

# EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) AND BEYOND

ANDREA LÖSCH (DFKI), STEFANIA RACIOPPA (DFKI), EILEEN MARRA (DFKI),  
THIERRY DECLERCK (DFKI)



# ELRC CONTEXT AND OBJECTIVES

---

## WHO WE ARE

### THE ELRC CONSORTIUM



CROSSLANG  
TRANSLATION AUTOMATION



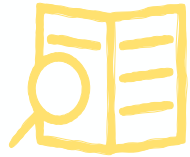
**STARTED**  
in April 2015

**CURRENT  
CONTRACT**  
until end of  
**2022**

### GOAL

To provide language resources to improve coverage and performance of automated translation solutions in the context of current and future CEF digital services

## ELRC OBJECTIVES



**Language data are key for machine translation /  
language technology!**

the  
publi  
SME  
dom

language data

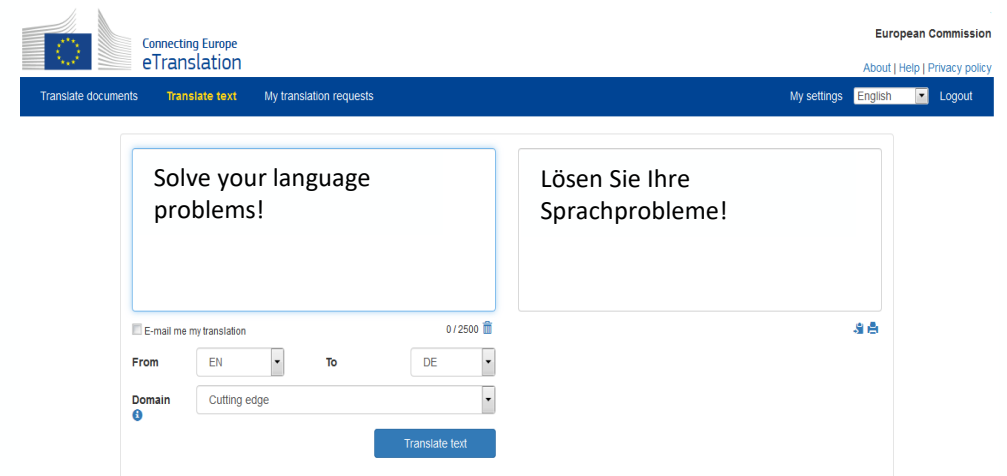
and SMEs

related to  
language data  
sharing

factory for  
language data in  
Europe

# ELRC: LANGUAGE DATA FOR eTRANSLATION

- CEF eTranslation helps public services being multilingual:
  - Neural machine translation tool provided by the European Commission to all EU bodies but also public services and public administrations across Europe
  - for 24 official languages of the EU (24 <-> 24), Russian, Chinese, Turkish, ...
  - Secure & scalable
  - Facts and Figures:
    - 70+ systems connected to eTranslation
    - 7.000+ SMEs registered
    - More than 240 Mio. pages translated
- Goal: ELRC provides language resources  
to make eTranslation work for different languages and domains!



The screenshot shows the eTranslation web interface. At the top, there is a navigation bar with the 'Connecting Europe eTranslation' logo and the European Commission logo. The navigation bar includes links for 'Translate documents', 'Translate text', and 'My translation requests'. On the right side of the navigation bar, there are links for 'My settings', a language dropdown menu set to 'English', and a 'Logout' button. The main content area is split into two columns. The left column contains the text 'Solve your language problems!' and the right column contains 'Lösen Sie Ihre Sprachprobleme!'. Below the text input area, there is a checkbox for 'E-mail me my translation' and a character count '0 / 2500'. The 'From' language is set to 'EN' and the 'To' language is set to 'DE'. The 'Domain' is set to 'Cutting edge'. A 'Translate text' button is located at the bottom right of the main content area.

# ELRC: LANGUAGE DATA FOR INCLUSION

- Additional dimension of ELRC: To provide language resources to overcome language barriers in public services and the European Digital Single Market in general
- Example: Digital Service Infrastructures (DSIs)



## BRIS

Citizens and business partners need legal certainty when doing business cross-border



## ODR

Citizens need to solve disputes online across borders



## eHealth

Citizens need to have access to medical information and their patient data when abroad



## EESSI

Citizens enjoy seamless exchange of social security information when abroad



It also includes

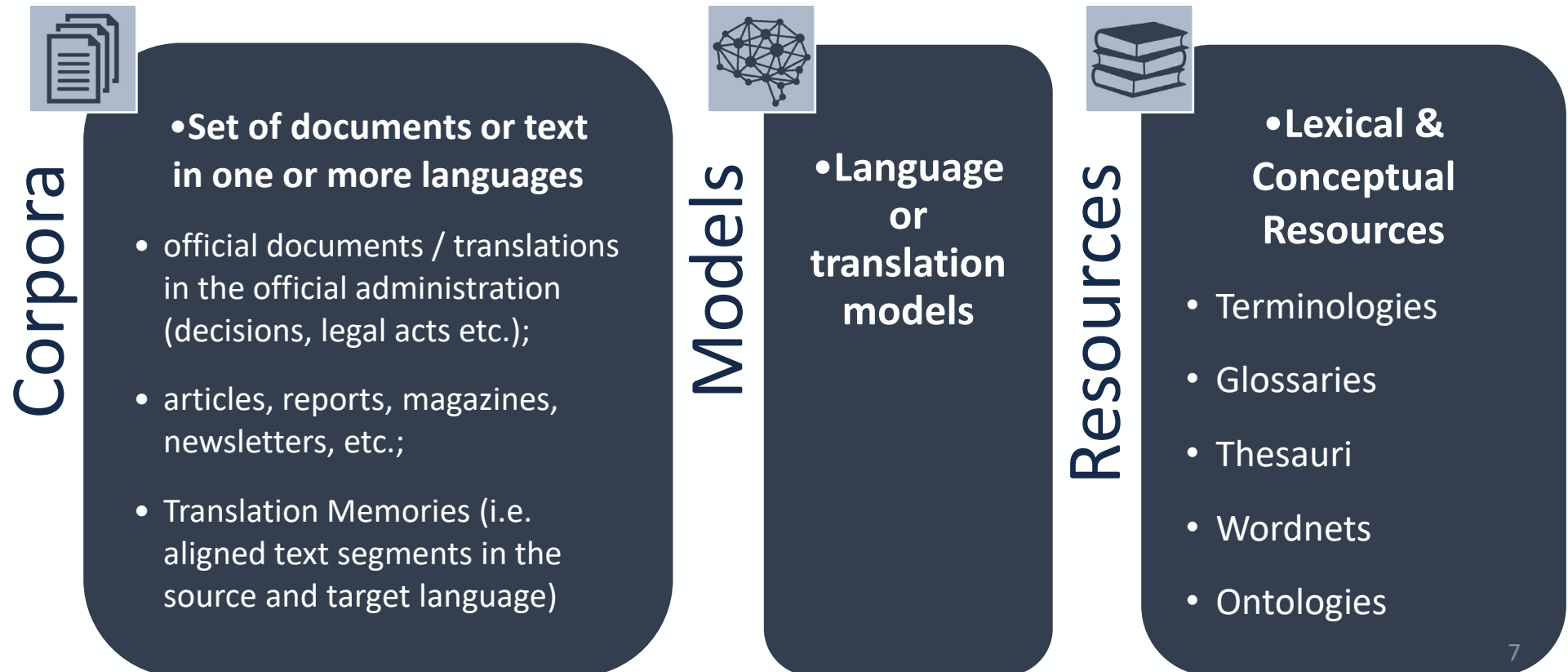
## e-Justice

## Open Data

## eProcurement

# LANGUAGE RESOURCES (LR) COLLECTED WITHIN ELRC

ELRC collects the following types of language data:



# ELRC NETWORK & OUTREACH ACTIVITIES

---



# ELRC NETWORK

National Anchor Points (NAP) in each country

[www.lr-coordination.eu/anchor-points](http://www.lr-coordination.eu/anchor-points)



# ELRC PROGRESS AND GROWTH

## PHASE 1:

- Initial Roadshow (Country Workshops)
- Set-up of ELRC Network (T-NAP plus P-NAPs)
- Set-up of data collection & validation workflow
- Technical and Legal Helpdesk

## PHASE 2:

- Sustainability, local ownership
- Extension of ELRC Network on national level (600 data donors)
- Extended data collection (crawling) & validation workflow (Generic Services Projects)
- Addition: On-site assistance

## PHASE 3:

- Sustainability, re-usability, opening (SME)
- Extension of ELRC Network (> 6.000 data donors and users)
- Extended data collection (speech, tools and services, ...)
- New languages (RU, CN, ...)

# TRADITIONAL ELRC OUTREACH AND NETWORKING ACTIONS



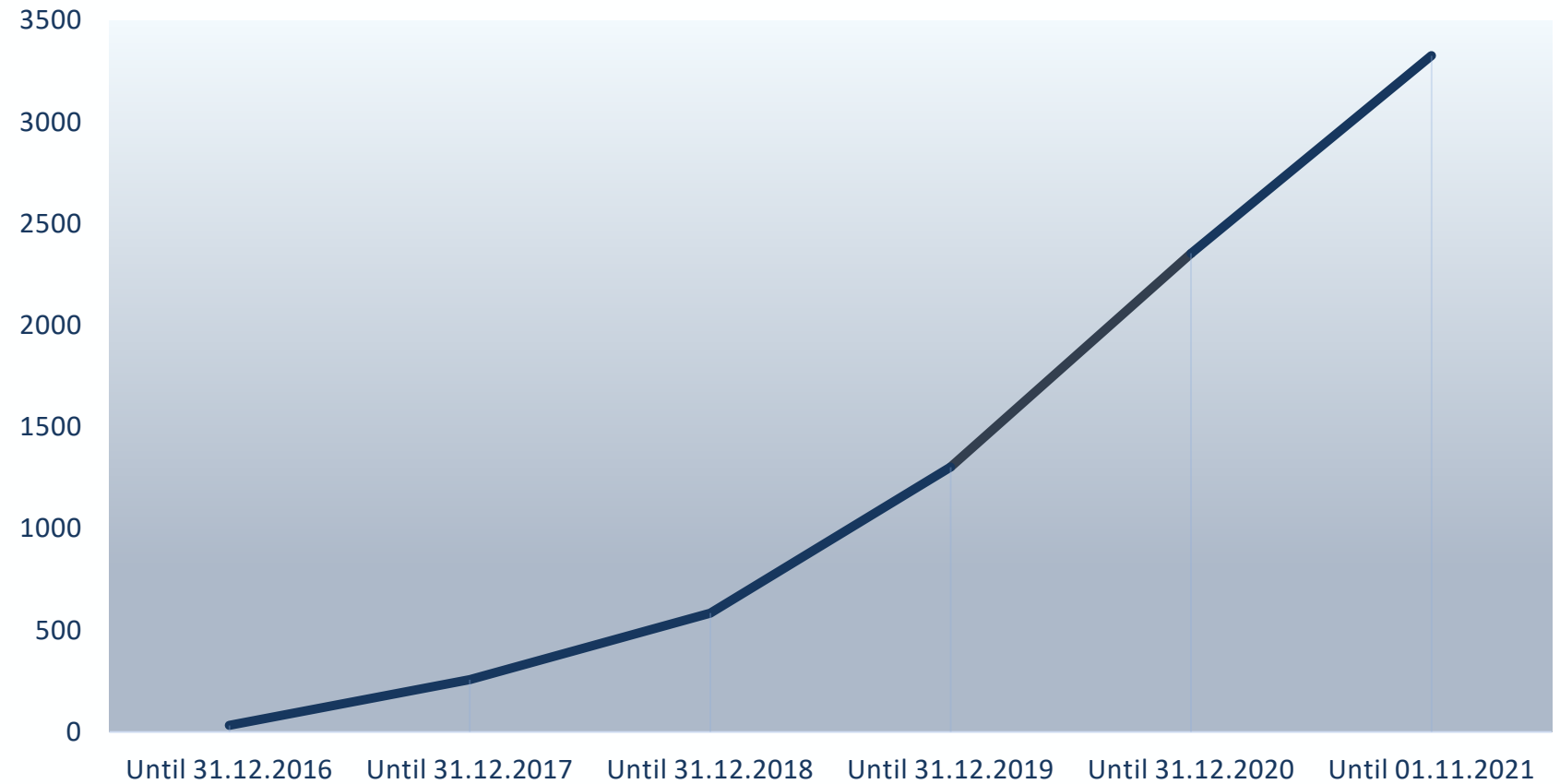


# ELRC: ACHIEVEMENTS

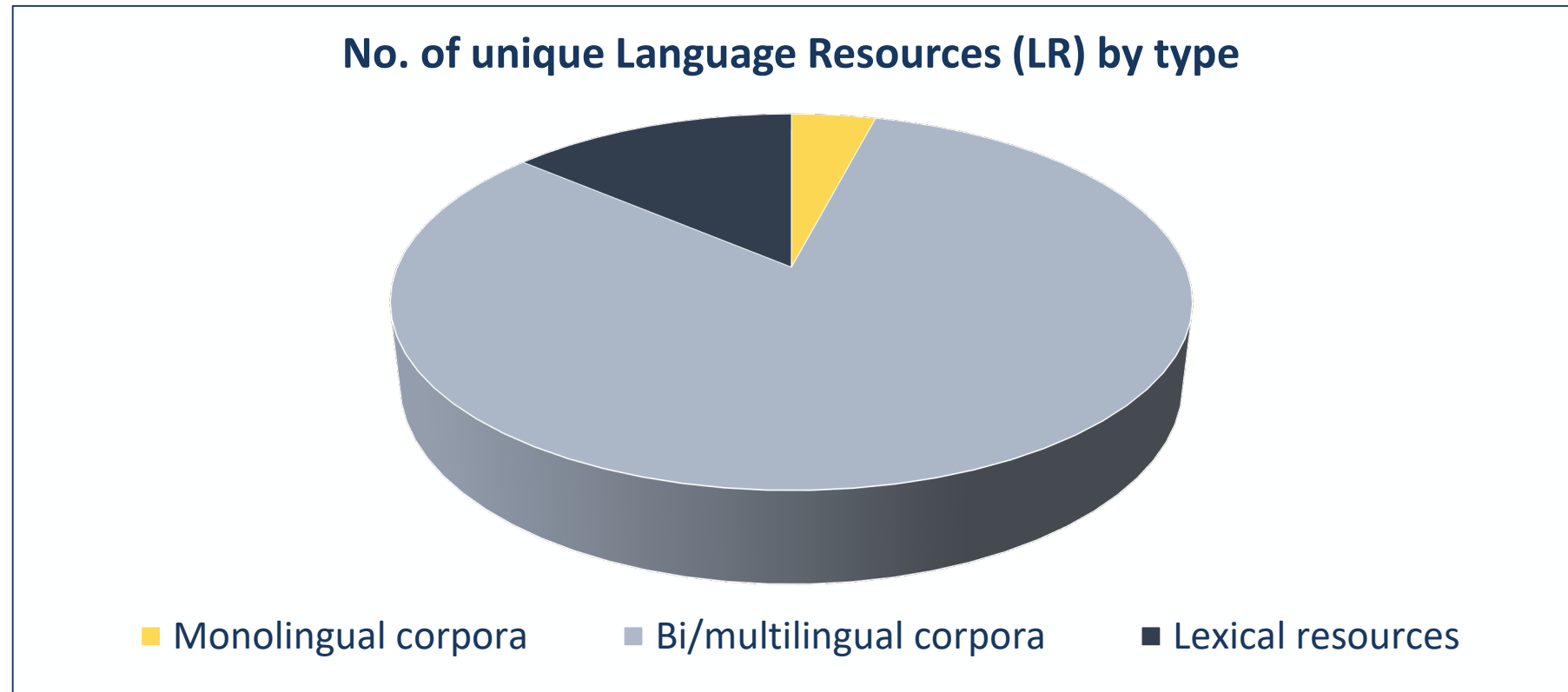
---

# ELRC ACHIEVEMENTS

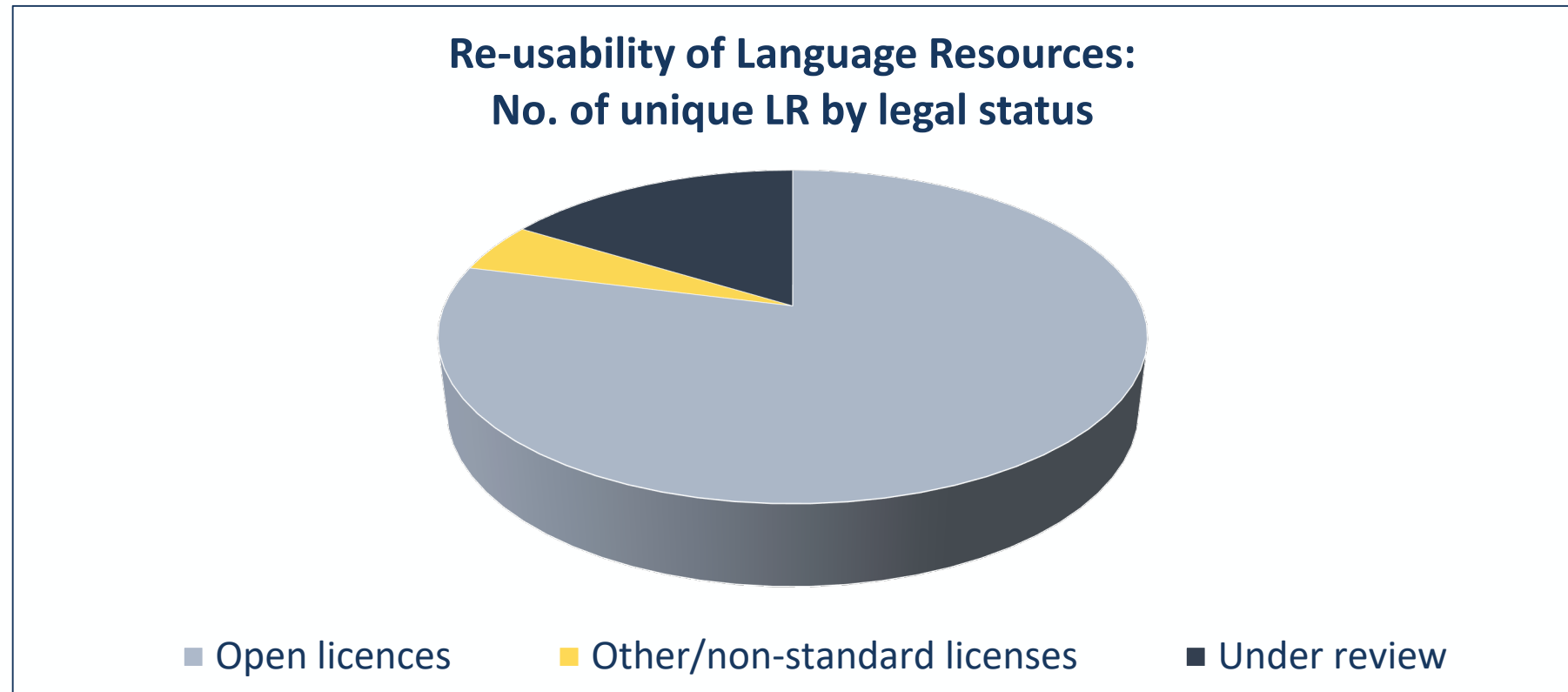
**No of unique LR on the ELRC-SHARE Repository**



