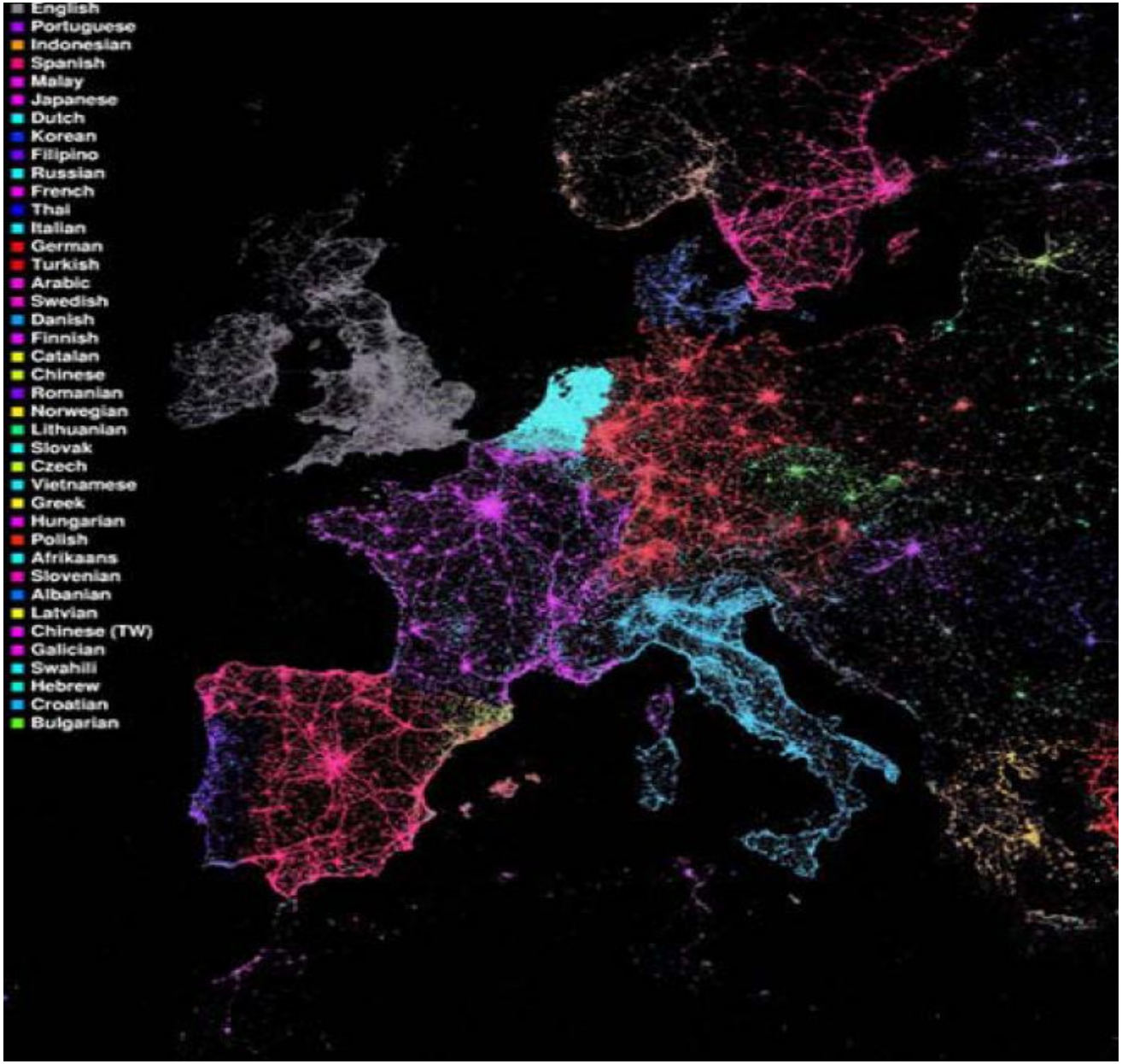
European Language Resource Coordination
Connecting Europe Facility

European Language Resource Coordination (ELRC)

Stelios Piperidis
Athena RC/ILSP


English
Portuguese
Indonesian
Spanish
Malay
Japanese
Dutch
Korean
Filipino
Russian
French
Thai
Italian
German
Turkish
Arabic
Swedish
Danish
Finnish
Catalan
Chinese
Romanian
Norwegian
Lithuanian
Slovak
Czech
Vietnamese
Greek
Hungarian
Polish
Afrikaans
Slovenian
Albanian
Latvian
Chinese (TW)
Galician
Swahili
Hebrew
Croatian
Bulgarian

European Language Resource Coordination
Connecting Europe Facility

Context

- Nobody speaks all languages
- Make sure languages do not create silos or invisible borders
- Restricting mobility of people, information, ideas, commerce

- No language discrimination but equal opportunities
- Languages as an asset, not a problem

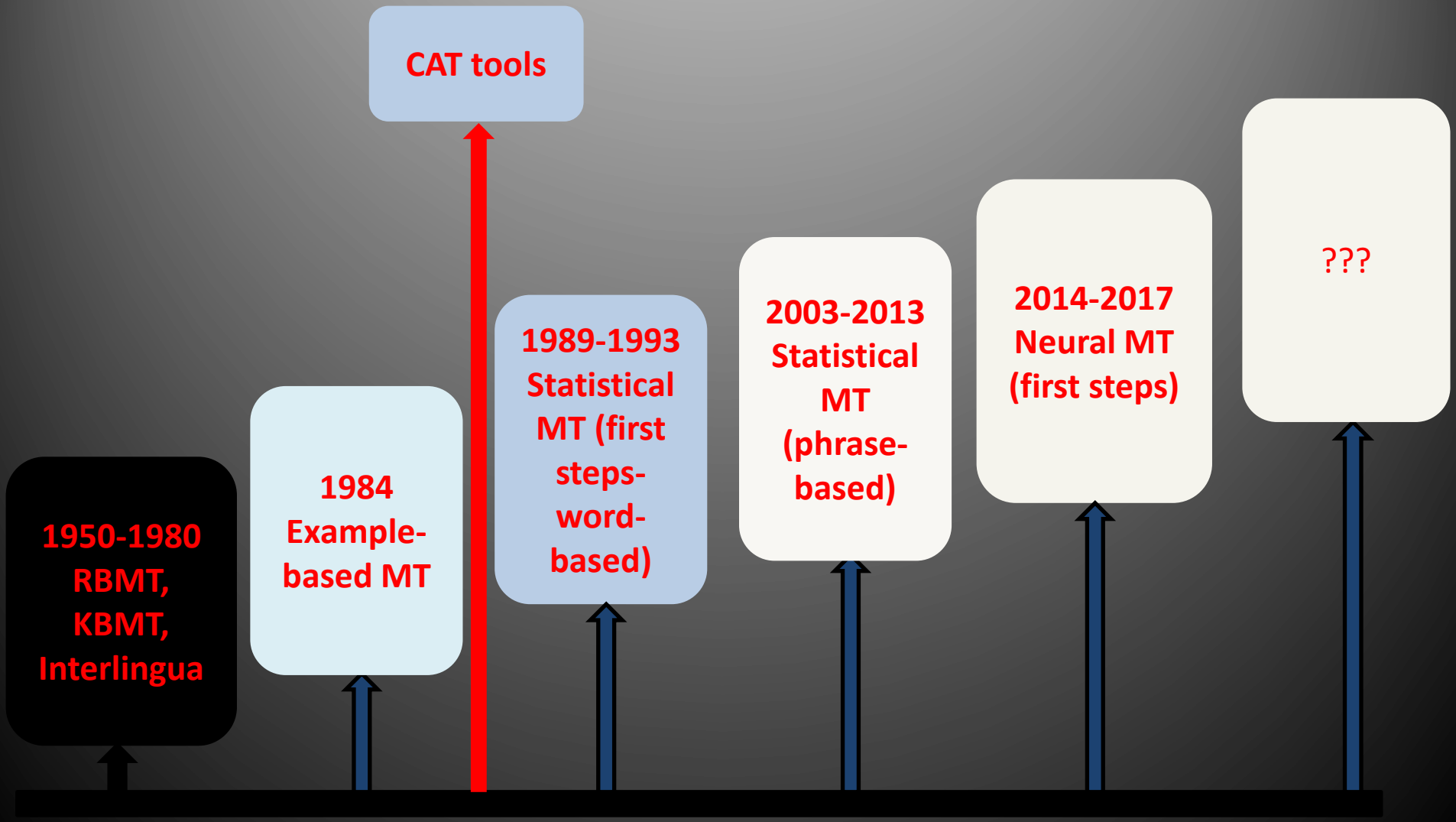
European Language Resource Coordination
Connecting Europe Facility

