EU and multilingualism &

How can public services benefit from CEF Automated Translation

Saila Rinne, European Commission
DG CONNECT, Data Value Chain Unit
Oslo, 8 June 2016

Outline

- EU policy context
- Need for multilingual pan-European public online services
- How can CEF countries contribute and benefit
- Role of machine translation
- How to achieve a multilingual enabler for the Digital Service Infrastructures of Connecting Europe Facility programme

EU and multilingualism

- The equality of the EU official languages is enshrined in the European legal basis
- EU is committed to supporting multilingualism
  - support for language technologies in R&I programmes
  - support for the use of Machine Translation in public administrations (MT@EC)
  - deployment of mature language technologies through the Connecting Europe Facility (CEF) Programme

Digital Single Market strategy

- A Digital Single Market can create up to €340 billion in additional growth, hundreds of thousands of new jobs, and a vibrant knowledge-based society
- European Digital Market is multilingual
  - **90% of European consumers prefer to browse websites in their own language;**
  - **82% of the 4000+ web shops are mono-lingual;**
  - **42% of multilingual internet users never purchase online products and services in languages other than their own.**
- Language barriers affect private and public services
- Digital public sector is part of the DSM strategy

Connecting Europe Facility – CEF

- established by REGULATION (EU) No 1316/2013
- **CEF Digital** = pan-European Digital Service Infrastructures (DSIs), for example, cross-border
  - **eJustice**
  - **eProcurement**
  - **eHealth**
  - **Europeana**
  - **Open Data Portal**
- CEF is a *deployment* programme – using mature technologies

Multilingual challenge of pan-European public services

- The European Digital Single Market is **multilingual**
- Pan-European public services address the whole Europe
- No lingua franca among public administrations in Europe
- 90% of EU web users prefer to use their own language in online services
- **Human translation** is **too expensive** and **too slow** with the intended text volumes
- Human translation is **not a solution** in all use scenarios
- Available online translators have **gaps in language coverage** and are **not secure**

→ multilingual support is needed

The role of CEF countries

- CEF.AT helps to create accessible and multilingual **pan-European** online public services
- Technology underlying CEF.AT needs linguistic data (**language resources**) in EU languages
- CEF countries should **take ownership** of their own language and make sure their language is adequately supported in CEF.AT
- CEF countries' involvement and help are essential for CEF.AT efforts to succeed!
- **European Language Resource Coordination** (ELRC) is a project, funded through CEF Digital, that facilitates this process

