



# building the **Automated Translation Platform**

Spyridon Pilos  
Head of language applications  
Directorate-General for Translation

*Riga, 29.04.2015*

## **Building the CEF automated translation platform**

- *MT@EC*
- *Online services and MT*
- *CEF Automated translation platform*



# MT@EC: a European Commission product

**Released :** 26 June 2013

**Languages:** All 24 EU official languages  
552 language pairs (*62 direct*)

**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](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)

# Who can use MT@EC today ?

## → *European institutions and bodies:*

- |  |   |
|--|---|
| <ul style="list-style-type: none"><li>▪ Commission</li><li>▪ Parliament</li><li>▪ Council</li><li>▪ Court of Justice</li><li>▪ Court of Auditors</li></ul> | <ul style="list-style-type: none"><li>▪ Economic and Social Committee</li><li>▪ Committee of the Regions</li><li>▪ European Central Bank,</li><li>▪ European Investment Bank<br/>etc.</li></ul> |
|--|---|

→ *Online services funded or supported by the EU*

→ *Public administrations in the EU Member States*

# Online services connected to MT@EC

Service	Description/URL/use
IMI	Internal Market Information System – an intranet for more than 7000 public administrations in the EU Member States <a href="http://ec.europa.eu/internal_market/imi-net/index_en.html">http://ec.europa.eu/internal_market/imi-net/index_en.html</a> → free-text message exchange
SOLVIT	An on-line problem solving network concerning misapplication of Internal Market law by public authorities. <a href="http://ec.europa.eu/solvit/">http://ec.europa.eu/solvit/</a> → free-text message exchange
nLex	A common gateway to National Law <a href="http://eur-lex.europa.eu/n-lex/">http://eur-lex.europa.eu/n-lex/</a> → translate results of queries
CircaBC	Communication and Information Resource Centre for Administrations, Businesses and Citizens (collaborative spaces) <a href="https://circabc.europa.eu/">https://circabc.europa.eu/</a> → Translate documents

# Existing online services

## → *preparing to connect to MT@EC*

Service	Description/URL/use
e-Justice portal	The future electronic one-stop-shop in the area of justice <a href="http://e-justice.europa.eu/">http://e-justice.europa.eu/</a> → Uploaded document (temporary)
EU Survey	Tool for creating multilingual online surveys <a href="http://ec.europa.eu/eusurvey/">http://ec.europa.eu/eusurvey/</a> → Interactive translation of survey questions
TED	Tenders Electronic Daily is the online version of the 'Supplement to the Official Journal of the European Union', dedicated to European public procurement <a href="http://ted.europa.eu/">http://ted.europa.eu/</a> → Free text translation
Joinup	Joinup is an open collaborative platform supporting interoperability in Europe <a href="https://joinup.ec.europa.eu/">https://joinup.ec.europa.eu/</a> → Uploaded documents

# Online services under construction

→ *interested in using MT*

Service	Description/URL
EESSI	Electronic Exchange of Social Security Information
ODR	Platform to facilitate the resolution of consumer disputes out-of-court (Alternative Dispute Resolution) <a href="http://ec.europa.eu/consumers/redress_cons/adr_en.htm">http://ec.europa.eu/consumers/redress_cons/adr_en.htm</a>

**but**

Can MT@EC serve fully  
any type  
of pan-European online service?



## MT@EC: Generic "EU" engines/data

- *725 million monolingual segments* [end 2014]  
→ **Growing at 2.6 million per month**
- *All 24 languages **but** different volumes*  
96 820 335 monolingual.en, 57 732 689 monolingual.fr  
2 798 524 monolingual.ga, 2 775 226 monolingual.hr  
→ The rest between 22 and 38 million
- *owned by the EU ("Euramis" managed by DGT)*  
**but** *"EU documents" (i.e. EU subjects and language)*

## Speed and scale

*(1 January 2014 to 31 December 2014)*

No requests: 611 597

Median (Documents): 4'21"

Median (Snippets): 17"

