

Despliegue de infraestructuras de **traducción automática** en la Administración Pública española, desarrolladas bajo el mecanismo de financiación “*Conectar Europa*”



Maite Melero

Barcelona Supercomputing Center
Secretaría de Estado de Digitalización e
Inteligencia Artificial



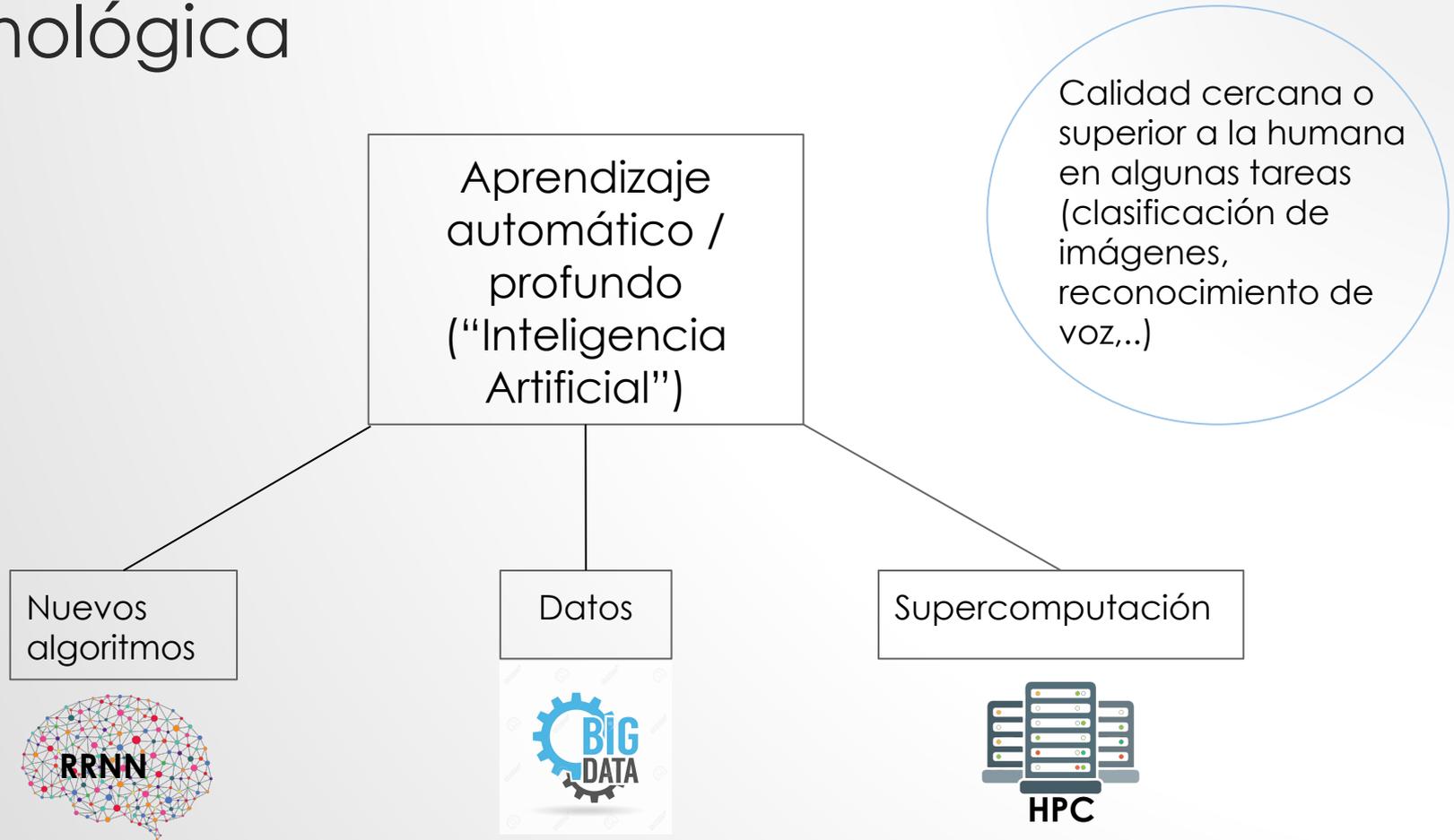
Jornada sobre Tecnologías de la Traducción
para la Administración Pública

Madrid, 2 de marzo 2020

Índice

1. La traducción, una industria en plena transformación
2. La TA, un componente transversal en los servicios digitales europeos
3. Participación española en proyectos CEF de TA para la Administración Pública