# ELRC ACHIEVEMENTS

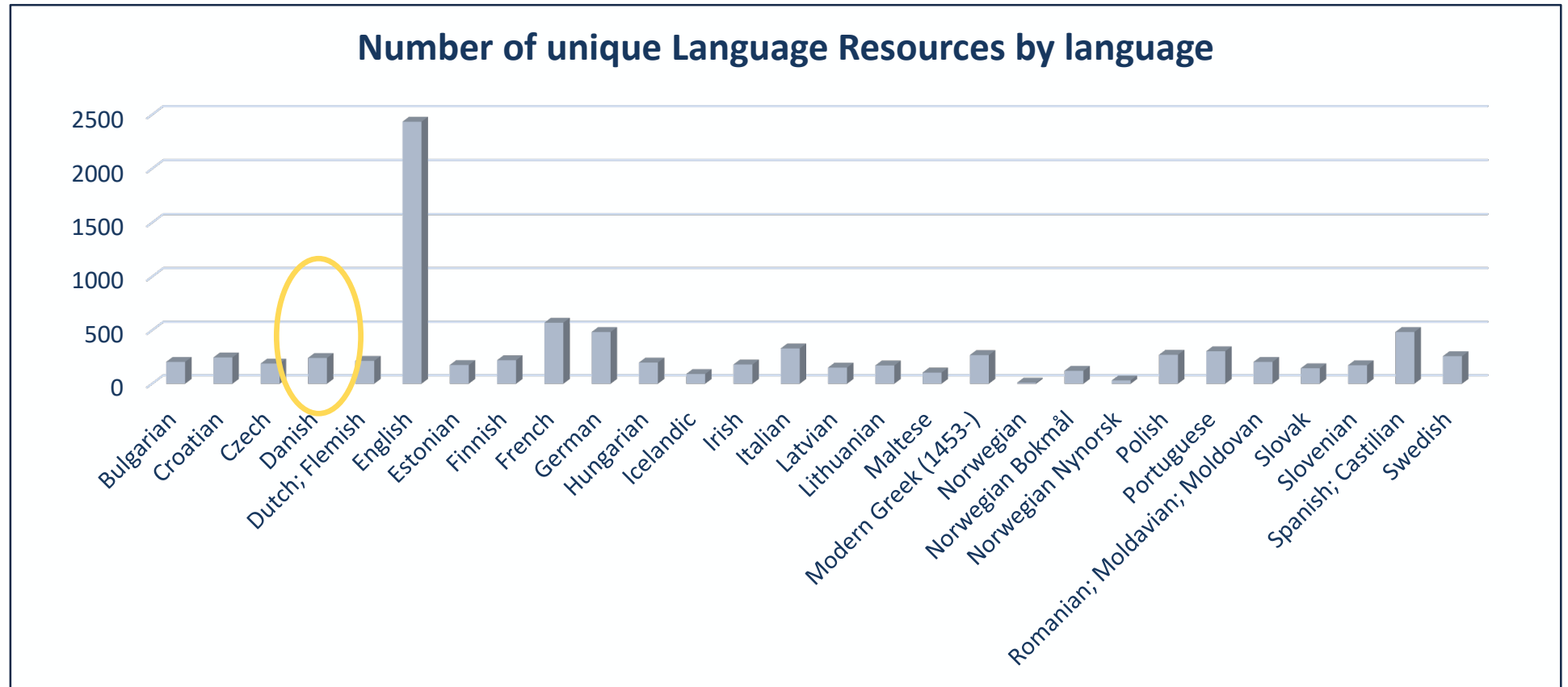


## ELRC ACHIEVEMENTS



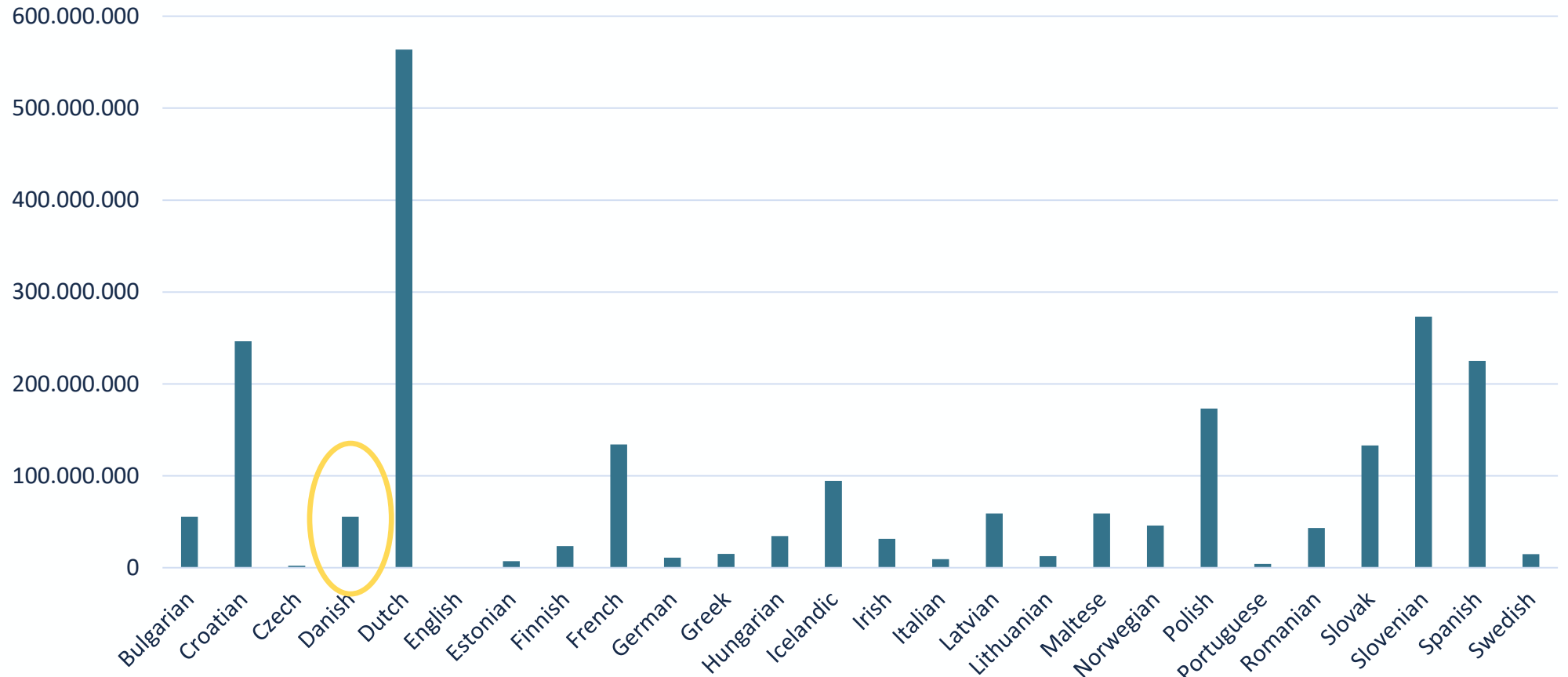


# CONTRIBUTIONS BY LANGUAGE



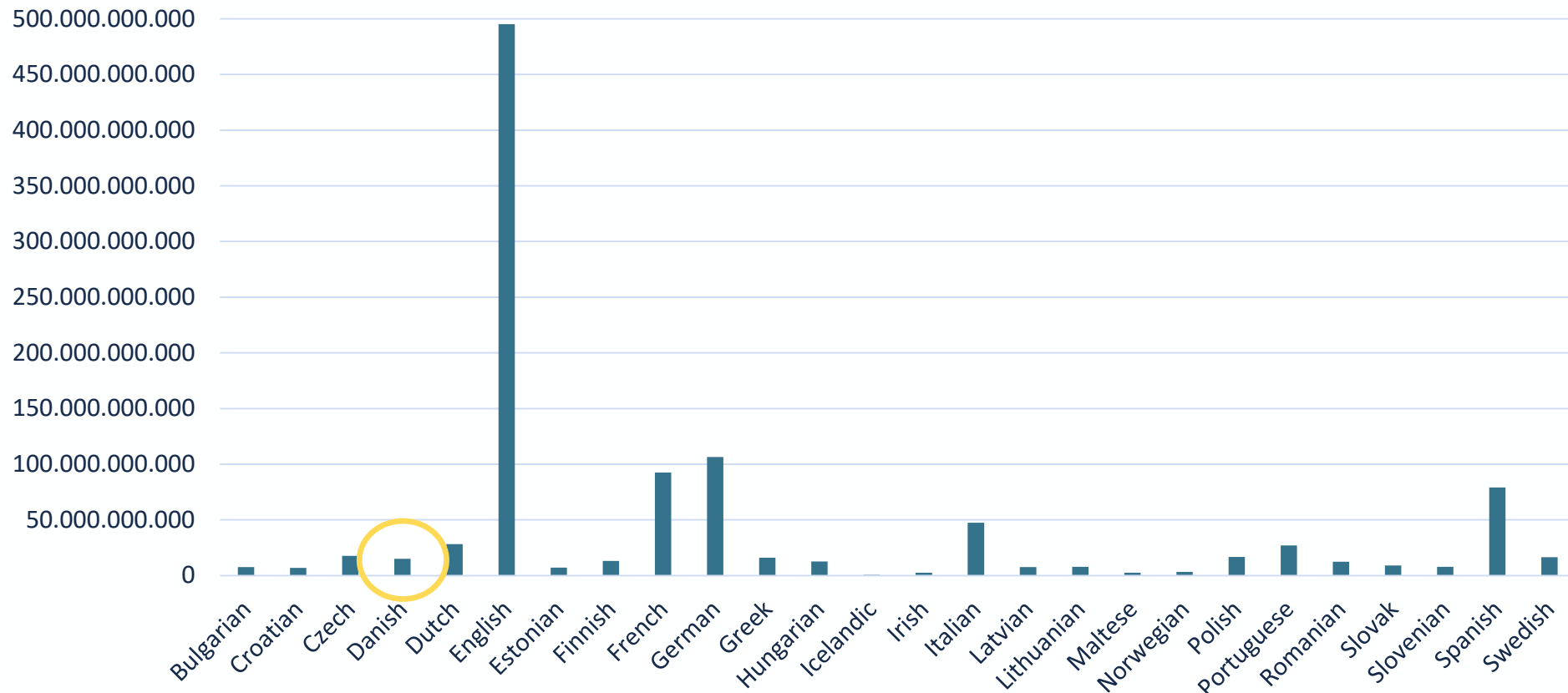
# COLLECTED WORDS IN MONOLINGUAL CORPORA

