EUROPEAN LANGUAGE RESOURCE COORDINATION (ELRC) UPDATE

ANDREA LÖSCH (DFKI)

ELRC Context and Objectives

ELRC OBJECTIVES

In

the public SMEs doma

language data

related to language data sharing

language data in Europe

Language data are key for machine translation / language technology!
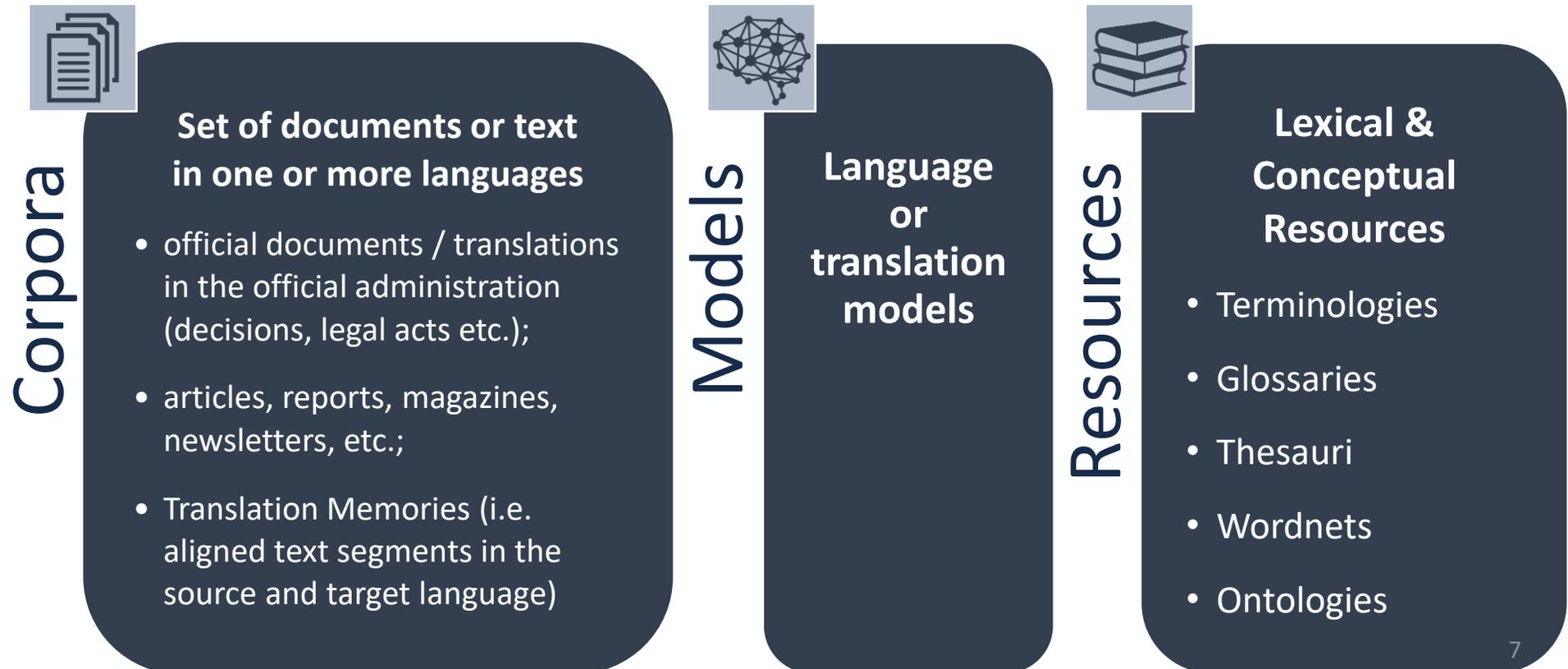
ELRC: Language Data for eTranslation

- **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:
    - 200+ systems connected to eTranslation
    - 7.000+ SMEs registered
    - More than 240 Mio. pages translated

- **Goal:** ELRC provides language resources to make eTranslation work for diffe-rent languages and domains!