Context

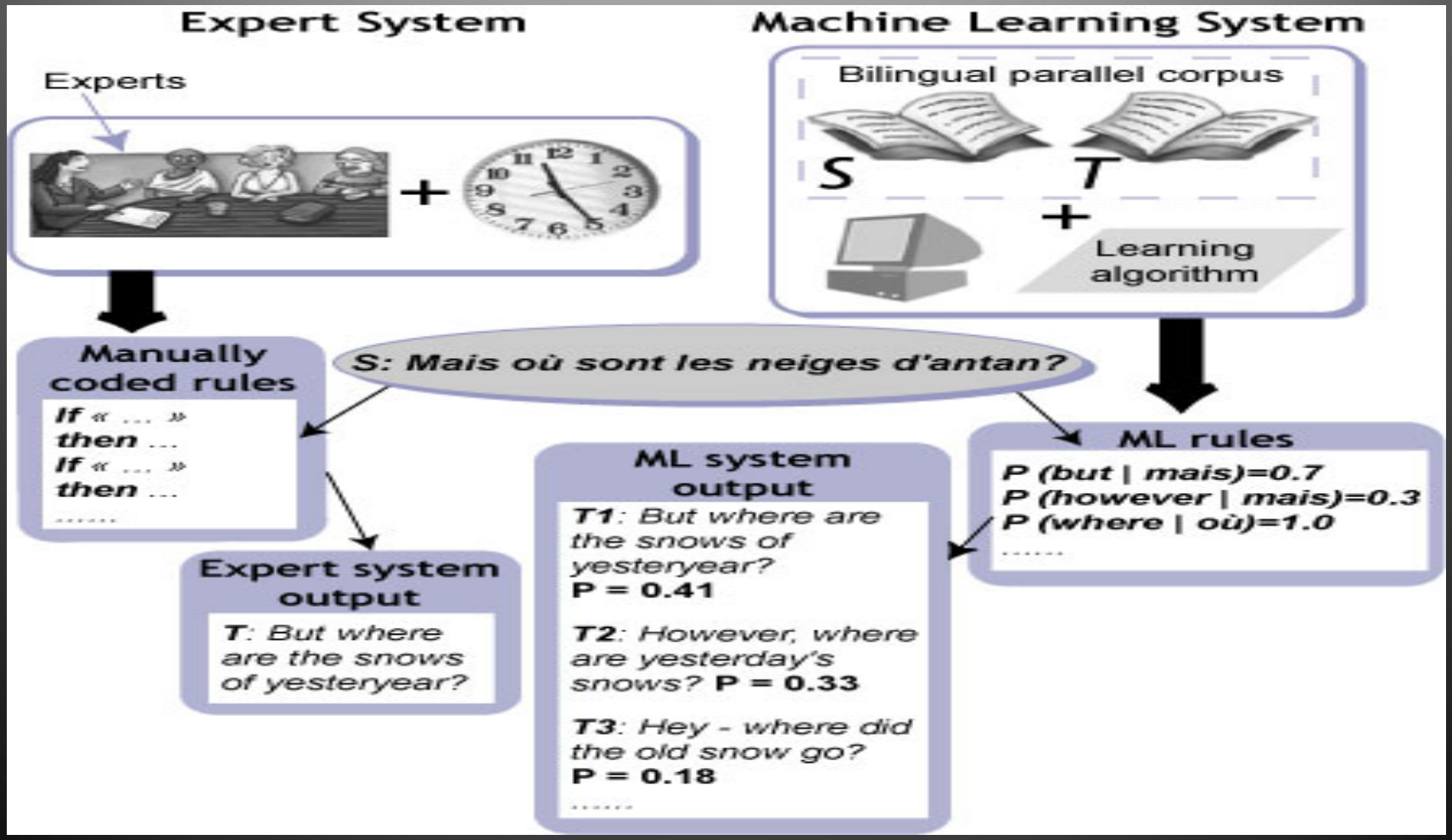
- How?
- Translation!
- Extremely skilled profession

- A lot to translate … in fact much too much
- Technology support
- Human translation supported by technology = Automated Translation (AT)

