



# **META-NORD**

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

**Project no. 270899**

## **Deliverable D2.4**

**Selection of resources, agreements, detailed work plan**

**Version No. 1.0**

**01/12/2011**

**Document Information**

Deliverable number:	D2.4
Deliverable title:	Selection of resources, agreements, detailed work plan
Due date of deliverable:	30.11.2011
Actual submission date of deliverable:	01.12.2011
Main Author(s):	Kadri Vider, Krista Liin
Participants:	All
Internal reviewer:	UHEL
Workpackage:	WP2
Workpackage title:	Analysis and selection of language resources
Workpackage leader:	UT
Dissemination Level:	PP
Version:	1.0
Keywords:	Resources, agreements, work plan

**History of Versions**

Version	Date	Status	Name of the Author (Partner)	Contributions	Description/ Approval Level
0.1	26/10/2011	Fishbone	UT	Kaili Mürisep, Kadri Vider	Draft
0.2	28/10/2011	First draft	UT	Kadri Vider, Krista Liin	Draft
0.3	30/11/2011	Stable draft	UT	All partners	Submitted to coordinator
0.4	30/11/2011	Prefinal	UT	Krista Liin	Submitted to coordinator
1.0	01/11/2011	Final	Tilde	Final review	Submitted to PO

**EXECUTIVE SUMMARY**

This report gives an overview of the planned activities regarding language resources until the completion of the project. It contains work plans for adding and including language resources for each of the 8 META-NORD project countries, describing milestones, the schedule for and format of uploading resources relative to the upload batches, as well as their access/reuse policies, current status of these resources, possible risks, and work to be done. A detailed work plan has been made for each partner describing intermediate and final outputs and the relevant human and technical resources.

The report also outlines the work plan for working groups in three horizontal actions (wordnets, treebanks, and terminology) as well as the upload nodes that will serve as the link between the META-NORD partners and the META-SHARE repository.

The plans in the current document are not final; they may and, probably, will change during the project.

## Table of Contents

<b>Abbreviations .....</b>	<b>4</b>
<b>1. Background.....</b>	<b>5</b>
<b>2. Selection of resources for upload batches.....</b>	<b>6</b>
2.1. Latvia (TILDE).....	6
2.1.1. Plans for uploads .....	6
2.1.2. Work plan .....	10
2.2. Denmark (UCPH).....	12
2.2.1. Plans for uploads .....	12
2.2.2. Work plan .....	15
2.3. Estonia (UT) .....	17
2.3.1. Plans for uploads .....	17
2.3.2. Work plan .....	22
2.4. Norway (UiB).....	26
2.4.1. Plans for uploads .....	26
2.4.2. Work plan .....	31
2.5. Finland (UHEL).....	35
2.5.1. Plans for uploads .....	35
2.5.2. Work plan .....	40
2.6. Iceland (HI).....	44
2.6.1. Plans for uploads .....	44
2.6.2. Work plan .....	48
2.7. Lithuania (LKI).....	50
2.7.1. Plans for uploads .....	50
2.7.2. Work plan .....	52
2.8. Sweden (UGOT).....	55
2.8.1. Plans for uploads .....	55
2.8.2. Work plan .....	58
<b>3. Horizontal Actions .....</b>	<b>60</b>
3.1. Horizontal Action on Treebanking .....	60
3.1.1. Objectives and overall plan.....	60
3.1.2. Action plan with subtasks and milestones .....	60
3.2. Horizontal Action on Wordnets.....	61
3.3. Horizontal Action on Multilingual Terminology.....	61
3.3.1. Content. New Terminology Resources .....	61
3.3.2. Integration. Interlinking of EuroTermBank with META-SHARE and with other external terminology resources.....	62
3.3.3. Collaboration .....	62
3.4. Partners establishing nodes.....	65
3.4.1. General .....	65
3.4.2. UGOT node .....	65
<b>4. Conclusions .....</b>	<b>66</b>
<b>5. List of tables.....</b>	<b>67</b>

## Abbreviations

Abbreviation	Term/definition
LRT	Language resources and tools
DoW	The META-NORD Description of Work document
CC	Creative Commons
TILDE	TILDE SIA (Latvia )
UCPH	Københavns Universitet (Danmark)
UT	Tartu Ülikool (Estonia)
UIB	Universitetet i Bergen Organisasjonsedd (Norway)
UHEL	Helsingin Yliopisto (Finland)
HI	Haskoli Islands (Iceland)
LKI	Lietuviu Kalbos Institutas (Lithuania)
UGOT	Göteborgs Universitet (Sweden)

**Table 1. Abbreviations**

## 1. Background

During the META-NORD project, 3 larger upload batches for resources are scheduled. This report gives detailed description of the plans of which resources a partner will include in which batch, as well as the work involved and agreements to be reached. The work plans are given first in regard to the partner involved and then further divided according to upload batches.

Each partner has composed a list of resources that are worth of further development, standardisation and/or distribution, based on the list in deliverable 2.2 “Report on resources (actually or potentially) available to the consortium” and results of deliverable 2.3 “Report on methodology and criteria followed for the selection of resources”. Taking into account that the resources can vary considerably among the partners, the reasons and criteria for including certain resources in specific batches are given by the partners who will be uploading them. Since according to META-SHARE it was only possible to upload unrestricted resources, excluding the tools for batch 1, this was a common restriction for all the partners. Otherwise, there were some guidelines, but no restrictions were given on which resources to upload and when. In some cases, short discussions were held about categories of language resources that are not included in our lists but would be considered important for a language, as described in deliverable 2.3. The partners have also defined milestones for their uploads by setting intermediate and final goals concerning the language resources, describing their further work until the project end.

A general plan has been made as to which batch the resource should be uploaded, including the nature of upgrades in later batches. Intended uploads of the resources are distinguished by format: due to availability or restrictions on a resource, it could be uploaded to META-SHARE as metadata only, an URL, or the resource itself. Descriptions have been provided for each resource regarding the state it should be in for upload, the licensing terms and work that needs to be done before the upload, as well as possible risks that could threaten it.

In the second part of the report, each META-NORD partner has given a detailed work plan describing the effort involved in bringing those resources to the state planned for their upload: a description of work to be done for each upload of the resource, an estimate of required time, as well as a timetable of scheduled work and the intended result. This part of the report is the most subject to change, especially since it is difficult to estimate the work needed and potential risks for resources scheduled for later upload batches. As such, detailed schedules for some resources will be specified at a later stage, when more information about the state of such resources becomes available.

## **2. Selection of resources for upload batches**

### ***2.1. Latvia (TILDE)***

#### ***2.1.1.Plans for uploads***

##### **1. batch (M10)**

Metadata upload: 7 (7 resources)

Resource/tool upload (either URL or resource itself): 7.

All of the 7 resources for the first batch intend to be as a first version of the resources. for next batches lexical resources (dictionaries and Eurotermbank) will be updated and available through META-SHARE API. Corpora that is prepared for the first upload is a part of complete set of corpora, in later stages corpora will be upgraded to agreed standards and will be extended with additional data.

##### **2. batch (M18)**

Metadata upload: 1 (1 resources)

Resource/tool upload (either URL or resource itself): 8

In second batch one new resource will be uploaded to META-SHARE repository. 7 resources from the first batch will have second version, resources will be updated with additional data and made available through (1) META-SHARE API created for specific purpose to share specific resources or (2) downloadable.

##### **3. batch (M24)**

Metadata upload: 2 (2 resources)

Resource/tool upload (either URL or resource itself): 3

It should be noted that table does not indicate all potential available resources as they are not fully known to be available yet. During preparation of the second batch, Tilde will start negotiations with resources owners for resources that are outcomes of FP7 projects where Tildes is involved, e.g., ACCURAT Corpora and tools, TTC data etc. Those resources belong to project consortiums and the IPR/ownership is ongoing discussion.

Beside negotiations with mentioned consortiums, Tilde will investigate cooperation possibilities with previous indicated third parties networks described in D2.5 *Description of third parties networks for each language covered in the project.*

During preparation of the second and third batch, Tilde will investigate potential resources that were not known before and will start negotiations with resources owners.

**Table 2.1.1 . Plan for including Latvian resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Eurotermbank	Tilde	1	1 (Web access)	2 ( access through META-SHARE API, improved documentation, appropriate META-SHARE license will be applied), 3 (integrated with other terminology resources)	Terms of use: ( <a href="http://www.eurotermbank.com/Disclaimer.aspx">http://www.eurotermbank.com/Disclaimer.aspx</a> )	Eurotermbank resources are available through META-SHARE API, documented and integrated with other available terminology resources	
Lithuanian-Latvian dictionary	Tilde	1	1 (Web access)	2 (access through META-SHARE API, improved documentation, appropriate META-SHARE license will be applied)	Proprietary	Resource is available through META-SHARE API and documented, ~60 000 entries	
Latvian-Lithuanian dictionary	Tilde	1	1 (Web access)	2 (access through META-SHARE API, improved documentation, appropriate META-SHARE license will be applied)	Proprietary	Resource is available through META-SHARE API and documented, ~43 000 entries	

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
				applied)			
Estonian-Latvian dictionary	Tilde	1	1 (Web access)	2 ( access through META-SHARE API, improved documentation, appropriate META-SHARE license will be applied), 3 (interlinked with other termbases)	Proprietary	Resource is available through META-SHARE API with documentation based on META-SHARE standards, ~26 000 entries	
Latvian-English legislation corpus of Republic of Latvia	Tilde	1	1	2 (updated with more data), 3 (updated with more data)	METASHARE COMMONS BY NC ND	Resource is upgraded to META-SHARE standards and documented, ~4 million words	
Multilingual dictionary of person names	Tilde	1	1 (Web access)	2 (access through META-SHARE API, improved documentation, appropriate META-SHARE license will be applied)	Proprietary	Resource is available through META-SHARE API and documented, ~4 million words	
Tilde's POS-tagger	Tilde	3	3	3 (API access, data input for tools , licensing type will be applied)	n/a	Tilde POS-tagger available through META-SHARE API and documented	
Corpus of Latvian literature	Tilde	1	1 (Web access)	2 ( improved documentation and content )	METASHARE COMMONS BY NC ND	Resource is available through META-SHARE API, documentation is based	

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
						on META-SHARE standards, ~15 000 pages	
EASTIN-CL multilingual ontology	EASTIN-CL Consortium	3	3	3(agreements with data owners on licensing, access policy and methods)	n/a	Access to EASTIN-CL multilingual ontology	
Latvian Russian person names and geo names glossary	Tilde	2	2	2(available through META-SHARE API, upgraded to the standards, documented, appropriate META-SHARE license will be applied)	METASHARE COMMONS BY NC ND	Resource is available through META-SHARE API with documentation based on META-SHARE standards	

META-SHARE node will be set at Tilde. Tilde has started work on setting up META-SHARE v1.1 and work is in progress. The planned time for the META-SHARE node to be set up is M12 of the project.

