



# **META-NORD**

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

**Project no. 270899**

## **Deliverable D5.3.1**

**Long-time viability and sustainability strategy**

**Version No. 1.0**

**31/07/2011**

**Document Information**

|  |   |
|--|---|
| Deliverable number:                    | D5.3.1  |
| Deliverable title:                     | Long-time viability and sustainability strategy                                     |
| 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:                     | LKI   |
| Workpackage:                           | WP5   |
| Workpackage title:                     | Outreach, awareness and sustainability  |
| Workpackage leader:                    | LKI   |
| Dissemination Level:                   | PU  |
| Version:                               | 1.0   |
| Keywords:                              | Long-time viability, sustainability, treebank, wordnet, terminology, infrastructure |

**History of Versions**

| Version | Date       | Status           | Name of the Author (Partner) | Contributions   | Description/ Approval Level |
|---------|------------|------------------|------------------------------|---|-----------------------------|
| V0.1    | 2011-05-24 | Fishbone         | HI                           |   | Fishbone approved           |
| V0.2    | 2011-06-28 | Draft            | HI, UIB, UHEL                | Paragraphs on treebanks and the clearinghouse model added                   | Approved                    |
| V0.3    | 2011-07-04 | Draft            | HI, UCPH, UHEL, UGOT         | Paragraphs on wordnets added and some other changes made                    | Approved                    |
| V0.4    | 2011-07-14 | Draft            | HI, Tilde                    | Paragraphs on terminology added and some other minor changes/additions made | Approved                    |
| V0.5    | 2011-07-19 | Draft            | HI                           | Some minor changes and addition of conclusion                               | Approved                    |
| V0.9    | 2011-07-26 | Prefinal version | HI                           | Executive summary added and some final refinements made                     | Approved                    |
| V1.0    | 2011-07-31 | Final version    | Tilde                        | Delivered to EC   | Delivered to EC             |

## EXECUTIVE SUMMARY

The META-NORD project is targeted to create a sustainable mechanism for cooperation and resource sharing as a part of the European open linguistic infrastructure. This will be achieved through cooperation with existing national and European projects and working with national and international stakeholders to launch uptake projects and funding. The continuation of the created infrastructure will be ensured by high visibility and the intrinsic usefulness of the project results on a national scale as well as active integration in META-NET activities.

After the two year pilot phase the META-NORD infrastructure will be established on a sustainable long-time basis, that is, by the end of the project, the META-NORD consortium will provide a complete and viable service with sustainable model for long-time running and further elaboration.

At this early stage in the project, the consortium is not in a position to develop a detailed plan for the post-EU funded phase. Thus, this document only presents the first version of a long-time viability and sustainability strategy for the META-NORD infrastructure. This strategy will be updated and refined as the consortium approaches the end of the project (Deliverable D5.3.2, M24).

## Table of Contents

|   |          |
|---|----------|
| <b>1. Background</b> .....  | <b>6</b> |
| <b>2. General sustainability strategies</b> .....                 | <b>6</b> |
| 2.1 META Coordination Council .....                               | 6        |
| 2.2 The clearinghouse model .....                                 | 6        |
| 2.3 META-CENTRES .....  | 7        |
| <b>3. Sustainability of the main types of resources</b> .....     | <b>7</b> |
| 3.1 Sustainability of treebank resources .....                    | 7        |
| 3.2 Sustainability of terminology resources.....                  | 7        |
| 3.3 Sustainability of wordnet resources .....                     | 8        |
| <b>4. Cooperation with other infrastructure initiatives</b> ..... | <b>8</b> |
| <b>5. Conclusion</b> .....  | <b>9</b> |

## Abbreviations

| Abbreviation | Term/definition   |
|--------------|---|
| CLARIN       | Common Language Resources and Technology Infrastructure       |
| DigHumLab    | Digitalt Humaniora Laboratorium                               |
| ERIC         | European Research Infrastructure Consortium                   |
| HPC          | High-performance Computing                                    |
| INESS        | Infrastructure for Syntactic and Semantic Exploration         |
| LFG          | Lexical-Functional Grammar                                    |
| LSP          | Language for Special Purposes                                 |
| LT           | Language Technology   |
| LTI          | Language Technology Infrastructure                            |
| META CC      | META Coordination Council                                     |
| NB           | Nasjonalbiblioteket – National Library of Norway              |
| NEALT        | Northern European Association for Language Technology         |
| SIG-INFRA    | Special Interest Group for Language Technology Infrastructure |