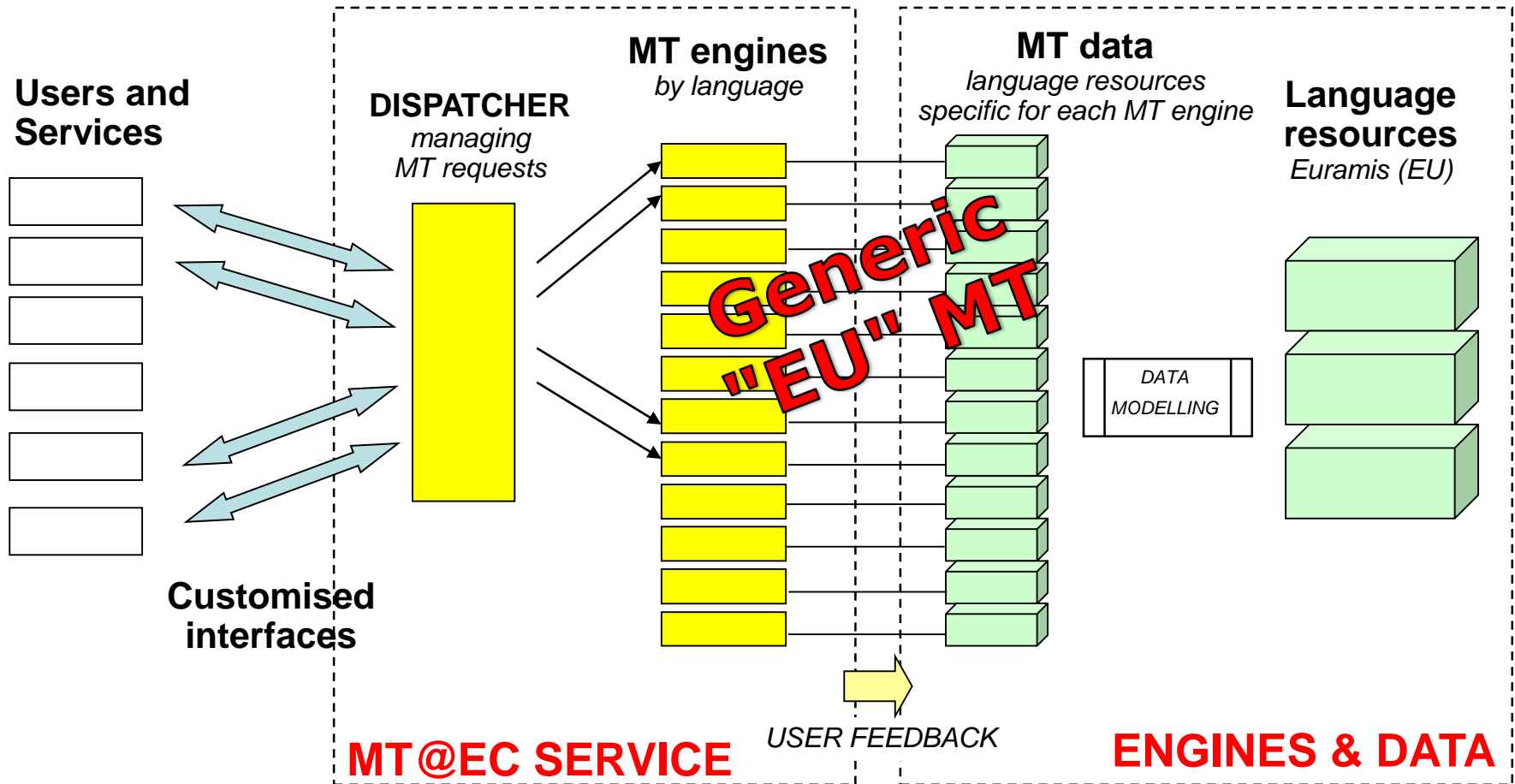
*OK for asynchronous translation  
**but** not for the "real-time" translation  
needed by most online services*