LANGUAGE RESOURCES (LR) COLLECTED WITHIN ELRC

ELRC collects the following types of language data:

Corpora

Set of documents or text in one or more languages

- official documents / translations in the official administration (decisions, legal acts etc.);
- articles, reports, magazines, newsletters, etc.;
- Translation Memories (i.e. aligned text segments in the source and target language)

Models

Language or translation models

Resources

Lexical & Conceptual Resources

- Terminologies
- Glossaries
- Thesauri
- Wordnets
- Ontologies

ELRC Network & Outreach Activities

ELRC Network

National Anchor Points (NAP) in each country

www.lr-coordination.eu/anchor-points

Traditional ELRC Outreach and Networking Actions

SINCE COVID-19 PANDEMIC: ELRC ON SOCIAL MEDIA

#languagedatamatters

ELRC: Achievements

ELRC Social Media Campaign

- Facebook: Page "likes" development since March 2020

Total Page likes: 639

ELRC SOCIAL MEDIA CAMPAIGN

More than 33.000 link clicks in October! (avg. 1.000)

Facebook user activities

- Shares
- Comments
- Likes/Reactions
- Page reach
- Post reach
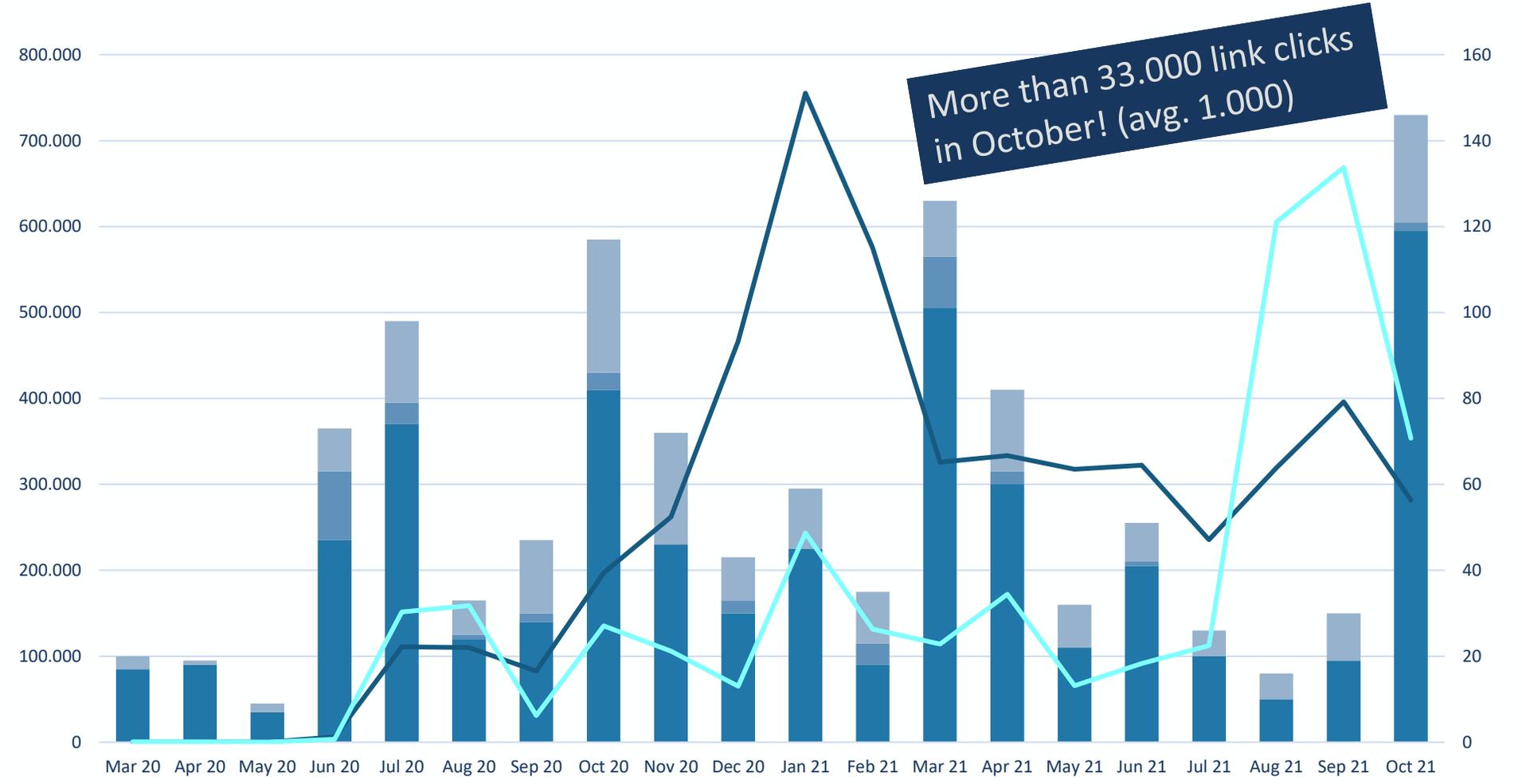
ELRC ACHIEVEMENTS