**Table 1 Abbreviations**

## 1. Background

The cornerstone of the long-time viability of the META-NORD service will be the involvement of the following main national and/or regional actors:

- producers, i.e. “competence centres”, typically public research centres, language institutes and academies, as well as private content owners like media companies and publishing houses;
- aggregators, i.e. “data centres” (repositories), usually supported by national and/or regional authorities;
- sponsors, i.e. public authorities, research agencies, language councils, companies with language resource needs.

## 2. General sustainability strategies

This section outlines the main sustainability strategies that will be utilised within META-NORD.

### 2.1 *META Coordination Council*

One of the main instruments for sustaining the service after the EU co-financed project ceases will be the META Coordination Councils (META-CC) that are being established in all the META-NORD countries. Through META-CC the key national stakeholders in the language technology area will be involved in META-CC activities from the very beginning, thus building awareness and support for continuation of activities through national and institutional funding instruments.

The META-CCs and their role are further described in Deliverable *D2.5, Description of third parties networks*.

### 2.2 *The clearinghouse model*

One possible business model that will be evaluated for the META-NORD infrastructure is the clearinghouse concept currently being developed by CSC in Finland which aims at cooperating on a long-time storage of resources with other similar service centres in other European countries using a cost-sharing principle. CSC is government-owned and serves the research community with funding from the Finnish Ministry of Education whereas commercial organisations can buy computing resources. The Language Bank of Finland will be physically located at CSC. The model for the Language Bank of Finland is a clearinghouse model in the sense that it will procure resources from resource providers by negotiating viable licenses for them and by offering them to the language technology community. The resources are offered to researchers for free and many will also be available to the industry.

### **2.3 META-CENTRES**

The principles for how to include new actors and countries in this business model will be further elaborated in the META-NORD project in cooperation with the best practices of other META-SHARE centres. Other business models will also be explored and it is possible that different types of service centres may arrive at different business models letting the content providers choose different service providers for different types of end-users.

To ensure sustainability and continuation of META-NORD activities, possibilities to create national META-CENTRES by the project partner institutions will be discussed. These centres could form sustainable backbone network for the European open linguistic infrastructure.

## **3. Sustainability of the main types of resources**

This section describes the initial plans and visions for the sustainability of the main types of resources that will be developed within META-NORD.

### **3.1 Sustainability of treebank resources**

After the end of the two year META-NORD project, treebank resources to be provided during the project Task 3.4 will be further maintained and disseminated through the INESS project, a nationally funded Norwegian project (financed by the Research Council of Norway) which will run until 2015 and to which UIB has declared the intention of long-time maintenance and support after 2015. The INESS project will host treebanks (including LFG treebanks, dependency treebanks and parallel treebanks) for any language on its HPC cluster where they will be searchable and downloadable.

### **3.2 Sustainability of terminology resources**

Sustainability of the META-NORD work on terminology resources will be ensured by the tight collaboration with the EuroTermBank<sup>1</sup> that is maintained and supported by the EuroTermBank Consortium and Tilde company. More than 100 terminology resources and several external databases have been consolidated and federated at the EuroTermBank database that has the necessary data access and sharing mechanisms to interlink with META-SHARE. For the users and providers of terminological data, either of the two methods will be offered – interlinking with existing live databases or storing users' terminological data at EuroTermBank (on user account basis) while sharing it with the community. The facility of periodically updating the data will be provided for static terminological data. By interlinking a one-stop access to users' terminology at EuroTermBank, the conformity with the standards and all the necessary interfaces will be provided including re-use scenarios for various LT industries.

---

<sup>1</sup> <http://www.eurotermbank.com>

### 3.3 *Sustainability of wordnet resources*

After the end of the two year META-NORD project, UCPH offers to maintain and disseminate the interlinked wordnet resources to be provided during the project Task 3.5 through the Danish research infrastructure, CLARIN (see Section 4). The monolingual wordnets which have been made available to META-NORD will be hosted nationally, and so the Danish wordnet, DanNet, will be hosted by the CLARIN platform, for instance.