Machine Translation Evolution



Methodological evolution



Rationalism Vs Empiricism

Rationalism vs. Empiricism

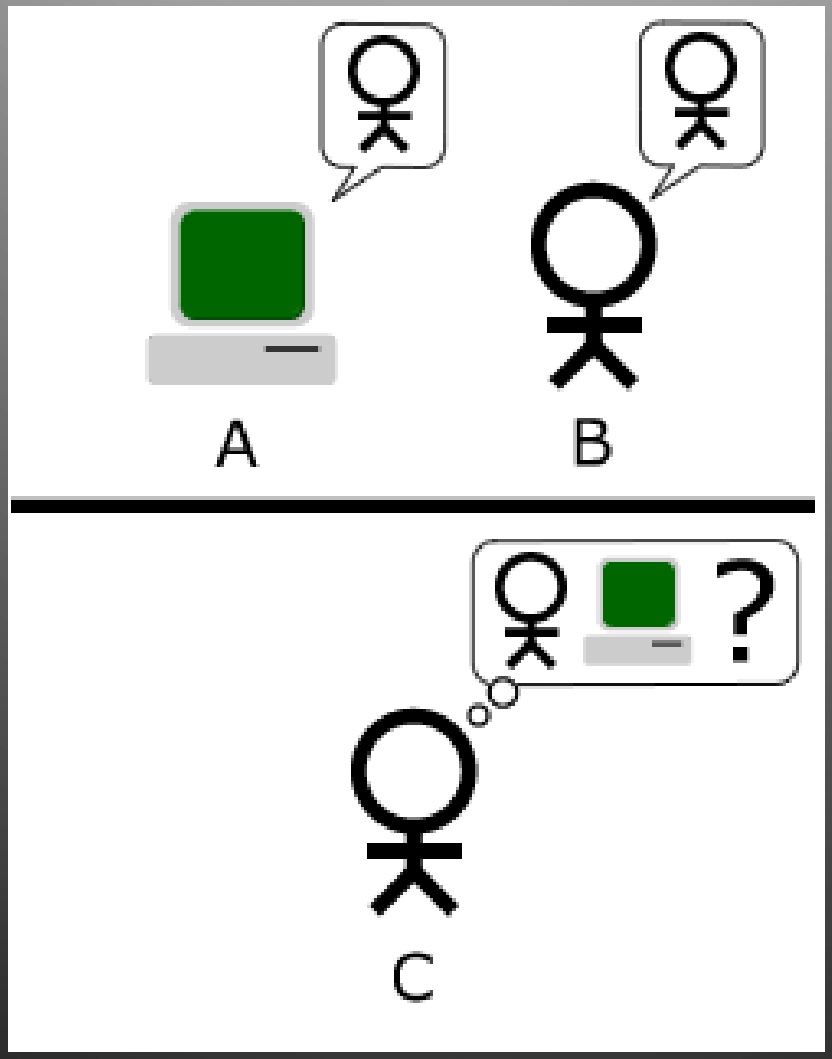
Rationalism

1. Emphasizes on *theory*
2. Assumes an "*innate language faculty*"
3. Aims at discovering the *language of the human mind* (linguistic competence)
4. Assigns *categories* to language units
5. Major advocates: Chomsky, Minsky

Empiricism

1. Emphasizes on *data*
2. Assumes all knowledge gathered only via *senses*
3. Aims at analysing *language as it actually occurs* (linguistic performance)
4. Assigns *probabilities* to language units
5. Major advocates: Shannon, Norvig

The Turing Test



European Language Resource Coordination
Connecting Europe Facility

Data

- Modern MT is all about data
- Learns to translate from human translations = data

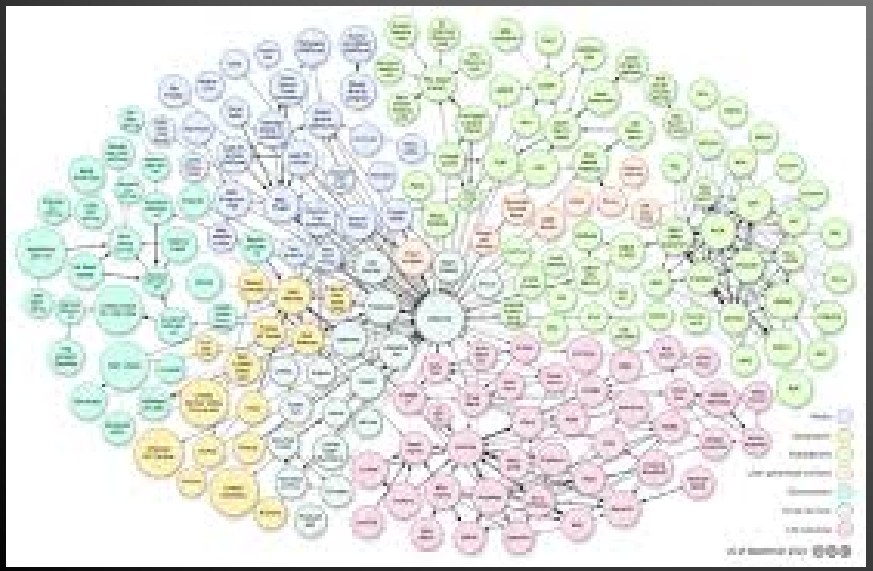
- To best support your needs MT needs the best data

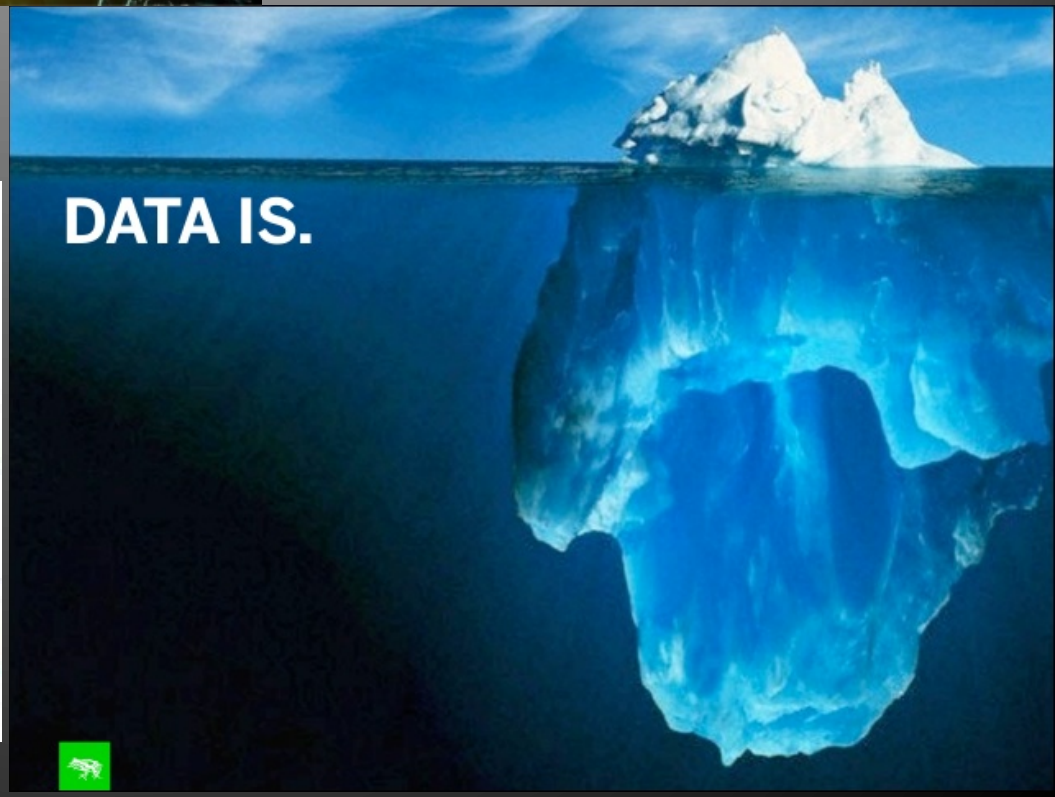
- Your data = the best data for MT for you

- Without the right data it won't work!