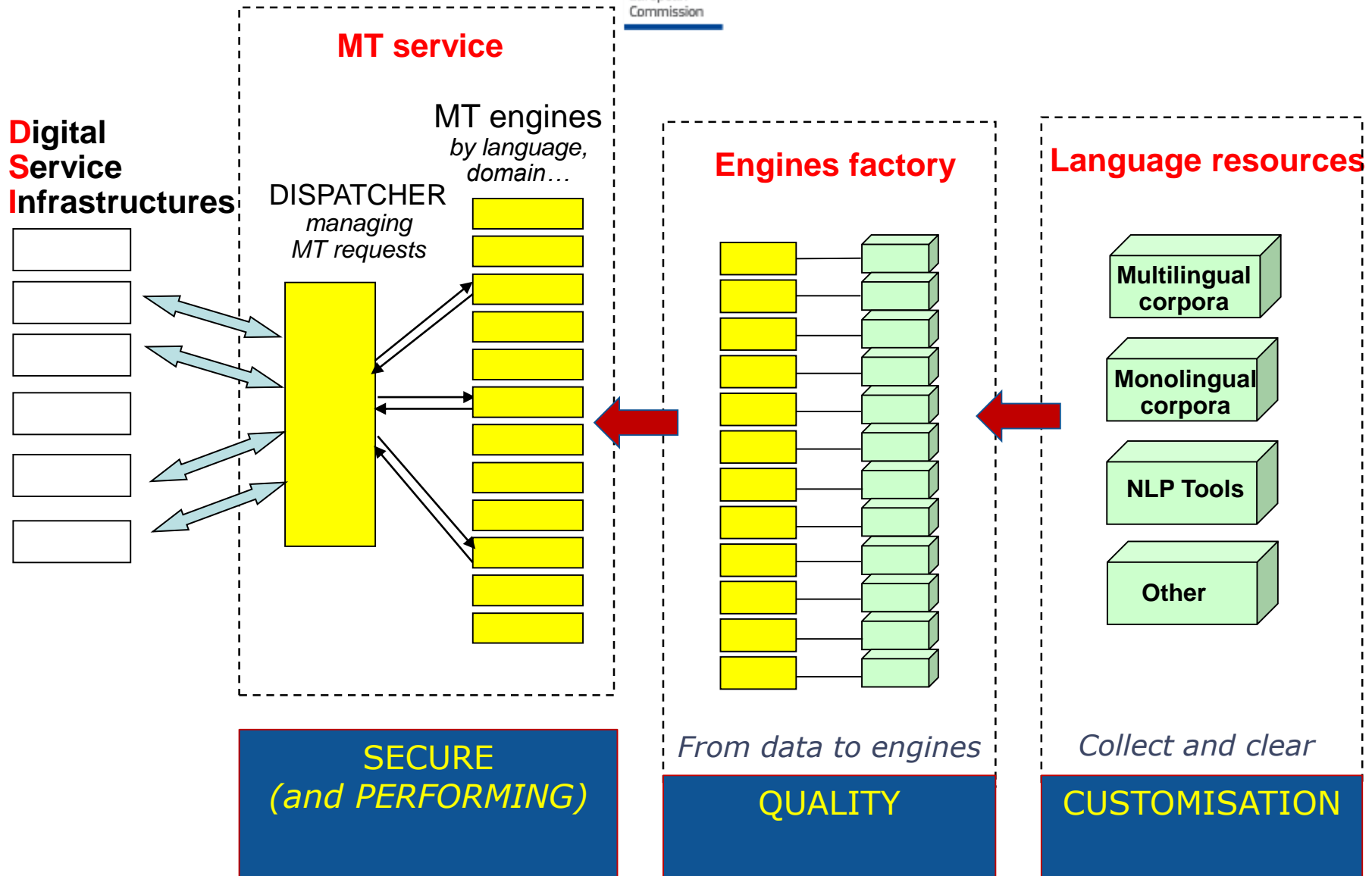
## So what next?

### from MT@EC... to the CEF automated translation platform

The **CEF AT platform** will:

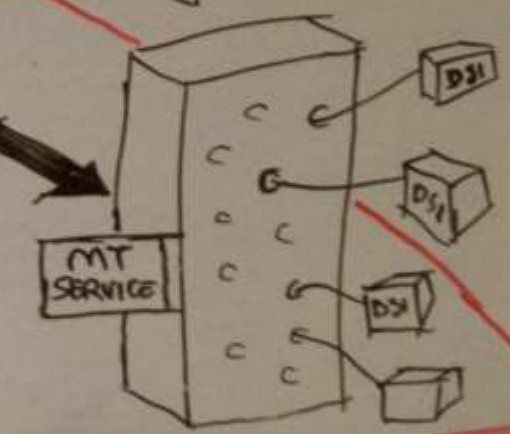
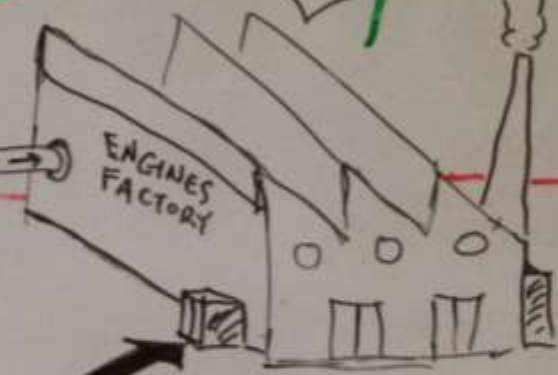
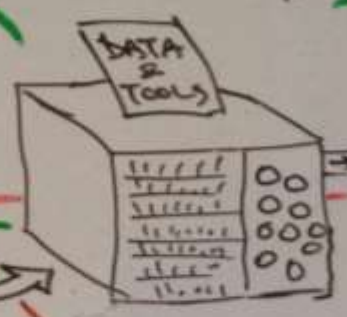
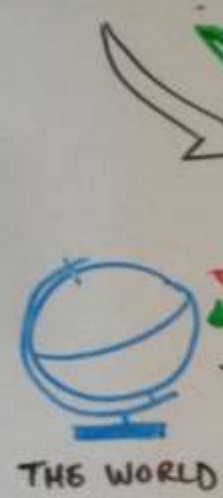
- build on the **existing** MT@EC service - but not be limited to it
- put emphasis on **secure, quality, customisable MT** for CEF online services (DSI) - but not be limited to them
- be a **multilingualism enabler** – not only MT