### 2.1.2. Work plan

**Table 2.1.2 Work plan for Latvian resources**

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Eurotermbank	1,2,3	Describe and update metadata according to the META-SHARE metadata schema, document, integrate with other available terminology resources, provide Web access, integrate in META-SHARE	3PM	M7	M24	Eurotermbank resources are available through META-SHARE API, documented and integrated with other available terminology resources
Lithuanian-Latvian dictionary	1,2	Describe and update metadata according to the META-SHARE metadata schema, document, provide Web access, integrate in META-SHARE	0,5PM	M9	M18	Resource is available through META-SHARE API and documented, ~60 000 entries
Latvian-Lithuanian dictionary	1,2	Describe and update metadata according to the META-SHARE metadata schema, document, provide Web access, integrate in META-SHARE	0,5PM	M0	M18	Resource is available through META-SHARE API and documented, ~43 000 entries
Estonian-Latvian dictionary	1,2	Describe and update metadata according to the META-SHARE metadata schema, document, provide Web access, integrate in META-SHARE	0,5PM	M9	M18	Resource is available through META-SHARE API with documentation based on META-SHARE standards, ~26 000 entries
Latvian-English legislation corpus of Republic of Latvia	1,2,3	Update content, upgrade to TMX, describe and update metadata according to the META-SHARE metadata	3PM	M7	M24	Resource is upgraded to META-SHARE standards and documented, ~4 million words

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
		schema, document, provide access				
Multilingual dictionary of person names	1,2	Describe and update metadata according to the META-SHARE metadata schema, document, provide Web access, integrate in META-SHARE	1PM	M9	M18	Resource is available through META-SHARE API and documented, ~4 million words
Tilde's POS-tagger	3	Describe metadata according to the META-SHARE metadata schema, document, provide access to according to META-SHARE requirements and specification	3PM	M16	M24	Tilde POS-tagger available through META-SHARE API and documented
Corpus of Latvian literature	1,2	Update content, upgrade to standards, describe and update metadata according to the META-SHARE metadata schema, document, provide access	0,5PM	M9	M18	Resource is available through META-SHARE API, documentation is based on META-SHARE standards, ~15 000 pages
EASTIN-CL multilingual ontology	3	Describe metadata according to the META-SHARE metadata schema, prepare in accordance with META-SHARE requirements and standards	~4PM	M16	M24	Access to EASTIN-CL multilingual ontology
Latvian Russian person names and geo names glossary	2	Describe and update metadata according to the META-SHARE metadata schema, document, provide access	1,5PM	M14	M18	Resource is available through META-SHARE API with documentation based on META-SHARE standards

## 2.2. Denmark (UCPH)

### 2.2.1.Plans for uploads

#### 1. batch (M10)

Metadata upload: 3

Resource/tool upload: 3

#### 2. batch (M18)

Metadata upload: 2

Resource/tool upload: 2

#### 3. batch (M24)

Metadata upload: 8

Resource/tool upload: 8

**Table 2.2.1 Plan for including Danish resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Danish wordnet, DanNet	UCPH & Det Danske Sprog- og Litteraturselskab	1,2,3	1,2,3	2,3	Princeton Wordnet CC BY v3.0	Completed	Minor corrections and full documentation in English
Cross-lingually linked resources (with FIN and SWE)	UCPH (with UGOT & UHEL)	2,3	2,3	2,3	CC BY	Completed pilot linking	Validation and development of a common web-interface for crosslingual resources

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
SprogTeknologisk Ordbase	UCPH	1, 2, 3	1, 2, 3	1 ( subpart of morphology in LMF) 2 ( 80,000 morphological entries in LMF with links to ISOcat) 3 ( 40,000 syntactic entries in LMF)	ACA+NC	Completed upgrade of STO to LMF with links to ISOcat.	Potential risk: 1)Problems with consistency in ISOcat linking. 2) Lack of overlap with SBLEX 3) Problems when linking STO and DanNet due to very different structuring principles
Copenhagen Dependency Treebanks	CBS	1	1	-	GPL v3.0	stable	-
The Copenhagen Danish-English Dependency Treebank	CBS	1	1	-	GPL v3.0	stable	-
Danish first encounters NOMCO corpus	UCPH	3	3	-	To be negotiated	Completed	Only tentative answer from NOMCO group. Availability to be negotiated
Reference corpus for Danish	Danish Language Council	-	-	-	-	Completed	No upload
Corpus of sublanguage texts (2000 – 2010)	University of Copenhagen - CST and Danish Language Council	-	-	-	-	Completed	Most presumably no upload
Danish XLE grammar	CBS/UCPH	-	-	-	-	Completed	No upload, used to build up treebanks

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
CstTokeniser	UCPH	3	3	3 (sentence segmentation + documentation)	ACA+NC	stable version	
CstNER	UCPH	3	3	3 (upgrade robustness + documentation)	ACA+NC	stable version	
CstTagger	UCPH	3	3	3(xml handling + documentation)	ACA+NC	stable version	
CstLemma	UCPH	3	3	3(xml handling + documentation)	ACA+NC	stable version	
CstKeyExt	UCPH	-	-	-	-	stable version	No upload
CstNP-Rec	UCPH	-	-	-	-	stable version	No upload
CstRep	UCPH	-	-	-	-	stable version	No upload
HPSG –grammar	UCPH	-	-	-	-	stable version	No upload

As can be seen, several of the resources from the table above will not be uploaded as part of Metanord either due to short-comings in coverage and/or stability or because of lack of accessibility. A few resources are still under consideration, such as NOMCO and Corpus of Sublanguage texts.

### 2.2.2. Work plan

**Table 2.2.2 Work plan for Danish resources**

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
DanNet	1	To finalize 65,000 synsets and link to 5,000 cores senses in Princeton WordNet.	2,5 PM	M3	M10	DanNet linked to Princeton core wordnet
DanNet	2	to finalize 30 page linguistic documentation in English. For crosslingual work, cf. section on horizontal action on wordnet.	2 PM	M12	M15	Documentation in English
DanNet	3	For cross-lingual milestone, cf. section on horizontal action on wordnet.	5 PM	M5	M24	Cf. Horizontal action on wordnets
STO upgraded to LMF:	1	About 80,000 morphological entries (nouns, adjectives and verbs) from STO converted to LMF.	2,5 PM	M5	M10	Conversion of part of STO morphological entries to LMF structure
STO upgraded to LMF:	2	81,500 morphological entries from STO. Features and values linked to ISOcat.	3 PM	M12	M17	All STO morphological entries upgraded to LMF structure, linking features and values with ISOcat. Coherence with the LMF version of SBLEX to the extent possible.
STO upgraded to LMF:	3	41,000 syntactic entries from STO added to the morphological entries of batch 2. Features and values linked to ISOcat. A number of syntactic entries linked to synsets of DanNet.	4 PM	M20	M24	Upgrade of the syntactic part of STO to LMF with feature and values links to ISOcat. Links of a number of syntactic entries of STO to the semantic synsets of DanNet.
Copenhagen Dependency Treebanks:	1	100000 words (Danish) + 95000 words (Danish – English)	0,5 PM	M9	M10	

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
parallel treebanks for the 1st chapter of Sofie's World	2		2 PM	M12	M17	Sentence aligned parallel treebanks for the first chapter of Sofie's World. Cf. also plan for horizontal action on treebanks.
Parallel treebanks for selected parts of the JRC-ACQUIS Multilingual Parallel Corpus	3	Phrase aligned parallel treebanks for selected parts of the JRC ACQUIS multilingual parallel corpus	3 PM	M18	M24	Cf. plan for horizontal action on treebanks.
CST Tools (CstTokeniser, CstNER, CstTagger, CstLemma):	3	Upgrade sentence segmentation functionality for CstTokeniser, robustness for CstNER, xml handling functionality for CstTagger and CstLemma. Complete documentation for the 4 tools.	2 PM	M18	M24	Improved CST tools

## **2.3. Estonia (UT)**

### **2.3.1.Plans for uploads**

#### **1. batch (M10)**

Metadata upload: 8

Resource/tool upload: 2

In the 1st batch will be uploaded unrestricted resources that are owned by UT. Since this is the first upload batch, two smaller written resources will be uploaded fully, the rest will have a metadata upload.

#### **2. batch (M18)**

Metadata upload: 6

Resource/tool upload: 5

2nd batch has the first uploads for the tools, as META-SHARE will be ready for them. Also included are less restricted written resources by other partners.

For uploads of actual resources in batch 2, there are new versions of batch 1 resources, as well as some corpora and written resources for which metadata had been uploaded in batch 1.

#### **3. batch (M24)**

Metadata upload:6

Resource/tool upload: 5

In batch 3 are included first uploads of larger resources and spoken resources with more complicated restrictions, also tools for which the licence negotiations will take longer or that will be more difficult to describe properly.

Also included are any new versions for resources uploaded in previous batches.

There are practically no multimodal resources for Estonian. Several of the tools included (e.g. speech technology tools) are not of very high quality and need to be developed further. Corpora with deeper linguistic annotation are small, the same can be said about the available cross-lingual corpora (parallel and comparable).

From the perspective of multilingual web there are several resources that could be useful - for example both named entity recognizer and automatic summarization tool have been made for Estonian, but such resources tend to lack standardization and detailed documentation. For the resources which are considered not of very high quality the effort to make them available through META-NET would probably take too long and as such they are not included in the tables.

