

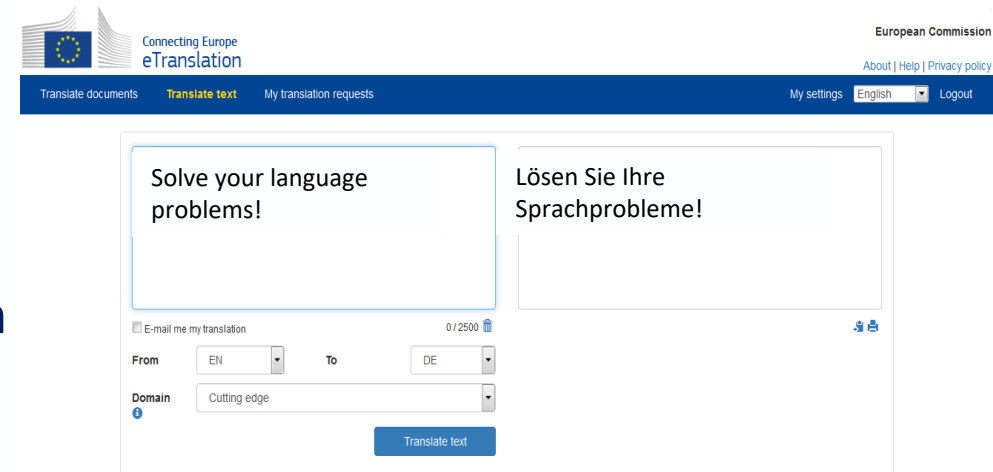
SPRACHDATEN ERSTELLEN, VERWALTEN UND VERFÜGBAR MACHEN: BESTEHENDE PRAKTIKEN UND HERAUSFORDERUNGEN

Andrea Lösch (DFKI)



WARUM SIND SPRACHDATEN SO WICHTIG?

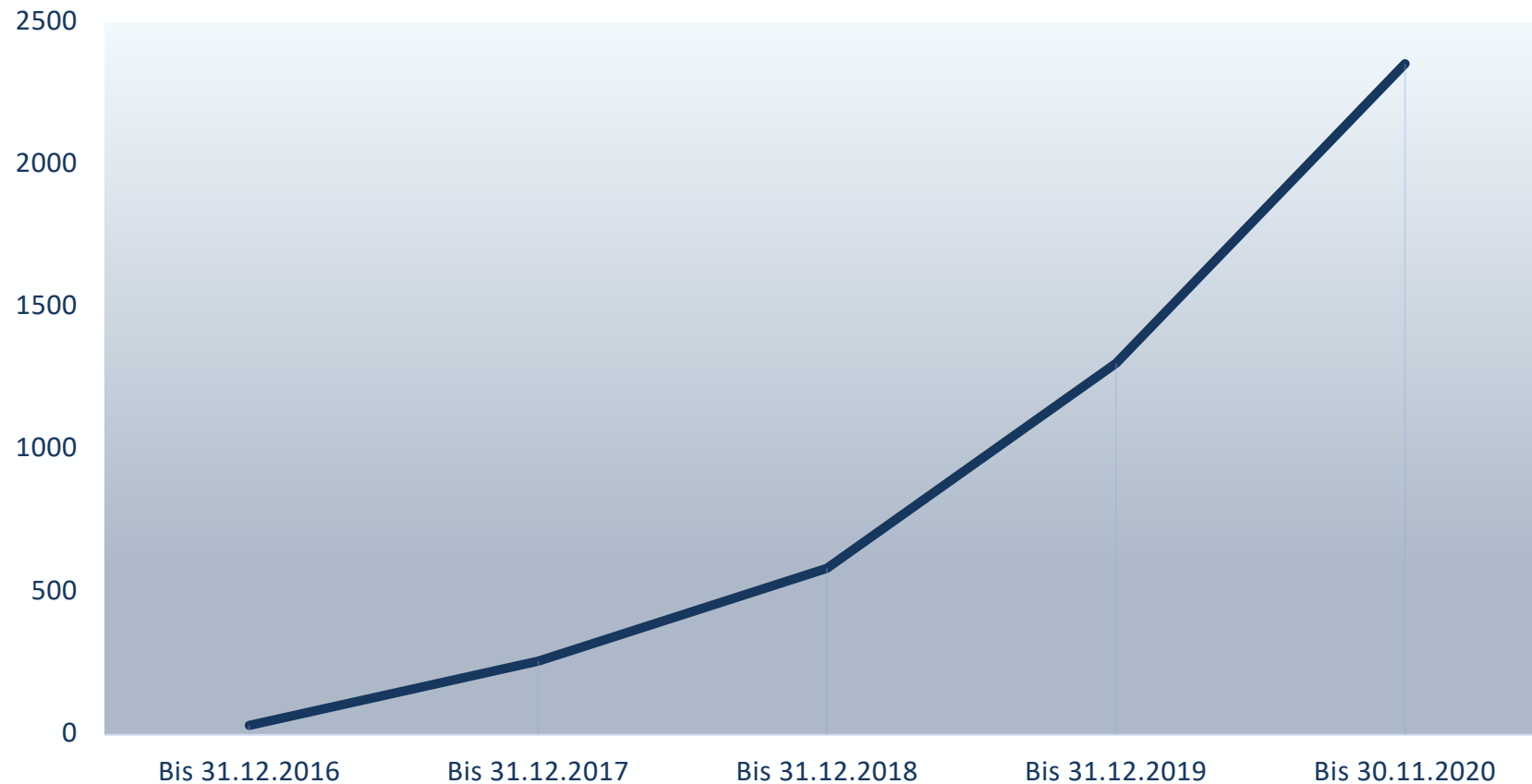
- CEF eTranslation hilft dem öffentlichen Dienst mehrsprachig zu agieren:
 - Das neuronale maschinelle Übersetzungstool der EK steht allen EU Organen, aber auch den Verwaltungen, dem öffentlichen und KMU in alle EU Mitgliedstaaten kostenfrei zur Verfügung
 - Es übersetzt alle 24 EU Amtssprachen (24 <-> 24), aber auch Russisch, Chinesisch, Türkisch, ...
 - Sicher und skalierbar
 - Unterstützt effektiv und effizient:
 - 70+ Systeme angebunden an eTranslation
 - 7.000+ SMEs registriert
 - Mehr als 40 Mio. Wörter übersetzt in 2020
- Damit eTranslation gute Übersetzungen den unterschiedlichen Sprachen und Domänen liefern kann, benötigt man entsprechende **Sprachdaten!**