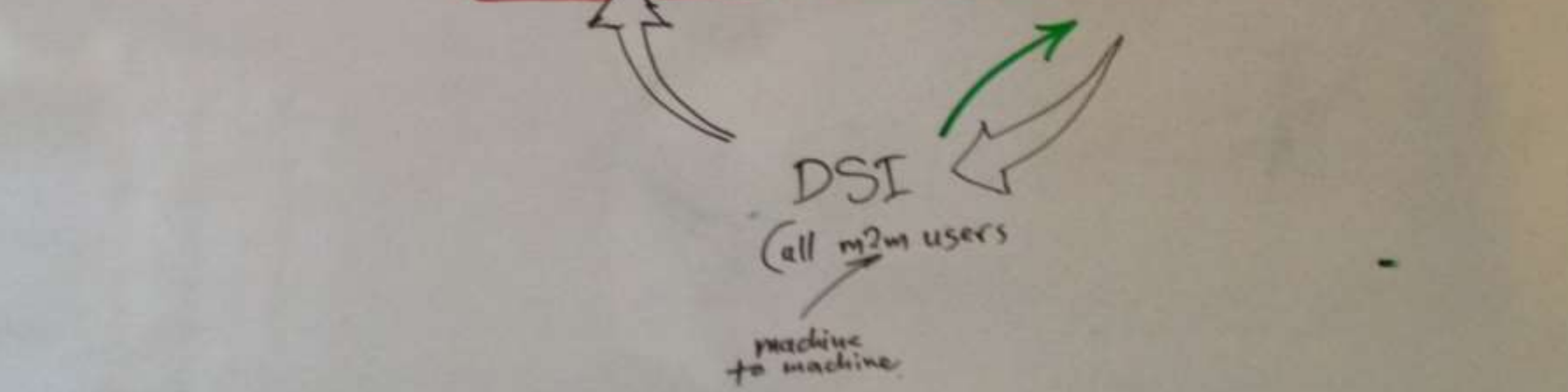
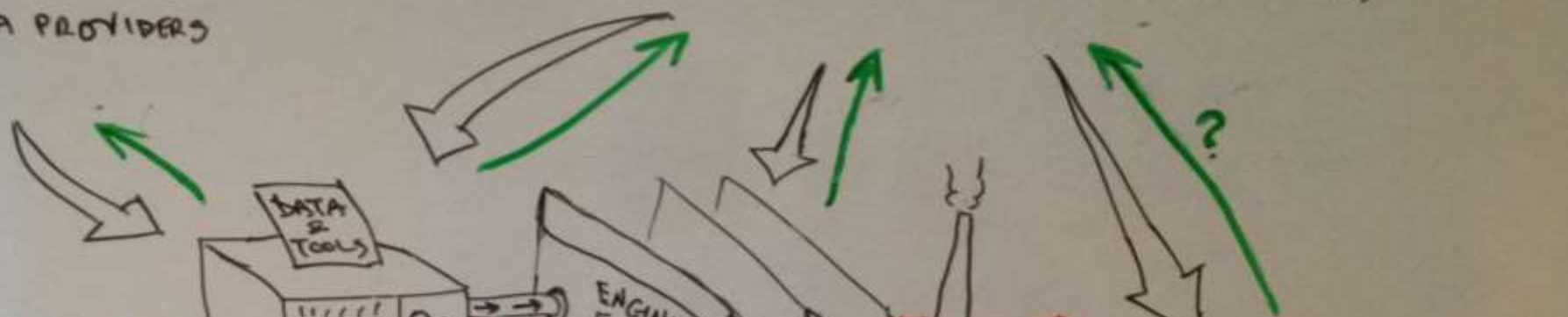