**Table 2.3.1 Plan for including Estonian resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>1</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
The Estonian Reference Corpus	UT	1	1 (URL)	2 (upload of the resource )	CC-0	Stable version, fully available	Copyright agreement with every text's author is impossible...
Treebank	UT	1	3 (metadata update, upload)	0	CC-0	Stable version, fully available	Work to be done: considerable extending, consistency checking, formal validating, translating the annotation manual into English
Estonian WordNet	UT	1	1	3 (with added new synsets)	Public for academic use	Stable version, available via internet	
BABEL Estonian Database	IOC	3	-	-	n/a	Resource fully available	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, confirm licensing and access terms
Corpora of morphologically disambiguated texts	UT	1	1 (upload)	2 (new version: correction of mistakes possibly)	CC-0 (batch 3)	Stable version, fully available	Work to be done: consistency checking of the corpus, formal validating of annotation

<sup>1</sup> Value is 0 if no upgrades are planned

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>1</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
				revealed while uploading and validating against the META-SHARE schema)			
Corpora with shallow syntactic annotation	UT	1	2 (upload)	2 (metadata update),	CC-0 (batch 3)	Stable version, fully available	Work to be done: consistency checking of the corpus, formal validating of annotation
Corpus of emotional speech	IEL	2	-	0	n/a	Stable version, fully available	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, confirm licensing and access terms
Corpus of Institute of Estonian Language	IEL	2	-	0	n/a	Stable version, fully available	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, confirm licensing and access terms
Corpus of Spoken Estonian	UT	3	-	0	RES	Stable version	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, confirm access restrictions
Cross-lingually linked resource	UT, UHEL	3	3	0	n/a	Stable version	Selecting, linking and validating the resource

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>1</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
Dictionaries of Estonian-English, Estonian-Russian,	IEL	2	-	0	n/a	Stable version, fully available	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, licensing and access terms
English-Estonian and Estonian-English parallel corpus	UT	1	3 (upload)	3 (metadata update)	CC-0	Stable version, fully available	Checking for possible alignment mistakes, changing the annotation from local standard to XML
Estonian Foreign Accent Corpus	IOC	3	-	0	n/a	Stable version	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, licensing and access terms
Monolingual dictionaries	IEL	2	-	0	n/a	Stable version, fully available	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, licensing and access terms
Semantically disambiguated corpus	UT	1	2 (upload)	2 (metadata update)	CC-0 (batch 3)	Stable version, fully available	Work to be done: consistency checking of the corpus, formal validating of annotation
The database of Estonian verbal multi-word expressions	UT	1	1 (upload)	2 (new version: correction of mistakes possibly revealed while uploading and	CC-0 (batch 3)	Stable version, fully available	formal validating

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>1</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
				validating against the META-SHARE schema)			
Estonian text-speech synthesizer	IEL/IOC	3	-	0	n/a	Stable version	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, licensing and access terms
Morphological analyzer	Filosoft	3	-	0	n/a	Stable version available for use	Need to negotiate with owner
Morphological analyzer	IEL	2	-	0	n/a	Stable version	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, licensing and access terms
Morph syntactic disambiguator and shallow parser	UT	2	3 (URL)	3 (metadata update)	NC	Stable version	Work to be done: Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards, Presumes a morphological analyzer

### 2.3.2. Work plan

**Table 2.3.2 Work plan for Estonian resources**

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Corpora of morphologically disambiguated texts	1	Describe metadata in META-SHARE format; Negotiate licence; Consistency checking of the corpus; Formal validation of annotations against the META-SHARE schema; trying to upload the resource in the SVN repository	~0,5 PM	M10	M10	Metadata uploaded; Resource ready for upload; but we will be prepared for the possible drawbacks while uploading the resource itself
Corpora of morphologically disambiguated texts	2	Continuing to negotiate the license; Correcting of the possible mistakes revealed while trying to upload the resource; Upload of the corrected version and formal validation of annotations against the META-SHARE schema; updating the meta-data, if needed	0,5 PM	M12	M17	License acquired; Any mistakes possibly revealed while uploading and validating against the META-SHARE schema have been corrected; meta-data up-to-dated, if needed
The Estonian Reference Corpus	1	Describe metadata in META-SHARE format; Negotiate licence, consistency checking, specifying URL for META-SHARE	0,4 PM	M10	M10	Metadata uploaded, URL given
The Estonian Reference Corpus	2	Negotiate licence, consistency checking, validating against META-SHARE schema, correcting the possible mistakes, uploading in META-SHARE SVN	3 PM	M04	M06	Metadata updated, resource uploaded
The database of Estonian verbal multi-word expressions	1	Describe metadata in META-SHARE format; Negotiate licence; converting the database to XML; Formal validation of annotations against the META-SHARE schema; trying to upload the resource in the SVN repository	1 PM	M10	M10	Metadata uploaded; Resource ready for upload; but we will be prepared for the possible drawbacks while uploading the resource itself
The database of	2	Continue to negotiate the license;	0,5 PM	M11	M13	License acquired;

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Estonian verbal multi-word expressions		Correcting of the possible annotation/coding mistakes revealed while trying to upload the resource; Upload of the corrected version and formal validation of annotations against the META-SHARE schema				Any mistakes possibly revealed while uploading and validating against the META-SHARE schema have been corrected; meta-data up-to-dated, if needed
The database of Estonian verbal multi-word expressions	3	Adding morphological information to the database, converting it to the META-SHARE format, uploading the new version of the database, updating the metadata.	2 PM	M16	M18	New version of metadata uploaded; new version of the resource uploaded
Semantically disambiguated corpus	1	Metadata uploaded and validated	-	M10	M10	Metadata uploaded
Semantically disambiguated corpus	2	Update the metadata; Consistency checking; Validate the resource against META-SHARE schema;	unknown	M11	M18	Resource in valid xml
English-Estonian and Estonian-English parallel corpus	1	Describe metadata in META-SHARE format; Negotiate licence;	0,4 PM	M10	M10	Metadata uploaded
English-Estonian and Estonian-English parallel corpus	2	Continue to negotiate the license; Consistency checking; converting the markup to XML; alignments checking	3 PM	M16	M18	Resource in valid xml
English-Estonian and Estonian-English parallel corpus	3	Update the metadata; validate the resource against META-SHARE schema; correct possible mistakes; upload in SVN	2 PM	M19	M22	Metadata updated, if needed; resource uploaded
Corpora with shallow syntactic annotation	1	Describe metadata in META-SHARE format; Negotiate licence;	0,4 PM	M10	M10	Metadata uploaded

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Corpora with shallow syntactic annotation	2	Continue to negotiate the license; Consistency checking; possibly converting the markup	2 PM	M16	M18	Accuracy and consistency of corpus annotation checked, possible errors fixed, documentation improved and translated.
Corpora with shallow syntactic annotation	3	Update the metadata; validate the resource against META-SHARE schema; correct possible mistakes; upload in SVN	1 PM	M19	M24	Metadata up-to-date Validated resource uploaded in SVN
Corpus of emotional speech	2	Describe metadata in META-SHARE format; Consistency checking, prepare resource and metadata in META-SHARE format negotiate licence	unknown	M11	M18	Metadata uploaded
Corpus of Spoken Estonian	3	Describe metadata in META-SHARE format; Negotiate licence; consistency checking;	unknown	M11	M24	Metadata uploaded Metadata uploaded, the status of the resource will depend on license
Estonian Treebank	1	Describe metadata in META-SHARE format; Negotiate licence; consistency checking; extending the resource	1 PM	M10	M10	Metadata documented in META-SHARE resource extended
Estonian Treebank	2	Continue to negotiate the license; Consistency and accuracy checking; fixing the possible errors; Extending the resource; Improving the documentation of the resource	4 PM	M11	M18	License acquired resource extended, Accuracy and consistency of treebank annotation checked, possible errors fixed, documentation improved and translated.
Estonian Treebank	3	Extending the resource;	3 PM	M19	M24	Metadata up-to-date;

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
		Update the metadata; validate the resource against META-SHARE schema; correct possible mistakes; upload in SVN				resource uploaded
Estonian WordNet	1	Extending the resource; Describe metadata in META-SHARE format; Converting the markup to XML; Linking to Princeton Core WordNet; Correct possible mistakes	6 PM	M7	M10	Metadata uploaded; Resource uploaded;
Estonian WordNet	2	Extending the resource; Checking possible mistakes; Helping develop a tool for extracting/viewing bilingual links with other languages; Selecting coverage for validation of linked resources (approx. 1000 synsets);	8 PM	M11	M18	Metadata updated, if needed; New version of resource uploaded
Estonian WordNet	3	Extending the resource; Correct possible mistakes; Linking wordnets (Estonian-Finnish); Valuation of bilingual experts (for Fin-Est);	6 PM	M19	M24	New version of resource uploaded; Uniform web interface for browsing monolingual and linked resources; Validation report of linked resources

## **2.4. Norway (UiB)**

### **2.4.1. Plans for uploads**

#### **1. batch (M10)**

Metadata upload: 12 (11 resources, 1 tool)

Resource/tool upload: 11 (10 resources, 1 tool).

Nine of the eleven uploaded resources/tools are considered to be completed by Batch M10. The remaining three will be followed up in the following batches: Norsk-Vietnamesisk ordbok is an existing, third party resource, which needs some formatting upgrading and the licence needs to be set. The Sofie treebank and the TRIS Spanish-German are still under development by a third party owner. We cannot fully control the progress for these three resources, although the plans and the current progress appears to be satisfactory.

#### **2. batch (M18)**

Please note that our Table 2.4.1 enumerates the *currently foreseen* resources and tools. On the one hand, further resources may still be added; on the other hand the licensing issues have not yet been fully negotiated for some of the currently listed resources.

Metadata upload for batch M18: 8 (7 resources, 1 tool)

Resource/tool upload: 9 (8 resources, 1 tool)

As mentioned under Batch (M10), three resources with a metadata upload in M10 are followed up in Batch M18 (for the Sofie treebank we expand from the Norwegian treebank to include parallel treebanks in selected languages; for the TRIS Spanish-German parallel corpus we will add metadata to each document while the third party developer expands the resource; for the Norwegian-Vietnamese digital dictionary we do file conversion, resolve licencing issues and upload the resource).

Nine resources are scheduled to be uploaded (i.e. upload of the resource itself) in Batch M18. We see a potential risk in that seven of them are third party resources (see comments in the table below). They all have a good potential for being made available and META-NORD has been in contact with the owners. However, META-NORD still needs to negotiate licences, and the exact status of the resources is still not clear as regards the need for resource upgrading (e.g. file format conversion).

### 3. batch (M24)

Once again, please note that Table 2.4.1 enumerates the *currently foreseen* resources and tools. On the one hand, further resources may still be added; on the other hand the licensing issues have not yet been fully negotiated for some of the currently listed resources.

Metadata upload: 6 resources

Resource/tool upload: 6 resources

These six resources are all third party resources, and their completion is therefore beyond our control. Four of them (in cooperation with the Norwegian National Library) are expected to be complete and uploadable by the time of Batch M24. As regards the other two resources, licensing issues are to be negotiated.

A beta version of the first Norwegian wordnet will be available in 2012 and the complete resource is scheduled to ready by January 2013. The Norwegian wordnet will be linked to concepts in the Danish wordnet, which in turn is linked to the Princeton WordNet. As of now, we do not currently foresee a need to explicitly add links directly from the Norwegian WordNet to Princeton WordNet.

The currently missing infrastructure for Norwegian terminology will probably be followed up in the upcoming project CLARINO.