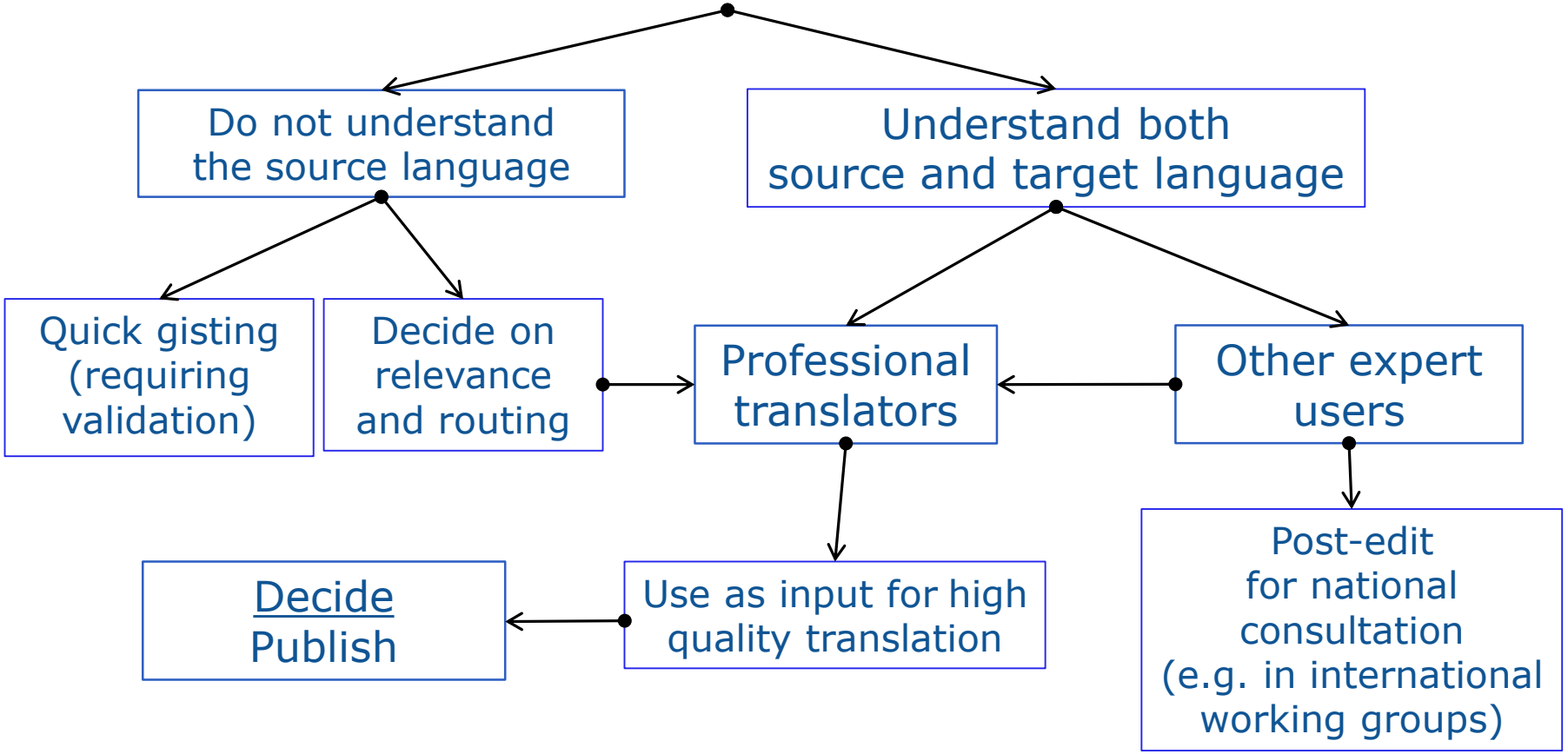
How can Machine Translation help?

MT is the only viable solution for:

- quick and cheap access to information in foreign languages.
- understanding information received in a foreign language that otherwise could not be used or would require substantial time and costs to translate.
- making multilingual use of websites possible
- facilitating cross-lingual information search and analytics.

That is why machine translation (MT) is a critically important technology for multilingual Europe

Machine Translation users



MT@EC: already there for you!

Released : 26 June 2013

Languages: All directions between
24 EU official languages + Norwegian and Icelandic

Technology: Statistical machine translation
using open source software Moses co-funded by EU Framework Programmes for research and innovation

Development by DGT between 2010-2013
co-funded by the ISA programme (action 2.8)
http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm

MT@EC description

Delivery: - web user interface (human to machine)
- web services (machine to machine)

