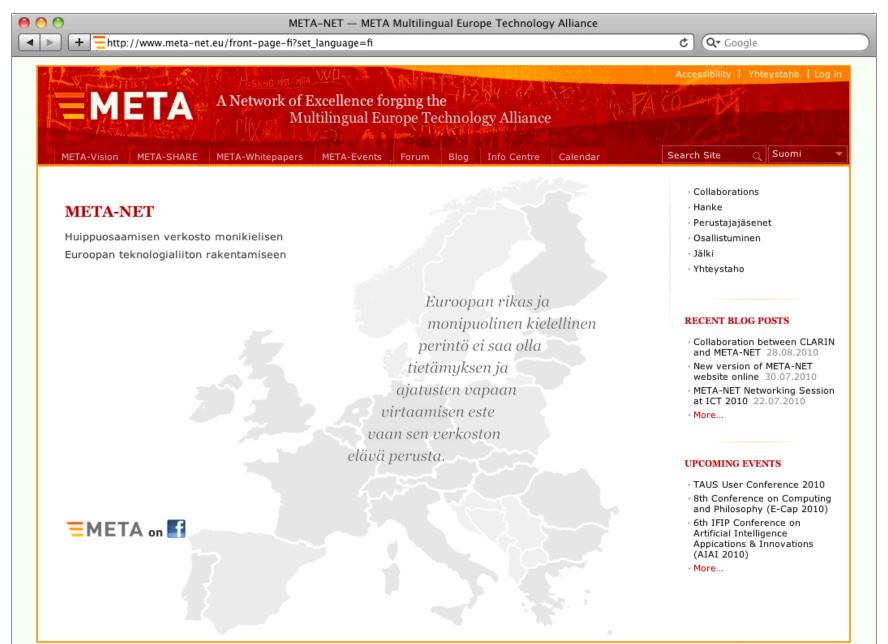
## **META-NET Website**





More...













More...









**META-NET** 

# **Extension of META-NET**

### **META-NET**



- T4ME is a funded project.
- META-NET is a Network of Excellence:
  - currently based on the initial T4ME consortium.
  - covers ten countries.
- Multiple European countries are not represented in META-NET.
- Of course we want to extend the Network of Excellence.

#### **Extension of the Network**



- Upcoming: three new consortia (ICT PSP call).
- Objectives (selection):
  - To collect, polish, upload, share LRs and LTs (via META-SHARE).
  - To carry out mobilisation and outreach activities in their countries.
- Start: early 2011
- META-NET will be extended by all partners of these consortia and also by individual researchers who support the initiative.
  - Goal: all EU member states and as many of the non-member states as possible to be covered.

## **Extension of META-NET**



Current and future composition of META-NET:

Year	Members	Countries	
2010	13	10	Current composition.
2011	39	28	Including the three new consortia.
2011	43	31	Including the three new consortia and a few additional contacts; <i>all</i> EU member states covered.

#### **Collaborations**

















#### **CLARIN and META-NET**





- Facilitates research by coordinating and making existing Language Resources and tools available and readily useable for the Social
- Offers resources and services to allow computer-aided language processing (e.g., querying data and complex processing of data sets).
- Focus on eResearch, eScience.

Sciences and Humanities.



- Building and offering results for Language Technology at large.
- Clear orientation towards ICT (development, innovation and services including commercial).
- Focus on the distribution of Language Resources (currently).
- End user: the European citizen.
- Goal: to address the problem of multilingualism in Europe.