**Table 2.4.1 Plan for including Norwegian resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>2</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
Acoustic database for Danish	The Norwegian National Library	1	1	0	CC0	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that the provider has not settled for a particular licence type yet.
Acoustic database for Norwegian	The Norwegian National Library	1	1	0	CC0	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that

<sup>2</sup> Value is 0 if no upgrades are planned

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>2</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
							the provider has not settled for a particular licence type yet.
Acoustic database for Swedish	The Norwegian National Library	1	1	0	CC0	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that the provider has not settled for a particular licence type yet.
Lexical database for Danish	The Norwegian National Library	1	1	0	CC0	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that the provider has not settled for a particular licence type yet.
Lexical database for Norwegian	The Norwegian National Library	1	1	0	CC0	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that the provider has not settled for a particular licence type yet.
Lexical database for Swedish	The Norwegian National Library	1	1	0	CC0	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that the provider has not settled for a particular licence type yet.
Norsk ordbank	The Norwegian National Library	1	1	0	CLARIN PUB	completed	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format; licence type may be renegotiated in the sense that the provider has not settled for a particular licence type yet.
Oslo-Bergen tagger	UIB	1	1	0	GPL	completed	-
SCARRIE lexicon	UIB	1	1	0	CC-BY	completed	-

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>2</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
Sofietrebanken	UIB	1	1	2[determine parallel languages; convert file formats; parallel alignment]  3[added language pairs; develop standoff annotation for parallel alignment at sentence level] tool itself and metadata available	CLARIN RES	completed	Norwegian quality-assured material is uploaded for the first batch, further languages follow in the two later batches. The conditions of use for this resource is not presently covered by a fixed licence but according to our information META-SHARE are developing licences that will fit. For now, we categorize the licence as a CLARIN RES.
TRIS Spanish-German parallel corpus	UIB	1	1	2[improved metadata per document]  3[complete encoding]	CLARIN RES	completed	Potential risk: Third party resource, still under development. The licence type may still be redefined
Norwegian-Vietnamese digital dictionary	UIB	1	2	2[convert from XML to LMF, upload]	CLARIN RES	completed	Potential risk: Third party resource. The resource exists as XML, needs format conversion and the licencing negotiations are not complete by D2.4

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>2</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
				resource]			
Acquis communautaire	UIB	2	2	0	CC0	completed	Describe metadata in META-SHARE format; Files must be selected and integrated in the INESS treebank with a satisfactory file format.
Leksikografisk bokmålskorpus	UIB	2	2	0	2	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats
Milterm	UIB	2	2	0	2	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats
NHH Termbase	UIB	2	2	0	2	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats
UHR's Termbase for Norwegian higher education institutions	UIB	2	2	0	2	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats
Stadsnamnsamlinga	UIB	2	2	0	2	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats
Translation Corpus Aligner 2	UIB	2	2	0	GPL	completed	licence may still be redefined; the resource as such is ready.
The Norwegian-Spanish Parallel Corpus	UIB/Uni Computing	2	2	0	CLARIN ACA	completed	Potential risk: Third party resource, still under development. The user rights may be renegotiated to become more open.
Det nynorske tekstkorpuset	UIB	3	3	0	3	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>2</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
International Computer Archive of Modern and Medieval English	UIB	3	3	0	3	completed	Potential risk: Third party resource. already existing resource, needs to negotiate availability and possibly convert file formats
n-grams for Norwegian bokmål and nynorsk	The Norwegian National Library	3	3	0	0	completed	Potential risk: resource is under development and rights are not yet cleared; Cooperate with the Norwegian National Library in describing metadata in META-SHARE format
Norwegian newspaper corpus	The Norwegian National Library	3	3	0	0	completed	Potential risk: rights are not yet fully cleared, Cooperate with the Norwegian National Library in describing metadata in META-SHARE format
Norwegian reference corpus for bokmål and nynorsk	The Norwegian National Library	3	3	0	0	completed	Potential risk: resource is under development and rights are not yet cleared; Cooperate with the Norwegian National Library in describing metadata in META-SHARE format
Norwegian wordnet	The Norwegian National Library	3	3	0	0	completed	Potential risk: resource is under development and rights are not yet cleared; Cooperate with the Norwegian National Library in describing metadata in META-SHARE format

### 2.4.2. Work plan

**Table 2.4.2 Work plan for Norwegian resources**

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Acoustic database for Danish	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	11, 2011	11, 2011	metadata and resource complete

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Acoustic database for Norwegian	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	2 hours	11, 2011	11, 2011	metadata and resource complete
Acoustic database for Swedish	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	3 hours	11, 2011	11, 2011	metadata and resource complete
Acquis communautaire	2	Describe metadata in META-SHARE format; Files must be selected and integrated in the INESS treebank with a satisfactory file format.	37,5 hours	11,2011	7,2012	complete
Det nynorske tekstkorpuset	3	Describe metadata in META-SHARE format; licencing conditions are to be negotiated, the database may need reformatting	15 hours	11,2011	1,2013	clarified terms of use; improved availability (final result depending on the licence agreement )
International Computer Archive of Modern and Medieval English	3	Describe metadata in META-SHARE format; licencing conditions are to be negotiated, the database may need reformatting	15 hours	11,2011	1,2013	clarified terms of use; improved availability (final result depending on the licence agreement )
Leksikografisk bokmålskorpus	2	negotiate licence, possibly reformatting	15 hours	12,2011	7,2012	complete
Lexical database for Danish	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	11, 2011	11, 2011	metadata and resource complete
Lexical database for Norwegian	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	2 hours	11, 2011	11, 2011	metadata and resource complete
Lexical database for Swedish	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	3 hours	11, 2011	11, 2011	metadata and resource complete
Milterm	2	Describe metadata in META-SHARE format; licencing conditions are to be negotiated, the database may need reformatting	22,5 hours	11,2011	7,2012	clarified terms of use; improved availability (final result depending on the licence agreement )
n-grams for Norwegian bokmål and nynorsk	3	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	01, 2013	01, 2013	metadata and resource complete
UHR's Termbase for	2	Describe metadata in META-SHARE format;	25 hours	12,2011	7,2012	clarified terms of use;

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Norwegian higher education institutions		negotiate licence.				improved availability (final result depending on the licence agreement )
NHH Termbase	2	Describe metadata in META-SHARE format; negotiate licence.	30 hours	12,2011	7,2012	clarified terms of use; improved availability (final result depending on the licence agreement )
Norsk ordbank	1	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	11, 2011	11, 2011	metadata and resource complete
Norwegian newspaper corpus	3	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	1,2013	01, 2013	metadata and resource complete
Norwegian reference corpus for bokmål and nynorsk	3	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	01, 2013	01, 2013	metadata and resource complete
Norwegian wordnet	3	Cooperate with the Norwegian National Library in describing metadata in META-SHARE format	1 hours	01, 2013	01, 2013	metadata and resource complete
Norwegian-Vietnamese digital dictionary	1	Describe metadata in META-SHARE format; clear the licence type, which is due after the submission date of D2.4.	15 hours	11,2011	11,2011	metadata ready, cleared licence
Norwegian-Vietnamese digital dictionary	2	convert from XML to LMF	15 hours	12,2011	7,2012	complete
Oslo-Bergen tagger	1	Describe metadata in META-SHARE format;upload	5 hours	11, 2011	11, 2011	tool itself and metadata available
SCARRIE lexicon	1	Convert from XML to LMF; add metadata	19 hours	11,2011	11,2011	metadata and resource complete
Sofietrebanken	1	Describe metadata in META-SHARE format; manual validation of sentence analysis; technical: make downloadable, convert to Tiger/Negra-XML file format.	520 hours	08, 2011	11, 2011	Norwegian sentence analyses quality-assured and uploaded.
Sofietrebanken	2	determine parallel languages; convert file formats;	520 hours	12,2011	07, 2012	Parallel treebank of Sofie

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
		parallel sentence alignment				texts available
Sofietrebanken	3	make experimental phrase alignment between Norwegian and Danish; add further parallel languages.	520 hours	8,2012	1,2013	Possibly: Parallel treebank of Sofie texts extended with further languages
Stadsnamnsamlinga	2	Describe metadata in META-SHARE format; licencing conditions are to be negotiated, probably the database needs reformatting	37,5 hours	11,2011	7,2012	clarified terms of use; improved availability (final result depending on the licence agreement )
The Norwegian-Spanish Parallel Corpus	1	Describe metadata in META-SHARE format;Negotiate licence with resource owner	268 hours	10, 2011	11, 2011	Metadata and resource( as much as planned for first batch) ready for upload
The Norwegian-Spanish Parallel Corpus	2	Add further books (scan, OCR-read, convert to XML), determine where to deposit the resource in the long term	37,5 hours	11, 2012	07, 2012	metadata and resource complete
Translation Corpus Aligner 2	2	Describe metadata in META-SHARE format; licence may still be redefined; the resource as such is ready.	7,5 hours	11,2011	7,2012	complete
TRIS Spanish-German parallel corpus	1	Describe metadata in META-SHARE format;clean and prepare xml code for aligned documents (austrian - Spanish),	173 hours	10, 2011	11, 2011	available parallel documents Austrian-Spanish
TRIS Spanish-German parallel corpus	2	Add xml-encoded documents for aligned German - Spanish; add xml metadata to each document, such as domain, source vs target language;	210 hours	12,2011	7,2012	improved metadata per document
TRIS Spanish-German parallel corpus	3	add sentence alignment (stand-off), add linguistic annotation (pos)	173 hours	8,2012	1, 2013	complete

## 2.5. Finland (UHEL)

### 2.5.1.Plans for uploads

#### 1. batch (M10)

Metadata upload: 7

Resource/tool upload: 7

#### 2. batch (M18)

Metadata upload: 12

Resource/tool upload: 9 resources and 3 tools, namely HFST, Lemmie and Open Source (Finnish) Morphology.

#### 3. batch (M24)

Metadata upload: 24

Resource/tool upload: 22 resources and 2 tools (TKK Voice Source Analysis and Parametrisation Toolkit and Morfessor). These resources and the previously mentioned tool are all provided by third parties, with whom we will have discussions on metadata and, if needed, on IPR-issues, too.