The Finnish wordnet, FinnWordNet, will be available through UHEL and the Language Bank of Finland and UHEL will provide a search interface in addition to the downloadable resource. UHEL will also provide occasional updates of the data and the interface after the end of the META-NORD project. The data will also be available for download from the Language Bank of Finland where it will have long-time backup and preservation.

The Swedish wordnet will be available for download (and browsing) at Språkbanken (UGOT), where it will be preserved, curated and further developed. Språkbanken could also host a copy of the linked wordnet.

Further to the wordnets which were available at the start of META-NORD, the National Library of Norway (Nasjonalbiblioteket; NB) has commissioned a wordnet for Norwegian, and the results, which will be made publicly available, are expected before the end of the META-NORD project.

## 4. Cooperation with other infrastructure initiatives

The participating partners will join the NEALT special interest group on resource infrastructures, SIG-INFRA, founded in 2007. The purpose of SIG-INFRA is to strengthen the cooperation regarding language technology infrastructure (LTI) between the countries and regions associated to NEALT, to coordinate activities related to LTI in the associated countries and to promote the development and use of joint LTI in the associated countries. META-NORD will also align itself with the activities of the CLARIN ERIC, if and when it comes into existence, in order to avoid duplication of work and promote long-time viability of the META-NORD results.

For the distribution of Norwegian language resources and tools, it is relevant that the National Library of Norway (Nasjonalbiblioteket; NB) has recently received a mandate to establish The Language Technology Resource Collection for Norwegian – Språkbanken. This implies that a unit at NB will be a natural centre for archiving and distributing language resources for Norwegian. Their mandate includes the aim to provide the language industries with basic resources and tools to develop better services and products for communication and access to information. The Norwegian META-NORD team has established cooperation with NB in the areas of meta-data for future cataloguing and archiving of information. In June 2011, NB made the first language data sets available on its website.

UCPH plans to distribute the Danish language resources and tools which will be upgraded in META-NORD in task 3.1, i.e. the Danish computational lexicon (STO) and the LSP corpus, through the Danish research infrastructure, CLARIN. CLARIN will become a part of the larger Danish infrastructure for Humanities (DigHumLab) and constitute part of the Danish membership contribution to the CLARIN European Research Infrastructure Consortium



(CLARIN-ERIC). The plan is to make the resources searchable and downloadable from the version of the Danish CLARIN platform which is operational now.

## 5. Conclusion

Even though the long-term viability and sustainability strategies are not yet fully developed the most important thing is that everyone agrees that there is a need to create a sustainable mechanism for cooperation and resource sharing as part of the European open linguistic infrastructure, and these plans are already well on the way.

One of the most important actions in the META-NORD sustainability plan is the forming of the META Coordination Councils (META-CC) that will involve the key national stakeholders in the language technology area and lead to the building of awareness and support for continuation of activities. Business models will be evaluated for the META-NORD infrastructure, such as the clearinghouse concept which aims at cooperation on a long-time storage of resources with other similar service centres in other European countries, using a cost-sharing principles. Inclusion of new actors and countries in the business model will be further elaborated in the META-NORD project in cooperation with the best practices of other META-SHARE centres. In order to ensure sustainability and continuation of META-NORD activities the possibilities will be discussed to create national META-CENTRES by the project partner institutions.

Several plans already exist for the sustainability of the main types of resources that will be developed within META-NORD. Treebank resources will be maintained and disseminated through the INESS project and further maintained and supported by UIB. Terminology resources will be ensured by the collaboration with the EuroTermBank, that is maintained and supported by the EuroTermBank Consortium and the Tilde company. The interlinked word-net resources will be maintained and disseminated by UCPH through the Danish research infrastructure, CLARIN.

The participating partners will join SIG-INFRA, the NEALT special interest group in order to strengthen cooperation and coordinate activities related to LTI and to promote the development and use of joint LTI in the associated countries.

At this early stage in the project a more detailed plan cannot be provided. However, the META-NORD partners are working on a long-time viability and sustainability strategy that will be presented towards the end of the project period (Deliverable D5.3.2, M24).