

META  **NET**

META  **NORD**

WELT  **MOB**

**Baltic and Nordic Branch of the European
Open Linguistic Infrastructure**

Project Goal

The META-NORD project aims to establish an open linguistic infrastructure in the **Baltic** (Estonia, Latvia and Lithuania) and **Nordic countries** (Denmark, Finland, Iceland, Norway and Sweden)

META-NORD Geography



Consortium

Partner	Country
Tilde	Latvia
University of Copenhagen	Denmark
University of Tartu	Estonia
University of Bergen	Norway
University of Helsinki	Finland
University of Iceland	Iceland
Institute of Lithuanian Language	Lithuania
University of Gothenburg	Sweden

Focus

- Focus on European languages with **less than 10 million speakers**
 - EU official languages – **Danish, Finnish, Swedish**
 - Languages of the recently accessed EU countries – **Estonian, Latvian and Lithuanian**
 - Languages of the European Economic Area – **Icelandic and Norwegian**
- For many META-NORD languages only **limited high-quality language resources** are currently available
- Non-textual resources have been created only for some META-NORD languages

Main objectives

- **Describe the national landscape** - language use, language-savvy products and services, language technologies and resources; main actors; public policies and programmes; prevailing standards and practices; main drivers and roadblocks
- **Collect resources in the Baltic and Nordic countries** and document, link and upgrade them to agreed standards and guidelines
- **Collaborate** with the META-NET network of excellence and other partner projects
- **Help build and operate** broad, non-commercial, community-driven, inter-connected **repositories**, exchanges, and facilities
- **Mobilize national and regional actors**, public bodies and funding agencies by raising awareness