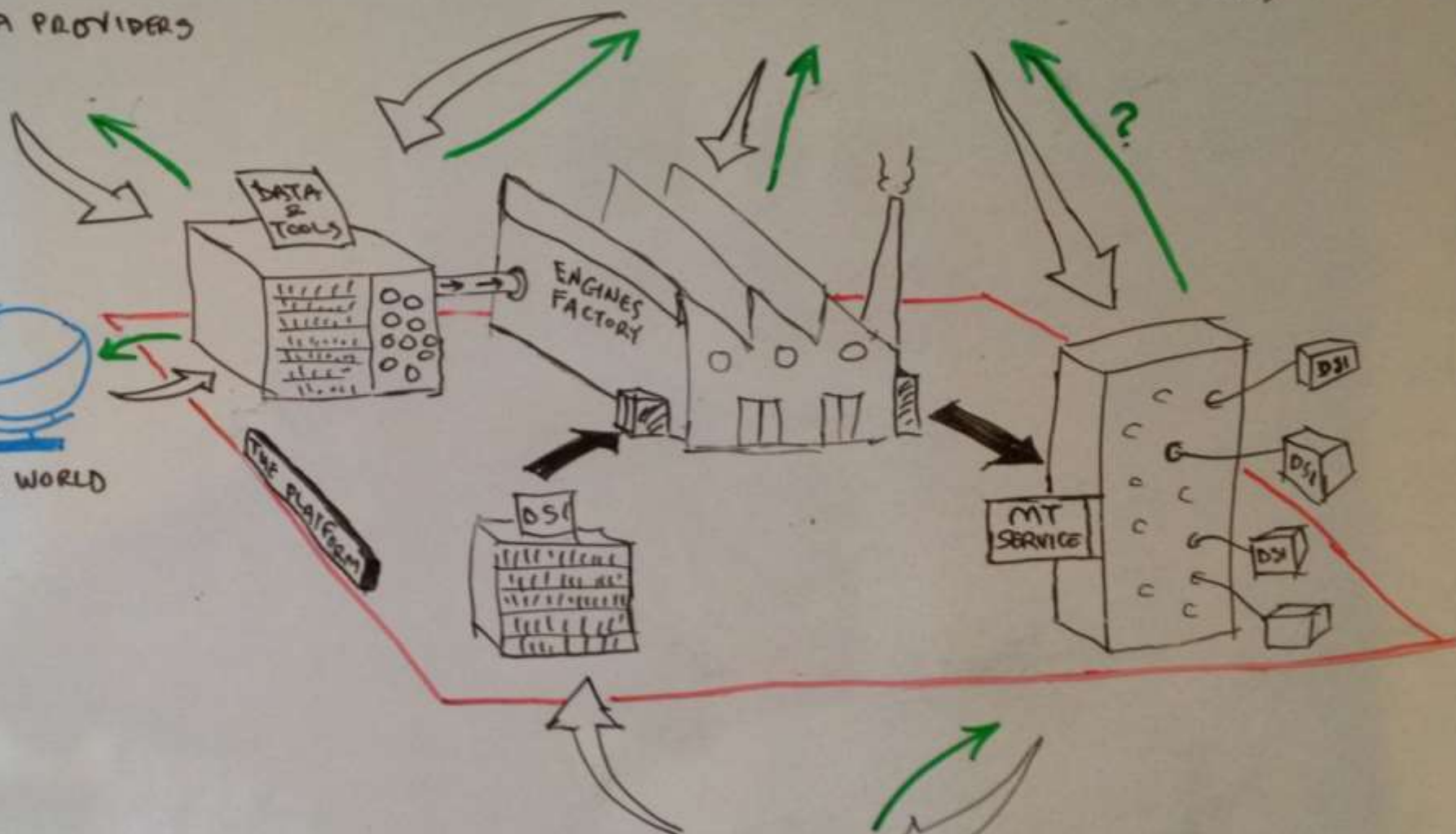
TECHNOLOGY PROVIDERS (OSS + COTS)