The screenshot shows the eTranslation web interface. At the top, there is a navigation bar with the European Commission logo and the text "Connecting Europe eTranslation". Below this, there are links for "Translate documents", "Translate text" (highlighted in yellow), and "My translation requests". On the right side of the navigation bar, there are links for "About | Help | Privacy policy", "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 has the heading "Solve your language problems!" and a large empty text input box. Below the input box, there is a checkbox for "E-mail me my translation" and a character count "0 / 2500". Below the input box, there are dropdown menus for "From" (set to "EN") and "To" (set to "DE"), and a "Domain" dropdown menu set to "Cutting edge". A blue "Translate text" button is located at the bottom right of the input area. The right column has the heading "Lösen Sie Ihre Sprachprobleme!" and a large empty text input box.

ELRC – BISHERIGE ERFOLGE

Sprachressourcen im ELRC-SHARE



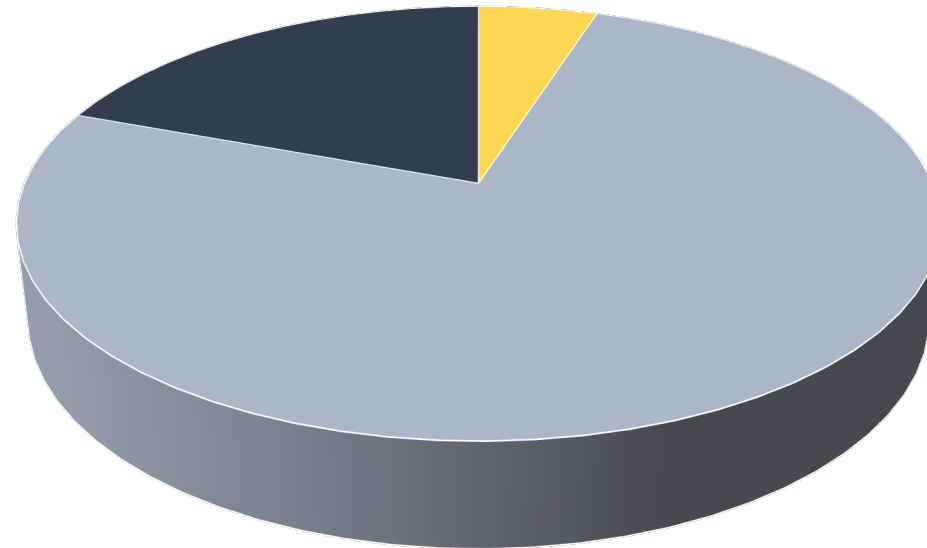
ELRC – BISHERIGE ERFOLGE

Sprachressourcen im ELRC-SHARE



ELRC – BISHERIGE ERFOLGE

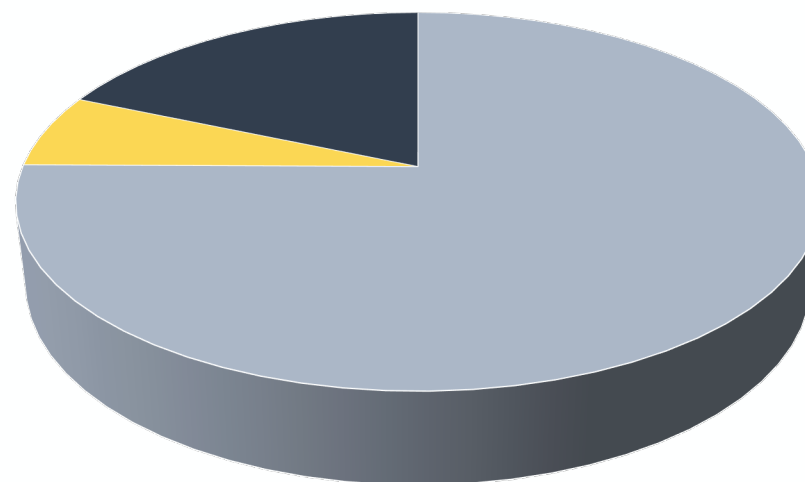
Anzahl einmaliger Sprachresources im ELRC-SHARE nach Typ



