



Co-financed by the European Union

Connecting Europe Facility

eTranslation TermBank

Language Resource Board meeting

April 19, Nice

Project: eTranslation TermBank

Agreement number: INEA/CEF/ICT/A2016/1331795

Action No: 2016-EU-IA-0122

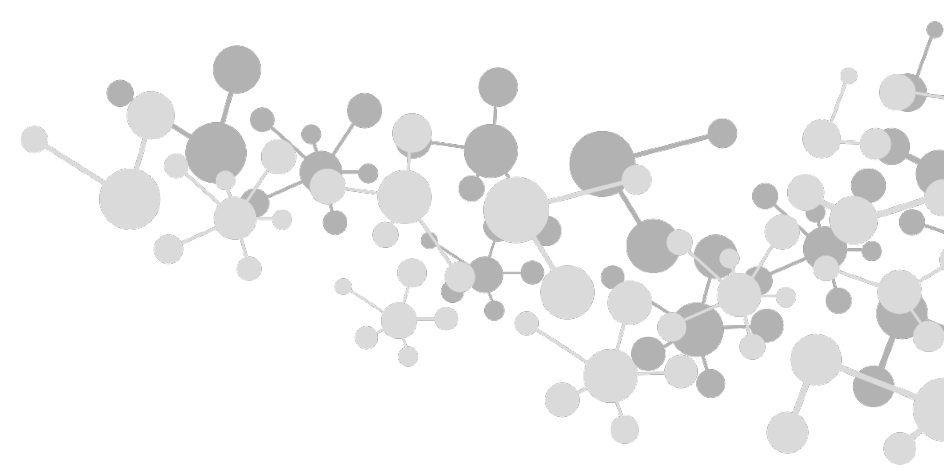




Language resource project: definition

...collect and provide **terminological resources of high quality, with clear licensing status, in specific domains appropriate for sector-specific DSIs, and in languages of interest**

in order to improve the quality and coverage of automated translation offered by the CEF eTranslation service, maintain and advance its improvement over time, based on mature operational and technical solutions implemented during the project, and thus boost the take-up of existing and future CEF eTranslation services



Why terminological resources for automated translation?



Terminology for automated translation

Current state of the art machine translation is based on **AI technologies**

neural machine translation (NMT) uses a large artificial neural network
performs better in most cases than previous paradigm of statistical machine translation (SMT)

while **SMT** does **better for out of domain** content, **NMT** likes things that are **domain-specific**

nevertheless, **terminology matters for both SMT and NMT**

consider **virus** in **medicine** and **computer science**

both systems need **terminology to disambiguate**

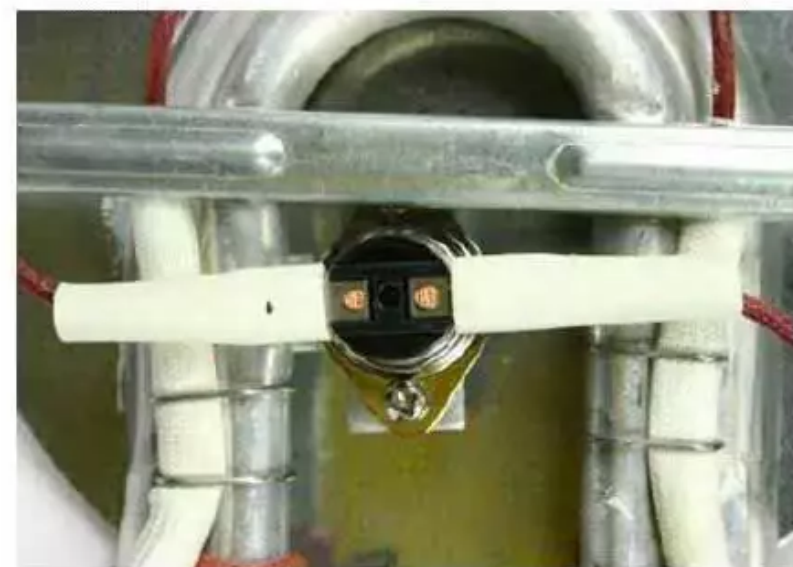
The [redacted] of [redacted]

The [redacted] is simply a [redacted], very similar to the [redacted] of a [redacted] or the [redacted] in an [redacted] that gets [redacted] when you [redacted] through it. In a [redacted] like this, the [redacted] is [redacted] in a [redacted] to make it more [redacted]. The [redacted] has two jobs:

- When you first put the [redacted] in the [redacted], the [redacted] [redacted] it.
- Once the [redacted] is made, the [redacted] keeps the [redacted] warm.