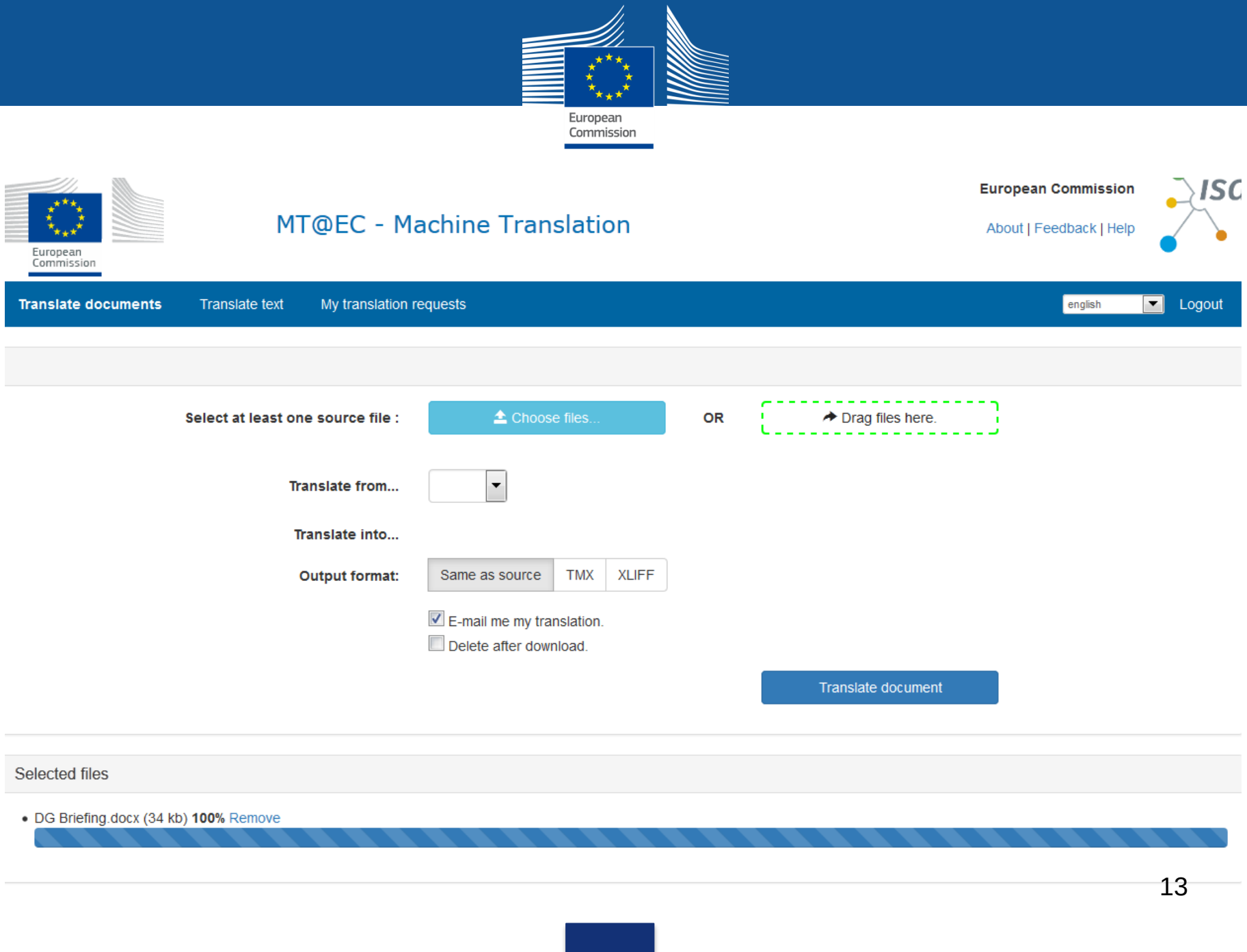
Special features:

- **User interface** in 24 languages
- Source document format/formatting **maintained** *[not for pdf]*
- Specific output **formats for translation:** tmx and xliff
- Translation can also be returned by **email**
- Can translate **multiple** documents to **multiple** languages
- Indication of **quality** for language pairs (using BLEU Scores)
- **Feedback** mechanism (using EU Survey)

MT@EC security

- Secure hosting in the **EC data centre**
- Access through **ECAS** (EC Authentication Service)
- Secure document transfers :
  - over **sTESTA***, a very secure private network between public administrations in the EU, separate from the internet
  - over the **internet** (through a secure **https** connection)

** You can check if your organisation has access to sTESTA on: https://portal.testa.eu/jetspeed/portal/homepage/about.psml.*


European Commission
European Commission
MT@EC - Machine Translation
European Commission
About | Feedback | Help
Translate documents
Translate text
My translation requests
english
Logout
Select at least one source file :
Choose files...
OR
Drag files here.
Translate from...
Translate into...
Output format:
Same as source
TMX
XLIFF
E-mail me my translation.
Delete after download.
Translate document
Selected files
DG Briefing.docx (34 kb) 100% Remove

MT@EC - Machine Translation

European Commission

About | Feedback | Help

Translate documents | **Translate text** | My translation requests

english | Logout

Text to translate :

Good morning to everyone!

