



# **META-NORD**

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

**Project no. 270899**

## **Deliverable D2.5**

**Description of third parties networks  
for each language covered in the project**

**Version No. 1.0**

**31/07/2011**

**Document Information**

Deliverable number:	D2.5
Deliverable title:	Description of third parties networks for each language covered in the project
Due date of deliverable:	2011-07-31
Actual submission date of deliverable:	2011-07-31
Main Author(s):	HI
Participants:	Tilde, UCPH, UT, UIB, UHEL, HI, LKI, UGOT
Internal reviewer:	UT
Workpackage:	WP2
Workpackage title:	Analysis and selection of language resources
Workpackage leader:	UT
Dissemination Level:	PU
Version:	1.0
Keywords:	Third parties' networks, META CC, LT industry, LT research community, society, government, funding bodies

**History of Versions**

Version	Date	Status	Name of the Author (Partner)	Contributions	Description/ Approval Level
V0.1	2011-05-23	Fishbone	HI	Fishbone with material from DoW	Fishbone approved
V0.2	2011-07-12	Draft	HI, All META-NORD partners	Draft version with contributions from all partners	Approved
V0.3	2011-07-14	Draft	HI, Tilde	A number of corrections and additions	Approved
V0.4	2011-07-20	Draft	HI	Some minor changes and additions	Approved
V0.9	2011-07-26	Prefinal version	HI, UGOT	Addition of Executive summary; final corrections and refinements	Approved
V1.0	2011-07-31	Final version	Tilde	Delivered to EC	Delivered to EC

## EXECUTIVE SUMMARY

Each of the eight countries participating in META-NORD (Denmark, Estonia, Finland, Iceland, Latvia, Lithuania, Norway, and Sweden) will form a network of third parties willing to cooperate with them in various ways and make their resources widely available through open non-commercial initiatives. The main actors in these networks come from the sectors of research, public decision making and the language technology industry. The modes of cooperation differ as some will mainly participate in awareness-raising or dissemination, whereas others might provide resources or tools, and yet others will cooperate in policy-making or coordination.

The composition of the third parties' networks planned in the participating countries will differ depending on the language technology landscape in each country. Norway is, for instance, particularly strong in the LT industry with 30 companies listed, more than the other seven countries combined, and Finland has a strong network of associations, federations and societies. Each third parties' network will, therefore, reflect its LT society.

The META Coordination Councils that will be established in each country to ensure a national network approach will be formed as an extension of the project partners, and its main roles will be to promote language resource sharing and the coordination of national efforts.

These third parties' networks are only in the preliminary stages as of now but in the next few months we will see the rise of eight new networks in all the Nordic and Baltic countries.

## Table of Contents

<b>1. Introduction .....</b>	<b>6</b>
<b>2. An overview of the networks .....</b>	<b>6</b>
2.1 Categories of actors .....	6
2.2 Modes of cooperation .....	8
2.3 META CC.....	9
<b>3. Third parties' networks in each participating country .....</b>	<b>9</b>
3.1. Latvia (TILDE).....	9
3.2. Denmark (UCPH).....	10
3.3 Estonia (UT) .....	10
3.4 Norway (UIB).....	11
3.5 Finland (UHEL).....	11
3.6 Iceland (HI).....	12
3.7 Lithuania (LKI).....	12
3.8 Sweden (UGOT).....	13
<b>4. List of tables .....</b>	<b>14</b>

## Abbreviations

Abbreviation	Term/definition
DoW	Description of Work
HLT	Human Language Technology
IT	Information Technology
LT	Language Technology
META CC	META Coordination Council
NLP	Natural Language Processing
R&D	Research and Development
SME	Small and Medium Enterprises

**Table 1 Abbreviations**

## 1. Introduction

The purpose of this document is to describe the networks of third parties that will be established in each of the META-NORD<sup>1</sup> participating countries. This introductory section is based on the project's Description of Work (DoW).

For each language represented in the consortium, the project partners will extend their network of third parties' contacts and approach (public and private sector) organisations willing to make their resources widely available through open non-commercial initiatives, including but not limited to META-NET<sup>2</sup>.

A META Coordination Council (META CC) will be established to ensure a national networking approach. META CC will consist of the representatives of such national organisations/institutions relevant to META-NORD. META CC will promote language resource sharing and the coordination of national efforts, ensuring that a greater number of language resource owners/producers/aggregators at a national level will make their resources available for the project and further their non-commercial use. META CC will also be involved as an outreach instrument to the relevant target groups within business and industry.