No of unique LR on the ELRC-SHARE Repository

ELRC Achievements

No. of unique Language Resources (LR) by type

■ Monolingual corpora   ■ Bi/multilingual corpora   ■ Lexical resources

ELRC ACHIEVEMENTS

Re-usability of Language Resources:
No. of unique LR by legal status

- Open licences
- Other/non-standard licenses
- Under review
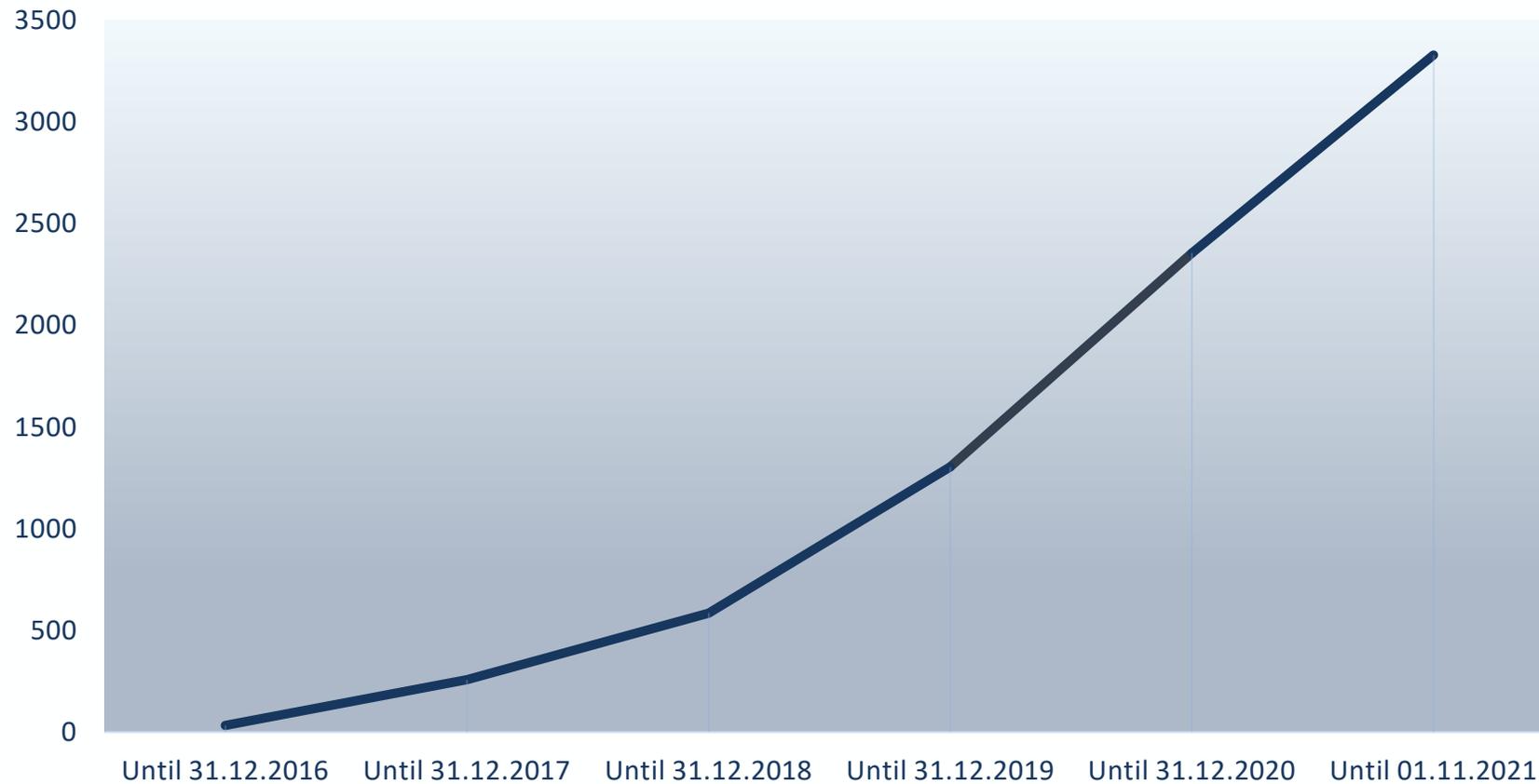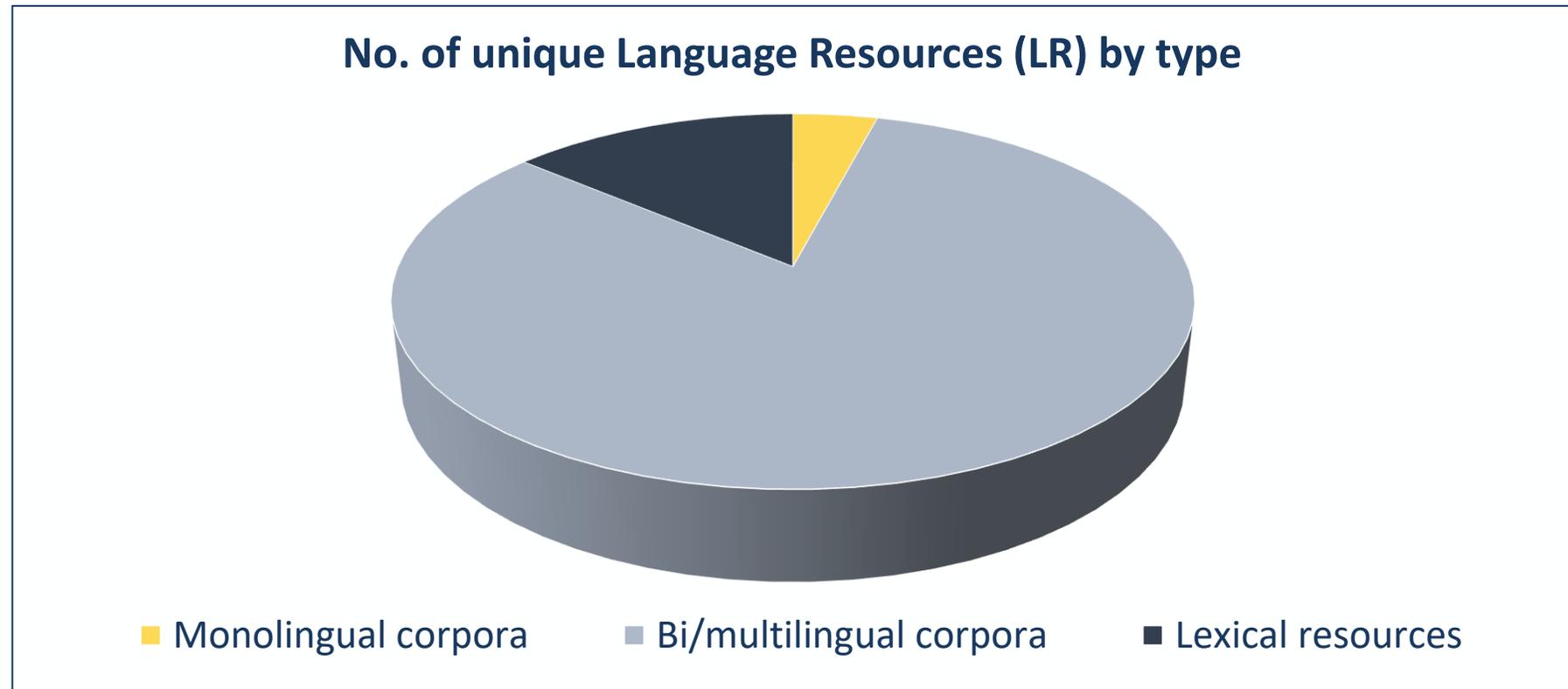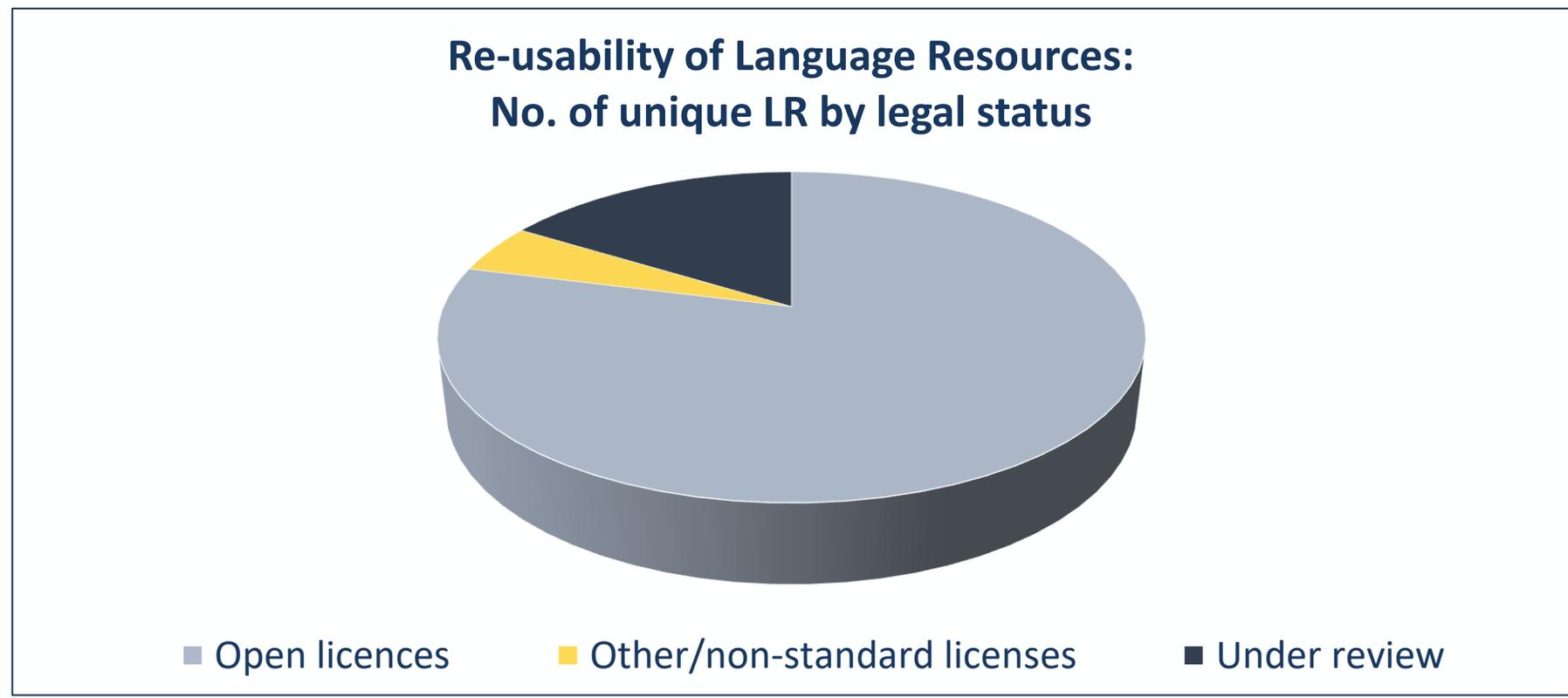
CONTRIBUTIONS BY LANGUAGE