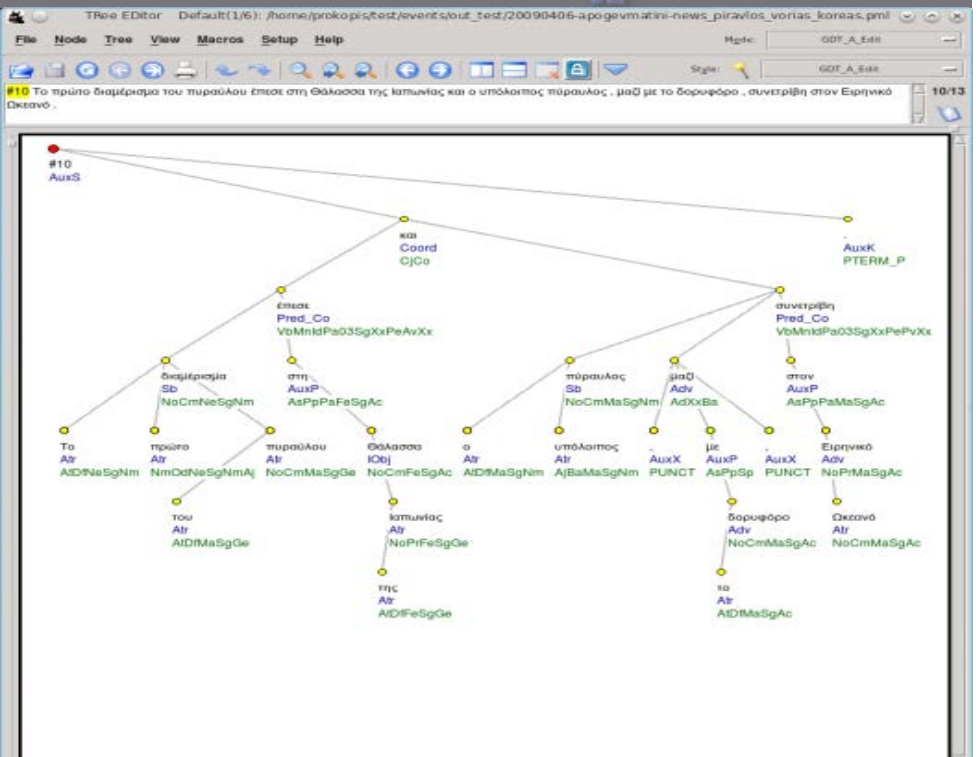
DATA is

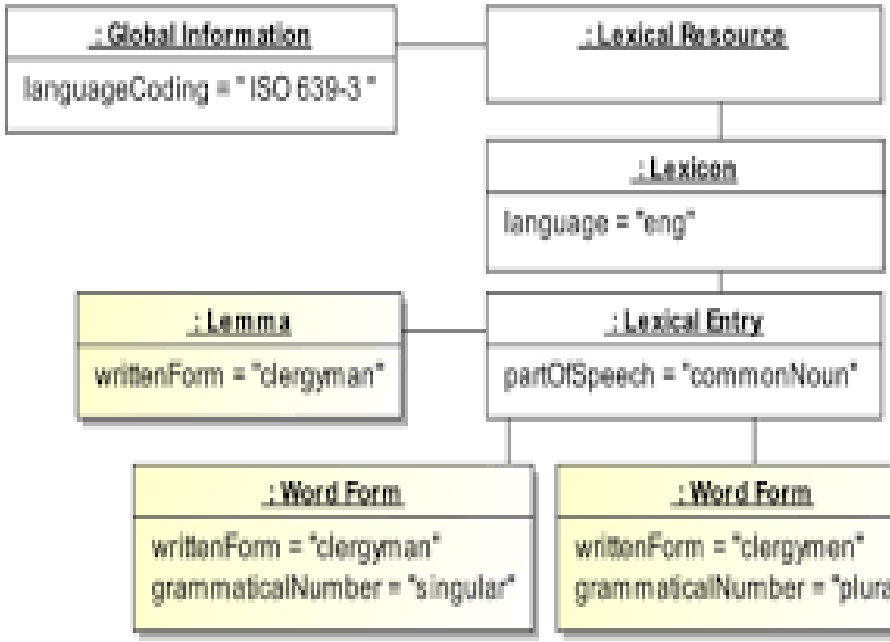
Everything





LANGUAGE
RESOURCES
FRANÇAIS
DEUTSCH
HRVATSKI
ITALIANO
ENGLISH
PORTUGUÊS
ESPAÑOL
TIẾNG VIỆT
TÜRKÇE
Русский
Български
Эрзянь
Молдовеняскэ
Словѣньскъ
Ming-dĕng-ngū
العربية
עברית
हिन्दी
ગુજરાતી
ಕನ್ನಡ
मराठी
தமிழ்
ภาษา
ქართული
Հայերեն
ئۇيغۇرچە
Моксень
Зэбек
Cwy


TRee EDitor Default(1/6): /home/prokopis/test/events/out_test/20090406-apogevmatini-news_piravlos_voreas_koreas.pml
File Node Tree View Macros Setup Help
#10 Το πρώτο διαμέρισμα του πυραύλου έπεσε στη Θάλασσα της Ιαπωνίας και ο υπόλοιπος πύραυλος , μαζί με το δορυφόρο , συνετρίβη στον Ειρηνικό Ωκεανό .
10/13
#10 AuxS
και Coord CjCo
. AuxK PTERM_P
έπεσε Pred_Co VbMnIdPa03SgXxPeAvXx
συνετρίβη Pred_Co VbMnIdPa03SgXxPePvXx
διαμέρισμα Sb NoCmNeSgNm
στη AuxP AsPpPaFeSgAc
πύραυλος Sb NoCmMaSgNm
μαζί Adv AdXxBa
στον AuxP AsPpPaMaSgAc
Το Atr AtDfNeSgNm
πρώτο Atr NmOdNeSgNmAj
πυραύλου Atr NoCmMaSgGe
Θάλασσα IObj NoCmFeSgAc
ο Atr AtDfMaSgNm
υπόλοιπος Atr AjBaMaSgNm
. AuxX PUNCT
με AuxP AsPpSp
. AuxX PUNCT
Ειρηνικό Adv NoPrMaSgAc
του Atr AtDfMaSgGe
Ιαπωνίας Atr NoPrFeSgGe
δορυφόρο Adv NoCmMaSgAc
Ωκεανό Atr NoCmMaSgAc
της Atr AtDfFeSgGe
το Atr AtDfMaSgAc