**Table 2.5.1 Plan for including Finnish resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Finnish TreeBank Grammar Definition Corpus	UHEL	1	1	3	CC-by	Completed/ stable version, fully available	This is the part of Finnish TreeBank that can be already uploaded in batch one
Finnish WordNet	UHEL	1	1	2	CC-by, Wordnet 3.0	Completed/ stable	

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
						version, fully available	
Written corpora of old literary Finnish (Vanha kirjasuomi)	KOTUS	1	-	n/a	EUPL version 1.1.	Completed/ stable version, fully available	only browsable
Corpus of early modern Finnish (Varhaisnykysuomen korpus)	KOTUS	1	-	n/a	EUPL version 1.1.	Completed/ stable version, fully available	only browsable
Finnish literature classics (Suomalaisen kirjallisuuden klassikoita)	KOTUS	1	-	n/a	EUPL version 1.1.	Completed/ stable version, fully available	only browsable
Up-to-date word list of modern Finnish (Ajantasainen nykysuomen sanalista)	KOTUS	1	-	n/a	GNU LGPL	Completed/ stable version, fully available	-
Frequency list of words in written Finnish (Kirjoitetun suomen kielen sanojen taajuuslista)	KOTUS	1	-	n/a	EUPL version 1.1.	Completed/ stable version, fully available	-
Kansainvälinen oppijansuomen korpus (ICLFI)	OY	2	-	n/a	2	Completed/ stable version, fully available	Possible risks: IPR issues
Corpus of Conversational Finnish (Keskusteluntutkimuksen arkisto)	UHEL	2	-	n/a	2	Completed/ stable version, fully available	Possible risks: IPR issues
Open Source (Finnish) Morphology	UHEL	2	2	0	GNU LGPL	Completed/ stable version, fully available	-
Oulu corpus (Language Bank Of Finland)	CSC	2	-	n/a	2	Completed/ stable version, fully available	-
Geographic Names Register of the National Land Survey	KOTUS	2	-	n/a	2	Completed/ stable version, fully available	-
Samples of Spoken Finnish	KOTUS	2	-	n/a	2	Completed/ stable	-

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
(Suomen kielen näytteitä)						version, fully available	
Helsinki Finite-State Transducer Technology	UHEL	2	2	0	GNU GPL	Completed/ stable version, fully available	
Finland-Swedish Text Collection (Kielipankki, Language Bank of Finland)	CSC	2	-	n/a	2	Completed/ stable version, fully available	
Finnish Text Collection (Kielipankki, Language Bank of Finland)	CSC	2	-	n/a	2	Completed/ stable version, fully available	
Lemmie	CSC	2	-	n/a	2	Completed/ stable version, fully available	
Helsinki Corpus	UHEL	2	-	n/a	2	Completed/ stable version, fully available	Possible risks: IPR issues
Finnish TreeBank	UHEL	2	2	0	CC-by	Completed/ stable version, fully available	
FinINTAS corpus (includes the FDC - Finnish Dialogue Corpus)	UHEL	3	-	n/a	3	Completed/ stable version, fully available	The name of the corpus was „Intas“ in the previous LR tables Possible risks: privacy issues
Proof Corpus	UHEL	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: privacy issues
UTA Cross-Language Information Retrieval System	UTA	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
ParRus: Russian-Finnish parallel corpus of literary texts	UTA	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
MultiJur: Multilingual Parallel	UTA	3	-	n/a	3	Completed/ stable	Possible risks: IPR issues

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Corpus of Legal Texts						version, fully available	
FiRuLex: Finnish-Russian Comparable Corpus of Legal Texts	UTA	3	-	n/a	3	Completed/ stable version, fully available	
Cross-lingually linked resource	under negotiation	3	3	n/a	3	Completed/ stable version, fully available	
Cross-lingually linked resource	under negotiation	3	3	n/a	3	Completed/ stable version, fully available	
Cross-lingually linked resource	under negotiation	3	3	n/a	3	Completed/ stable version, fully available	
International Corpus of Learner Finnish (Kansainvälinen oppijansuomen korpus)	OY	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: privacy issues
Other Speech corpora	UHEL; UEF; JYU; OY; UTA	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR and privacy issues
Several written corpora	JYU	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
The Finnish Broadcasting Company Corpus of Subtitles (YLE-korpus)	UEF	3	-	n/a	3	Completed/ stable version, fully available	

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Corpus of translated Finnish (Käännösuomen korpus)	UEF	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
The Tampere Bilingual Corpus of Finnish and English	UTA	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
Corpus of Spoken Southwestern Finnish	UEF	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: privacy issues
Finnish Telegraphese Corpus	UEF	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
Emotional speech (Emootiopuheen aineisto)	Aalto	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: privacy issues
Speech and EGG (electroglottography) simultaneous recordings (Puheen ja EGG:n samanaikaiset tallenteet)	Aalto	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: privacy issues
Morfessor	Aalto	3	-	n/a	GNU GPL	Completed/ stable version, fully available	
National Semantic Web Ontology Project in Finland	Aalto, UHEL	3	-	n/a	3	Completed/ stable version, fully available	
TKK Voice Source Analysis and Parametrisation Toolkit	Aalto	3	-	n/a	GNU LGPL	Completed/ stable version, fully available	
ParFin: Finnish-Russian parallel corpus of literary texts	UTA	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: IPR issues
Tieteen kansallinen termipankki (The Bank of Finnish Terminology in Arts and Sciences, BFT)	UHEL	3	-	n/a	3	Completed/ stable version, fully available	Possible risks: work on the terminology database has just started, so it might be that it will

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
							not be ready by the end of 2013

### 2.5.2. Work plan

Table 2.5.2 Work plan for Finnish resources

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Finnish TreeBank Grammar Definition Corpus	1	Describe metadata in META-SHARE format	5 hours (MD described and checked for Göteborg)	M10	M10	Metadata and resource ready for upload
Finnish WordNet	1	Describe metadata in META-SHARE format	5 hours	M10	M10	Metadata and resource ready for upload
Written corpora of old literary Finnish (Vanha kirjasuomi)	1	Describe metadata in META-SHARE format	2 hours	M10	M10	Metadata ready for upload
Corpus of early modern Finnish (Varhaisnykysuomen korpus)	1	Describe metadata in META-SHARE format	2 hours	M10	M10	Metadata ready for upload
Finnish literature classics (Suomalaisen kirjallisuuden klassikoita)	1	Describe metadata in META-SHARE format	2 hours	M10	M10	Metadata ready for upload
Up-to-date word list of modern Finnish (Ajantasainen nykysuomen sanalista)	1	Describe metadata in META-SHARE format	2 hours	M10	M10	Metadata ready for upload
Frequency list of words in written Finnish (Kirjoitetun suomen kielen sanojen)	1	Describe metadata in META-SHARE format	2 hours	M10	M10	Metadata ready for upload

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
taajuuslista)						
Kansainvälinen oppijansuomen korpus (ICLFI)	2	Describe metadata in META-SHARE format; Negotiate licence	3 days	M11	M18	Metadata ready for upload
Corpus of Conversational Finnish (Keskusteluntutkimuksen arkisto)	2	Describe metadata in META-SHARE format; Negotiate licence	7 days	M11	M18	Metadata ready for upload
Open Source (Finnish) Morphology	2	Describe metadata in META-SHARE format	5 hours	M11	M18	Metadata and resource ready for upload
Oulu corpus (Language Bank Of Finland)	2	Describe metadata in META-SHARE format	10 hours	M11	M18	Metadata ready for upload
Geographic Names Register of the National Land Survey	2	Describe metadata in META-SHARE format	7 hours	M11	M18	Metadata ready for upload
Samples of Spoken Finnish (Suomen kielen näytteitä)	2	Describe metadata in META-SHARE format	5 days	M11	M18	Metadata ready for upload
Helsinki Finite-State Transducer Technology	2	Describe metadata in META-SHARE format	5 hours	M11	M18	Metadata and resource ready for upload
Finland-Swedish Text Collection (Kielipankki, Language Bank of Finland)	2	Describe metadata in META-SHARE format	7 hours	M11	M18	Metadata ready for upload
Finnish Text Collection (Kielipankki, Language Bank of Finland)	2	Describe metadata in META-SHARE format	7 hours	M11	M18	Metadata ready for upload
Lemmie	2	Describe metadata in META-SHARE format	7 hours	M11	M18	Metadata ready for upload
Helsinki Corpus	2	Describe metadata in META-SHARE format; Negotiate licence	4 days	M11	M18	Metadata ready for upload
Finnish TreeBank	2	Describe metadata in META-SHARE format	5 hours	M11	M18	Metadata and resource ready for upload

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
FinINTAS corpus (includes the FDC - Finnish Dialogue Corpus)	3	Describe metadata in META-SHARE format	3 weeks	M18	M24	Metadata ready for upload
ProoF Corpus	3	Describe metadata in META-SHARE format	3 weeks	M18	M24	Metadata ready for upload
UTA Cross-Language Information Retrieval System	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
ParRus: Russian-Finnish parallel corpus of literary texts	3	Describe metadata in META-SHARE format; Negotiate licence	2 weeks	M18	M24	Metadata ready for upload
MultiJur: Multilingual Parallel Corpus of Legal Texts	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
FiRuLex: Finnish-Russian Comparable Corpus of Legal Texts	3	Describe metadata in META-SHARE format	3 days	M18	M24	Metadata ready for upload
Cross-lingually linked resource	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata and resource ready for upload
Cross-lingually linked resource	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata and resource ready for upload
Cross-lingually linked resource	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata and resource ready for upload
International Corpus of Learner Finnish (Kansainvälinen oppijansuomen korpus)	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata ready for upload
Other Speech corpora	3	Describe metadata in META-SHARE format; Negotiate licence	2 weeks	M18	M24	Metadata ready for upload
Several written corpora	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
The Finnish Broadcasting Company Corpus of Subtitles	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata ready for upload

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
(YLE-korpus)						
Corpus of translated Finnish (Käännössuomen korpus)	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
The Tampere Bilingual Corpus of Finnish and English	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
Corpus of Spoken Southwestern Finnish	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
Finnish Telegraphese Corpus	3	Describe metadata in META-SHARE format; Negotiate licence	1 week	M18	M24	Metadata ready for upload
Emotional speech (Emootiopuheen aineisto)	3	Describe metadata in META-SHARE format	3 days	M18	M24	Metadata ready for upload
Speech and EGG (electroglottography) simultaneous recordings (Puheen ja EGG:n samanaikaiset tallenteet)	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata ready for upload
Morfessor	3	Describe metadata in META-SHARE format	5 hours	M18	M24	Metadata ready for upload
National Semantic Web Ontology Project in Finland	3	Describe metadata in META-SHARE format	3 days	M18	M24	Metadata ready for upload
TKK Voice Source Analysis and Parametrisation Toolkit	3	Describe metadata in META-SHARE format	2 days	M18	M24	Metadata ready for upload
ParFin: Finnish-Russian parallel corpus of literary texts	3	Describe metadata in META-SHARE format; Negotiate licence	2 weeks	M18	M24	Metadata ready for upload
Tieteen kansallinen termipankki (The Bank of Finnish Terminology in Arts and Sciences, BFT)	3	Describe metadata in META-SHARE format	1 week	M18	M24	Metadata ready for upload

## **2.6. Iceland (HI)**

### **2.6.1.Plans for uploads**

#### **1. batch (M10)**

Metadata upload: 6

Resource/tool upload: 6 resources (3 text corpora, 2 corpora containing text and speech, 1 lexical resource). Metadata will be uploaded for all the resources together with URL for download.

#### **2. batch (M18)**

Metadata upload: 11

Resource/tool upload: 11 resources (3 tools, 3 text corpora, 3 corpora with text and speech, one language description, one terminological resource), eventually upgrades of resources previously uploaded. Metadata will be uploaded for all the resources together with URL for download.