Number of unique Language Resources by language
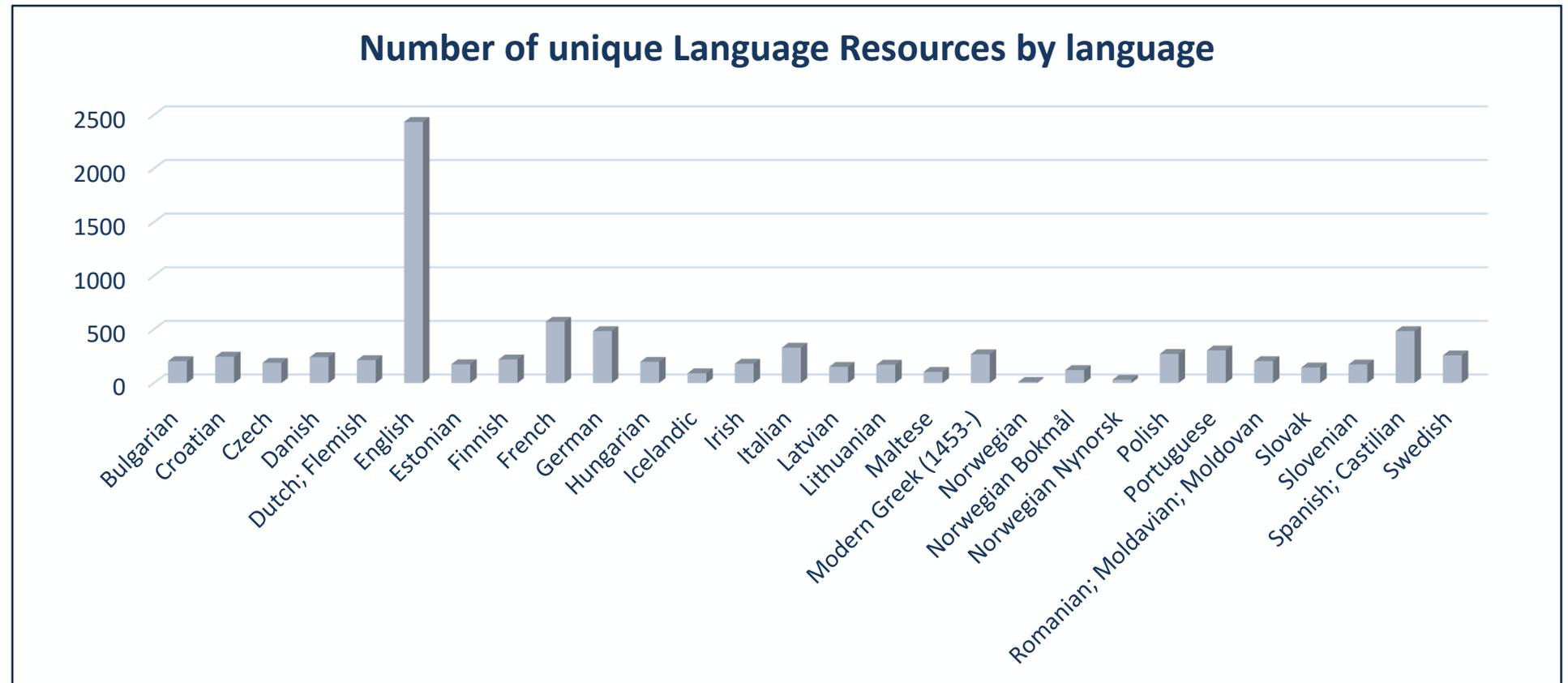
ELRC Achievements