The consortium members and META CC will meet the representatives of industry (companies of different sectors and sizes) to discuss the benefits of LT services and applications, and the associated pre-conditions and constraints. As an establishment that ensures national networking and resource sharing facilitator, META CC will be the one who will perform independent quality assurance of project outcomes, e.g. project deliverables and other outputs.

## 2. An overview of the networks

This section gives an overview of the main categories of participants in the LT networks.

### *2.1 Categories of actors*

The main actors in the LT networks can be divided roughly into three categories:

- the research community
- the language technology industry
- the authorities or public decision makers

How these three groups are further divided depends on many factors, such as their background, their cooperative role and their main area of expertise. Some sub-groups might even belong to two or even all three of the major categories. In this chapter we will, therefore, discuss the sub-groups independently without categorizing them further. These groups are associations, education, federation, funding bodies, government, industry – language services, industry – vendor, integrator, press, research and society.

---

<sup>1</sup> <http://www.meta-nord.eu>

<sup>2</sup> <http://www.meta-net.eu>

Several associations play an important role in the LT network, particularly for policy making and awareness-raising, but also for dissemination. This group is particularly prominent in the Finnish network with nine representatives, but also figure in the networks from Denmark, Iceland and Norway. Closely related to the associations are the two federations and two societies from the Finnish network, all of which play the role of awareness-raising.

The categories of education and research cannot always be easily distinguished as both mainly consist of universities and their cooperation role overlaps. However, the research community is particularly a provider of resources and tools whereas the education community also plays a role in dissemination, awareness-raising and coordination. It also seems to depend somewhat on countries whether the organisations are categorised as education or research. It is particularly telling that while Latvia lists seven organisations as education and Lithuania lists five, Finland and Sweden list only one each but no other country refers to the category of Education. Most seem to prefer using the category of research and are listed four by Estonia, six by Denmark, Iceland and Sweden, seven by Latvia and Lithuania and nine by Finland and Norway. The cooperation role of the research community is first and foremost as provider of resources (37 organisations) but also as provider of tools (7 organisations), awareness-raising (4 organisations), coordination (3 organisations), and dissemination (4 organisations).

The category named funding-body is relatively small and represented only by three countries, Finland with seven organisations, Norway with two and Iceland with one. Their role is first and foremost in policy-making and awareness-raising.

Related to the category of funding-body is the government category, which includes both ministries and various agencies and institutes. The ministries listed are usually ministries of education, research, science, culture and foreign affairs but in this category are also national libraries, state language commissions and languages councils. Of these organisations, one comes from Iceland, two from Sweden, three from Norway and Estonia, four from Finland, five from Lithuania and fifteen from Latvia.

The industry category mainly consists of language services providers and industry vendors even though one company is categorised as corporate user and two are not categorised further than as industry. These companies and institutes have a varied role even though the one as providers of tools is by far the most prominent. Other roles include the fields of coordination, dissemination, awareness-raising and providers of resources. Norway has by far the biggest LT industry of the eight Nordic and Baltic countries with 30 companies and institutes falling into the industry field – in fact, more than the other seven countries combined. Finland has eight LT companies in the industry sector, Sweden six, Denmark and Iceland four, Latvia two and Estonia and Lithuania one each. Majority of these companies specialises in NLP but other areas of expertise include speech, interpretation and translation.

A small group represents the press which main role is dissemination. In the Nordic and Baltic network there is one representative of the press from Iceland, Latvia and Norway each and two from Lithuania.

Another small group is that of integrators, which is represented in two countries, Norway and Iceland, with the roles of dissemination and coordination.

It is striking when looking at the networks of each of the eight META-NORD countries how differently prominent the various sectors really are in these third parties' networks. For instance, associations, federations and societies play an important role in the Finnish network while they are not at all a part of the networks from the Baltic countries and Sweden. In Latvia, seven organisations are listed as education while none is categorised as such in Estonia, Iceland and Norway. Only three countries (Finland, Iceland and Norway) list funding-bodies,

and only four (Iceland, Latvia, Lithuania and Norway) list the press as a member. It is not clear whether there really is such a big difference between the kinds of organisations that participate or whether this can be explained by different categorisation of the organisations. Some universities, for instance, are categorised as education where most are categorised as research and similarly some associations are categorised simply as associations where others are categorised as education.