■ Monolinguale Korpora ■ Bi/multilinguale Korpora ■ Lexikalische Ressourcen

ELRC – BISHERIGE ERFOLGE

**Wiederverwendbarkeit der Sprachressourcen im ELRC-SHARE:
Anzahl einmaliger Sprachressourcen nach Rechtsstatus**

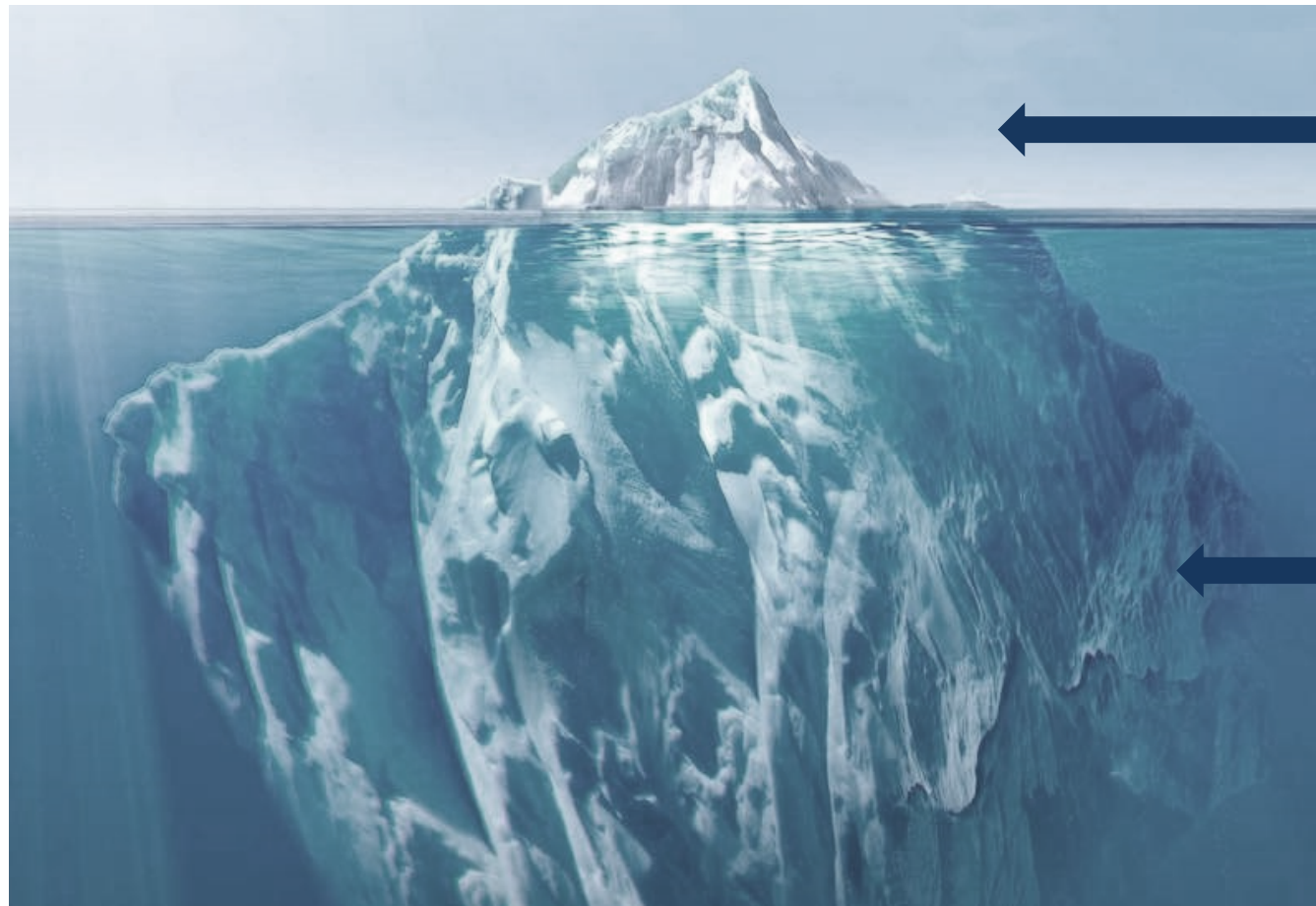


■ Offene Lizenzen

■ Andere Lizenzen

■ In Prüfung

ELRC - AKTUELLER STAND



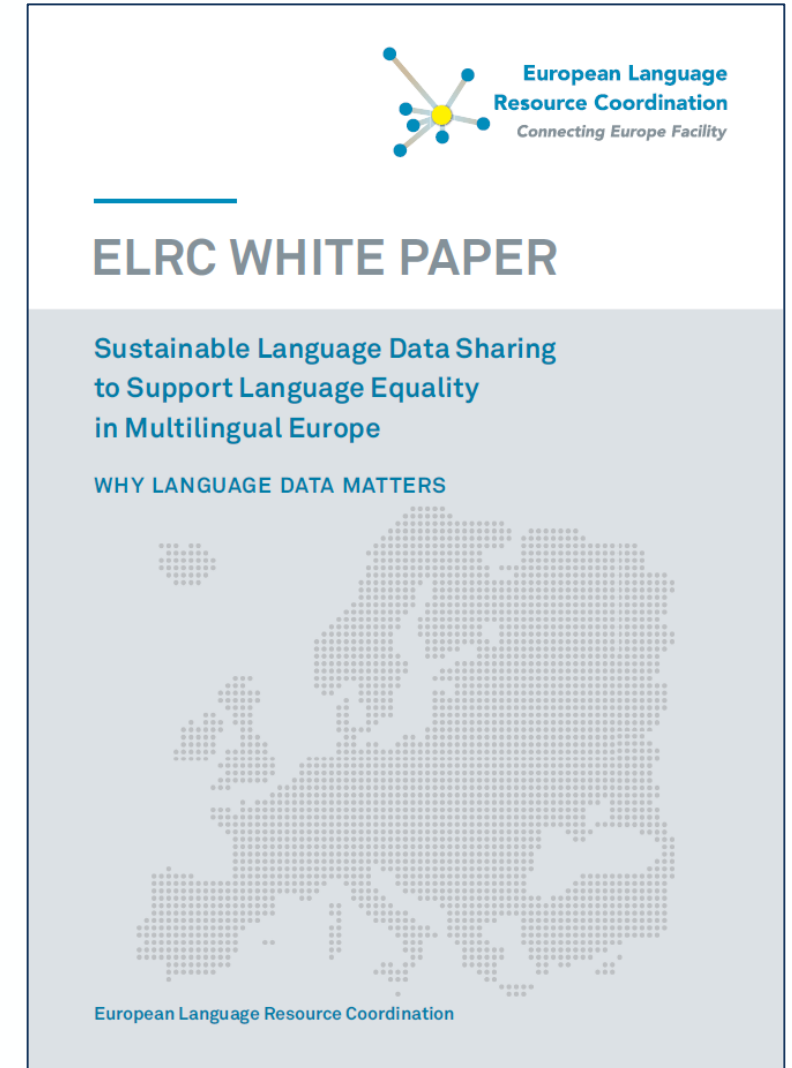
← Daten, die wir bis
jetzt gesammelt
haben ...