Estimated #words in monolingual corpora



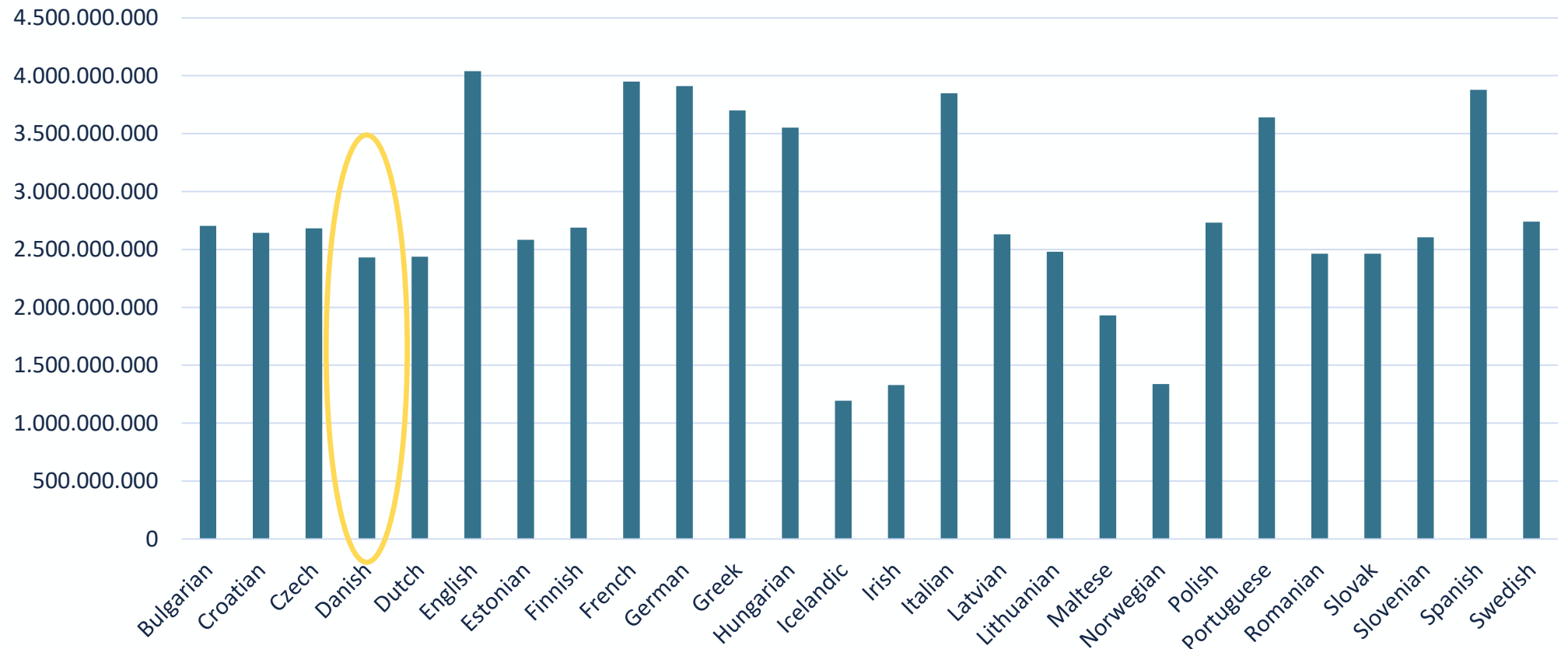
# COLLECTED WORDS IN BILINGUAL CORPORA

Estimated #words in bilingual corpora



# COLLECTED WORDS IN MULTILINGUAL CORPORA

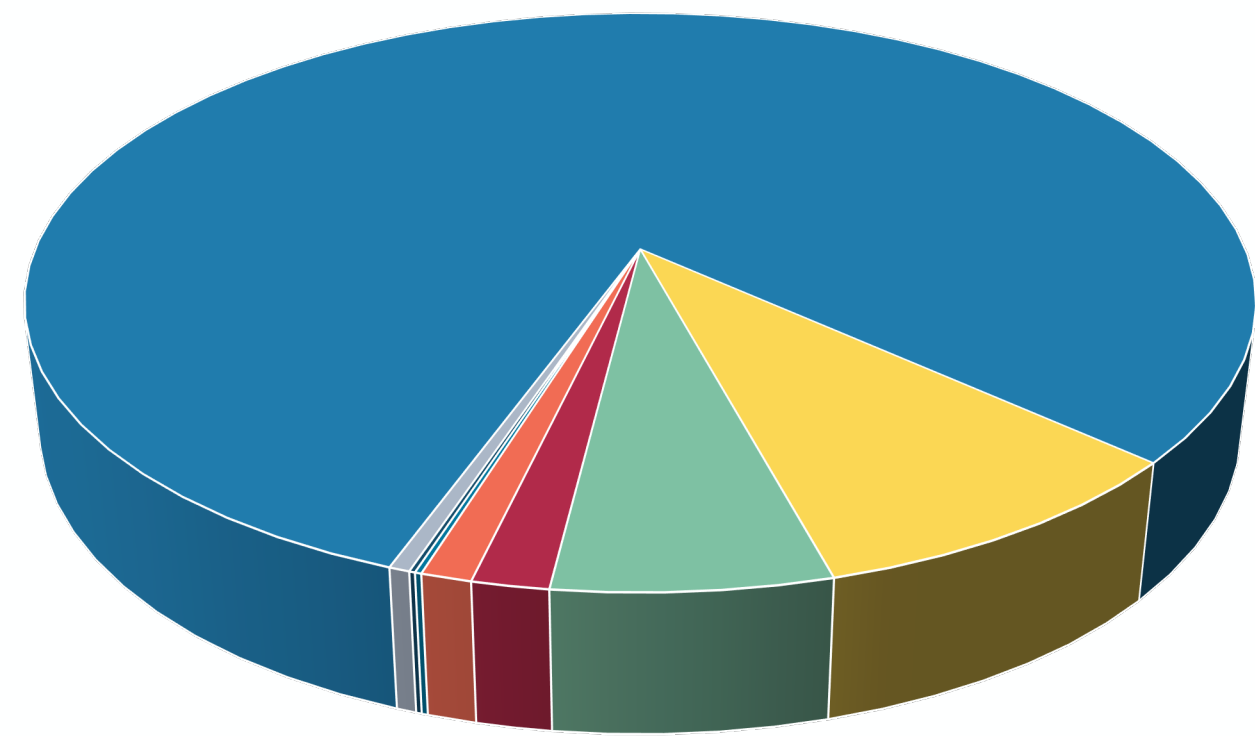
Estimated #words in multilingual corpora



