



# Language Resources needs of CEF Automated Translation platform

**Aleksandra Wesolowska**  
**Programme Officer - EU policies**  
**Data Value Chain Unit, DG CONNECT**

*LTi Summit 2015, 25 June 2015*



# What is CEF Automated Translation?

- Secure and adaptable **Automated Translation platform** that plugs into public online services and makes them multilingual
- Builds on MT@EC system but it is a broader concept: eventually platform that offers **tools and services for multilingual enablement of public services**
- **Underlying technologies:** machine translation, CAT, CMS, terminology management, semantic interoperability...
- **Public services:** Online Dispute Resolution (ODR), Open Data portal, Europeana, eJustice, Electronic Exchange of Social Security Information (EESSI), Safer Internet service infrastructure, Interconnection of Business registers ... eProcurement, eHealth

## What LR do we need for CEF.AT?

- *all types of data useful for MT training*
  - aligned parallel corpora: human translations (preferably TMs), monolingual textual data, dictionaries and other lexical resources, translation/language models, grammars...
- *resources for semantic interoperability*
  - taxonomies, ontologies ... ?
- *general domain and domain specific LR*
  - general domain parallel corpora and large monolingual corpora , e.g. official corpora of national languages;
  - parallel corpora, term collections, taxonomies in the domains of consumer affairs, culture, important domains of data included in the Open Data Portal, legal domain, social security domain...
- *open and proprietary*

# How do we satisfy the LR demand of CEF.AT?

- April 2015: launch of a pan-European **Language Resource Coordination effort (ELRC)**
- service contract with a consortium of European players (DFKI, Tilde, ILSP and ELDA + TAUS)
  - functional **network of stakeholders**
  - a **series of workshops** in all EU Member States and countries associated to CEF programme
  - a **technical and legal helpdesk**
  - **Language Resource Board**

## Next steps

- **website** of the LR coordination effort to be launched soon
- **workshops** in all EU Member States and CEF Associated countries: September-December 2015
- autumn 2015: **call for tenders** to extend and deepen the LR coordination network
  - move from training and networking to **systematic collection** of language resources
- **second CEF conference** in Q2 of 2016

## Challenges in LR collection for CEF. AT

- identification of the **right stakeholders**
- identification and obtaining of the **valuable resources** for all languages (open and licensed)
- **awareness raising** about the value of language resources at all levels
- spreading **best practice in LR management**
- **convincing** the stakeholders **to contribute** the resources to CEF.AT
- explaining/clarifying **the legal and technical aspects** for LR providers



**Thank you!**

***Email: [aleksandra.wesolowska@ec.europa.eu](mailto:aleksandra.wesolowska@ec.europa.eu)***