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
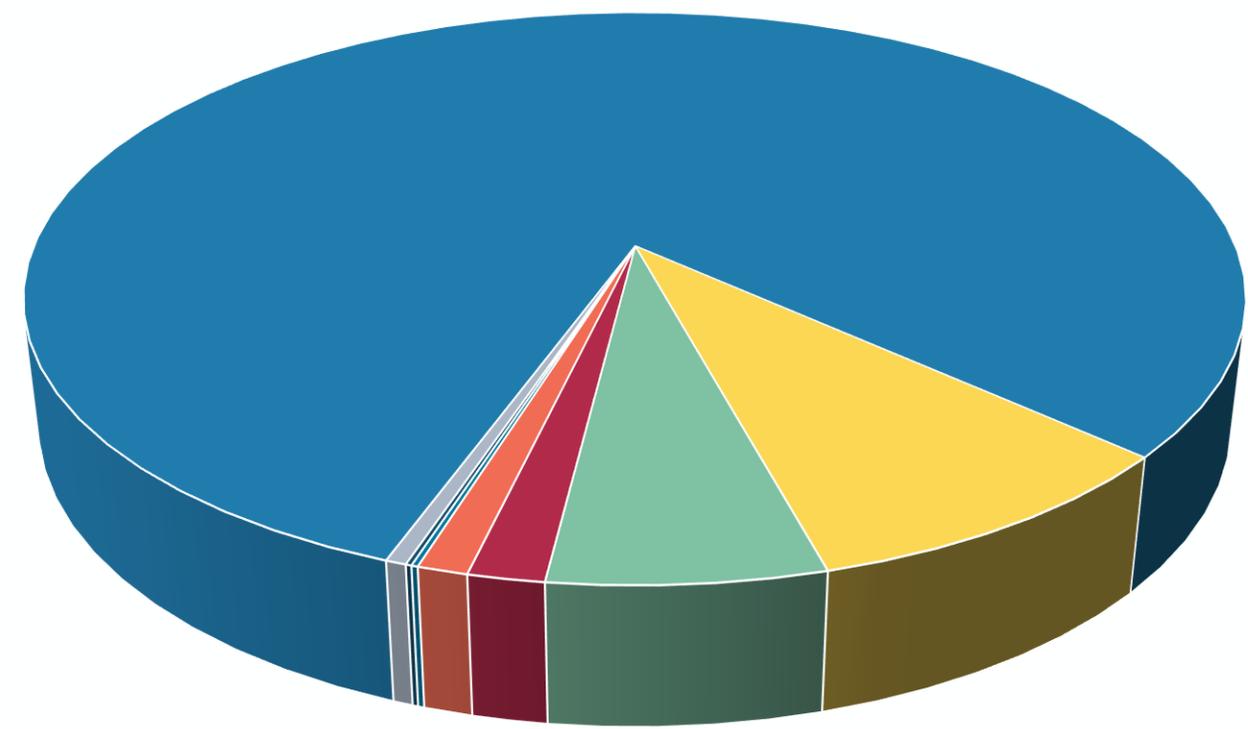