In the picture above, you can see how the [redacted] is [redacted] between the [redacted] and the [redacted]. The [redacted] directly against the [redacted] of the [redacted] and [redacted] makes sure the [redacted] efficiently. You find this [redacted] in all sorts of [redacted], including [redacted] -- pretty much anything that has to [redacted].

The [redacted] turns [redacted] to the [redacted] on and off. To keep the [redacted] from [redacted], there are also [redacted] such as [redacted] and [redacted]. In [redacted], [redacted] if the [redacted] is getting too hot and cut off the [redacted]. Then, when it cools down, they turn the [redacted] back on. By [redacted] like this, they keep the [redacted] at an even [redacted]. [redacted] simply cut the [redacted] if they [redacted] too high a [redacted]. They're there for [redacted] reasons, in the event that the [redacted].



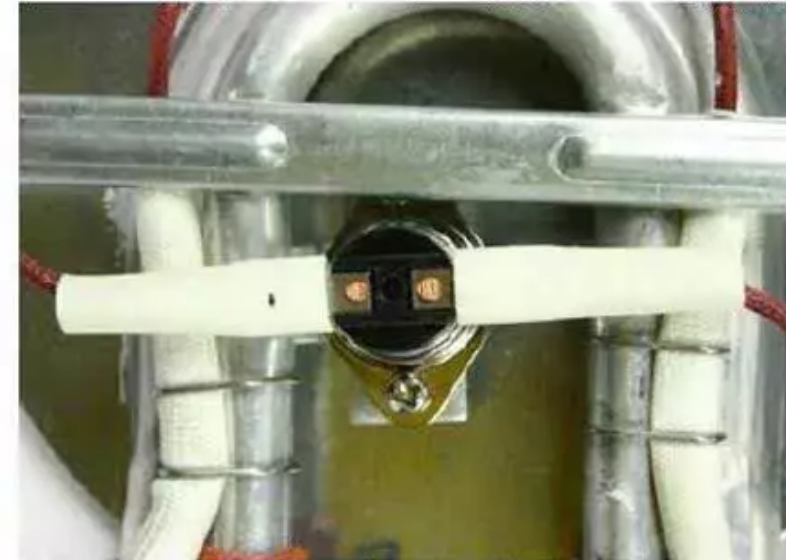
[redacted] Heating Element [redacted] Coffee Makers

[redacted] resistive heating element [redacted] coiled wire, [redacted] filament [redacted] light bulb [redacted] element [redacted] electric toaster [redacted] hot [redacted] run electricity [redacted] resistive element [redacted] coil [redacted] embedded [redacted] plaster [redacted] heating element [redacted].

- [redacted] water [redacted] coffee maker, [redacted] heating element heats [redacted].
- [redacted] coffee [redacted], [redacted] heating element [redacted] coffee [redacted].

[redacted] resistive heating element [redacted] sandwiched [redacted] warming plate [redacted] aluminum water tube. [redacted] resistive heating element presses [redacted] underside [redacted] warming plate, [redacted] white, heat-conductive grease [redacted] heat transfers [redacted] grease [redacted] devices, [redacted] stereo amplifiers, power supplies -- [redacted] dissipate heat.

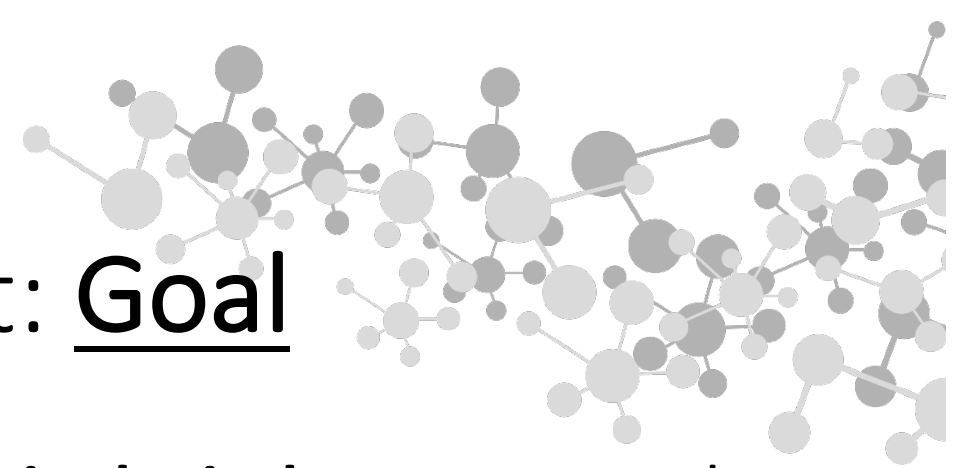
[redacted] coffee maker's switch [redacted] power [redacted] heating element [redacted], [redacted] heating element [redacted] overheating, [redacted] components [redacted] sensors [redacted] fuses, [redacted] coffee makers, sensors detect [redacted] coil [redacted] current, [redacted] cycling on and off [redacted] coil [redacted] current [redacted] temperature. Fuses [redacted] power [redacted] sense [redacted] temperature. [redacted] safety [redacted], [redacted] main sensor fails.