Maximum length : 4000 characters.
Remaining : 3975 characters.

Clear

Translate from...

EN

Translate into...

BG	CS	DA	DE
EL	ES	ET	FI
FR	GA	HR	HU
IT	LT	LV	MT
NB	NL	PL	PT
RO	SK	SL	SV

E-mail me my translation.
Delete after download.

Translate text

Translated text :

Buongiorno a tutti!

Select all

Who can use MT@EC today ?

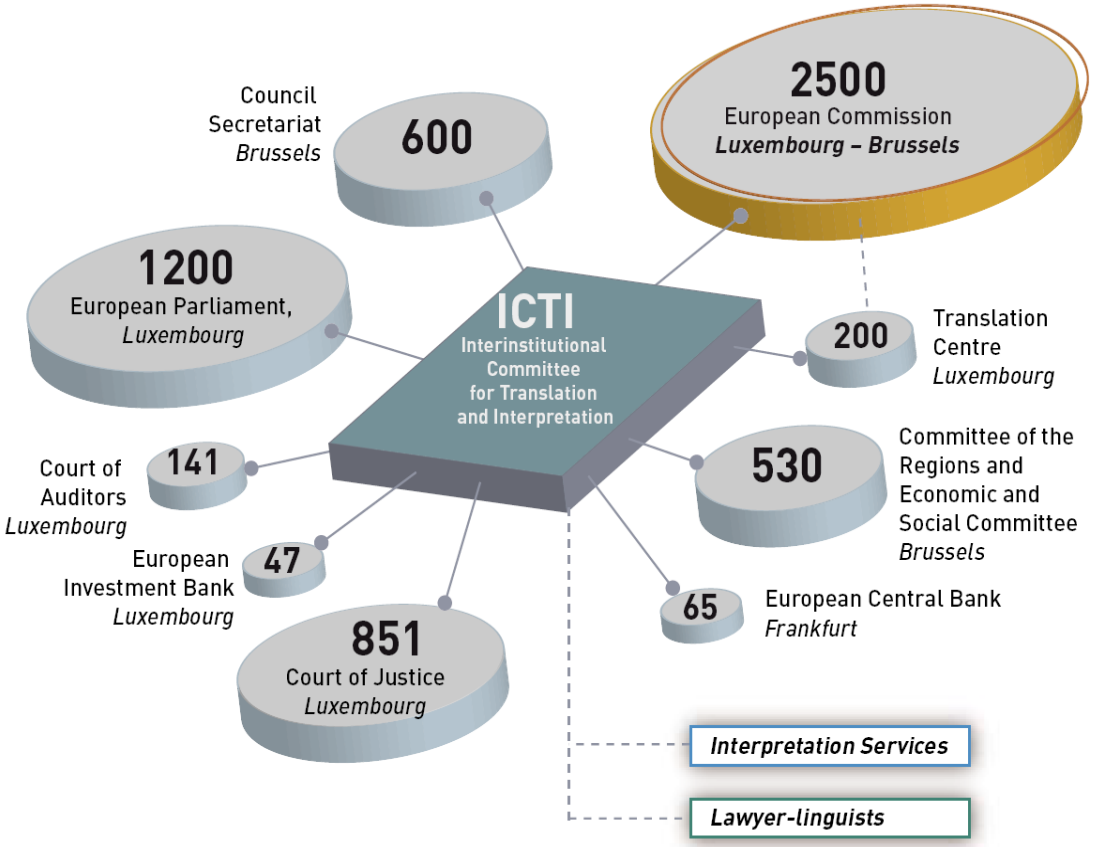
→ ***European institutions and bodies:***

▪ Commission ▪ Parliament ▪ Council ▪ Court of Justice ▪ Court of Auditors	▪ Economic and Social Committee ▪ Committee of the Regions ▪ European Central Bank ▪ European Investment Bank, etc.