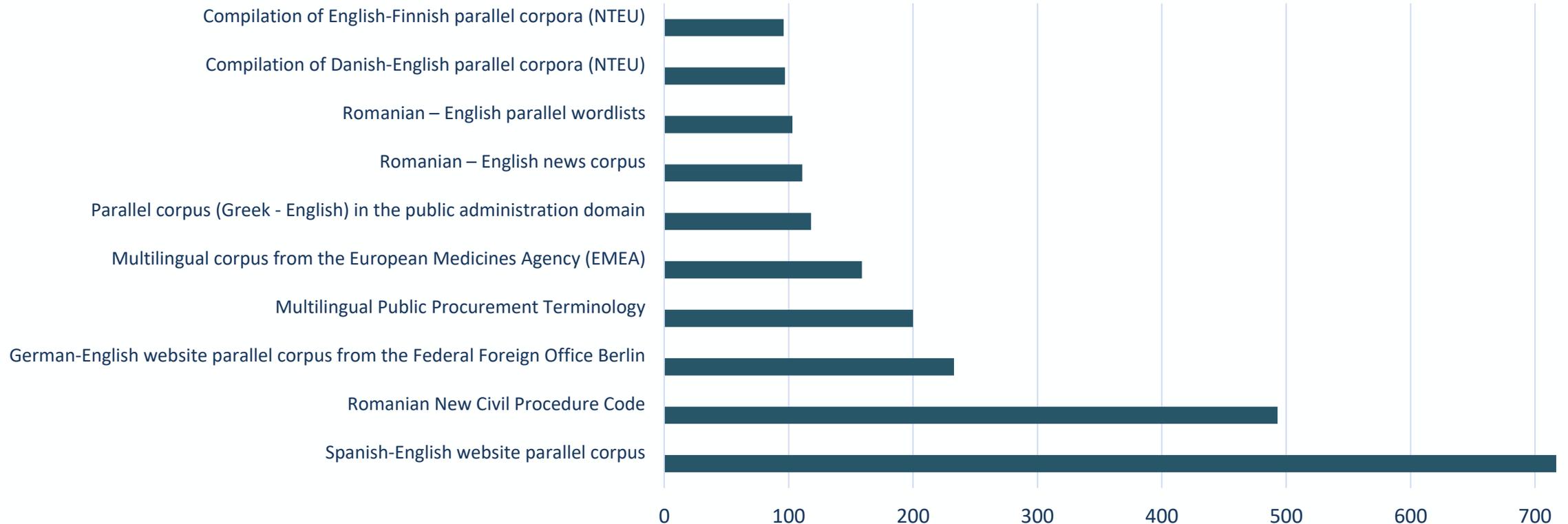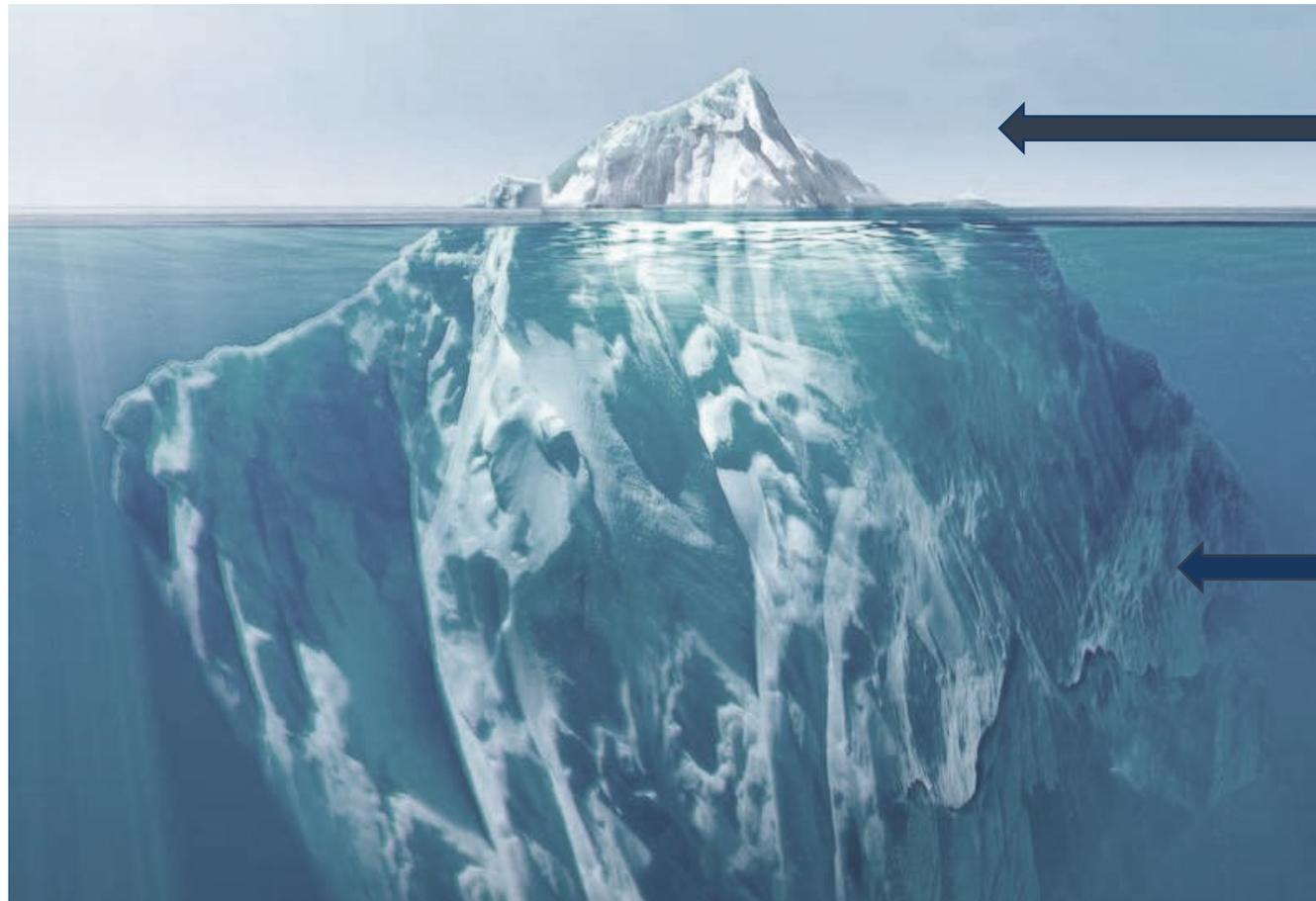