# ELRC ACHIEVEMENTS

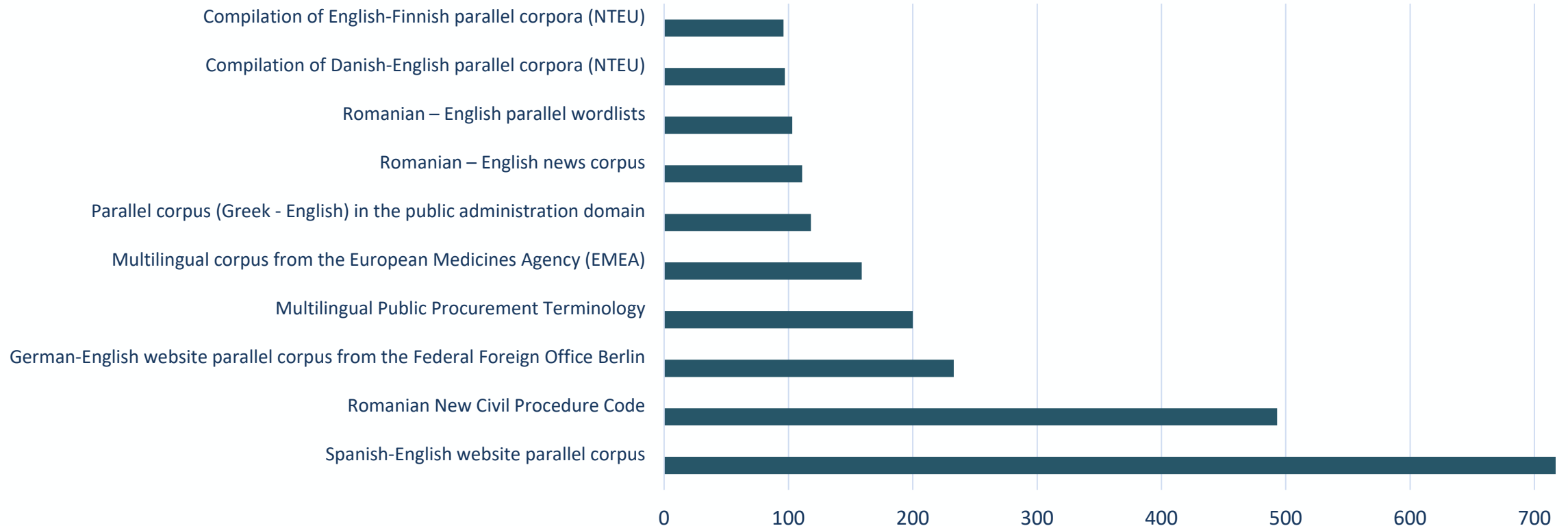
## Page visits by Continent

- Europe
- Asia
- North America
- Africa
- South America
- Central America
- Oceania
- Unknown

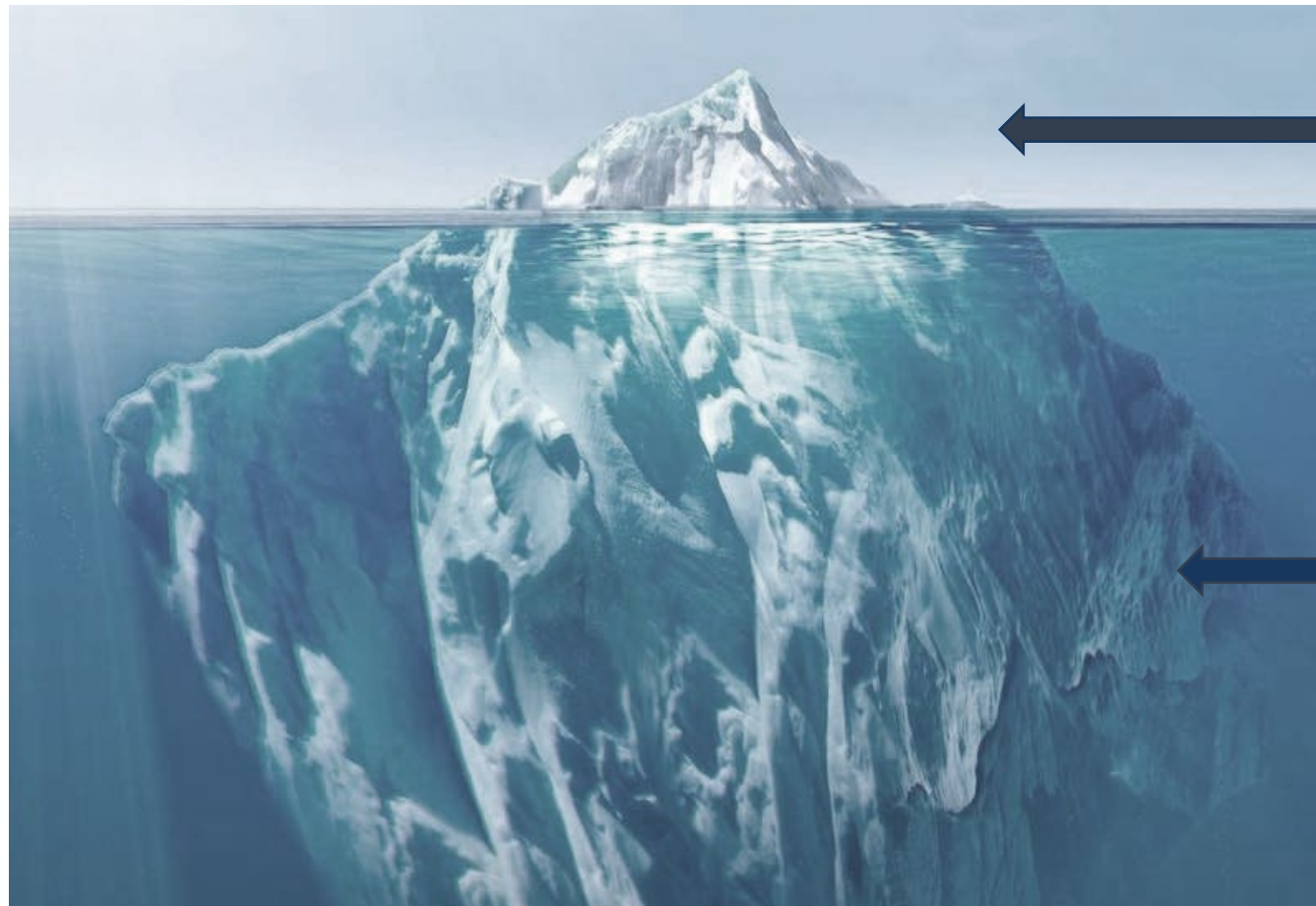


# ELRC ACHIEVEMENTS

## Top 10 Downloads



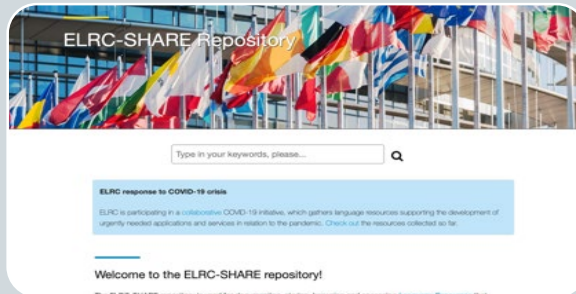
## STATUS QUO: THE TIP OF THE ICEBERG



← What we have...

← ... and what there is!

# DONATING LANGUAGE DATA TO ELRC



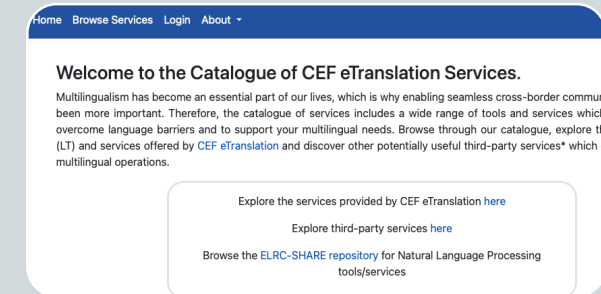
**ELRC-SHARE**  
ELRC Language Resource  
Repository

[www.elrc-share.eu](http://www.elrc-share.eu)



**ELRC Helpdesk**  
for any questions on technical  
or legal aspects related to the  
use, production, collection,  
processing and sharing of LR.