## ***2.2 Modes of cooperation***

The modes of cooperation can be divided into six main categories:

- awareness-raising
- coordination-council
- dissemination
- policy-making
- providers of resources
- providers of tools

The organisations most responsible for awareness-raising come from various aspects of the social spectrum, such as federations, government, education, research and funding-body. A huge majority of these awareness-raising companies come from Finland or 24 in total, followed by Latvia with nine. However, only one organisation is listed as being mainly responsible for awareness-raising in Estonia, Iceland and Sweden whereas none come from Denmark, Lithuania and Norway. This does not mean that no awareness-raising takes place in these countries, nor that only very little awareness-raising takes place in Estonia, Iceland and Sweden – only that in these countries there are no, or only a few companies, in the third parties' network whose main cooperative role is awareness-raising. That role might, in fact, fall mostly on the hands of those organisations that also play an important role as providers of resources or tools.

Only a handful of companies have coordination-council as their main role in the third parties' network, and these are mostly from Iceland, or seven in total. Two are from Sweden and one from Latvia. These organisations come from research, education, the industry and integrators. Most focus on NLP but localisation, speech and interpretation are also represented.

The organisations that cooperate in dissemination are obviously some members of the press but also representatives from the industry, associations, education, integrators, government, and research. Most of these organisations reside in Lithuania, or thirteen, with nine in Finland and eight in Norway. One representative comes from Iceland and one from Latvia but no organisations in the Danish, Estonian and Swedish third parties' networks have the main role of dissemination.

The organisations involved in policy-making are first and foremost governments of the different countries, as well as some associations and funding-bodies. Here the Latvian government is the most prominent with 12 different groups within the government taking part, followed by six Norwegian funding-bodies and governmental groups. Two policy-making partners come from Denmark, Estonia and Finland and one from Iceland.

The biggest group in the third parties' network are, unsurprisingly, providers of resources. Most of them come from the research community but a few are from the industry, government or education. The biggest network of providers of resources exists in Latvia with fifteen



members, followed closely by Norway with eleven and Finland and Denmark with nine each. Much smaller networks exist in Iceland with four, Estonian and Lithuania with three and Sweden with two. The resources provided by these organisations are of various kinds, although NLP is prominent. Other areas of expertise count, for instance, history, IT, literature, speech, translation, terminology, language policy, interpretation, philosophy, and localisation. Just like with providers of resources, the countries vary hugely when it comes to providers of tools. Norway has, for instance, 23 and Sweden eleven, but all the other countries have fewer than five providers of tools in their third parties' network. Most of these are in Lithuania, or four, Latvia and Finland have three each, Denmark and Estonia two each and Iceland only one. Most of these tools provided are within the area of NLP but some exist also for speech, translation, phonetics, localisation and language policy. Most of the providers come from the industry sector although some are from research, government or education.

### **2.3 META CC**

In each country a META Coordination Council (META CC) will be established to ensure a national networking approach. The META CC will be an extension of the project partners and consists of public and private organisations willing to make their language resources widely available through open non-commercial initiatives, including but not limited to META-NET. The idea is that this META CC network will promote language resource sharing and the coordination of national efforts. It will also be involved as an outreach instrument to the relevant target groups within business and industry. The consortium members and META CC will meet the representatives of business and industry to discuss with them the benefits of LT services and applications, and the associated pre-conditions and constraints.

The META-NORD Consortium aims to collect resources not only in partner institutions but also outside the consortium. In some countries, such as Estonia and Iceland, partner organisations and collaborators already own most of the resources that exist for particular language. In other countries the consortium will make every effort to convince other national resource providers to make their resources available through the META-NORD infrastructure. Here the META CC network plays an important role.

The META CCs have not yet been formally established in any of the eight META-NORD countries but plans thereof are in the works and discussions have started. The goal is for there to be eight META CCs operating in the Nordic and Baltic countries before the end of 2011.

## **3. Third parties' networks in each participating country**

This section provides a brief description of the networks in each participating country.

### **3.1. Latvia (TILDE)**

The third parties' network in Latvia can be divided in three main categories: the research community, society and government and LT industry.

The HLT research community in Latvia is represented by ten institutions: Institute of Mathematics and Computer Sciences of the University of Latvia, Faculty of Computing of the University of Latvia, Terminology Commission of Latvian Academy of Sciences, Riga Teacher Training and Educational Management Academy, University of Liepaja, Rezekne Higher Education Institution, Daugavpils University, Latvian Language Institute, Ventspils University College, and National Library of Latvia Society.

Society and government are represented by Saeima (Parliament), Ministry of Education and Science, Ministry of Culture, State Language Commission, State Language Center, Latvian Language Agency, and Cultural Information Systems Agency.

The following LT industry players in Latvia will be addressed: Tilde, Datorzinību Centrs, Latvian Information and Communications Technology Association (LIKTA), Algorego company, Trident MT.

The more detailed description of the abovementioned institutions can be found in *D5.1.1 Action plan*, Section 2 *Target groups*. Table 4.1 in section 4 lists the main contacts (professionals), their areas of expertise and contact information.