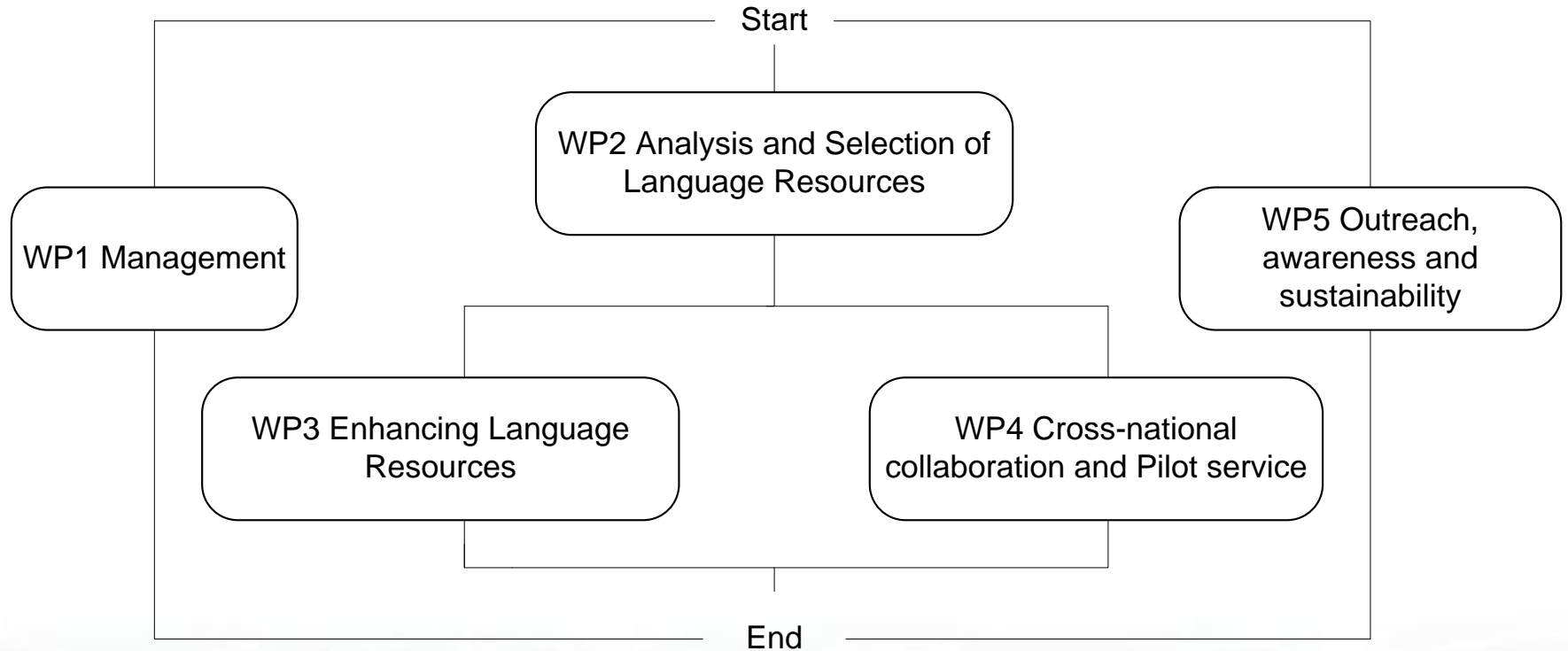
Specific targets

- **Provide expertise** to the META-NET in the fields where META-NORD partners have outstanding expertise: **treebanks/syntax databases, terminology resources, wordnets** and **finite-state techniques**
- **Develop and document methodologies** for building language resources for under-resourced languages with focus on **semi-automatic/machine assisted resource generation**
- Facilitate **availability of BLARK resources** for META-NORD languages
- Facilitate **knowledge transfer** between CLARIN and META-NORD, especially on standards and intellectual property rights (IPR)

Target Language Resources

- **WordNets:** monolingual WordNets and cross-linked pilots
Danish, Estonian, Finnish and Icelandic
- **Treebanks:** treebanks integrated on a uniform platform and linked across languages using parallel multilingual treebanking
Danish, Estonian, Finnish, Icelandic and Norwegian
- **Terminology collections:** distributed terminology resources across languages and domains will be consolidated
META-NORD languages
- **Corpora**
Danish, Estonian, Finnish and Icelandic
- **Tools:** Morphological analyzers, taggers, parsers
Latvian, Lithuanian, Estonian, Finnish, Swedish
- **Lexicons:** dictionaries, thesaurus
Latvian, Lithuanian, Estonian, Swedish

Chosen approach



PERT Chart at a WP Level

Major Milestones

- **May'11: National scene charts**
language community landscapes for the project languages
- **Jul'11: Language resources charts**
available resources for the project languages
- **Nov'11: Selection of resources**
methodology and criteria for the selection of resources,
agreements and data
- **Nov'11, Jul'12, Jan'13: Uploads of language resources**
- **Jan'13: Parallel treebanks**
- **Jan'13: Linked wordnets**
- **Feb'13: Multilingual terminology**
- **Jul'12: META-NORD national workshops**

Key Mobilization Activities

- META-NORD national workshops
- Targeted meetings with the representatives of business and industry
- Joint activities with META-NET Network of Excellence
- Collaboration with other LT R&D projects
- Collaboration with CLARIN project
- Mobilisation of the research community
 - national, regional and international scientific conferences, forums and end-user and public events
 - showcase in educational scenarios
 - professional e-mail lists
- Enhancing awareness in society and government

Key results

- **national scene charts** describing language community and the role of language in the respective country, research community, language service and language technology industry, use of language technology by business and administration, legal provisions
- **language resources charts** of actually or potentially available resources to the META-NORD consortium
- **treebanks** for relevant languages accessible through a uniform web interface and state-of-the-art search tool and linked across languages using a parallel multilingual treebanking
- **wordnets** upgraded to agreed standards and used for creation of pilot multi-lingual lexicons for IR purposes using cross-language synset linking
- **monolingual and bilingual terminology collections** integrated into multilingual terminology bank with elaborated terminology data access and sharing mechanisms
- **language resources batches** upgraded to agreed standards, extended and linked across different sources, aligned across languages and populated into digital exchange platform for pilot operation

Expected results - usability

- META-NORD **specific types** of language resources are prerequisites for future development of language technology products and multilingual services
- **Monolingual corpora** are used for creation of language models of statistical MT systems
- Many META-NORD languages are **heavily inflected languages**, it is difficult to retrieve sufficient information from raw text, and the existence of **tagged and parsed corpora** is a prerequisite for MT
- **Monolingual analyzers** are used in rule-based as well as statistical MT systems

META-NORD

The work within the project **META-NORD** has received funding from the ICT
Policy Support Programme as part of the
Competitiveness and Innovation Framework Programme

Grant agreement no **270899**

Project duration: February 2011 – January 2013

Contact information:

Andrejs Vasiljevs

Andrejs[at]tilde.lv

Tilde, Vienibas gatve 75a, Riga

LV1004, Latvia