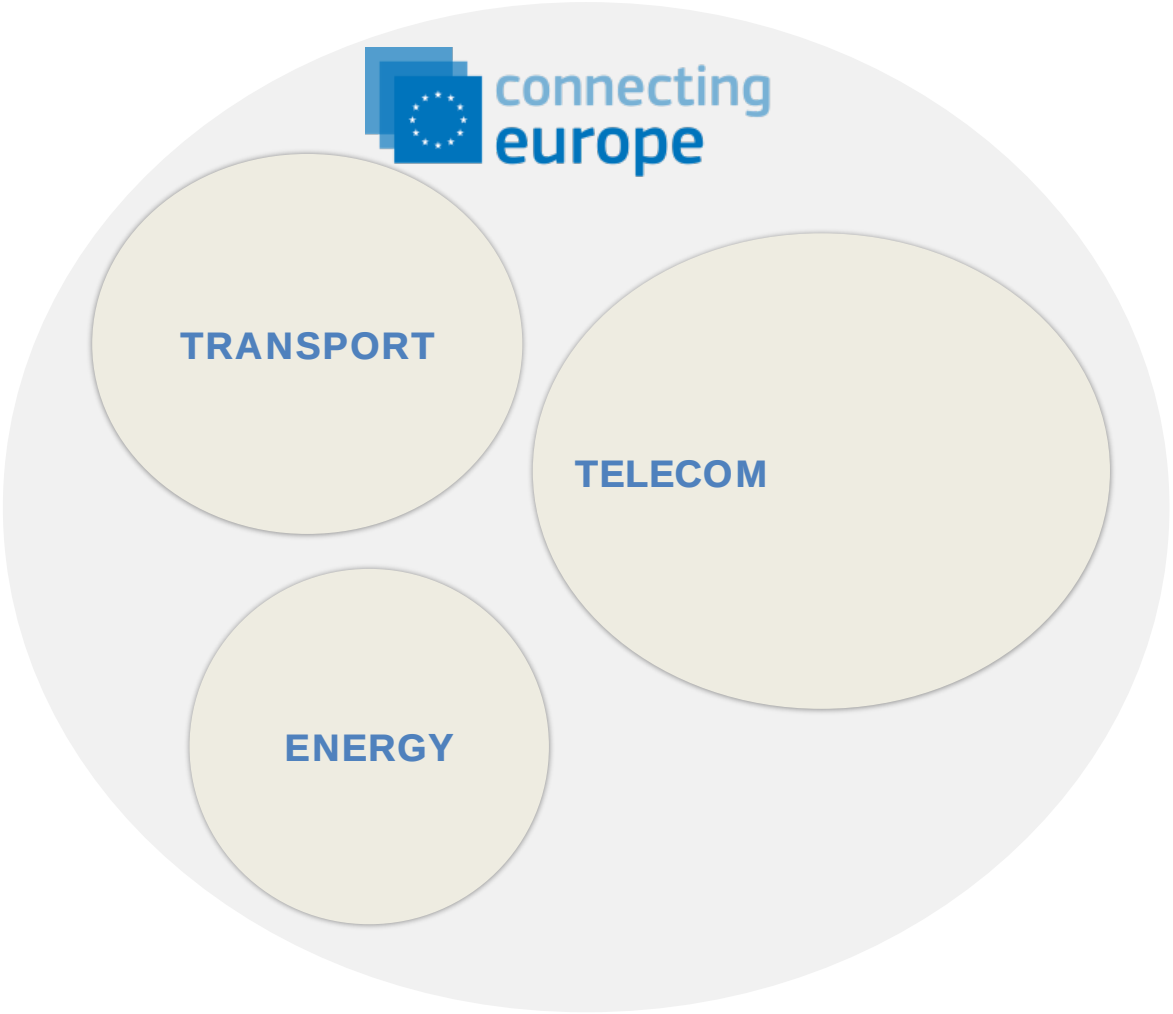

: Global Information
languageCoding = " ISO 639-3 "
: Lexical Resource
: Lexicon
language = "eng"
: Lemma
writtenForm = "clergyman"
: Lexical Entry
partOfSpeech = "commonNoun"
: Word Form
writtenForm = "clergyman"
grammaticalNumber = "singular"
: Word Form
writtenForm = "clergymen"
grammaticalNumber = "plural"

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

The ELRC initiative

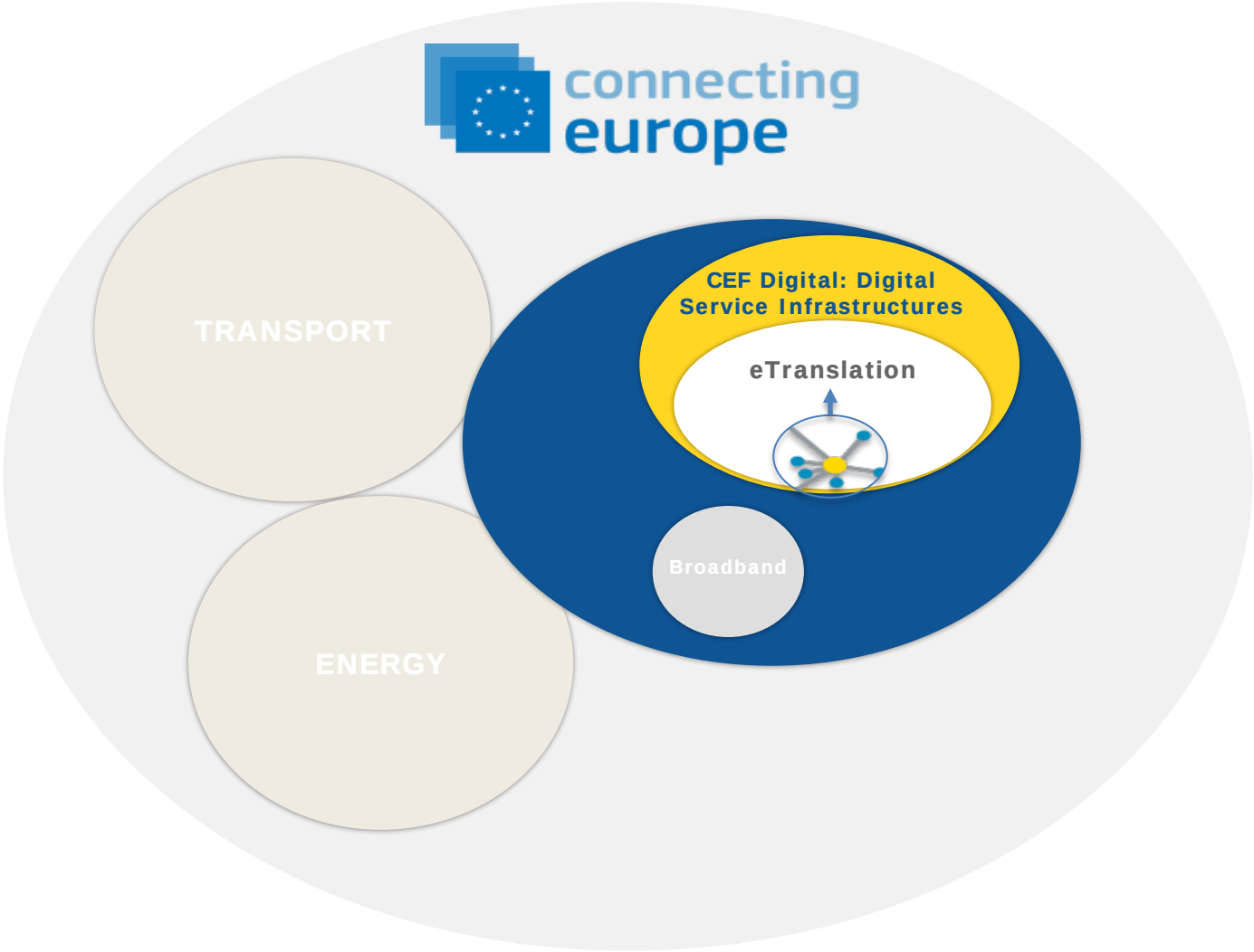
European Language Resource Coordination
Connecting Europe Facility

The ELRC context



European Language Resource Coordination
Connecting Europe Facility

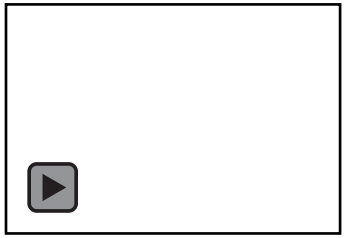
The ELRC context

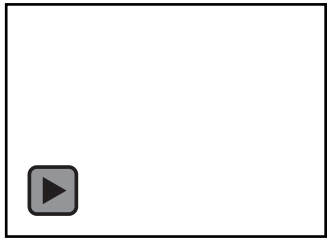


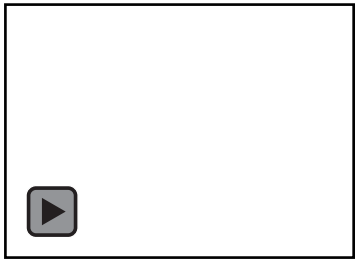
Heads & organisations behind the ELRC

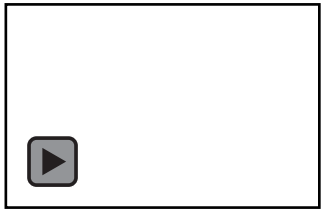
- The ELRC Consortium:

German Research Center for Artificial Intelligence









Heads & organisations behind the ELRC

- 30 ELRC Technological National Anchor Points
- 30 ELRC Public Services National Anchor Points

European Language Resource Coordination
Connecting Europe Facility

What does ELRC do?

