



# NEC TM Data &



— **Evento Diseminación Madrid** —



# La importancia de los datos que generamos

El productor de datos nº1 en cada Estado Miembro es....



CIKLOPEA

# La importancia de los datos que generamos

## Fuentes de datos

- Administraciones Públicas
- Organizaciones
- Empresas
- Ciudadanos

*Google*

*Microsoft*

*Amazon*

*Baidu*

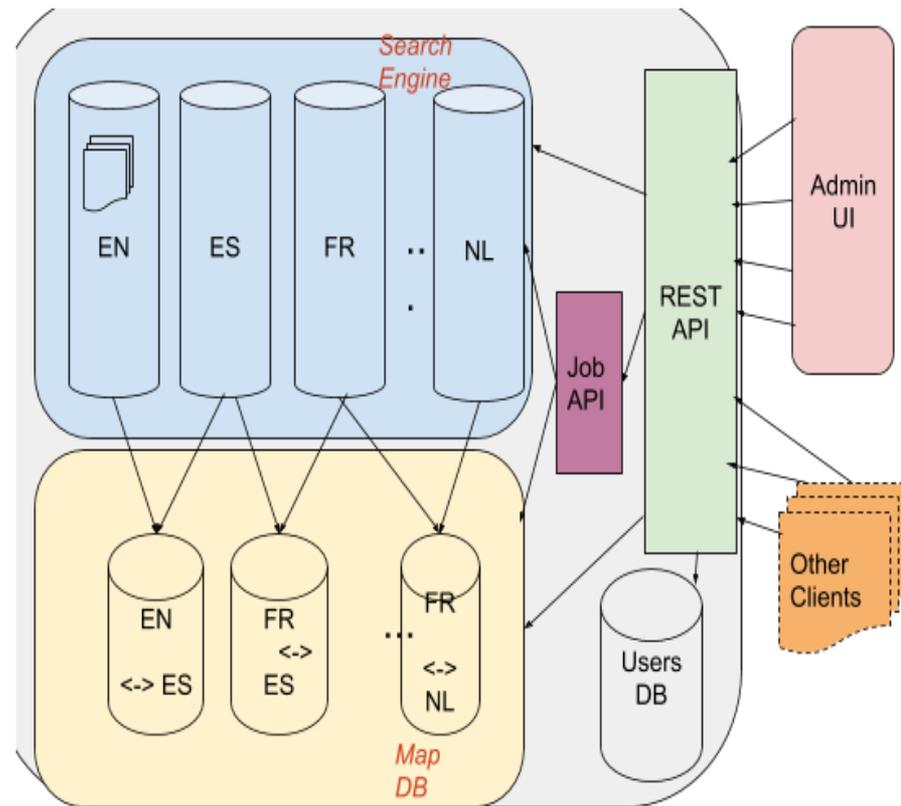
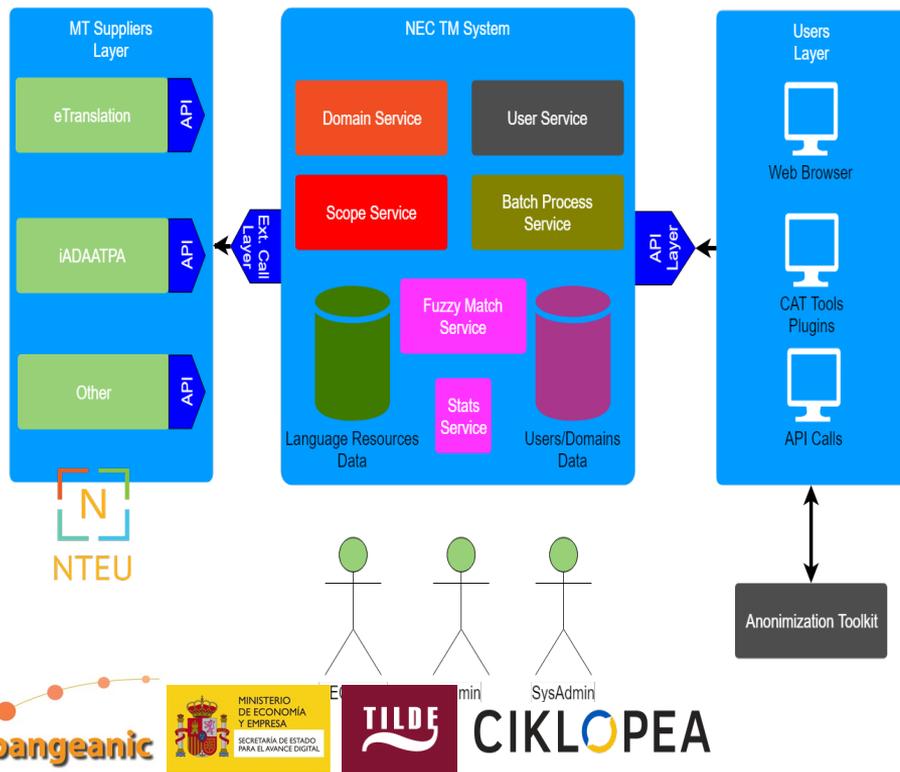
*Yandex*

# Objetivos

Incrementar los volúmenes de datos paralelos disponibles de modo organizado en los Estados Miembro mediante:

- **Adquisición, centralización, re-utilización de textos bilingües pagados por las AAPP.**
- Promover el flujo de memorias de traducción de las empresas contratistas (de traducción) a las AAPP
- Ahorro de costes para las AAPP (*fuzzy match*)
- Facilitar el trabajo de los equipos traductores del sector público
- Desarrollar la plataforma eTranslation de CEF

# Sistema NEC TM



# Acciones concretas



- Estudio del valor de los contratos de traducción a nivel europeo para establecer las relaciones contratista-Administración
- Analizar y publicar el tamaño del mercado público de cada Estado Miembro ([www.nec-tm.eu](http://www.nec-tm.eu))
- Desarrollar la bbdd NEC TM: aspectos técnicos. Open Source
- Diseminación (3 conferencias, Zagreb, Varsovia, Madrid + cursillos en ELRC, formación online)
- Implementación a nivel nacional (España, Letonia... Polonia, Malta, etc.)



CIKLOPEA

# Inversión nacional en servicios de traducción – datos paralelos para ML

El estudio por país y el Libro Blanco establece las relaciones entre los contratos de traducción, los contratistas, las cantidades y los compradores. Se establecen valores globales desde 2015, por año y por sector.

<https://www.nec-tm.eu/country-reports-nec-tm/>

Se ha utilizado el BOE, boletines autonómicos/regionales (EU), municipales, TED, etc.

Gracias a este informe podemos saber el tamaño de los datos, potencial, ahorro pero sobre todo **la mina que Universidades, centros de investigación, empresas tecnológicas y la Administración pueden explotar.**

# Inversión nacional en servicios de traducción – datos paralelos para ML 2015-2018

Country	Expenditure (million)	No. of contracts
Spain*	€77	261
France	€55	531
The Netherlands	€49	45
Portugal	€13	348
The Republic of Ireland	€15.5	108
Belgium	€30	31
Bulgaria	€8.37	44

\* El que mejor reporta todo en webs oficiales

# Inversión nacional en servicios de traducción – datos paralelos para ML 2015-2018

Country	Expenditure (million)	No. of contracts
Malta	€0.37	27
Croatia	€2.7	54
Hungary	€ 7.1	17
Slovenia	€8.4	110
Romania	€0.9	110
Slovakia	€1.37	117
Czech Republic	€10.05	78

# Inversión nacional en servicios de traducción – datos paralelos para ML 2015-2018

Country	Expenditure (million)	No. of contracts
Germany	€54.9	534
Finland	€22.1	424
Sweden	€20.03	151
Denmark	€17.2	40
Poland	€16.7	727
Lithuania	€9.7	50
Estonia	€6.5	171
Latvia	€5,05	2.200

# Inversión nacional en servicios de traducción – datos paralelos para ML 2015-2018

2019 EU institutions expenditure = €115,989,747.00

2018 EU institutions expenditure = €47,592,807.27

2017 EU institutions expenditure = €22,090,000.00

2016 EU institutions expenditure = €21,745,740.00

2015 EU institutions expenditure = €140,661,049.00

# Resumen

(provisional):

- Las Administraciones Públicas europeas gastan aproximadamente **€300M** en servicios de traducción tomando una postura muy conservadora.
- Sumando instituciones europeas (CE/Parlamento/BCE, etc.) **€400M**
- **No se incluyen los gastos de personal, que se estiman en unos 100M (nacional + UE)**

**500M € de datos de valor y casi 1.000M € si incluimos interpretación**



CIKLOPEA

**La mayor granja de motores de calidad casi human basada en redes neuronales.**

**Combinación directa idioma-idioma, sin pasar por inglés.**

**Todos los idiomas oficiales de la UE.**

- +500 combinaciones! HR into PT, ES, FR, DE, NL, IT, GA, DA, SV, FI, LT, LV, ET, HE, BG, RO, CZ, SLO, SLK, HU, PO, MT...
- Suministrar datos de entrenamiento **limpios** a Estados Miembro para su reutilización, via ELRC, para uso por parte de las AAPP.
- Disponible bajo acuerdo (no terceros países). Máquinas entrenadas con 15M alineaciones para idiomas con recursos. Idiomas de bajos recursos 10-12M; recursos ultrabajos 10M alignments.
- Adopción de la TA a nivel nacional y europeo. Facilitar el intercambio de información entre AAPP de toda Europa.

# NTEU



NTEU

	BG	CZ	DA	DE	ES	ET	FI	FR	GA	HE	HR	HU	IT	LT	LV	MT	NL	PO	PT	RO	SL	SLK	SV	lang combos	
BG	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
CZ	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
DA	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
DE	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
ES	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
ET	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
FI	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
FR	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
GA	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	22
HE	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	22
HR	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	22
HU	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	22
IT	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	*	22
LT	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	*	22
LV	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	*	22
MT	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	*	22
NL	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	*	22
PO	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	*	22
PT	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	*	22
RO	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	*	22
SL	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		*	*	22
SLK	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		22
SV	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	22

Total NMT Engines.....506



# ¡Gracias!

Hvala vam! Thank you! Ευχαριστώ! Hvala! благодаря! Köszönöm!  
Paldies! Ďakujem! Vă mulțumesc! Dziękuję!

Info:



Manuel Herranz / Amando Estela / Mercedes García  
[info@nec-tm.eu](mailto:info@nec-tm.eu)