[www.lr-coordination.eu  
/helpdesk](http://www.lr-coordination.eu/helpdesk)



**Catalogue of CEF  
eTranslation Services**  
676 language tools/services  
offered by  
[CEF eTranslation](https://cef-at-service-catalogue.eu) and  
third-party services

[https://cef-at-service-  
catalogue.eu](https://cef-at-service-catalogue.eu)

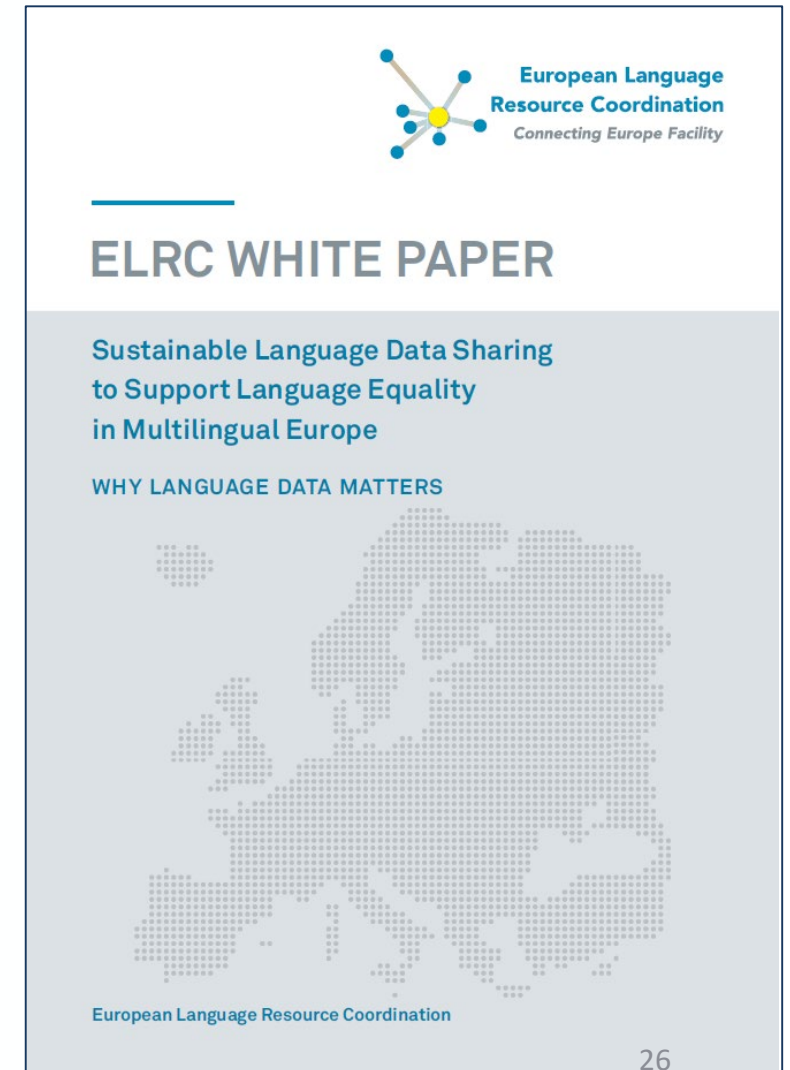


# ELRC: CHANCES, CHALLENGES AND OUTLOOK

---

# THE ELRC WHITE PAPER

- Stakeholders
  - Infrastructure for LR sharing
  - Challenges for LR sharing
  - Recommendations & Actions
  - Annex: Country Profiles
- Please visit: [www.lr-coordination.eu](http://www.lr-coordination.eu)  
➔ Helpdesk ➔ Infopoint



## MAJOR CHANCES

- Great number of potential data holders!
- More than 4.000 individuals from several hundred organisations
- But again: This is only the tip of the iceberg!



## MAJOR CHALLENGES

- Lack of appreciation of the value of language data
- Structural challenges

**Updated ELRC White Paper in 2022!  
Initial findings to be presented at upcoming LRB  
Meeting on 25 November!**

- Legal concerns

Scan for  
more info!



## DEVELOPMENTS AND OUTLOOK

### Extending Country Profiles and White Paper through Country Workshops

- 16 ELRC Country Workshops conducted under the current contract since end of 2020, another 13 are in the planning for end of 2021/2022
- Dedicated session on “Language Data creation, management and sharing”
- Define ways of overcoming obstacles in each country and across Europe

### Increasing language coverage

- Initially only EU official languages plus Norwegian and Icelandic (as CEF affiliated countries)
- Recently also Russian (end of 2019), Chinese (mid 2020), and other languages added based on increasing demand for new languages

### Opening of eTranslation for SMEs

- eTranslation is available to SMEs as well since March 2020

## DEVELOPMENTS AND OUTLOOK

### Extending collection of LR

- Relevant language tools for all EU official languages, Icelandic and Norwegian are available in the ELRC-SHARE (<https://elrc-share.eu/>)
- Speech and multimodal data are being considered too for inclusion

### Adjusting outreach activities to COVID-19 pandemic limitations

- ELRC Social Media Campaign since summer 2020 with help of the LRB


### Defining and developing Language Technology Specifications









- The purpose is to support the creation of technical specifications or best technical practices for the language technology applications in CEF eTranslation.
- Anonymisation API specification ready for implementation

# SPEECH-TO-TEXT TRANSCRIPTION

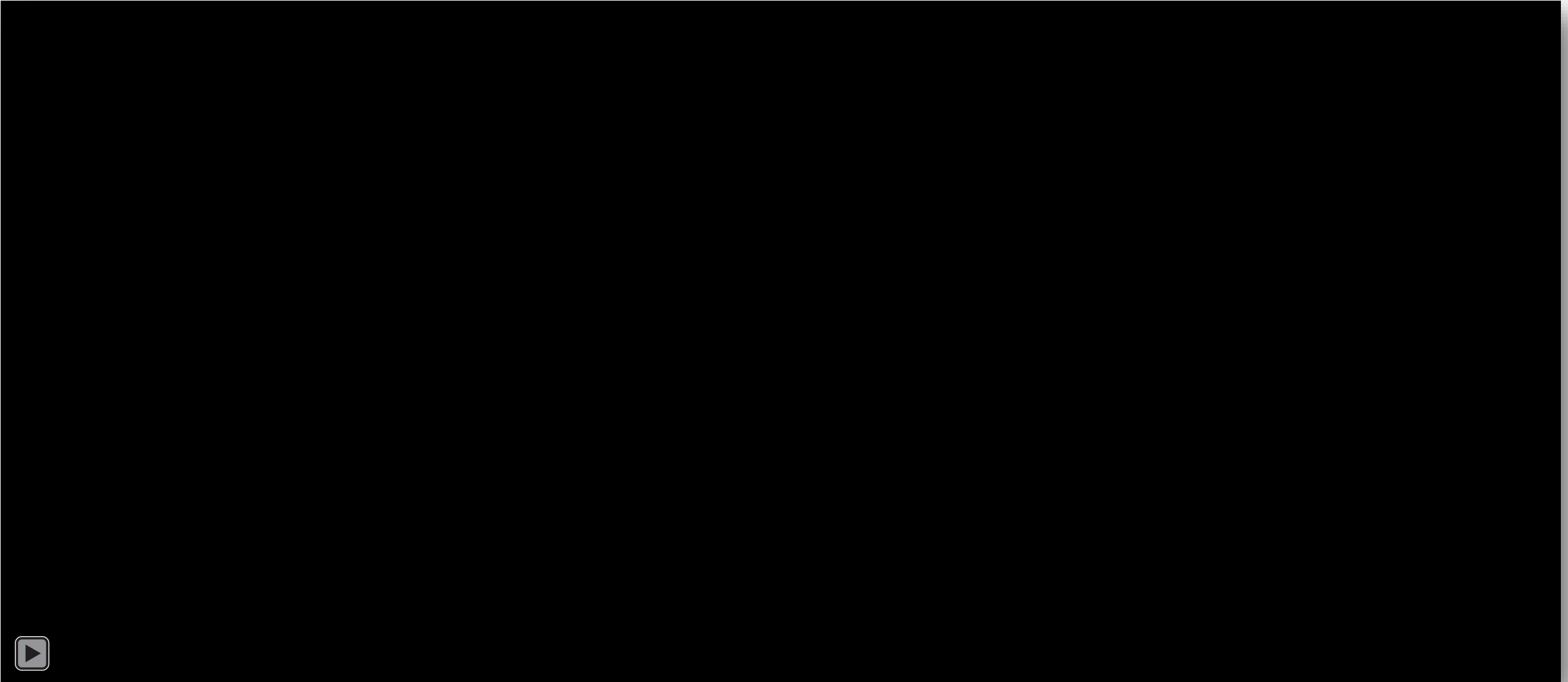
 <https://language-tools.ec.europa.eu/SpeechServices/Transcription>

