

# SPRÅKBANKENSAM

A CLARIN knowledge centre  
for the languages of Sweden

Rickard Domeij and Gunnar Eriksson

The Language Council of Sweden at the Institute for Language and Folklore

[Rickard.Domeij/Gunnar.Eriksson@sprakochfolkminnen.se](mailto:Rickard.Domeij/Gunnar.Eriksson@sprakochfolkminnen.se)

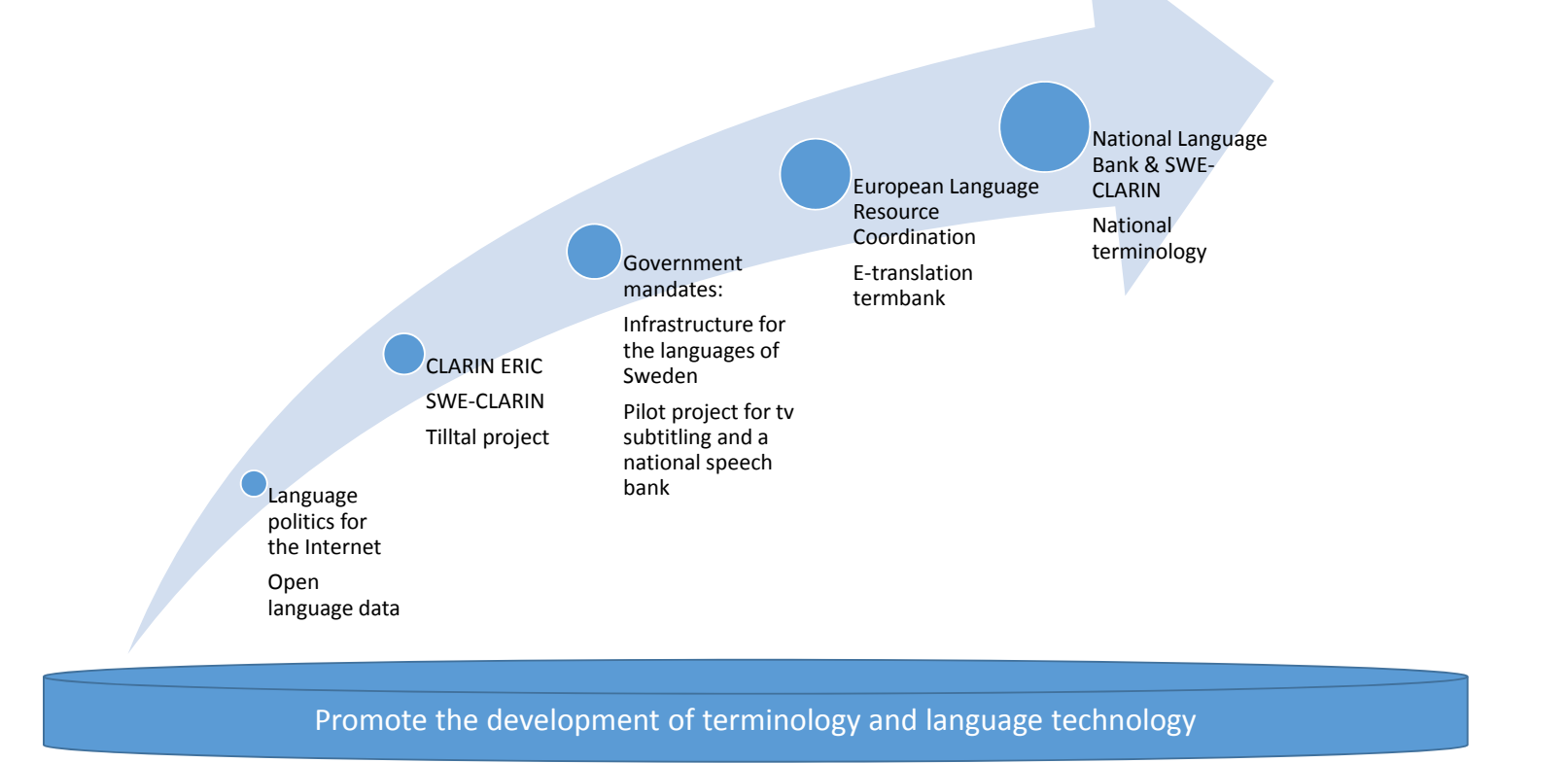
NATIONELLA  
SPRÅKBANKEN



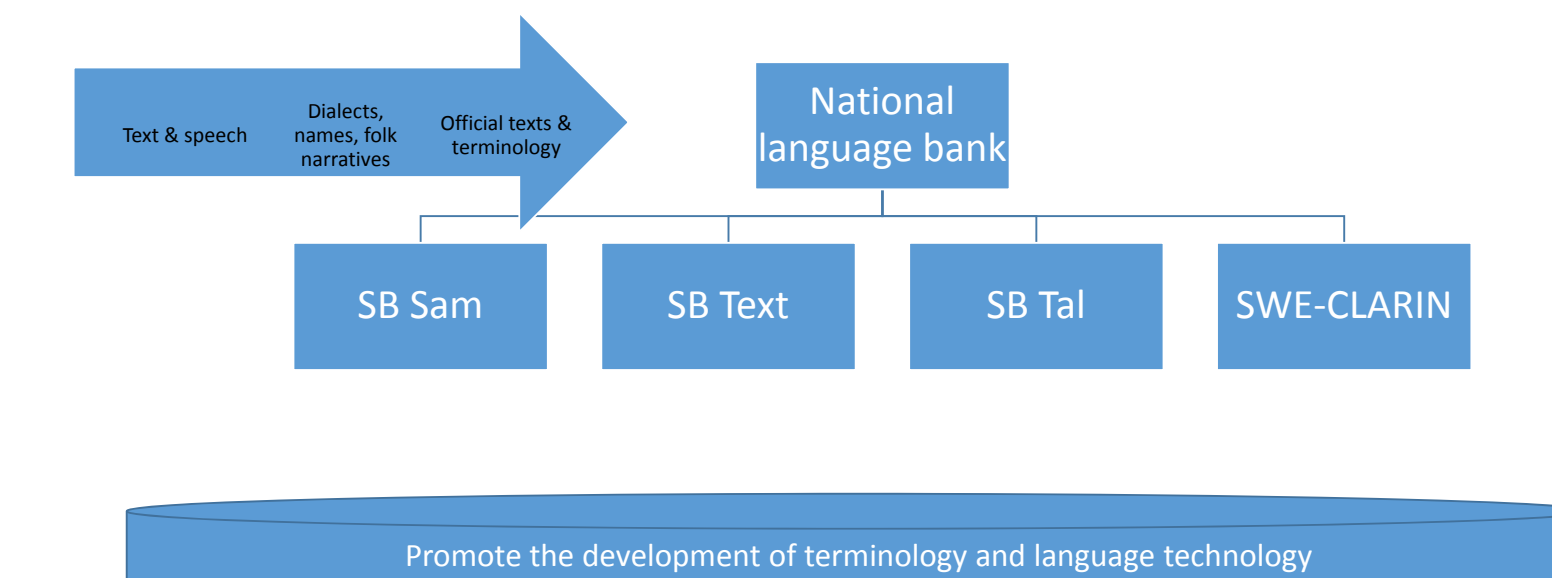
## SWELANG: K-centre for the languages of Sweden

- Information service at SB Sam offering advice on the use of digital language resources and tools for the Swedish language, minority languages in Sweden, the Swedish sign language, Swedish dialects and other parts of the intangible cultural heritage of Sweden in text and speech, as well as language policy and planning.
- Run by the Language Council of Sweden (Stockholm), in cooperation with the other sections of the Institute of Language and folklore in Uppsala and Göteborg. The institute has an assignment from the government to promote the development of terminology and language technology in Sweden.