### **3.2. Denmark (UCPH)**

The Danish META CC is a forum which hasn't been formally established yet but will have been by early autumn of 2011. The groups of people that will be included in this forum are listed in table 4.2 in section 4 below. All the people on the list are co-authors and/or contributors to the Danish White Paper on Language Technology and thus, collaboration among the members has already been successfully initiated and initial meetings have been held.

The list comprises the main actors in LT and language politics in Denmark, i.e. main research institutions and LT companies as also described in the Danish White Paper. With respect to cooperation role, we have assigned the role of 'Provider' to several of the private company contacts even if these have not yet agreed to provide their tools and resources to the consortium. Some of them may be willing to do so under some license, but this is still to be negotiated.

### **3.3 Estonia (UT)**

The third parties' network in Estonia may be divided into two main groups: institutions related to government, which are the main policy-makers, and research institutions. Also, there is one active private company (Filosoft OÜ) in the field of LT in Estonia. Other possible private companies are rather weakly related to language and speech technology. They are mostly focused on providing web interfaces to dictionaries, localisation or interpretation services.

The Ministry of Education and Research and Estonian Language Council are the main institutions which shape the development strategy of Estonian language and LT policies.

The research activities of the Institute of Estonian Language include the development of mono- and bilingual dictionaries, thesauri and also tools for speech synthesis.

The main research fields of the Institute of Cybernetics at Tallinn University of Technology are Estonian phonetics and speech technology.

Estonian Literary Museum has a thorough collection of dictionaries and corpora related to Estonian folklore.

### **3.4 Norway (UIB)**

The third parties' network in Norway can be divided into three groups: LT and service providers in the private sector, researchers in the University sector and public decisions makers on language policies. Up until now the cooperation has taken place along the following axes:

- Third parties have contributed to the work on WP 2.1 and 2.2 (Language report and Report on resources). The contributions have focused on the situation for Norwegian LT in terms of existing resources and the need for new ones.
- Third parties have contributed by transferring materials for use by META-NET in WP 3.4.
- Third parties have been contacted for dissemination of META-NORD activities. This contact has taken place at seminars and conferences and during the work on WP 2.1.
- Third parties have been encouraged to join META-NET, and several actors have already done so.

In Norway, the work on establishing the Meta CC has not been finalised, but we are working closely with the National Library and consider them a likely partner. This partnership will focus on the development of meta-data descriptions, a platform for freely available resources provided by the National library and dissemination.

### **3.5 Finland (UHEL)**

The third parties network in Finland is slowly emerging. The network will be changing its shape as the project progresses. The actors recorded in table 4.5 in section 4 include several types of third parties with different priorities.

An important group for META-NORD can be characterised as the Finnish LT research community. It includes research institutions, universities, research projects, individual researchers and students at the master's level, and PhD students within the doctoral programme LANG-NET. Students of language studies will benefit from META-NET in the form of assistance in questions regarding the best practices in collecting, annotating and distributing resources collected for research purposes. META-NORD can offer seminars or short courses on these questions and the junior researchers will know how to deposit the resources and tools. Research projects will benefit in the questions regarding deposition, maintenance and compliance to different standards. The accomplishments of META-NORD, i.e. the resources and tools made available will be the most important issue in the dissemination in Finland.

LT projects within the FIN-CLARIN consortium universities will be an important source of resources potentially available for META-NORD, as are the libraries, such as the National Library, and the media, such as the Finnish Broadcasting Company, and publishers such as Sanoma. Also other than FIN-CLARIN universities, departments offering language studies, and other research institutions are important players in this group providing resources and tools for the later uploads, and for the long-time sustainability of the project.

It is very typical for Finns to operate in associations and foundations. These actors will offer not only a discussion forum on the importance of LT resources, tools and applications, but also a way of disseminating information on META-NORD to their members and to the public. The associations can also offer the information on the needs of special groups to META-NORD, such as the blind or the deaf, benefiting from LT applications. Industry vendors and LT companies are an important group in motivating the project and its goals to offer resources also for commercial LT product development.

Political decision-makers and funding-bodies provide insight in how the society is willing to invest in the language technologies and the proposed projects. The role of the funding-body is also important in specifying the application procedure so that the projects funded will automatically be required to make the resources created within the project available.

### **3.6 *Iceland (HI)***

The Icelandic third parties' network can roughly be divided in three: the research community in human language technology and related domains, the LT industry and the public decision makers.