Collect	• Language resources
Identify	• Needs of public services
Engage	• With the public sector in the identification of LR
Help	• With any technical or legal issues
Act	• Observatory for language resources across Europe

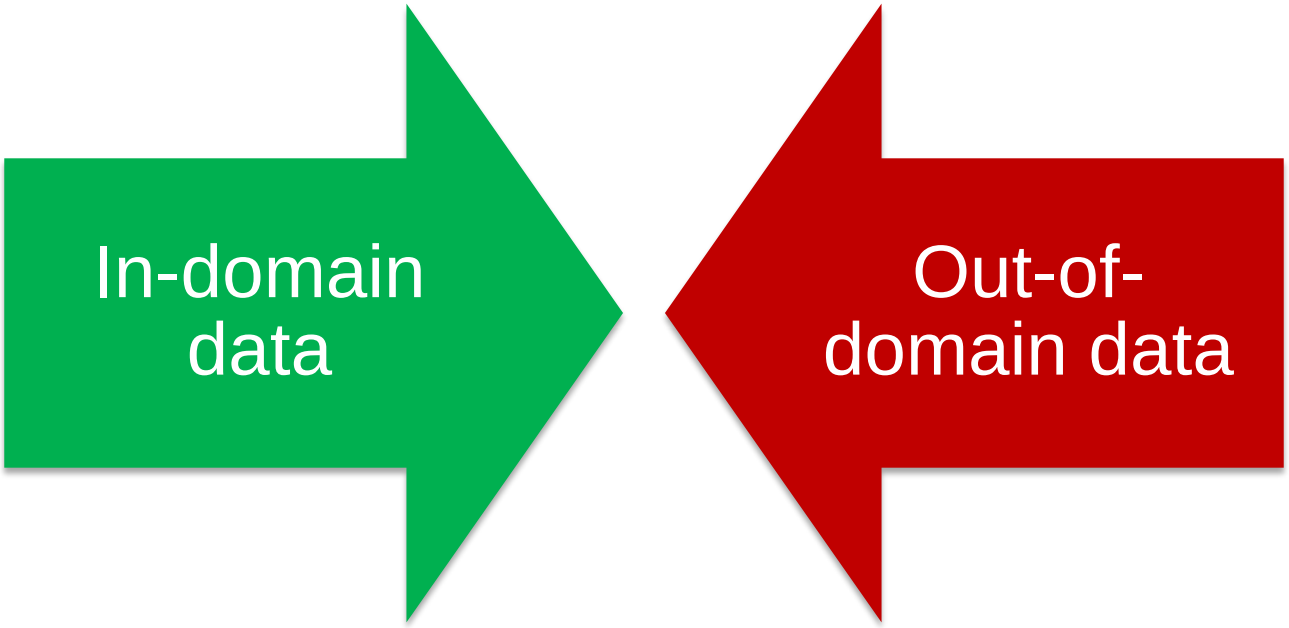
European Language Resource Coordination
Connecting Europe Facility

Why ELRC…?

- To facilitate multilingual communication and exchange of information in key public service scenarios
- How? With the right language resources!
- Examples of public services scenarios to which eTranslation may be applied:
  - Consumer Rights (ODR),
  - Social Security (EESSI),
  - Justice (eJustice),
  - Commerce (BRIS),
  - Open Data (EDP),
  - …

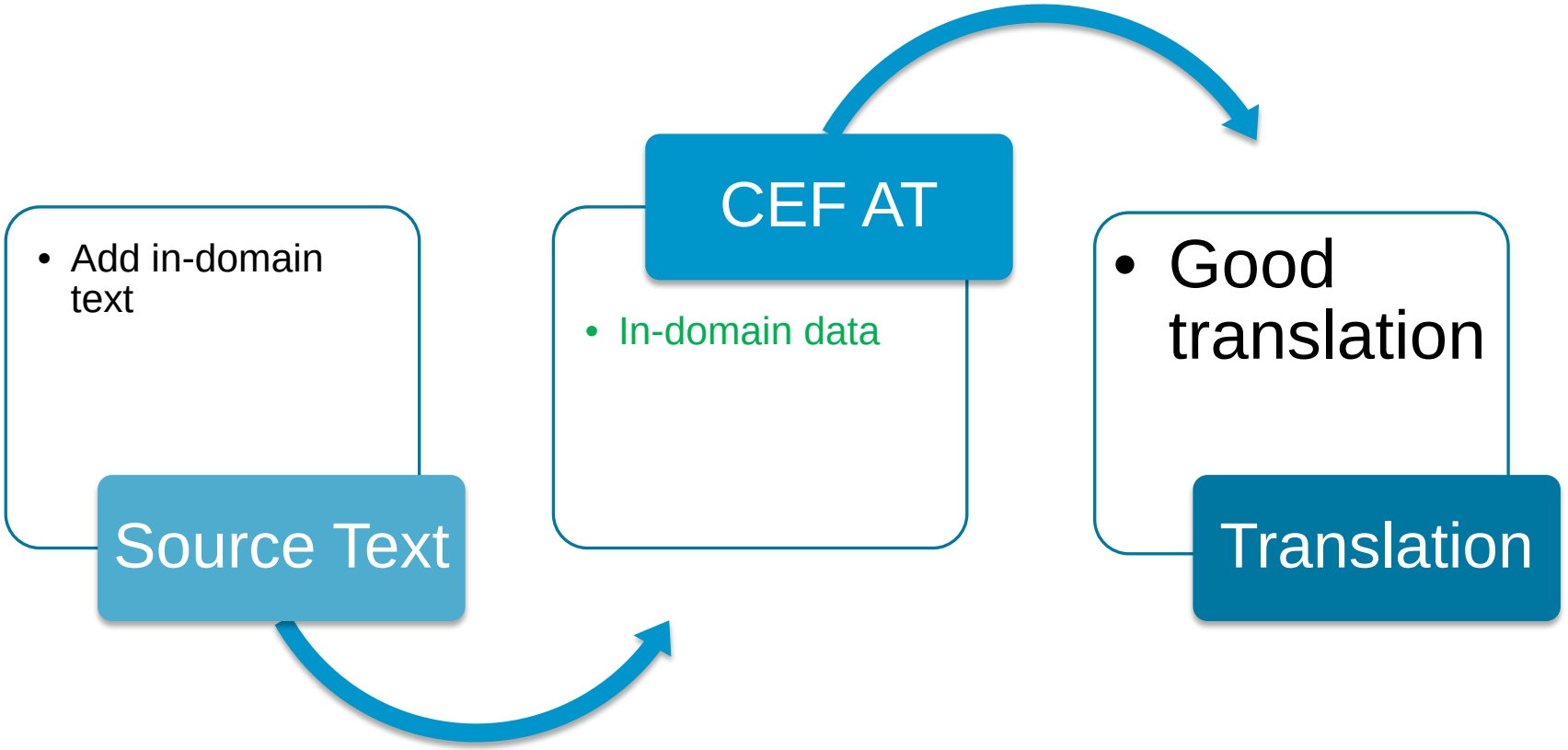
European Language Resource Coordination
Connecting Europe Facility

How to make it work…?