← ... und Daten, die es
tatsächlich gibt!

DAS ELRC WHITE PAPER

Überblick über:

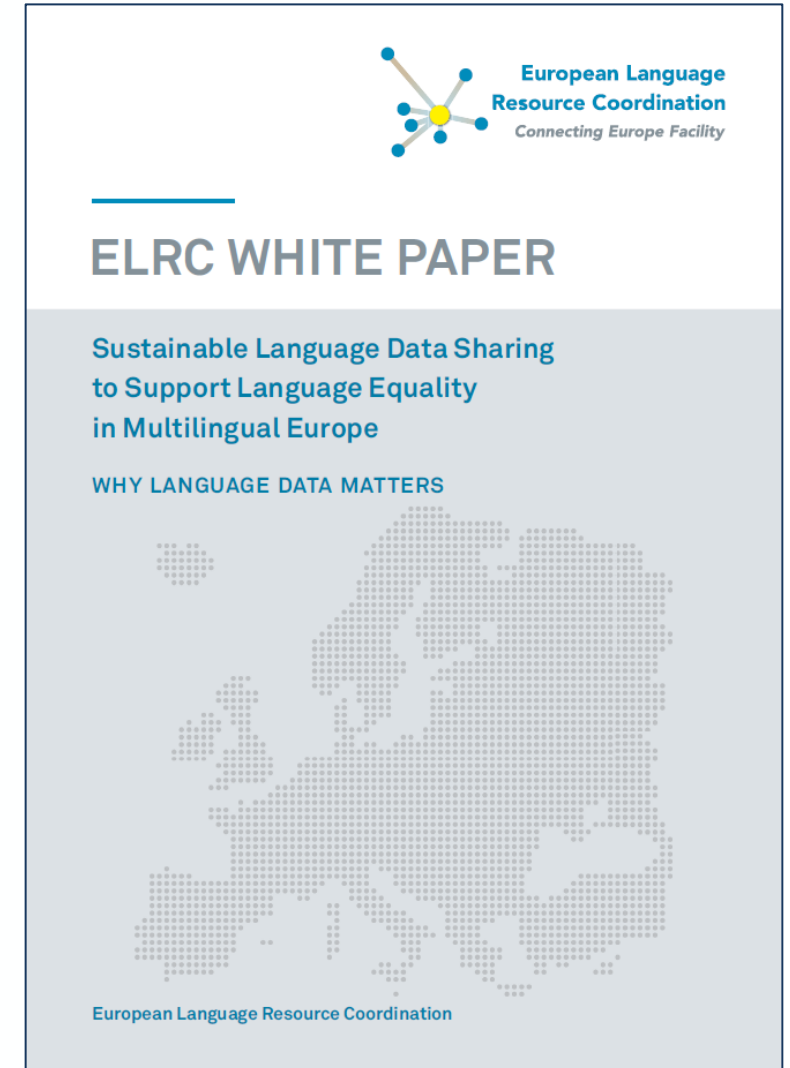
- Akteure im Bereich Sprachdaten
 - Infrastrukturen für Sprachdaten
 - Herausforderungen beim Teilen von Sprachdaten
 - Handlungsempfehlungen
 - Anhang: Länderprofile
- Online verfügbar unter:
www.lr-coordination.eu → Helpdesk
→ Infopoint