An official website of the European Union How do you know? ▾

 Connecting Europe  
Language Tools

 <b>eTranslation</b>	 <b>Multilingual Tweet</b>	 <b>Speech-to-Text</b>	 <b>NLP Tools</b>
 <b>Interactive Terminology for Europe</b>	 <b>European Language Resource Coordination (ELRC)</b>	 <b>Catalogue of services</b>	 <b>CEF Building Block Information</b>

Access to some of these tools requires registration. EU staff are pre-registered.  
Please visit the registration page: <https://webgate.ec.europa.eu/etranslation/public/welcome.html>.  
For any other issues, please contact [help@cef-at-tools-services.eu](mailto:help@cef-at-tools-services.eu).





# CATALOGUE OF SERVICES

 <https://cef-at-service-catalogue.eu/>

Choose from 700 Language Tools and Services  
available in and for Europe!



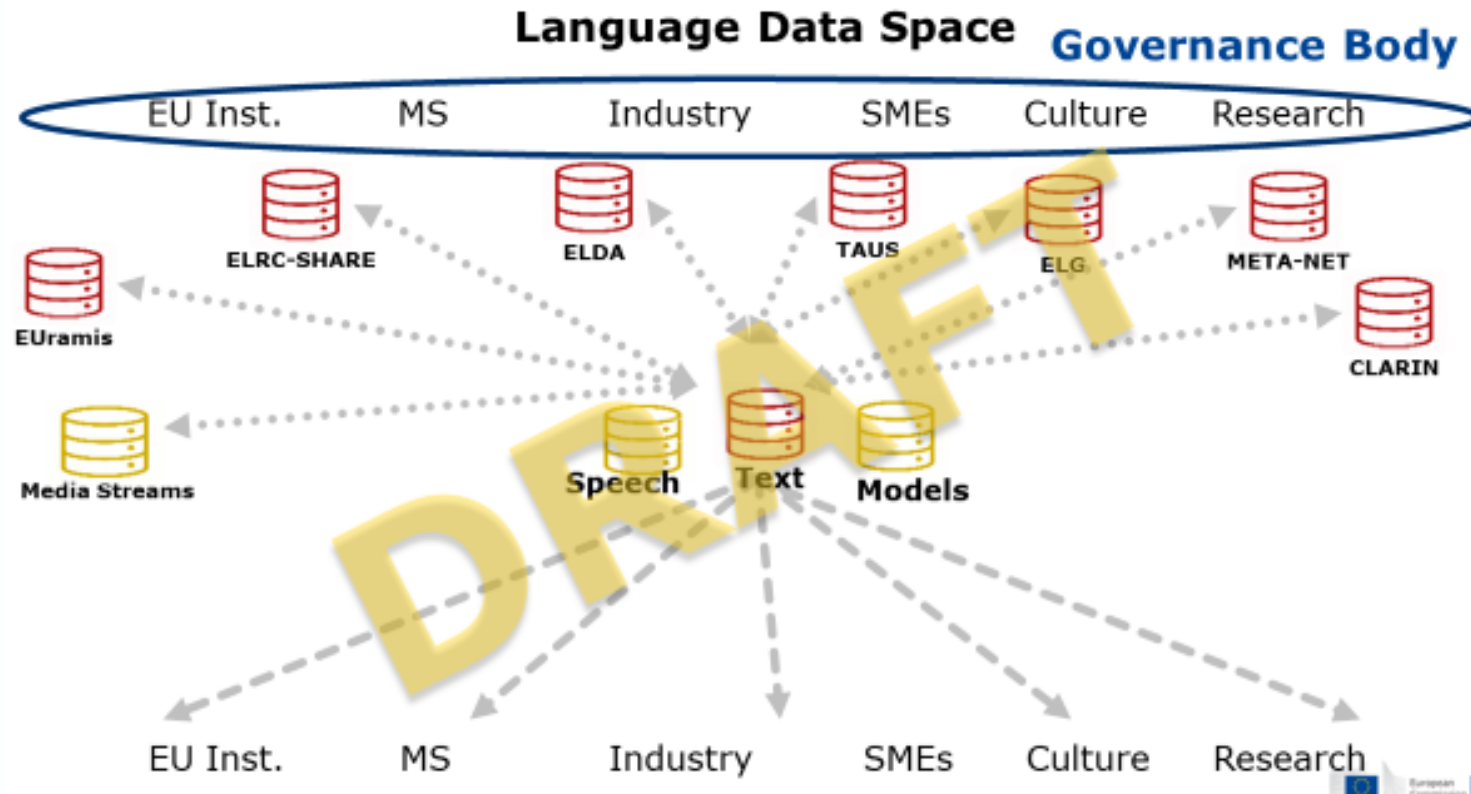
The screenshot shows the 'eTranslation services' website interface. The top navigation bar includes 'Home', 'Browse Services', 'Login', and 'About'. A search bar with a 'Search' button is on the right. Below the navigation, it displays 'Results 682 (1/35)'. A 'Filter By:' section is visible on the left, with a 'MORE' button. The main content area shows a search result for 'itl Institute for Technical Literature AG' with the following details:

- Function:** Language Technologies
- Task:** Authoring Support, Grammar Checking, Spell Checking, Terminology Management
- Provided by:** itl Institute for Technical Literature AG

On the left side of the screenshot, there are filter options for 'Provider's country' with checkboxes for German (125), Spanish; Castilian (118), and Germany (141).

# SOON TO COME UNDER DIGITAL EUROPE

👉 DIGITAL's Language Data Space

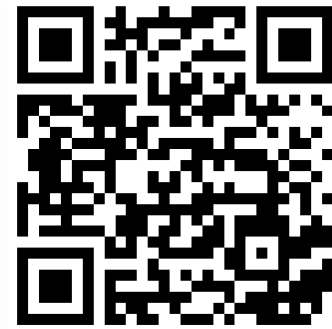


# STAY TUNED!

## Follow us on Social Media



EuropeanLanguage  
ResourceCoordination



LRCoordination



LR\_Coordination



# THANK YOU FOR YOUR ATTENTION!

Website: [www.lr-coordination.eu](http://www.lr-coordination.eu)

Twitter: @LR\_Coordination

Email: [info@lr-coordination.eu](mailto:info@lr-coordination.eu)