What we have…

… and what there is!

DONATING LANGUAGE DATA TO ELRC

ELRC-SHARE

ELRC Language Resource Repository

www.elrc-share.eu

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
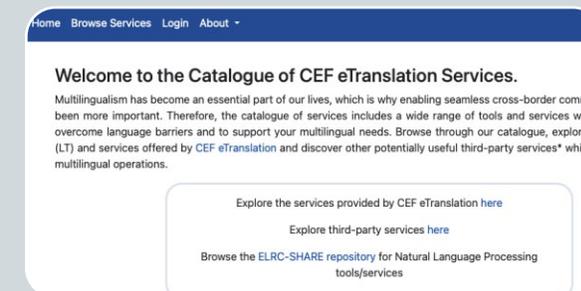
Catalogue of CEF eTranslation Services

676 language tools/services offered by CEF eTranslation and third-party services

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

ELRC: Chances, Challenges and Outlook

THE ELRC WHITE PAPER

- Stakeholders
- Infrastructure for LR sharing
- Challe...

...visit: www.lr-coordination.eu
➔ Helpdesk ➔ Infopoint

Updated version of the ELRC White Paper in 2022!

ELRC WHITE PAPER

...uage Data Sharing
...age Equality
...rope

...MATTERS

European Language Resource Coordination

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
- Today: Language Technologies in AI – Synergies, Collaboration Opportunities and Best Practices

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**

- Goal: To support the creation of technical specifications or best technical practices for the language technology applications in CEF eTranslation. Today: **Three LT specifications for review developed by ELRC**
- Anonymisation API specification soon ready for implementation

STAY TUNED!

Follow us on Social Media

EuropeanLanguage
ResourceCoordination
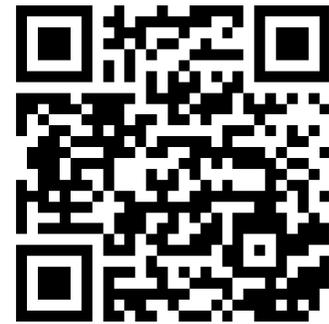
LRCoordination

LR_Coordination

THANK YOU FOR YOUR ATTENTION!

Website: www.lr-coordination.eu
Twitter: @LR_Coordination
Email: info@lr-coordination.eu