SPOTLIGHT LUXEMBURG

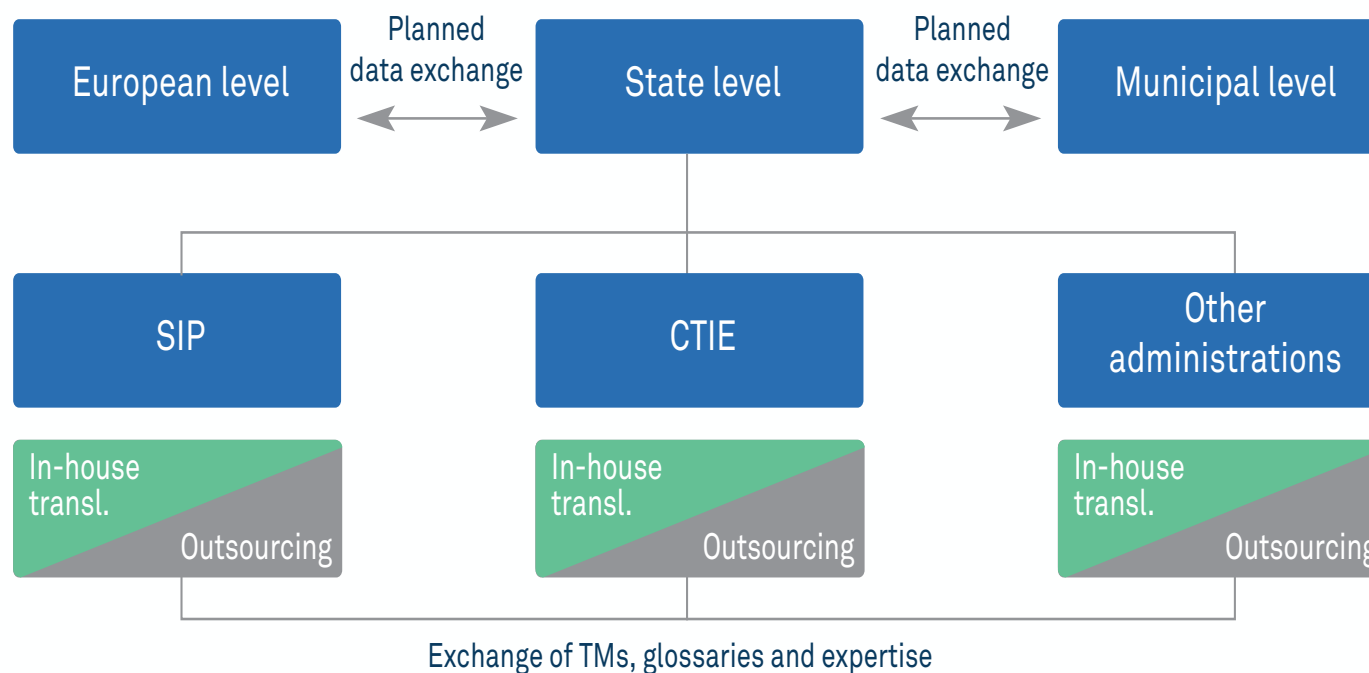
Übersetzungen im öffentlichen Dienst:

- Drei Amtssprachen (FR, DE, Luxemburgisch)
- Zusätzliche Übersetzungen in andere Sprachen notwendig (EN).
- Digitalisierung öffentlicher Dienste: Guichet.lu (CTIE – Zentrum für Staatliche Informationstechnologien) mit eigenem Übersetzungsdienst



