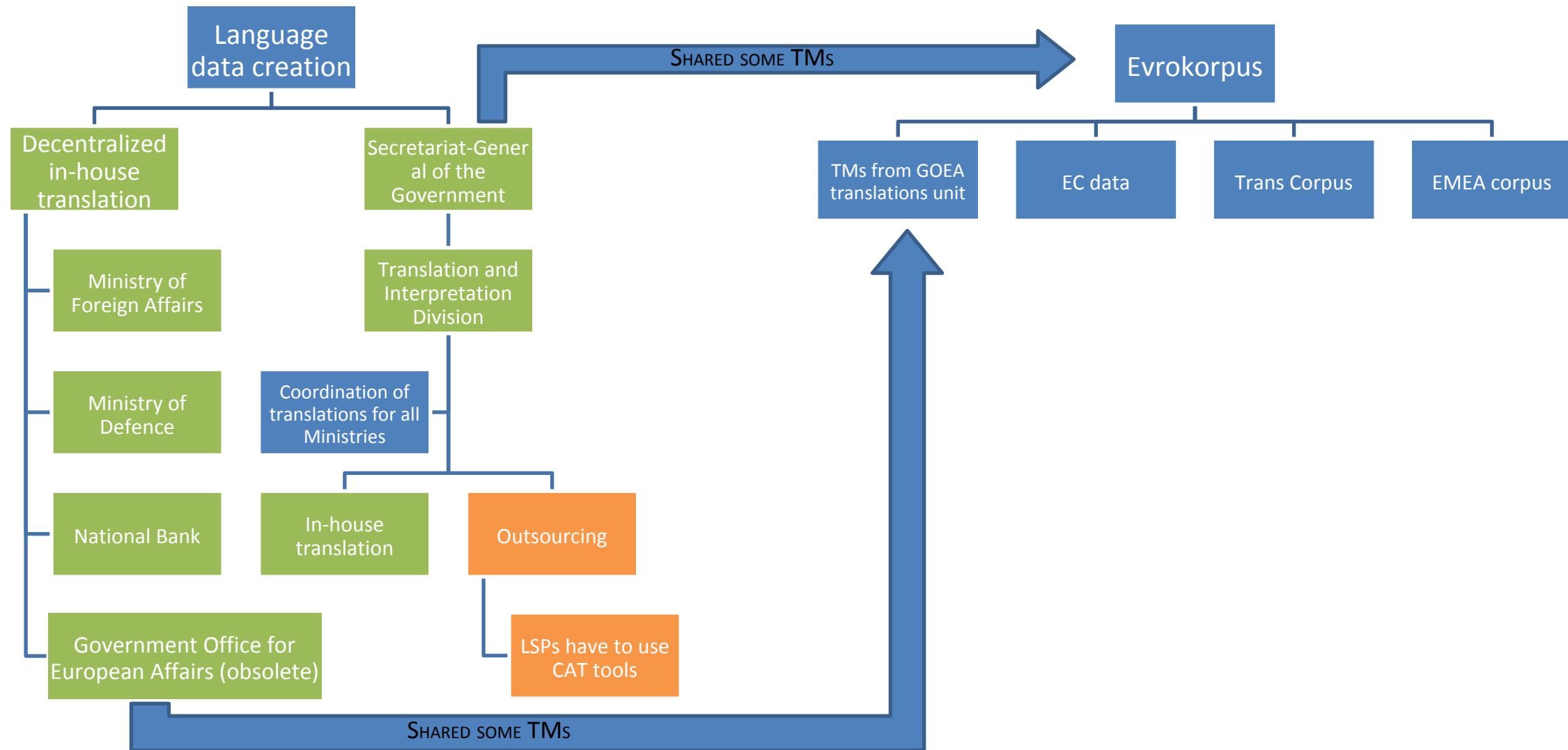


# Country profile – Slovenia

7th LRB Meeting, Paris

Andraž Repar, JSI Institute

# Infrastructure of language data and use on national level in Slovenia v2



# Evrokorpus

- 2002: EN-SL Trados TMs from the Slovenian Government Office for European Affairs (GOEA)
  - Later: DE-SL, FR-SL, IT-SL, ES-SL
- Also includes: some EC data (2008), TRANS corpus, EMEA data
- Current size EN-SL: 2 billion words
- Maintained by Secretariat-General of the Government of the Republic of Slovenia from 2005 on

# Terminology and other resources

## Terminology

- Evroterm
  - Internal terminology and many other glossaries created by various public service organizations
- Termania
  - Around 70 dictionaries
- Fran
  - Monolingual and bilingual dictionaries

## Other

- ccKres, ccGigafida
- Clarin network

# Data collection in Slovenia

## Main problem

- Sensitive information
  - Resource owners reluctant to share data without anonymization

## Two most common concerns

- I am afraid people will introduce mistakes and I will be blamed
- What's in it for me?
  - Some resource owners don't see the value of sharing
  - Demonstrating the benefits, e.g. for MT - “a chicken and egg problem”

Thank you!