Bundesministerium
des Innern, für Bau
und Heimat

# Erstellung, Management und Teilen von Sprachdaten

## Sprachdaten für den EU Council Presidency Translator



#LanguageDataMatters

# Hintergrund

- Ende 2019: Aufruf des AA an alle Sprachendienste, Daten für den EUCPT zur Verfügung zu stellen
- Sprachen: EN, FR, IT, ES, PL
- Größter Bestand in allen Sprachen: AA
- Meist nur Daten für EN und FR vorhanden
- 15 Behörden haben sich beteiligt

# Welche Art von Daten haben wir gesammelt?

- Format
  - bilinguale Dateien (xliff, tmx)
  - Word-Dateien (docx)

- Inhalte
  - Alle Themengebiete
  - Gezielt: Übersetzungen zu Finanzen und Corona-Pandemie
  - Nicht vertraulich

# Woher stammen die Daten?

- Mit Kolleg:innen überlegen, welche Berichte, Gesetze usw. gut geeignet wären
- Filtern von Kandidaten aus dem Auftragsverwaltungssystem (AVS) nach Textsorten
- Filtern im CAT-Tool anhand der gefundenen Auftragsnummern und Export
- Manuelles Herunterladen von Word-Dokumenten aus dem AVS
  ⇒ Entwicklung eines Tools nur für diesen Zweck
- BMI hat überwiegend Daten für DE-EN geliefert, ca. 80.000 Segmente.

# Wie wurden die Daten weiterverarbeitet?

- Wenig Aufbereitung im Sprachendienst
- Austausch über USB-Stick und BSCW-Server
- Vertraulichkeitsvereinbarung mit DFKI/Tilde:
  Daten wurden dort aligniert, bereinigt und anonymisiert
- Sprachendienste haben bereinigte Daten zurückbekommen

# Welche Fragen haben sich mir gestellt?

- Wie genau ist eine „Domäne“ definiert?
  Die Themen der Behörden sind sehr breit gefächert.
- Wie müssten Metadaten aussehen, um leichter aus Datenbanken exportieren zu können?
- Könnten die Sprachendienste ihre Daten mit wenig Aufwand selbst bereinigen oder ist immer externe Unterstützung notwendig?
- Wie kann die spätere Verwendung der Daten für das MÜ-Training im Übersetzungsprozess mitgedacht werden?

# Vielen Dank für Ihre Aufmerksamkeit!

## Kontakt

Alexandra Soska
Bundesministerium des Innern, für Bau und Heimat
ELRC Public Service National Anchor Point, Deutschland

alexandra.soska@bmi.bund.de

Bundesministerium des Innern, für Bau und Heimat