**META-NET** 

# Wrapping up ...

# Participate in META!

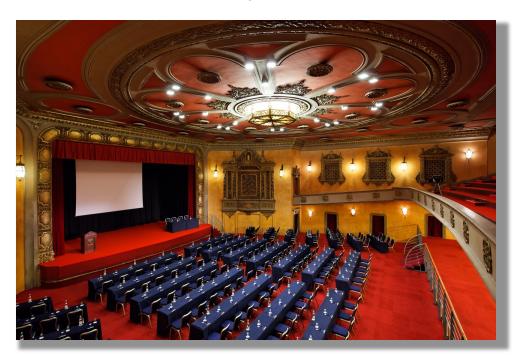


- Let's shape the future of Europe's linguistic and also Language Technology landscape and boost LT support significantly!
- Researchers, professionals, administrators developing, providing or using LT: please contribute ideas.
- Educators, journalists, politicians, public administrations and organisations participate in the public discussions on the visions and steps towards a truly multilingual information society.
- RTD projects can provide and use resources via META-SHARE.
- Propose projects that contribute to the multilingual digital information space (ICT FP7 Call 7 and SME-DCL; ca. 85 mio. EUR)

# http://www.meta-forum.eu



- **META-FORUM 2010** Challenges for Multilingual Europe.
- November 17/18, Hotel Le Plaza, Brussels, Belgium.
- Outreach and mobilisation event primarily aimed at LT industry.
- Programme includes EC info-session on Call 7 and Call SME-DCL.
- Ideal networking opportunity.
- Register soon: interest is *very* high, we only have limited capacity.



# Q/A



Ļoti pateicos.

Tänan väga.

Labai ačiū.

Thank you very much.

office@meta-net.eu

http://www.meta-net.eu

http://www.facebook.com/META.Alliance

#### **Important META-NET Events in 2010**

- ♦ META-NET Kick-off-Meeting, Feb 22/23, Berlin
- ◆ Language Technology Days, March 22/23, Luxembourg
- ◆ LREC 2010, May 19-21, Malta
- theMETAnk 2010, June 4/5, Berlin
- ♦ Translingual Europe 2010, June 7, Berlin
- ◆ ICT 2010, Sep 27-29, Brussels: Networking Session "Language Technology for a Multilingual Europe"
- Vision Groups (first round of meetings):
   July 23 (Berlin); Sep 10 (Paris; double feature)
- Vision Groups (second round of meetings):
   Sep 29 (Brussels); Oct 5 (Prague); Oct 15 (Barcelona)
- ♦ META-FORUM 2010 Challenges for Multilingual Europe, Nov 17/18, Brussels

# META-NET and Research Communities

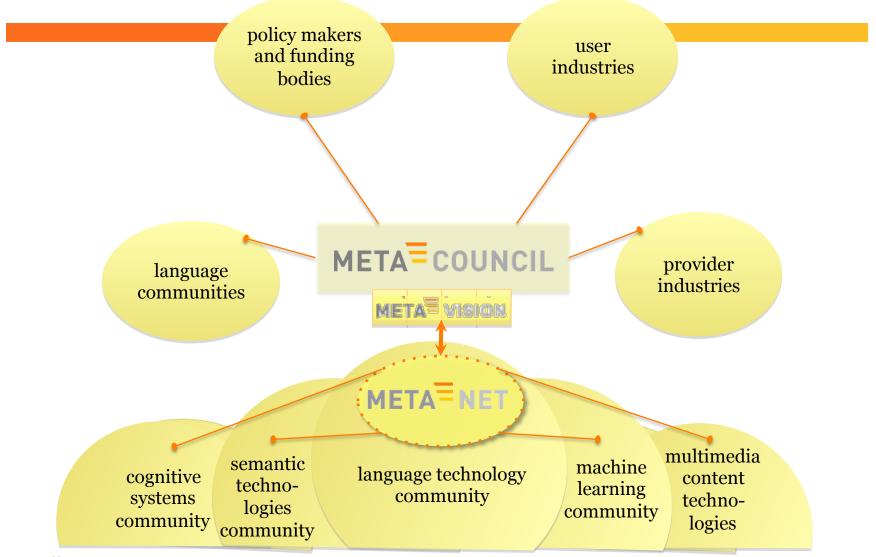


50



# **Community Building**





#### The META Alliance





user industries



language communities



META<sup>=</sup> VISION

provider industries

cognitive systems lo

semantic technologies community

language technology community

META NET

machine learning community

multimedia content technologies