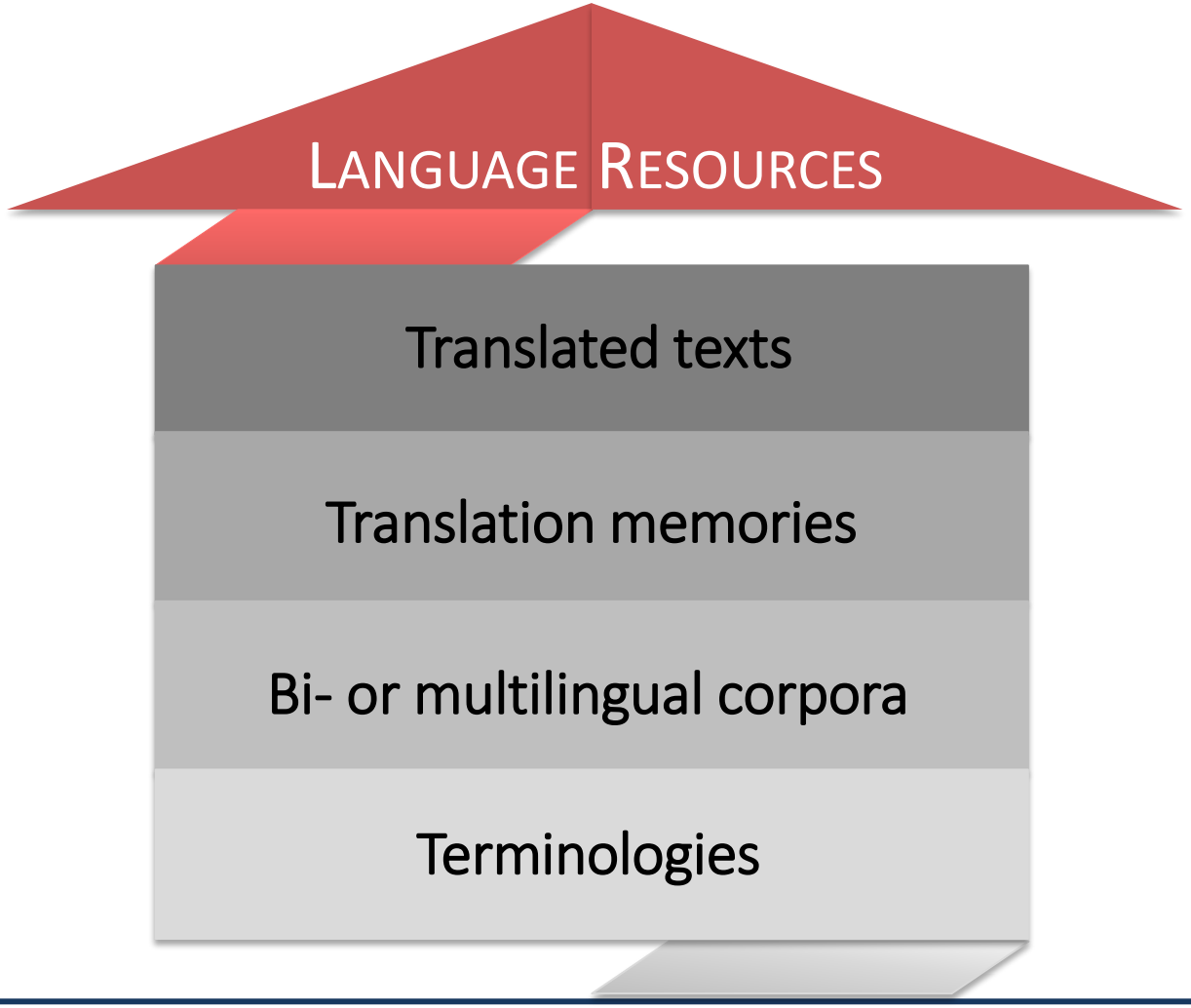
➔ **language resources of the public sector make a difference!**

How to make it work…?



European Language Resource Coordination
Connecting Europe Facility

Language Resources of interest



European Language Resource Coordination
Connecting Europe Facility

What has been achieved so far…?

Collection of 225 language resources overall

More than 2 billion words in all EU official languages, Norwegian and Icelandic

More than 91 language resources to be used by you!

Over 450.000 terms

More than 2 million translation units

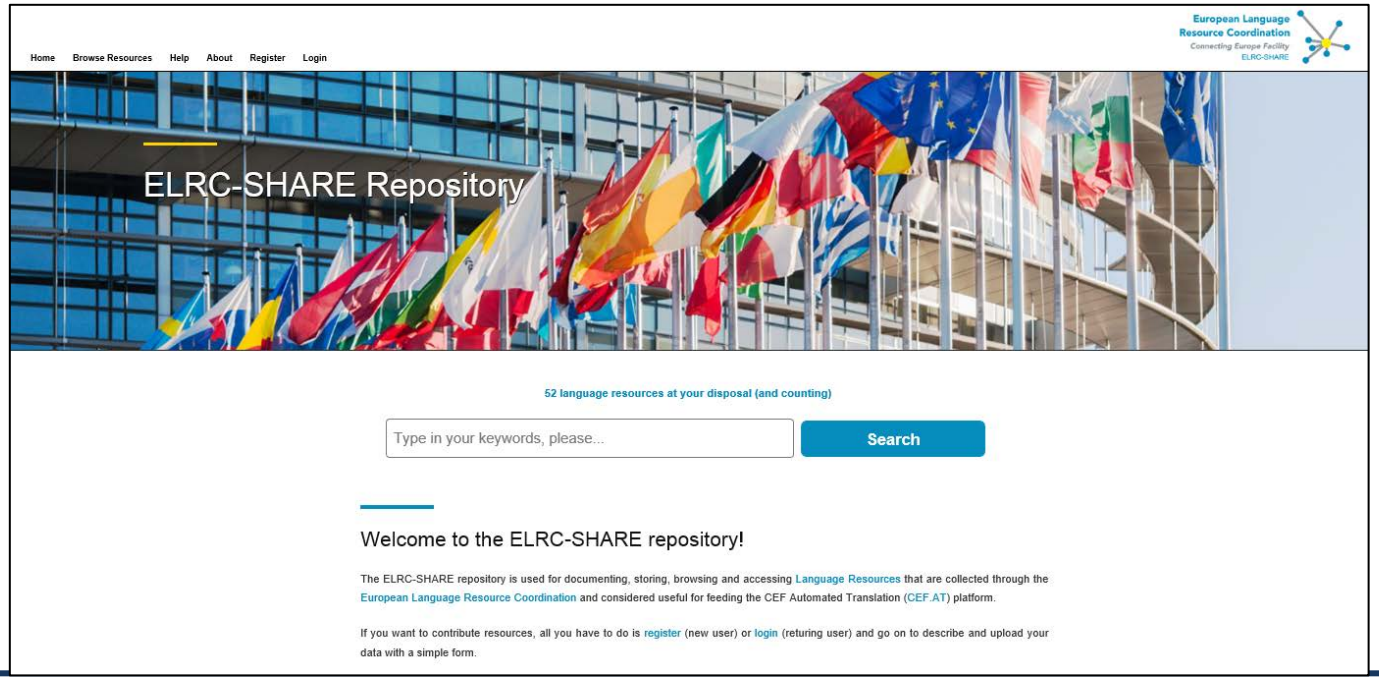
European Language Resource Coordination
Connecting Europe Facility