#### **3. batch (M24)**

Metadata upload: 6

Resource/tool upload: 5 (3 semantic relations corpora, one lexical resource, one terminological resource), eventually upgrades of resources previously uploaded. Metadata will be uploaded for all the resources together with URL for download.

Included in the three upload batches are resources and tools for Icelandic which we consider most important for Icelandic language technology. Bigger treebank and bigger and more accurately tagged text and speech corpora would be beneficial. More accurate NLP-tools would also be helpful, e.g. a more accurate lemmatizer and tagger. Work is progressing on tools for spelling correction, especially on the correction of context-sensitive spelling errors.

**Table 2.6.1 Plan for including Icelandic resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>3</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
Icelandic Parsed Historical Corpus	University of Iceland	1	1	2 (annotation improved) 3 (annotation improved, more texts)	LGPL	Upgraded continually	Resource will be continually upgraded in other projects, no risks involved
Icelandic Frequency Dictionary Corpus	The Arni Magnusson Institute for Icelandic Studies	1	1	0	Own licence for download	Completed	Agreements have been made with copyright holders, no risks involved
Parliament Speech Corpus	The Arni Magnusson Institute for Icelandic Studies	1	1	3 (annotation improved)	CLARIN PUB	Annotation will be upgraded at later stages	No copyright agreements necessary, no risks involved
Hjal Speech Corpus	University of Iceland	1	1	0	CLARIN PUB	Completed	No copyright agreements necessary, no risks involved
Pronunciation Dictionary for Icelandic	University of Iceland	1	1	0	CLARIN PUB	Completed	No copyright agreements necessary, no risks involved
The Saga Corpus	University of Iceland	1	1	3 (annotation improved)	CLARIN PUB	Annotation will be upgraded at later stages	No copyright agreements necessary, no risks involved

<sup>3</sup> Value is 0 if no upgrades are planned

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>3</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
CombiTagger	Reykjavík University	2	2	0	GPL	Completed	No risks involved
IceNLP	Reykjavík University	2	2	0	LGPL	Completed	No risks involved
Apertium-is-en Translation System	Reykjavík University	2	2	3 (improved dictionaries etc.)	GPL	Continually upgraded	No risks involved
Tagged Icelandic Corpus	The Arni Magnusson Institute for Icelandic Studies	2	2	3 (annotation upgraded)	Own licenses	Stable version, fully available, annotation upgraded at later stages.	Copyright agreements have been secured, no risks involved.
Database of Modern Icelandic Inflections	The Arni Magnusson Institute for Icelandic Studies	2	2	Upgraded continually	Own license	Upgraded continually	No risks involved.
Icelandic Term Bank	The Arni Magnusson Institute for Icelandic Studies	2	2	Upgraded continually	Own license	Upgraded continually	Need to secure agreements from copyright holders
Íslenskur orðasjóður	Deutscher Wortschatz, Leipzig University	2	2	0	Own license	Completed	Users have to contact copyright holders to gain access for download.
The Jensson Corpus	Tokyo Institute of Technology	2	2	0	CC-BY-NC-SA 3.0	Completed	No risks involved
The Thor Corpus	Tokyo	2	2	0	CC-BY-NC-SA	Completed	No risks involved

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>3</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
	Institute of Technology				3.0		
The Broadcast News RUV-1 Corpus	Tokyo Institute of Technology	2	2	0	CC-BY-NC-SA 3.0	Completed	No risks involved
GOLD	The Arni Magnusson Institute for Icelandic Studies	3	3	0	Own license	Annotations need to be improved	No risks involved
Database of Semantic Relations	University of Iceland	3	3	0	GPL	Not quite clear yet.	No risks involved
Icelandic WordNet	University of Iceland	3	3	0	Own license	Not quite clear yet.	No risks involved
Íslenskt orðanet - Thesaurus	The Arni Magnusson Institute for Icelandic Studies	3	3	0	Own license	Continually upgraded	Need to secure consent from provider
ISLEX	The Arni Magnusson Institute for Icelandic Studies	3	3	0	Own license	Continually upgraded	Need to secure consent from copyright holders
Ministry for Foreign Affairs - Translation Centre – Dictionary	Ministry for Foreign Affairs	3	?	0	Own license	Continually upgraded	Need to secure consent from ministry

## 2.6.2. Work plan

**Table 2.6.2 Work plan for Icelandic resources**

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Comments
Icelandic Parsed Historical Corpus	1	make a number of corrections	1,5PM	M10	M24	Resource will be continually updated in other projects, no risks involved
Icelandic Frequency Dictionary Corpus	1	bring the resource to the required standard and make web-interface for download available	0,5PM	M10	M10	Agreements have been made with copyright holders, no risks involved
Parliament Speech Corpus	1	bring the resource to the required standard and make web-interface available	0,25PM	M10	M10	No copyright agreements necessary, no risks involved
Hjal Speech Corpus	1	prepare web page for download	0,2PM	M10	M10	No copyright agreements necessary, no risks involved
Pronunciation Dictionary for Icelandic	1	prepare web page for download	0,2PM	M10	M10	No copyright agreements necessary, no risks involved
The Saga Corpus	1	bring the resource to the required standard and make web-interface for download available	0,25PM	M10	M10	No copyright agreements necessary, no risks involved
CombiTagger	2		0			No risks involved
IceNLP	2		0			No risks involved
Apertium-is-en Translation System	2		0			No risks involved
Tagged Icelandic Corpus	2	bring the resource to the required standard and make web-interface for download available	6PM	M13	M18	Copyright agreements have been secured, no risks involved.
Database of Modern Icelandic Inflections	2	prepare documentation	0,25PM	M12	M18	No risks involved.
Icelandic Term Bank	2	obtain permission from copyright holders bring the resource to the required standard and make web-interface for download available.	3PM	M12	M18	Need to secure agreements from copyright holders
Íslenskur	2	prepare documentation	0,25PM	M12	M18	Users have to contact copyright holders

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Comments
orðasjóður						to gain access for download.
The Jenson Corpus	2	prepare web page for download	0,2PM	M12	M18	No risks involved
The Thor Corpus	2	prepare web page for download	0,2PM	M12	M18	No risks involved
The Broadcast News RUV-1 Corpus	2	prepare web page for download	0,2PM	M12	M18	No risks involved
GOLD	3	bring the resource to the required standard and make web-interface for download available	4PM	M12	M24	No risks involved
Database of Semantic Relations	3	prepare web page for download and write documentation	1PM	M13	M24	No risks involved
Icelandic WordNet	3	translate the Core WordNet and link it to other resources	4PM	M4	M24	No risks involved
Íslenskt orðanet - Thesaurus	3	negotiate with owners	1PM	M13	M24	No risks involved
ISLEX	3	needed to make an agreement with owners, bring the resource to the required standard and make web-interface for download available	6PM	M13	M24	Need to secure consent from copyright holders
Ministry for Foreign Affairs - Translation Centre – Dictionary	3	needed to make agreement with provider of resource, bring the resource to the required standard and make web-interface for download available.	1PM	M13	M24	Need to secure consent from ministry

## ***2.7. Lithuania (LKI)***

### ***2.7.1.Plans for uploads***

#### **1. batch (M10)**

Metadata upload: 5

Resource/tool upload: -

#### **2. batch (M18)**

Metadata upload: 3

Resource/tool upload: -

There will be uploaded metadata of three resources and possible upgrades/updates for resources from batch 1.

#### **3. batch (M24)**

Metadata upload: 3

Resource/tool upload: 1

There will be uploaded metadata of three resources and upgrades/updates for resources from previous batches.

Currently there is no Treebank and WordNet for Lithuanian language. It is planned to create/develop these resources during the project (at least first draft version). More detailed and comprehensive information about volume and quality will be available till the end of December 2011.

**Table 2.7.1 Plan for including Lithuanian resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>4</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
Database of the Lexicon of Standard Lithuanian	LKI	2	-	n/a	Meta-Share-NC	Completed/ stable version, fully available	
Modern Lithuanian Dictionary	LKI	1	3	2,3 (correction of data)	CLARIN PUB	Completed/ stable version, fully available	
Database of a historical ethnic place names	LKI, co-authored with the Institute of Mathematics and Informatics	1	-	n/a	CLARIN ACA + NC	Completed/ stable version, fully available	
Database of Neologisms	LKI	1	-	2 (addition of new data)	CLARIN ACA + NC	Completed/ stable version, fully available	
Database Synonyms of Lithuanian Terms	LKI	2	-	n/a	Meta-Share-NC	Completed/ stable version, fully available	
Database of proper names	LKI	3	-	0	CLARIN ACA + NC	Completed/ stable version, fully available	
Morphological analyser, lemmatiser and synthesiser for Lithuanian	LKI	3	-	0	2	TBD	License issues
The Dictionary of Lithuanian	LKI	1	-	0	CLARIN ACA + NC	Completed/ stable version, fully available	
Term Bank of the Republic of Lithuania	The State Commission of Lithuanian	3	-	0	2	TBD	More information will be available after planned

<sup>4</sup> Value is 0 if no upgrades are planned

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade <sup>4</sup>	Agreements resource upload as listed	Final status of resource in M24	Comments
	Language						meetings
Term Base of the Lithuanian Standards Board	Lithuanian Standards Board	3	-	0	2	TBD	More information will be available after planned meetings
Term Base of the Lithuanian Language	Institute of Mathematics and Informatics	3	-	0	2	TBD	More information will be available after planned meetings
Science language parallel corpus	Vilnius University	X*	-	0	2	TBD	More information will be available after planned meetings

### 2.7.2. Work plan

**Table 2.7.2 Work plan for Lithuanian resources**

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Database of Neologisms	1	Create REST API for getting data in different formats (XML, JSON, LMF)	2 PM	M7	M14	Metadata is uploaded. Resource corrected and updated
Modern Lithuanian Dictionary	1	* Create web application for accessing the data; * Create REST API for getting data in different formats (XML, JSON, LMF)	2 PM	M7	M12	Metadata and resource is uploaded. Resource corrected and updated
Standard Lithuanian Dictionary	2	*Create new database schema; * Parse data from old DB, convert to a new format and	4 PM	M11	M14	Metadata is uploaded. Resource corrected and