(Source: <http://home.howstuffworks.com/coffee-maker2.htm>)

Non-modular
cemented
zementfrei





eTranslation TermBank project: Goal

eTranslation TermBank aims at providing **new terminological resources** to the language resource pipeline of CEF eTranslation for **three specific domains**

Health

Business legislation

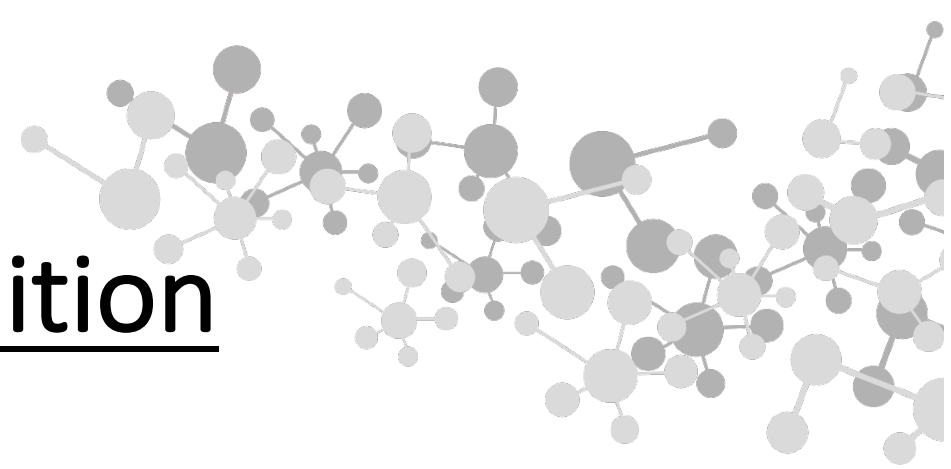
Customer protection

to support the three sector-specific **CEF DSIs**

eHealth

e-Justice

Online Dispute Resolution



Terminological resource: definition

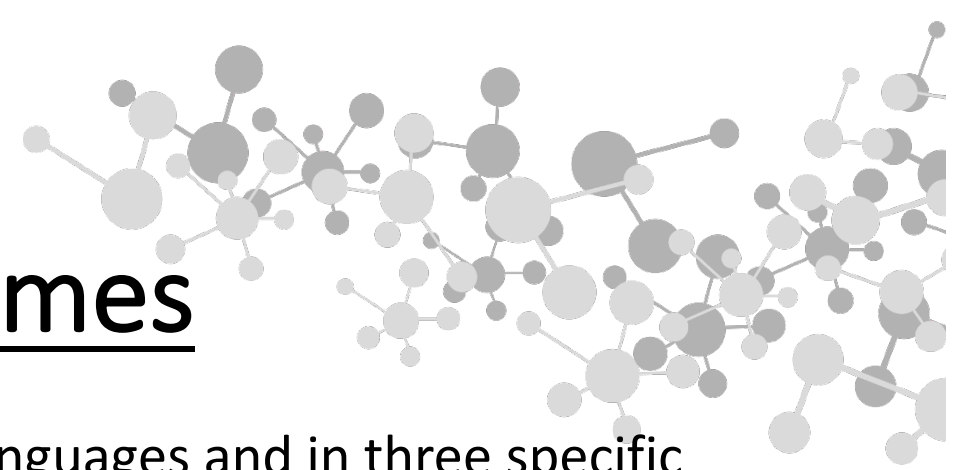
New terminological resources covering languages of interest – **all official languages** of the European Union plus **Norwegian and Icelandic**

a **terminological resource** is a set of **monolingual, bilingual, or multilingual** terminological metadata and data in a machine-readable form,
new terminological resources within the project are terminological resources,
which are **not available via the ELRC-SHARE language resource repository** at
the time and are **appropriate for sector-specific CEF DSIs** covered by the
project