DATA PROVIDERS



THE PLATFORM

DSI  
(all m2m users)  
machine to machine



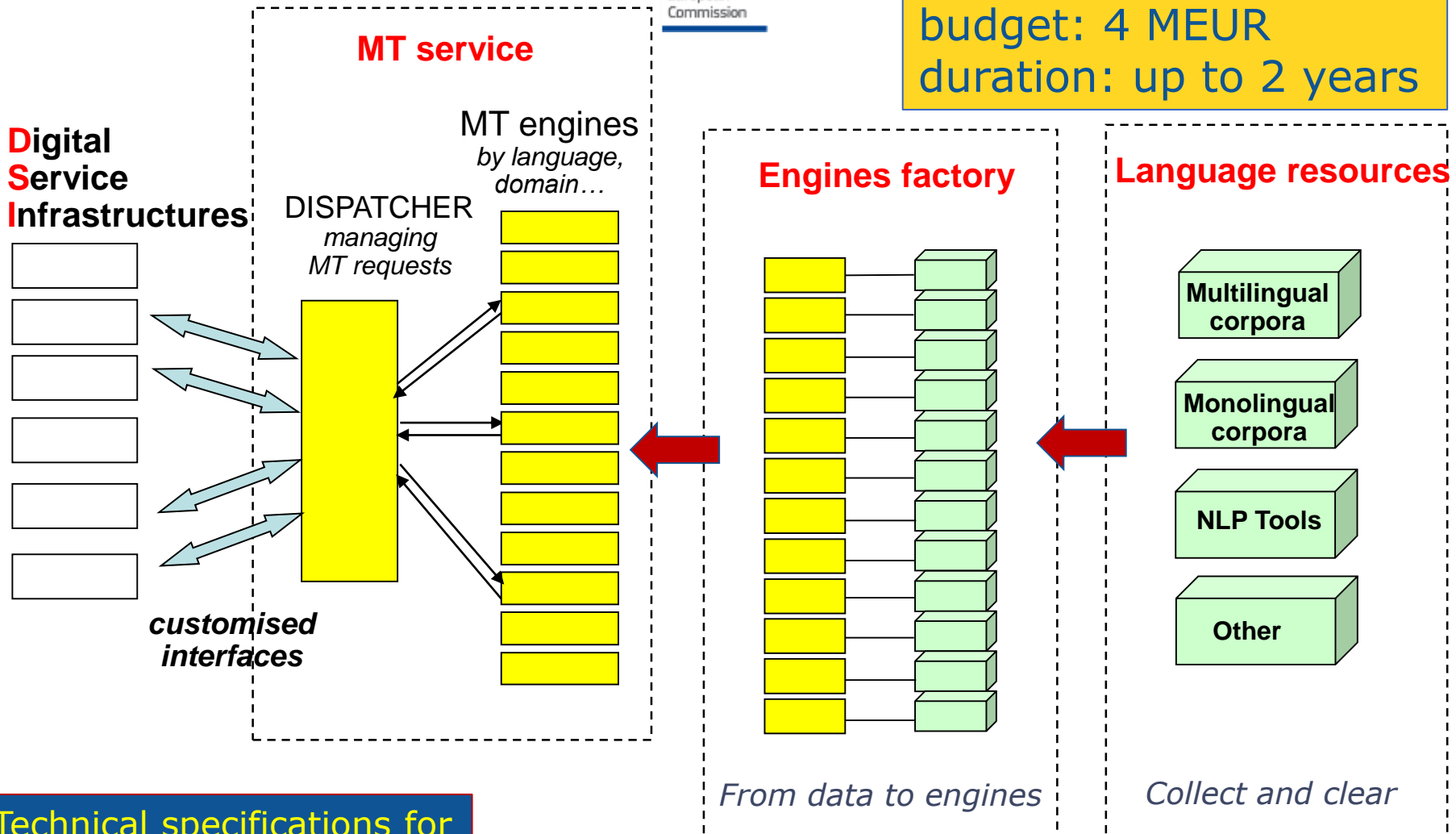
# CEF.AT platform

## 2014 Work programme



# Preparing for the platform

budget: 4 MEUR  
duration: up to 2 years



Technical specifications for the full AT platform (scale, performance etc.)

POC domain adaptation for 2-3 **DSIs** using MT@EC

Setup coordination & networking

# Language resources: the key to success

## Type

- Domain specific language resources
- Better lexical coverage for all languages
- More for all languages
- Different use conditions
- Different technical conditions

## Action

→ Find → Assess → Clear for use  
and then → Prepare for translation automation





*CEF.AT work started already with some CEF.DSIs...*

*... to specify their needs, and*

*... to identify domain specific language resources*

*... to build customised solutions ("engines")*

*... to run a proof-of-concept under MT@EC.*

*DGT and its' **Translating Europe** initiative, aimed to create a community of interest on translation across the EU, will support the CEF language resource coordination and networking activity*

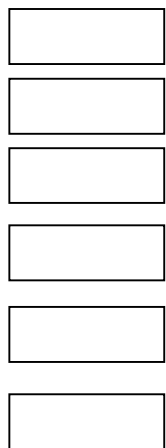
*[http://ec.europa.eu/dgs/translation/programmes/translating\\_europe/index\\_en.htm](http://ec.europa.eu/dgs/translation/programmes/translating_europe/index_en.htm)*



European Commission

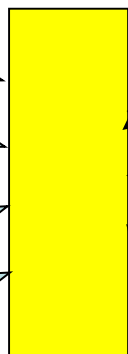
budget: 8 MEUR  
duration: up to 3 years

Digital Service Infrastructures

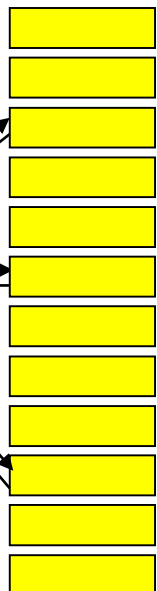


customised interfaces

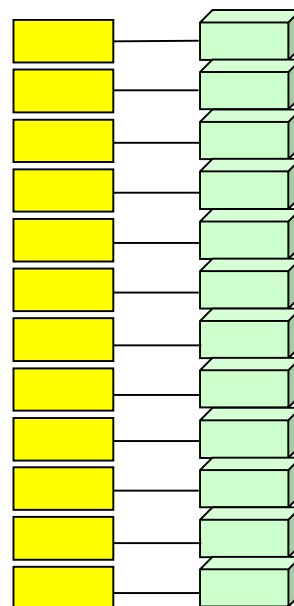
DISPATCHER  
managing  
MT requests



MT engines  
by language,  
domain...

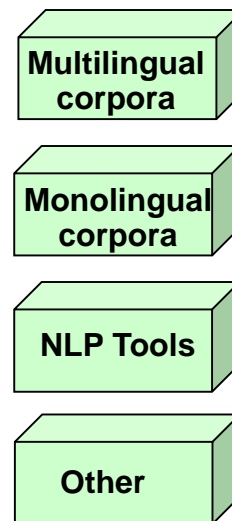


Engines factory



From data to engines

Language resources



Collect and clear

Acquire/set-up infrastructure and technologies to scale up from MT@EC

Dom. adaptation for more DSIs  
Prepare factory automation

Setup repository  
Acquire LR

## For CEF.AT we will be looking for:

- *even better **language technologies***  
*both open source and commercial products*
- *more quality **language resources***  
*both open and licenced data*
- *even better **ways to enable multilingualism***  
*both from research and market*
- *appropriate infrastructure to **scale up***

The time for making Digital Europe multilingual  
is **NOW**.