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
		<p>populate it to the new DB;</p> <ul style="list-style-type: none"> <li>* Create web application for accessing the data;</li> <li>* Create REST API for getting data in different formats (XML, JSON, LMF)</li> </ul>				updated
Database of the Synonyms of Lithuanian Terms	2	<ul style="list-style-type: none"> <li>* Create new database schema;</li> <li>* Convert data from old DB to a new format and populate it to the new DB;</li> <li>* Create web application for accessing the data;</li> <li>* Create REST API for getting data in different formats (XML, JSON, LMF)</li> </ul>	3 PM	M12	M15	Metadata is uploaded. Resource corrected and updated
Database of Historical Ethnic Place Names	2	<ul style="list-style-type: none"> <li>* Create new database schema;</li> <li>* Convert to a new format and populate it to the new DB;</li> <li>* Create web application for accessing the data;</li> <li>* Create REST API for getting data in different formats (XML, JSON)</li> </ul>	3 PM	M16	M18	Metadata is uploaded. Resource corrected and updated
The Dictionary of Lithuanian	3	<ul style="list-style-type: none"> <li>* Create web application for accessing the data;</li> <li>* Create REST API for getting data in different formats (XML, JSON, LMF)</li> </ul>	2 PM	M15	M17	Metadata is uploaded. Resource corrected and updated
Database of proper names	3	<ul style="list-style-type: none"> <li>* Create web application for accessing the data;</li> <li>* Create REST API for getting data in different formats (XML, JSON)</li> </ul>	3 PM	M18	M21	Metadata is uploaded. Resource corrected and updated
Morphological analyser, lemmatiser and synthesiser for Lithuanian	3	Negotiate with the owner, analyze resource quality and availability	n/a	n/a	n/a	Metadata is uploaded
Science language parallel corpus	3	Negotiate with the owner, analyze resource quality and availability	n/a	n/a	n/a	Metadata is uploaded. Resource corrected and updated
Term Base of the Lithuanian Standards Board	3	Negotiate with the owner, analyze resource quality and availability	n/a	n/a	n/a	Metadata is uploaded. Resource corrected and updated

Resource	Batch	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Expected results
Term bank of Lithuanian Republic	3	Negotiate with the owner, analyze resource quality and availability	n/a	n/a	n/a	Metadata is uploaded. Resource corrected and updated
Term Base of the Lithuanian Language	3	Negotiate with the owner, analyze resource quality and availability	n/a	n/a	n/a	Metadata is uploaded. Resource corrected and updated

## 2.8. Sweden (UGOT)

### 2.8.1.Plans for uploads

#### 1. batch (M10)

Metadata upload: 16 (16 resources)

Resource/tool upload: 16.

#### 2. batch (M18)

Metadata upload: approx. 60 (approx. 60 resources)

Resource/tool upload: approx. 60

#### 3. batch (M24)

Metadata upload: approx. 30 (approx. 30 potential resources)

Resource/tool upload: approx. 30

**Table 2.8.1 Plan for including Swedish resources**

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Saldo	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Saldo morphology	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Saldo examples	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Swedish Framenet	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Swesaurus	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Parole lexicon	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Simple	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Swedish LWT list	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Swedish Kelly list	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Dalin	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Dalin morphology	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	

Resource name	Provider	Batch for metadata	Batch for upload resource itself	Batches for upgrade	Agreements resource upload as listed	Final status of resource in M24	Comments
Schlyter	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Söderwall	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Söderwall supplement	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Old Swedish morphology	Språkbanken	1	1 (URL)	1 (upgraded to LMF, harmonization and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	
Swedish Wikipedia corpus	Språkbanken	1	1 (URL)	1 (annotation and documentation)	CC_BY_3.0, LGPL_3.0	Available through META-SHARE with Full documentation based on META-SHARE standards	

### 2.8.2. Work plan

**Table 2.8.2 Work plan for Swedish resources**

Resource name	Batch for metadata	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Final status of resource in M24
15 lexical resources and 1 corpus	1	Describe metadata in META-SHARE format; update resources to LMF format	5 PMs	M1	M10	Metadata and resource ready for upload
Approx. 60 lexical resources and corpora	2	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	10 PMs	M1	M18	Metadata and resource ready for upload
Approx. 30 potential resources: lexical, corpora, speech, and tools.	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	10 PMs	M12	M24	Metadata and resource ready for upload
Citation corpora	2	Describe metadata in META-SHARE format, upgrade to agreed standard, should be made available through META-SHARE	5 PM	M10	M18	Resources are available through META-SHARE with full documentation based on META-SHARE standards
More lexical resources	2	Describe metadata in META-SHARE format, upgrade to agreed standard, should be made available through META-SHARE	5 PM	M10	M18	Resources are available through META-SHARE with full documentation based on META-SHARE standards
CLT Toolkit and CLT Cloud	3	Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	1 PM	M6	M24	Full documentation based on META-SHARE standards

Resource name	Batch for metadata	Short description of work that needs to be done	Approximation of amount of work involved	Start of work	End of work	Final status of resource in M24
Litteraturbanken	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	2 PM	M10	M24	Resources are available through META-SHARE with full documentation based on META-SHARE standards
Lexin dictionaries	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	2 PM	M10	M24	Resources are available through META-SHARE with full documentation based on META-SHARE standards
SweDia	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	2 PM	M10	M24	Resources are available through META-SHARE with full documentation based on META-SHARE standards
Humanities Laboratory Data	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	1 PM	M10	M24	Resources are available through META-SHARE with full documentation based on META-SHARE standards
Gothenburg Spoken Language Corpus	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	1 PM	M10	M24	Resources are available through META-SHARE with full documentation based on META-SHARE standards
Speech databases	3	Negotiation with copyright owners; Describe metadata in META-SHARE format, prepare in accordance with META-SHARE requirements and standards	1 PM	M10	M24	Resources are available through META-SHARE with full documentation based on META-SHARE standards

## 3. Horizontal Actions

### 3.1. Horizontal Action on Treebanking

Treebanks are among the most highly valued language resources. Applications include development and evaluation of text classification, word sense disambiguation, multilingual text alignment, indexation and IR, parsing and MT systems. Especially multilingual parallel treebanks are useful for developing hybrid MT systems and other multilingual applications.

Treebanks are unfortunately scarce for the languages covered by META-NORD and available treebanks are difficult to access and exploit. Linking of treebanks across languages is still mostly at the development stage. Treebanks are very expensive to produce and it is not possible to develop full size treebanks in the current project, but this horizontal action will improve and link existing resources, and promote access to them.

#### 3.1.1. Objectives and overall plan

The first objective of this horizontal action is to make treebanks for relevant languages accessible through a uniform web interface and state-of-the-art search tool. In cooperation with the INESS project, an advanced server-based solution will be provided for parsing and disambiguation, for uploading of existing treebanks, indexing, management, and exploration. Existing treebanks available in the consortium will be integrated on this platform.

A second objective is to link treebanks across languages, mostly based on existing treebanks. This objective has more of a demonstration value than an exploitation value due to the preliminary stage of the technology and the limited number of available resources of this kind.

#### 3.1.2. Action plan with subtasks and milestones

- a. *Selection of multilingual corpora* (June 30, 2011). The following materials have been selected for linking across languages.
  1. The JRC-Acquis, which provides source and translation materials for all EU languages. The META-NORD partner in Norway has secured access to a partial translation to Norwegian, provided by the EEA-secretariat of the Ministry of Foreign Affairs in Norway. This material has been selected because it is publicly available and covers a variety of domains even if the genre is linguistically narrow.
  2. The Norwegian novel "Sofies verden" by Jostein Gaarder and its translations. META-NORD has cleared rights for the Norwegian version with Aschehoug. Rights will need to be negotiated separately for the translations with publishers in the relevant countries. This material has been selected because it is a linguistically more varied genre than the EU/EEA texts.
- b. *Linking across languages* (Nov. 30, 2012). The materials selected for linking will be preprocessed, converted and uploaded to a linking environment on the INESS treebank server system. Sentences will be aligned, annotated and quality-assured. The linking will be exported and made downloadable in an XML-based standoff annotation format. A parallel search interface will also be provided. Linguistically motivated interactive linking with XPAR technology will be performed for LFG-based parsebanks which support f-structure linking. Danish, Norwegian and English will be used in the first pilot, based on the multilingual Sofie-corpus. An extension of linking to other dependency treebanks will be attempted, e.g the Finnish treebank, using technology from FIN-CLARIN. Combining

these technologies, a pilot parallel treebank is planned for at least Norwegian, Danish, Finnish and English.

- c. *Other treebanks, uploads*, in different batches. Available treebanks will be collected, licensed and uploaded in different batches, see above under section 2. This includes the monolingual treebanks of the different language versions of "Sofies verden" as well as other treebanks available to the consortium.

### **3.2. Horizontal Action on Wordnets**

#### ***First batch:***

**Milestone for first batch:** was to have all existing wordnets linked to core wordnet (SE, DA, FIN, EST).

#### **Nature of upload: c) resources**

**Coverage:** All monolingual wordnets (DA, FIN, EST, SE) including at least 5000 links to Princeton Wordnet core senses. For details on each wordnet, cf. monolingual sections. Estonian wordnet should be adjusted to agreed standards. Norwegian Wordnet (which is a third party resource being developed at a private company "Kaldera") is currently being developed. It is expected that a public final version will be available before the end of the META-NORD project which will then be included in the third batch upload of resources. Norwegian wordnet will link to Princeton Core via DanNet.

#### ***Work plan and milestone for second and third batch:***

February 2012: Meeting for detailed planning in Copenhagen February 2012.

March-June 2012: Develop tool for extracting/viewing bilingual links (**Milestone 1**)

June 2012: Select coverage for validation of linked resources (approx. 1000 synsets)

July 2012- Nov. 2012: Validation report of linked resources (we need bilingual experts for Fin-Est, Fin-Da, Da-Swe) (**Milestone 2**)

January 2013: Upload of Norwegian Wordnet. Upload of linked wordnets.

Uniform web interface for browsing monolingual and linked resource (**Milestone 3**).

### **3.3. Horizontal Action on Multilingual Terminology**

Task 3.6 Horizontal Action on multilingual terminology will integrate EuroTermBank and other terminology resources discovered during the project with META-SHARE. The horizontal action has two main parts – content and integration. All partners involve in the content part – identify and contact terminology resource holders willing to contribute. Integration will be based on the API from META-SHARE. Based on that Tilde will do analysis, specifications, development, and will complete the integration.

#### **3.3.1. Content. New Terminology Resources**

Identification of Terminology Resources in partners' countries

Evaluation and selection of identified Resources

Meta Data and formats of new terminology data => META-SHARE batch 2  
Strategy for integration - decide and agree for each resource  
Import of static terminology resources into EuroTermBank  
Linking of external terminology resources with EuroTermBank  
Prepare and sign resource licensing contracts  
Terminology Data integration complete => META-SHARE batch 3

### ***3.3.2.Integration. Interlinking of EuroTermBank with META-SHARE and with other external terminology resources.***

#### **Specifications**

**Milestone:** Receiving META-SHARE API for integration of terminology resources with META-SHARE, commenting and accepting of the API

Examination of EuroTermBank architecture in relation with integration with META-SHARE over API

Specification of API for data exchange with META-SHARE, EuroTermBank side

Specification of API for external data – live linking

Analysis of EuroTermBank - requirements to adaptation of EuroTermBank in relation to the interlinking with META-SHARE and for consolidated multilingual representation of monolingual and bilingual terminology entries

**Milestone:** All requirements and specifications available

#### **Development**

Development of import tools for static terminology data

Implementation of API for external data – linking EuroTermBank with external terminology resources

Development of a bridge between META-SHARE and EuroTermBank

Implementation of the mechanisms for consolidated multilingual representation of monolingual and bilingual terminology entries

#### **Deployment**

META-SHARE to ETB interlinked

### ***3.3.3.Collaboration***

Sharing of expertise to other 6.1 pilot projects

Possibly – integration of terminology resources originating countries other than those represented by META-NORD.