SPOTLIGHT LUXEMBURG

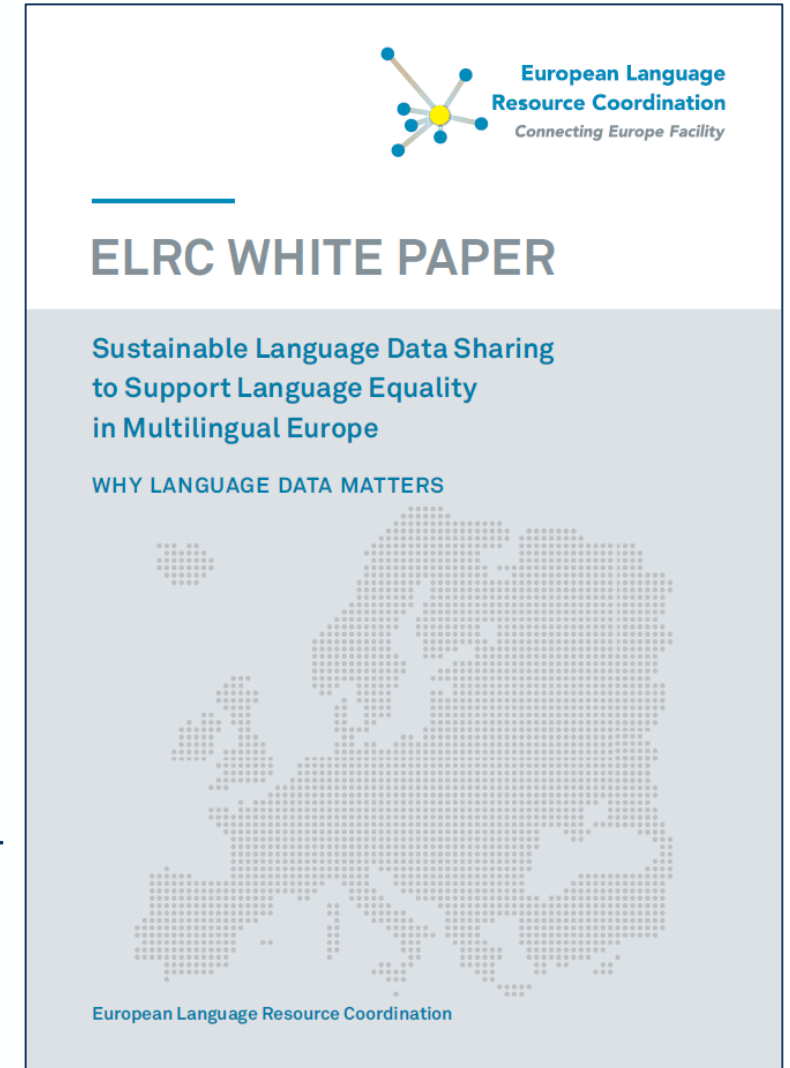
Infrastrukturen für das Erstellen von Übersetzungen:



SPOTLIGHT LUXEMBURG

Das Teilen von Sprachdaten:

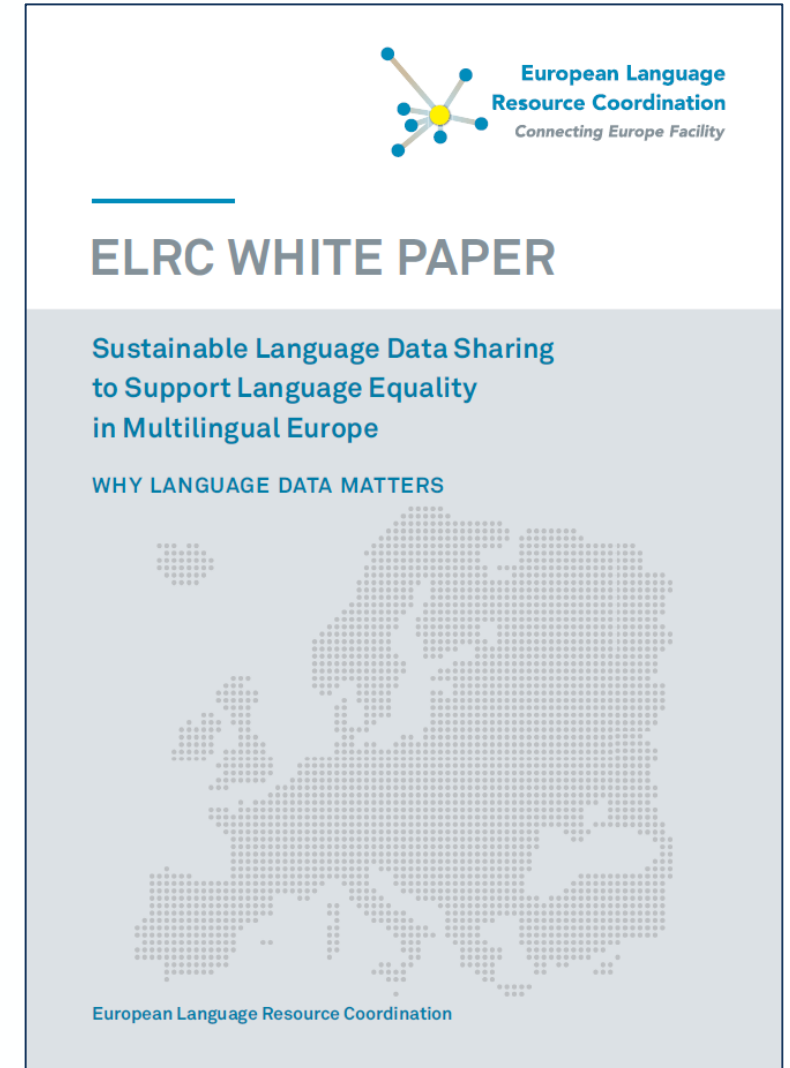
- Guichet.lu teilt Glossare mit dem Presse- und Informationsdienst (SIP) und auch dem nationalen Open Data Portal.
- Open Data Portal Luxemburg:
<https://data.public.lu/en/>
- Das Teilen von Translation Memories (TMs) zwischen Ministerien und Guichet.lu sollte ausgeweitet werden, z.B. auf die Agentur für Entwicklung und Arbeit (ADEM), den Informations- und Presse Service (SIP), den Luxemburger Stadtverwaltungen, und auch dem luxemburgischen Open Data Portal