## Background: Language infrastructure projects

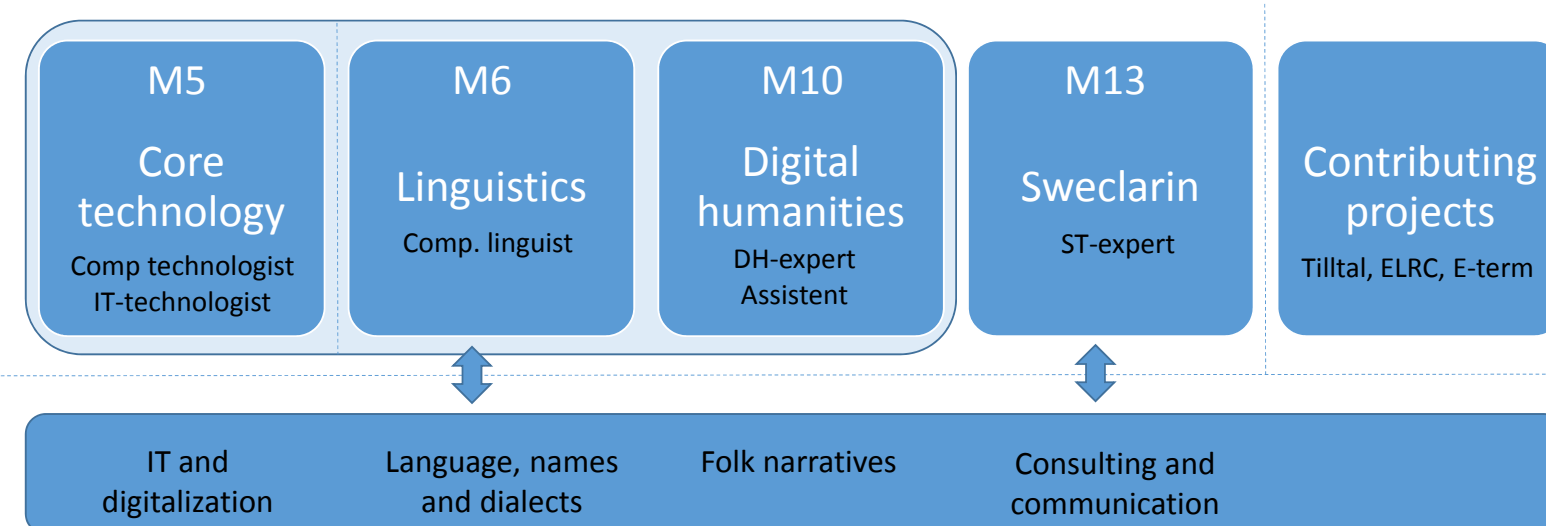


## Språkbanken and Sweclarin 2018–2024



## SB Sam

Head of dept



## Focus on two types of material

- Official texts and terminology for research in official communication and social conditions. The material is multilingual with parallel texts in Swedish and translations into easy-to-read/plain language, national minority languages, and minority languages.
- Folk narratives, as well as other text and speech material from the dialect and folklore archives. The material consists of inventories, dialect glossaries, letters, recordings, transcriptions, etc. It is interesting both in terms of content and linguistic quality, with a great geographical, social, and stylistic variation.

## Development of tools and resources at SB SAM

- Develop digital methods for collecting:
  - dialect and folklore data, e.g. through web-based questionnaires, crowdsourcing and automatic speech recognition
  - official texts and terminology in Swedish and minority languages through web crawling and harvesting through DCAT-AP
- Develop the means to annotate, process and explore text and speech data in the archives through language technology-based tools in close collaboration with SB Text, SB Tal and the SWE-CLARIN centres .
- Develop methods to manage and making available contextualized digital archive material through a map based research interface called *Digitalt kulturarv* presented on the SWE-CLARIN blog 2018-02-09.
- Develop dictionary infrastructure to store, edit, integrate and make available dialect and official terminology in collaboration with SB Text.

## Tiltal – an interdisciplinary example for collaboration

- In the project Tiltal we examine how speech technology methods can make the historical speech recordings more accessible to research in cooperation between data holders, researchers and speech and language technologists.
- Three case studies examine how speech technology can be used to investigate issues within different subject fields (ethnology, linguistics and conversation analysis).
- An activity-theoretical perspective is applied to investigate how the archive material is used for research. Considering the needs of researchers, language technological solutions are suggested and their practical usefulness is assessed through the case studies.
- Tiltal is a collaboration project involving SB Sam, SB Tal, Riksarkivet and SWE-CLARIN. The project is supported by Riksbankens Jubileumsfond (SAF16-0917:1).

Berg, J; Domeij, R; Edlund, J; Eriksson, G; House, D; Malisz, Z; Nylund Skog, S. & Öqvist, J: 2016. *Tiltal – making cultural heritage accessible for speech research*. Paper for the Proceedings of the CLARIN Annual Conference 2016 in Aix-en-Provence, France, 26–28 oct. 2017. *Involving users and collaborating between disciplines in making cultural heritage accessible for research*. Paper presented at the CLARIN Annual Conference 2017 in Budapest, September 2017.

## Long term goals for SB SAM

- At present the workflow collection–processing–knowledge sharing is largely analogue at ISOF – after project completion it should be completely digital.
- We have significantly improved the possibilities of searching and navigating the collections by methods from language technology and digital humanities.
- We have developed an interface for collecting, processing and making data available, to which researchers contribute to a great degree both as co-designers and contributors.
- In cooperation with SB Text, SB Tal and other SWE-CLARIN centres we have created integrated, secure processes and interfaces for how speech and text can be processed and made available through methods and tools from other centres.
- The new national infrastructure is well-known and used by linguists, as well as folklorists, ethnologists and other humanistic and social scientists in a way that gives them new possibilities for exploring text and speech in their research.

## Acknowledgement

- The research infrastructure Språkbanken och SWE-CLARIN is supported by the Swedish Research Council (2017-00626), which will provide funding from 2018 to 2024 in a first stage.
- The new infrastructure collaboration builds on years of collaboration within SWE-CLARIN and other research and infrastructure projects.