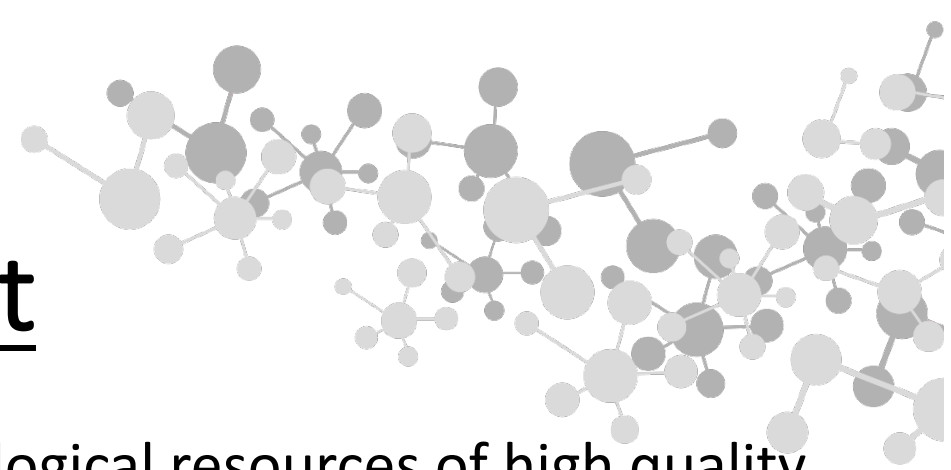
Fatal mistake





eTranslation TermBank: Outcomes

- A minimum of **150 terminological resources** for all CEF languages and in three specific domains for CEF eTranslation
- Stakeholder network “**Terminology for Europe**” for sustainable cooperation with CEF eTranslation
- **Pipeline of terminological resources flow** to CEF eTranslation
- **Case study and terminology requirement analysis** of trans-European and national CEF Digital Service Infrastructures
- The **directory of entities dealing with terminology** on European, regional, and national levels and **directory of existing terminological resources** in the three specific domains for CEF eTranslation
- The **methodology for preparation and application of terminological resources** for CEF eTranslation:
 - selection, harmonisation, and consolidation
 - methodology to integrate terminological resources into automated translation



eTranslation Termbank: Impact

...to **stimulate the collection and provision** of terminological resources of high quality and with clear licensing status in specific domains and languages of interest to sector-specific CEF DSIs

in order to **improve the quality and coverage of automated translation offered by CEF eTranslation**, maintain and advance its **improvement over time** based, on **mature operational and technical solutions** implemented during the proposed action, and boost the take-up of **existing and future CEF eTranslation services**

...to raise awareness about the importance of terminological resources to **improve the quality and coverage of automated translation offered by CEF eTranslation**



