

Estonian language in European Language Resource Infrastructures

Kadri Vider

Kadri.vider@ut.ee

Center of Estonian Language Resources
University of Tartu, Estonia



Baltic HLT 2016, Riga

Center of Estonian Language Resources (CELR, ee.clarin.eu)

- The research infrastructure of national importance in the Estonian Research Infrastructures Roadmap
 - To **make Estonian language data and tools available** to everyone working with digital language materials
 - To coordinate and organise the **documentation** and **archiving** of the resources
 - To develop and promote LR **standards in Estonia**
 - To draw up necessary legal contracts and **licences** for **different types of users**: public, academic, commercial, etc
- **CELR performs the obligations of Estonia as the member in CLARIN ERIC.**



Riga Summit 2015



Euroopa Liit
Euroopa
Regionaalarengu Fond



Eesti tuleviku heaks

3. Center of Estonian Language Resources (CELR) has an obligation to manage and to deposit all resources and tools developed within NPELT for preservation and long-term access (<http://keeleressursid.ee/en/>).

- CELR is an infrastructure of national importance in the Estonian Research Infrastructures Roadmap and will provide access to language resources and technologies for all researchers.

- CELR will perform as organisational framework for coordinating and implementing the obligations of Estonia as the member in CLARIN ERIC.

Center of Estonian Language Resources (CELR)

- Consortium of 4 partners: UT, TTU, IEL and ELM
- SSH research infrastructure, [CLARIN ERIC](#) representative in Estonia
- Archives and makes available LR and LT, incl. all NPELT outcomes
- [Catalogue](#) and META-SHARE [Registry](#) of Estonian LRs and LTs
- Provides know-how and expertise about Estonian LR and LT, but also supports research with data and (web) tools in Digital Humanities

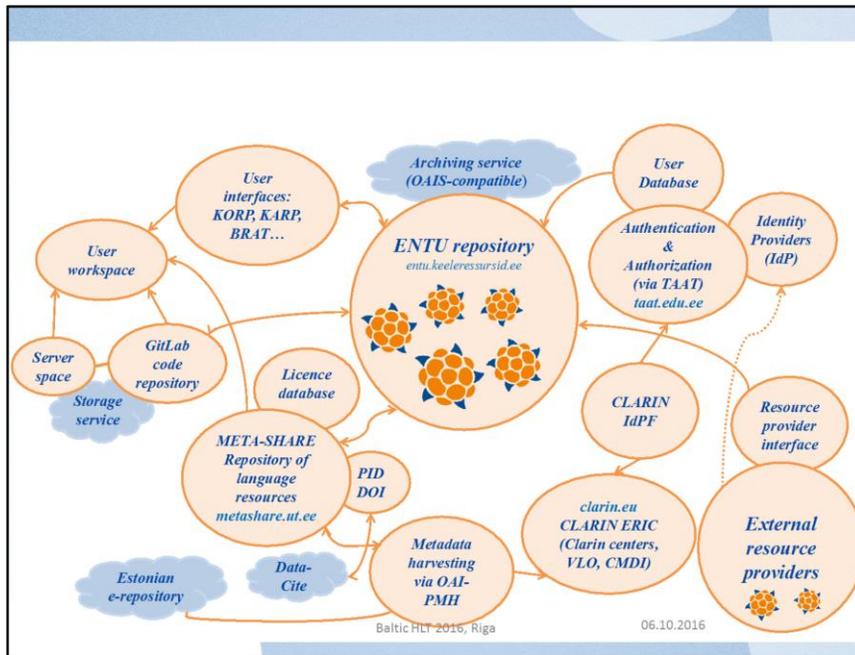


ee.clarin.eu

Baltic HLT 2016, Riga

06.10.2016

The goal of the Centre of Estonian Language Resources (CELR) is to create and manage an infrastructure to make Estonian language digital resources (dictionaries, corpora – both text and speech –, various language databases) and language technology tools (software) available to everyone working with digital language materials. The centre provides language technology resources in the form of a web-based service that uses archived data. Thereby the centre facilitates the use of existing resources and their combination in the creation of new added value, while assuring the maintenance and safekeeping of the resources created. The centre is also responsible for archiving and enabling access to the outcomes of the projects of the National Programme for Estonian Language Technology.





- www.keeletehnoloogia.ee
- Financed from state budget
 - 2006-2010: total 3,4 M€
 - 2011-2017: **765 K€** per year
- Managed by Steering Committee
- Results of projects are declared public domain.
- CELR is required to deposit all such resources and tools for preservation and long term access.



Riga Summit 2015

In Estonia, the National Programme for Estonian Language Technology (NPELT) was launched in 2006. NPELT is financed from state budget, in 2006–2010 ca 54 MEEK (3,4 Me) in total. For period 2011–2017 it has planned to fund different projects and activities with **765 K€** every year.

NPELT management is carried out by a Steering Committee including HLT experts and representatives of the ministries and ICT industry.

All results of the NPELT projects (language resources and software prototypes) are released as public domain. The Center of Estonian Language Resources has a duty to deposit all such resources and tools for preservation and long term access.

The aim of the programme is to ensure a situation where the Estonian language technology support is on equal level with that of other languages in countries with advanced language technology (e.g. the Nordic countries) in directions that require the software developments and applications oriented to the Estonian language. It has also been noted for languages with a more advanced language technology that many language technology tasks do not have as easy a solution as previously estimated, especially the creation of dialogue systems for machine translation and spoken communication have proven to be much more difficult tasks than previously assumed.

Planned impacts of the programme:

Applications that process language materials adequately have found a broad use in society; it is possible to use natural speech synthesis, speech recognition created for specific domains, widespread technical aids for text editors, man-machine-dialogue systems and machine translation prototypes;

the Estonian language shall be recognised as a language with advanced language technology, Estonia belongs to the countries with most advanced language technology.