Legal Framework

- Public Sector Information (PSI) Directive 2013/37/EU
  - Amending Directive 2003/98/EC
  - Emphasis on open data
  - Clear cost rules
  - Including Cultural Institutions (Museums, Libraries and Archives) within the scope of the Directive
  - Introduce a clear regime for exclusive agreements
  - Emphasis on standard and machine readable licences

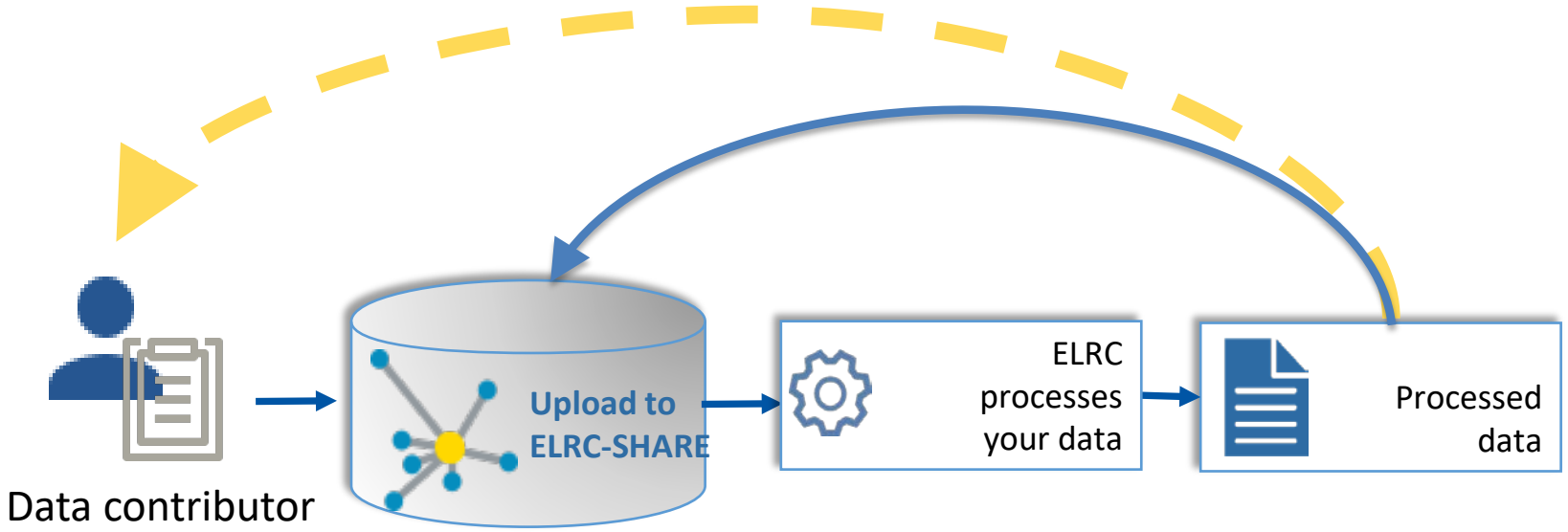
ELRC for you…

- **The ELRC-SHARE Repository**
  - Access to, sharing and contribution of language resources
  - Access to tools and services catalogue (upcoming)
  - Visit https://elrc-share.eu



European Language Resource Coordination
Connecting Europe Facility

What happens to your data?



Data processing before delivery to EC

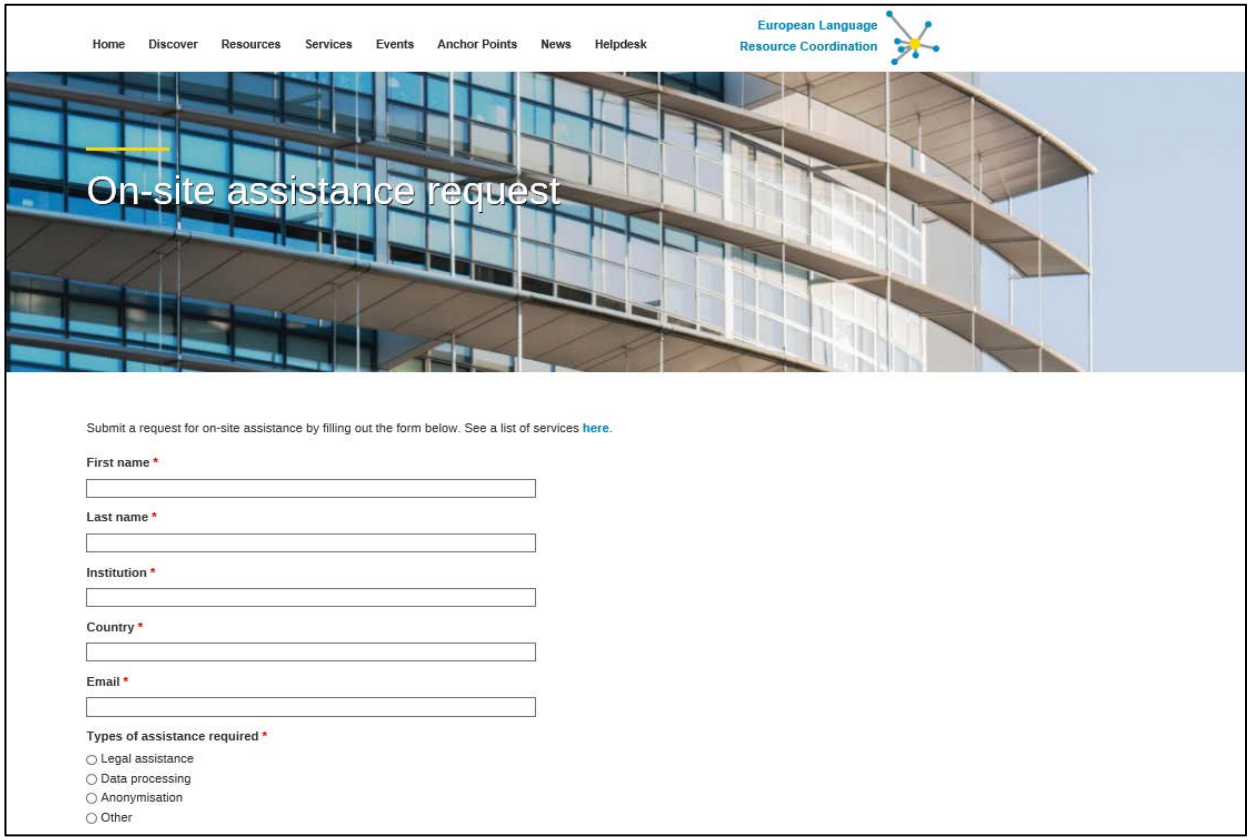
- All datasets are processed to result in tmx/tbx/txt files
- Data will indicatively undergo the following processing:
  - cleaning
  - format conversion
  - sentence alignment
  - metadata completion

ELRC for you…

- **ELRC Technical and Legal Helpdesk**
  - Continuous support for potential data donors (phone, email, …)
  - Accessible online via http://www.lr-coordination.eu/helpdesk

ELRC for you…

- **ELRC On-site assistance**
  - Submit request online at http://www.lr-coordination.eu/services



European Language Resource Coordination
Connecting Europe Facility

Thank you for your attention!

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