→ ***Online services*** *funded or supported by the EU*

→ ***Public administrations*** *in the EU/EEA countries*

MT@EC... built on EU data



Interinstitutional cooperation on translation and interpretation

Euramis

940 million sentences
[end 2015]

Growing at 2.6 million per month

All 24 languages
[different volumes]

"EU documents"
[EU subjects and language]

Free access to MT@EC for public administrations staff in CEF countries

- Get an individual **ECAS** user name and password (self-registration) using your work email address.
  [go to https://webgate.ec.europa.eu/cas/eim/external/register.cgi and follow the instructions]
- Send an **email** to *DGT-MT@ec.europa.eu* asking for the activation of access to the service.
- DGT will activate your access and inform you by email.

Online services connected to MT@EC

(status: January 2016)

Service	Description/URL/use
IMI	Internal Market Information System – an intranet for more than 7000 public administrations in the EU Member States http://ec.europa.eu/internal_market/imi-net/index_en.html ➔ *free-text message exchange*
SOLVIT	An on-line problem solving network concerning misapplication of Internal Market law by public authorities. http://ec.europa.eu/solvit/ ➔ *free-text message exchange*
nLex	A common gateway to National Law http://eur-lex.europa.eu/n-lex/ ➔ *translate results of queries*
TED	Tenders Electronic Daily is the online version of the 'Supplement to the Official Journal' of the EU (European public procurement). http://ted.europa.eu/ ➔ *c*ontract notices
CircaBC	Communication and Information Resource Centre for Administrations, Businesses and Citizens (collaborative spaces) https://circabc.europa.eu/ ➔ Translate documents

What next?

from MT@EC… to the CEF Automated Translation platform

CEF.AT will:

- build on the **existing** MT@EC service - but not be limited to it
- put emphasis on **secure**, **quality**, **customisable** MT for pan-European online services - but not be limited to them
- be a **multilingualism enabler** – not only MT

Customisation for online services

connecting to the CEF.AT platform (starting from MT@EC)

Service	Description/URL/use
Europeana	The digital European Library, common, multilingual access point to digital resources of European heritage.
ODP*	The pan-European **O**pen **D**ata **P**ortal for accessing open data infrastructures distributed over a EU and MS data repositories.
EESSI	The **E**lectronic **E**xchange of **S**ocial **S**ecurity **I**nformation, a platform between 32 countries (EU+EFTA).
ODR*	The **O**nline **D**ispute **R**esolution platform for resolution of online contractual disputes between consumers and traders, linking all national Alternative Dispute Resolution (ADR) entities.
e-justice*	A portal which is a single point of access to law, enabling EU judicial cooperation.
SaferInternet	Services to make Internet a trusted environment for children.

** already connected to MT@EC (generic EU service)*

CEF.AT in 2014 Work Programme

(4 MEUR)

a) **In-house development & deployment of the core MT platform (MT@EC)**
   - domain adaptation pilots for most "urgent" CEF DSIs
   - preparing for more automation of MT trainingS

b) **Requirements study:** what are the linguistic needs of CEF DSIs?

c) **European Language Resource Coordination (ELRC)**: setting up LR networking between 30 Member States of CEF, country workshops, raising awareness, promoting best practices

CEF.AT in 2015 WP (8 MEUR)

a) Call for tender SMART 2015/1091

- LOT1: language technology evaluation for CEF DSIs
- LOT2: language resource coordination, repository, deepening collaboration with CEF countries
- LOT3: purchasing "deficient" language resources

b) Cloud infrastructure: Framework Contract concluded, trials ongoing in-house

c) Support activities: Framework Contracts for common portal, awareness, publicity, governance...

CEF.AT in 2016 WP (9.5 MEUR)

a) Core platform – deployment (procurement, 3 M)

- engines factory
- up-scaling, operations, programme support
- details of implementation remain to be confirmed

b) Generic services (grants, 6.5 M)

Call for proposals **CEF-TC-2016-3** opens 13/9/2016, with two action lines:

1. Collaborative projects on language resource collection with/by/in CEF countries
2. Collaborative projects to **integrate** CEF automated translation into (multilingual, cross-border) digital services

Questions on MT@EC?

dgt-mt@ec.europa.eu

Thank you!