The research community, which is in many ways responsible for providing the necessary resources, coordinating projects and raising awareness of LT can first and foremost be found in three institutes: the Árni Magnússon Institute of Icelandic Studies, the University of Iceland, and Reykjavik University.

The language industry sector has mostly focused on NLP, speech and localisation and has produced some hugely important applications for LT in Icelandic. These are companies such as Clara, Frisk Software, Mobilitus, and Skjal Translation Service. The public decision makers also play an important role for Icelandic LT and here we need to mention the Icelandic Language Council and the Science and Technology Policy Council. However, the Ministry for Foreign Affairs also plays a role as a resource provider because of their translation service.

It is important that these groups form a functioning LT network where issues can be discussed, ideas can be developed and real actions taken. Collaboration does not have to be extensive in order to be fruitful.

### **3.7 *Lithuania (LKI)***

One of the target groups of the META-NORD project in Lithuania consists of scientists, investigators, LT specialists, science and research staff, association members and staff, public agencies, and employees of budgetary institutions.

The potential stakeholders on the society level with relevance to the project in Lithuania are: the Lithuanian State Language Commission, Association of Lithuanian Language and Information Technology, Institute of Mathematics and Informatics of the Vilniaus University, Vytautas Magnus University, and Kaunas University of Technology.

National, regional and local government bodies fall under authorities. They shape the politics of the state, and hence the needs of the whole of the public as well. The level of information available to them as well as their attitude may affect the project's success. The project is in-

evitably linked with the authorities. The following authorities might have relevance to this project: Parliament of the Republic of Lithuania, Government of the Republic of Lithuania, Information Society Development Committee under the Ministry of Transport and Communications, Ministry of Transport and Communications of the Republic of Lithuania, and the Republic of Lithuania Ministry of Education and Science.

The social groups that are important for the dissemination of this project are union, councils, associations, societies, and forums, such as Association of Lithuanians, the Lithuanian Association of Literary Translators, and Lithuanian Union of Journalists.

The LT industry that might have relevance to this project includes TILDE IT, Fotonija, and Electronic Publishing House. All these institutions work for the purpose of integrating the Lithuanian language into IT.

### 3.8 Sweden (UGOT)

The Swedish third parties' network currently consists of five – overlapping – types of organisations:

Research organisations – This segment comprises mainly universities, but there is at least one public research institute with LT R&D activities.

Educational organisations – This group largely overlaps with the research organisations, but since its goals and target audience are different, it should also be seen as a separate venue for dissemination and awareness raising with respect to META-NORD's activities, particularly since Sweden has had specialised (bachelor and master level) study programs in LT since 1984.

Industry – There is a thriving LT industry in Sweden, especially in the speech domain, but there are also companies working in terminology, machine translation, text authoring tools, and intelligent information extraction. Traditionally there are strong ties between academia and industry, since SMEs tend to be spinoffs from academic research, and developers in larger companies often come from the academic LT study programs.

Public sector – This category refers to the public sector outside the universities, which of course also are mainly public bodies. The 'early adopters' of LT in the public sector in Sweden are found in the health care system, with a number of R&D collaborations, in public service television (where there is strong interest in cost-effective subtitling solutions) and – perhaps surprisingly – in the Swedish Language Council, which has appointed a national advisory group on LT, where questions of language policy and language planning are discussed in relation to LT. In addition, public cultural heritage institutions, such as the National Library and the National Archives seem to be slowly beginning to realise that LT may offer attractive solutions for effective access to digitised collections of cultural heritage objects.

Research funding bodies – The Swedish Research Council and Vinnova (the Swedish Governmental Agency for Innovation Systems) have from time to time had special LT R&D funding programs. The research council is the national agency responsible for channeling Swedish involvement in European and other international research infrastructure initiatives.

Looking at the larger picture, the network at present consists mostly of research organisations and companies. These parties almost always represent providers of LT resources and tools, and are essential for the advancement of Swedish LT. However, table 4.8 in section 4 clearly shows that further efforts are needed to bring in representatives from the educational and pub-

lic sectors, as well as the funding bodies, as their participation will be necessary for sustaining and promoting the status of LT in Sweden.

#### 4. List of tables

The network participants in all the META-NORD countries are listed in an accompanying Excel file (*D2.5\_Tables*). The lists for all the partners are combined in Table 4.0, whereas the network participants in each country are listed in Tables 4.1-4.8, respectively.

- Table 4.0      META-NORD
- Table 4.1      Latvia
- Table 4.2      Denmark
- Table 4.3      Estonia
- Table 4.4      Norway
- Table 4.5      Finland
- Table 4.6      Iceland
- Table 4.7      Lithuania
- Table 4.8      Sweden