SPOTLIGHT LUXEMBURG

Das Teilen von Sprachdaten:

- Grundsatz: Das Teilen von Daten aus dem Public Sector ist erlaubt (Public Sector Information Directive 2003/98/EC (modified in 2013 by the Directive 2013/37/UE)!
- Mögliche Einschränkungen:
 - Sind Daten evtl. Urheberrechtlich geschützt (Copyright?)
 - Wenn ja, können Sie den Urheber oder Autor identifizieren (ggf. Lizenz zur weiteren Verbreitung erwirken?)
 - Sind die Daten unter offenen Lizenzen verfügbar (z.B. cc licenses, NCGL 1.0, OGL 3.0 etc.)
 - Gibt es evtl. entsprechende Nutzungsbedingungen in dem Dokument / Datensatz zur Wiederverwertbarkeit?



SPOTLIGHT LUXEMBURG

Das Teilen von Sprachdaten:

- Grundsatz: Das Teilen von Daten aus dem Public Sector ist erlaubt (Public Sector Information Directive 2003/98/EC (modified in 2013 by the Directive 2013/37/UE)!
- Mögliche Einschränkungen:
 - Sind Daten evtl. Urheberrechtlich geschützt (Copyright?)
 - Wenn ja, können Sie den Urheber oder Autor identifizieren (ggf. Lizenz zur weiteren Verbreitung erwirken?)
 - Sind die Daten unter offenen Lizenzen verfügbar (z.B. cc licenses, NCGL 1.0, OGL 3.0 etc.)
 - Gibt es evtl. entsprechende Nutzungsbedingungen in dem Dokument / Datensatz zur Wiederverwertbarkeit?

➔ **ELRC Technical and Legal Helpdesk:** <https://www.lr-coordination.eu/helpdesk>



SPOTLIGHT LUXEMBURG

Das Teilen von Sprachdaten:

- Grundsatz: Das Teilen von Daten aus dem Public Sector ist erlaubt (Public Sector Information Directive 2003/98/EC (modified in 2013 by the Directive 2013/37/UE)!

Ihre Meinung zählt!

- Mögliche Fragen für Kinder:
 - Sind Daten evtl. Urheberrechtlich geschützt (Copyright?)
 - Wenn ja, können Sie den Urheber oder Autor identifizieren (ggf. Lizenz zur weiteren Verbreitung erwirken?)
 - Sind die Daten unter offenen Lizenzen verfügbar (z.B. cc licenses, NCGL 1.0, OGL 3.0 etc.)
 - Gibt es evtl. entsprechende Nutzungsbedingungen in dem Dokument / Datensatz, die Aussagen zur Wiederverwertbarkeit machen?

ELRC WHITE PAPER

Sustainable Language Data Sharing
to Support Language Equality
in Multilingual Europe

WHY LANGUAGE DATA MATTERS

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!

Webseite: www.lr-coordination.eu

Twitter: @LR_Coordination

Email: info@lr-coordination.eu