Co-financed by the European Union

Connecting Europe Facility

Terminology for Europe

Workshop

July 13, Vienna

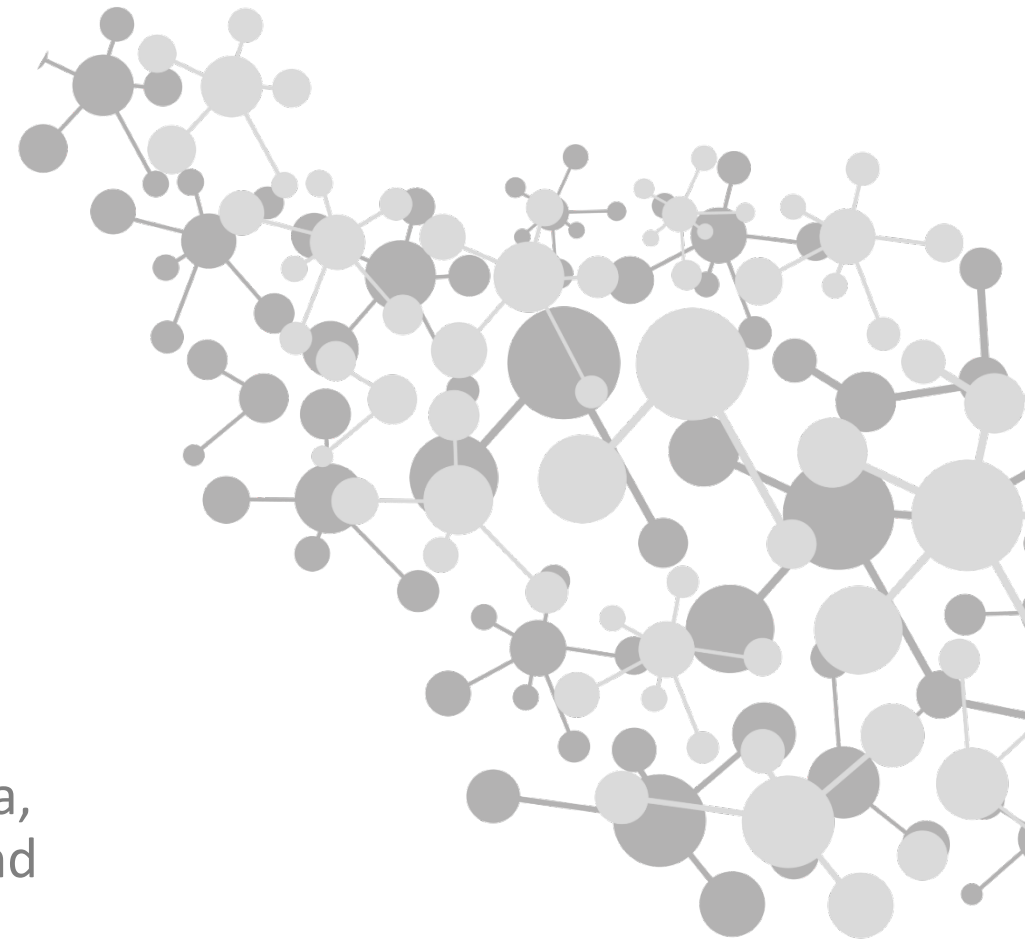
National use cases on terminology management and application in public administration services in Austria, Germany, Sweden, Iceland, Spain, Portugal, Latvia, and other CEF countries

Project use cases on terminology collection for eHealth, e-Justice, and ODR – CEF DSIs covered by the project

Project: eTranslation TermBank

Agreement number: INEA/CEF/ICT/A2016/1331795

Action No: 2016-EU-IA-0122



eTranslation TermBank

Stimulates the collection and provision of **terminological resources**
in sector-specific domains and languages of interest
to sector-specific digital public services



Tilde, Latvia (coordinator)

University of Copenhagen, Denmark

Árni Magnússon Institute for Icelandic Studies, Iceland

Institute of the Estonian Language, Estonia

Jožef Stefan Institute, Slovenia

International Network for Terminology, Austria

Institute for Language and Folklore, Sweden

Swedish Centre for Terminology, Sweden

Institute of the Lithuanian language, Lithuania



Co-financed by the European Union

Connecting Europe Facility

eTranslation TermBank

Stimulates the collection and provision of **terminological resources**
in sector-specific domains and languages of interest
to sector-specific digital public services



Andrejs Vasiljevs andrejs@tilde.com

Tatjana Gornostaja tatjana.gornostaja@tilde.com



Co-financed by the European Union

Connecting Europe Facility

eTranslation TermBank

Stimulates the collection and provision of **terminological resources** in sector-specific domains and languages of interest to sector-specific digital public services

Sector-specific domains and digital public services in focus:

- Health to support CEF DSI **eHealth**
- Business legislation to support CEF DSI **e-Justice**
- Customer protection to support CEF DSI **ODR**

Outcomes:

- 150 terminological resources for **official languages of the EU** plus **Norwegian and Icelandic**
- "Terminology for Europe" network for **sustainable cooperation**
- Pipeline of **terminological resources flow to CEF.AT**
- Directory of entities dealing with **terminology on national level**
- Directory of existing **terminological resources in the three sector-specific domains** for CEF.AT
- Methodology for **preparation and application of terminological resources** for CEF.AT

Impact:

- Improve the **quality and coverage of CEF.AT**

Partners:

Tilde, Latvia (coordinator)
University of Copenhagen, Denmark
Árni Magnússon Institute for Icelandic Studies, Iceland
Institute of the Estonian Language, Estonia
Jožef Stefan Institute, Slovenia
International Network for Terminology, Austria
Institute for Language and Folklore, Sweden
Swedish Centre for Terminology, Sweden
Institute of the Lithuanian language, Lithuania



Institut "Jožef Stefan", Ljubljana, Slovenija



Co-financed by the European Union
Connecting Europe Facility

Project eTranslation TermBank is co-financed
by the European Union's Connecting Europe Facility