Upon completion of this deliverable the partners of META-NORD have identified the following terminological resources in their respective countries:

No.	Basic Information			Administrative Information				Technical Information				Content Information				
	META-NORD Partner	Resource Name	Resource Type	Resource Owner	License Type	Contact person	Format	Online browsing	Downloadable	Languages	Size (Languages total)	Size (Concepts/entries)	Size (terms)	Size (Definitions)	Size (pictures)	Domain Classification Type
1	TILDE	EuroTermBank	Terminology Data Base	Tilde / EuroTermBank Consortium	Terms of Use at <a href="http://www.eurotermbank.com">www.eurotermbank.com</a>	Roberts Rozis < <a href="mailto:roberts.rozis@tilde.lv">roberts.rozis@tilde.lv</a> >	N/A	YES	NO	EU BG HR CS DA NL EN ET FI FR DE EL HU IT LA LV LT MT NO PL PT RO RU SK SL ES SV	27	625345	230000	223568	0	EUROVOC
2	TILDE	Terminology Database of Assistive Technology	Glossary	Tilde / EASTIN-CL Project Consortium	TBD	Roberts Rozis < <a href="mailto:roberts.rozis@tilde.lv">roberts.rozis@tilde.lv</a> >	TBX	NO	YES	EN DE IT DK LV LT EE	7	12650	88550	0	0	NONE
3	UHEL	Tieteen kansallinen termipankki (The Bank of Finnish Terminology in Arts and Sciences, BFT)	Terminology Data Base	The Department of Finnish, Finno-Ugrian and Scandinavian Studies, UHEL		Lea Laitinen < <a href="mailto:lea.laitinen@helsinki.fi">lea.laitinen@helsinki.fi</a> >	N/A	NO	NO	N/A	N/A	N/A	N/A	N/A	N/A	N/A
4	LKI	Synonymy of Lithuanian Terms	Terminology Data Base	LKI	TBD	Robertas Stunžinas < <a href="mailto:robertas.stunzinas@gmail.com">robertas.stunzinas@gmail.com</a> >	FoxPro database	NO	NO	LT EN DE RU FR LA	6		33800		0	NONE
5	LKI	Term Bank of the Republic of Lithuania	Terminology Data Base	The State Commission of Lithuanian Language	TBD	Audra Ivanauskienė < <a href="mailto:audra.ivanauskienė@vlkk.lt">audra.ivanauskienė@vlkk.lt</a> >	TBX	YES	NO	EN ES IT LV PL PT FR RU DE LA	10		300000		0	EUROVOC
6	LKI	Term Base of the Lithuanian Standards Board	Terminology Data Base	Lithuanian Standards Board	TBD	Lilijana Balčiauskienė < <a href="mailto:lilijana.balciauskienė@lsd.lt">lilijana.balciauskienė@lsd.lt</a> >	TBX	YES	NO	LT EN DE RU FR LV	6		38000		0	ISO

No.	Basic Information				Administrative Information				Technical Information				Content Information			
	META-NORD Partner	Resource Name	Resource Type	Resource Owner	License Type	Contact person	Format	Online browsing	Downloadable	Languages	Size (Languages total)	Size (Concepts/entries)	Size (terms)	Size (Definitions)	Size (pictures)	Domain Classification Type
7	LKI	Term Base of the Lithuanian Language	Terminology Data Base	Institute of Mathematics and Informatics	TBD	Valentina Dagienė <valentina.dagiene@mii.vu.lt>	TBX	YES	NO	LT EN DE RU FR LA	6		280000		0	NONE
8	HI	Icelandic Term Bank		The Arni Magnusson Institute for Icelandic Studies	Open for search, for download with special license (to be negotiated)	Ágústa Þorbergsdóttir <agustath@hi.is>	TBX	YES	Most likely	eng, isl, fin, fao, fra, deu, lat, nor, spa, swe, dan	13	186859	?	?	?	NONE
9	HI	Ministry for Foreign Affairs - Translation Centre - Dictionary		Ministry for Foreign Affairs	Open for search, permission for download is to be negotiated	Sigrún Þorgeirsdóttir <sigrun.thorgeirsdottir@utn.stjr.is>	N/A	Yes	Not known	is, en, dan, nor, swe, fra,, deu, lat	8	55000	?	?	?	EU-terminology
10	UIB	Milterm	Terminology Data Base	Standard Norge	Open for search, permission for download is being negotiated	desmedt@uib.no	TBX	YES	Most likely	nob,eng	2	8350	19150	unknown	N/A	Environment
11	UIB	UHR's Termbase for Norwegian higher education institutions	Terminology Data Base	UHR	Open for search, permission for download is being negotiated	desmedt@uib.no	TBX	YES	Most likely	nob,nno,eng	3	unknown	2000 in each language	unknown	N/A	Education

### ***3.4. Partners establishing nodes***

#### ***3.4.1. General***

META-NORD partners have been looking for alternatives to using the META-SHARE metadata editor and uploading their resources directly to META-SHARE. One possibility to send data to the META-SHARE repository is for each partner to establish their own repository where the metadata and resources provided by this partner will be made available for META-SHARE. An alternative used among the META-SHARE contributors is to establish one node that several contributors can upload their data to.

In order to facilitate uploading metadata and language resources from META-NORD partners to the META-SHARE repository, UGOT has established a META-SHARE node where all partners can upload their resources to. The data submitted is validated against the META-SHARE schemas, further interaction with META-SHARE and data gathering goes automatically through the UGOT node, so the individual partners will not have to deal with those technical aspects. Currently the UGOT node offers different methods for uploading data: metadata can be submitted either in the form of Microsoft Excel spreadsheets that have been used in the project or as XML. Both formats will be converted into XML in the node and forwarded to META-SHARE as such. UGOT node will be managing a SVN repository and upload environment for META-NORD resources throughout the project.

There are also plans for establishing META-SHARE nodes at UHEL and Tilde for more specialized effort in mediating and managing the language resources from META-NORD. However at the moment the work has not been started and those nodes' timeline and subtasks can be defined only after open source versions of META-SHARE editor become available.

#### ***3.4.2. UGOT node***

The initial release of META-SHARE's editor tool was set up first as a node at UGOT for all META-NORD partners, which hopefully will reduce the total workload overall.

Work plan for establishing the UGOT META-SHARE node:

- **M9 (done):**
  - received alpha software from META-SHARE to set up node,
  - populated the node with data eventually received from META-SHARE,
  - testing
- **M10 (done):**
  - reported bugs to META-SHARE,
  - fixed some bugs locally on the node,
  - testing
- **M10 (todo):**
  - resolving saving issues, search issues, and schema issues
  - exploring xml import/export functionality
- **M18 (second batch):**
  - upgrading META-SHARE node with latest software version

- synchronization and conversion of metadata according to latest META-SHARE metadata schema
- upgrading content with resources for the second batch
- **M24 (third batch):**
  - upgrading META-SHARE node with latest software version
  - make sure that restricted access limitations actually work
  - upgrading content with resources for the third batch

## 4. Conclusions

As stated in the “Description of Work” (DoW), one of the main objectives of the META-NORD project is to contribute to a pan-European digital resource exchange facility by identifying and collecting resources in the Baltic and Nordic countries and by documenting, processing, linking and upgrading them to agreed standards and guidelines. This report follows deliverables 2.2 “Report on resources (actually or potentially) available to the consortium” and 2.3 “Report on methodology and criteria followed for the selection of resources.” and based on those earlier results gives a list of resources that have been selected to be made accessible through META-NET, as well as detailed plans on how and when each project partner will make their resources available. For each resource there is information on the state of licensing agreements, the work that needs to be done before the resource can be made publicly available and an approximate schedule for both intermediate and final version of that resource as it will be included in META-SHARE.

The amount of resources listed as well as the principle for uploads varies considerably between project partners. Some partners will be mostly uploading metadata, leaving access to the resources in their own servers, others will give an access URL or upload the resource itself in META-SHARE repository for almost every language resource listed. It should be noted that the resources in this report are those selected to be made accessible through META-NET from the list of available resources. Altogether is planned to upload 47 actual resources in the first batch as well as 67 metadata descriptions are registered in D4.3. “First upload of language resources”. For the two later upload batches the amount of metadata descriptions will be around one hundred each, the number of actual resource uploads will grow somewhat compared to the first batch. From the descriptions of selection criteria it can be seen that resources in first batch will have new versions in later batches as the resources are being upgraded or expanded during the project. Also it is common to upload metadata for several resources earlier and add uploads of actual resources in later batches, using a resource uploads from previous batches as reference.

Most of the language resources included in this report will be provided by project partner institutions, but there are also a considerable number by other institutions in partner countries who are willing to make their resources accessible through META-NET. Information concerning resources by third parties is considered less certain and as such has not yet been planned in much detail, since a lot depends on the other parties. Those resources have been mostly scheduled for later upload batches, third batch will see the first upload for many resources for which the licence agreements are still being settled.

As mentioned in the deliverable, the plans included in this report are not to be considered final, especially the work plans detailing amount of work needed and time schedules are liable to change as new information becomes available or problems arise. In addition, as LKI has shown in adding new resources to those found in deliverable 2.3, it is highly possible that new resources by third parties will become available to META-NORD in the course of the project.

The work in the three horizontal actions is well under way, future work plans are outlined in this report. At the time there is one functioning META-SHARE node in META-NORD – at UGOT, however there are plans to establish two more: at UHEL and TILDE.

## 5. List of tables

Table 1	Abbreviations
Table 2.1.1	Plan for including Latvian resources
Table 2.1.2	Work plan for Latvian resources
Table 2.2.1	Plan for including Danish resources
Table 2.2.2	Work plan for Danish resources
Table 2.3.1	Plan for including Estonian resources
Table 2.3.2	Work plan for Estonian resources
Table 2.4.1	Plan for including Norwegian resources
Table 2.4.2	Work plan for Norwegian resources
Table 2.5.1	Plan for including Finnish resources
Table 2.5.2	Work plan for Finnish resources
Table 2.6.1	Plan for including Icelandic resources
Table 2.6.2	Work plan for Icelandic resources
Table 2.7.1	Plan for including Lithuanian resources
Table 2.7.2	Work plan for Lithuanian resources
Table 2.8.1	Plan for including Swedish resources
Table 2.8.2	Work plan for Swedish resources