Estamos en medio de una revolución tecnológica



Estamos en medio de una revolución tecnológica

Traducción Automática Neuronal (NMT)

```
graph TD; A[Traducción Automática Neuronal (NMT)] --- B[Nuevos algoritmos (Transformer)]; A --- C[Datos: genéricos y específicos de dominio]; A --- D[Supercomputación]; E((Calidad cercana a la humana para algunos pares de lenguas))
```

Nuevos algoritmos (Transformer)

Código abierto favorece avances rápidos

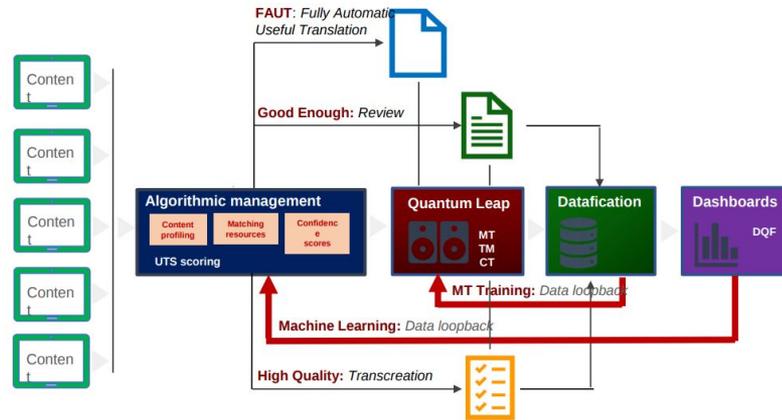
Datos: genéricos y específicos de dominio

Supercomputación

Calidad cercana a la humana para algunos pares de lenguas

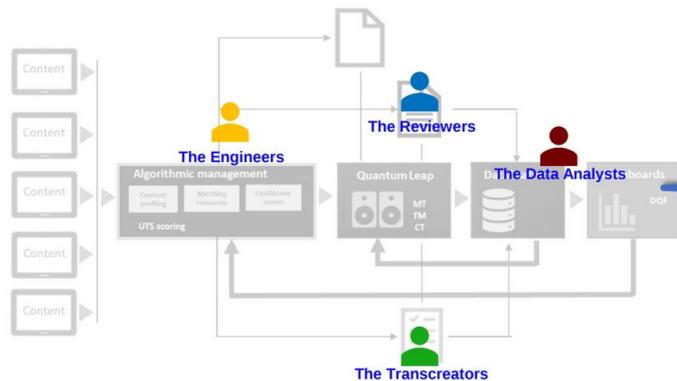
La traducción, una industria en plena transformación

The Modern Translation Pipeline



La traducción, una industria en plena transformación

... and where are the jobs?!



A search for global multilingual talents to undertake a plethora of new value-added services:

- Brand Validation
- Multilingual Digital Marketing
- eDiscovery
- Usability Testing
- Campaign Management
- Content Curation
- Community Development

La traducción, una industria en plena transformación

MT Thousands Time Bigger than HT



300 trillion words

VS.

**200 billion words translated
by professionals**

TA permite un *Mercado Único Digital* por encima de barreras lingüísticas

Infraestructuras de Servicios Digitales



BRIS

Garantiza seguridad jurídica a las empresas cuando hacen negocios en otros países europeos



ODR

Permite a los ciudadanos resolver conflictos en países distintos del propio



eHealth

Permite acceder a la historia clínica en el extranjero



EESSI

Permite a los ciudadanos disfrutar de sus derechos de Seguridad Social en cualquier país



Otros servicios digitales:

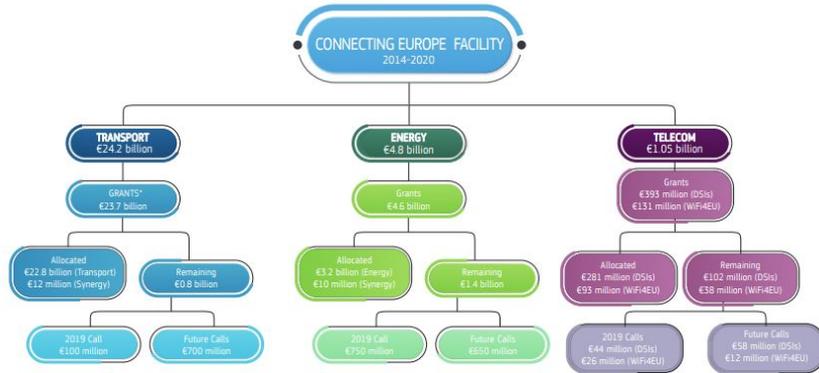
**eJustice
Europeana
Open Data**



CEF - a success story

The Connecting Europe Facility provides dedicated financing for infrastructure actions. This is so that important investments can be made across the European Union in the trans-European (TEN) transport, telecommunications and energy network infrastructure - to help Europe reach its full potential in terms of growth and cohesion. The CEF has been tailored to support actions in different ways – grants for studies and/or works and financial instrument support.

For the entire programming period, the CEF plans to finance actions in three sectors for a total of more than €30 billion, out of which an indicative amount of €28.8 billion is in the form of grants. 92% of this budget is currently allocated to actions selected under the 2014-2018 calls for proposals – with a remaining budget of €2.4 billion.



*Including internally assigned revenues and revenues from predecessor programmes.

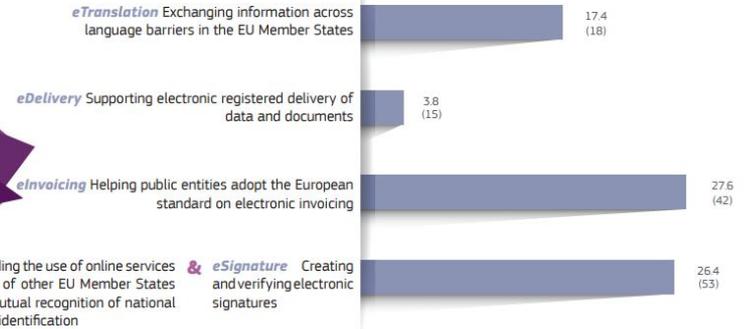
DSI building blocks

CEF Digital actions aim to be mutually reinforcing and complementary. There are many potential links between the digital services – in particular the building blocks that are reused in more complex DSIs. This possibility not only contributes to creating economies of scale but also speeds up implementation of more complex actions.

Aside from the development of the respective Core Service Platforms with a dedicated budget of €92 million, the Commission supports their deployment through dedicated calls for proposals. In 2014-2018, 16 dedicated calls were launched for these building blocks, allocating €75.3 million in CEF Telecom funding to 128 actions*.

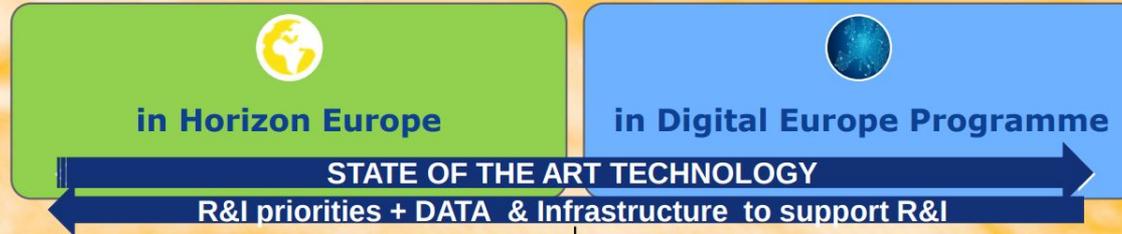
CEF GRANT FUNDING PER DSI BUILDING BLOCK

€ million
(number of actions)



*At the moment of drafting the brochure no Generic services were funded through CEF yet for Blockchain, eArchiving, Context Broker or Big Data test Infrastructure.

LANGUAGE TECHNOLOGIES & MULTILINGUALISM IN THE NEXT MFF



RESEARCH & INNOVATION

- **Cultural Heritage:**
Support Language Preservation
- **Next Generation Internet:**
Support research to achieve digital language transparency

CAPACITY BUILDING

- *Artificial Intelligence*
- *Data Spaces*
- *High Power Computing*

DEPLOYMENT & BEST USAGE

- *Cloud infrastructures*
- *Deployment of modern public services*

Widening the best use of Digital Technologies

- *Language Technologies*



Prioridades del programa de financiación **Conectar Europa (CEF)**

- Superar las **barreras del idioma**
- Preservar la **diversidad lingüística**
- **Modernizar** las Administraciones Públicas
- Aumentar la **competitividad** de Europa y su **independencia** de proveedores tecnológicos externos.



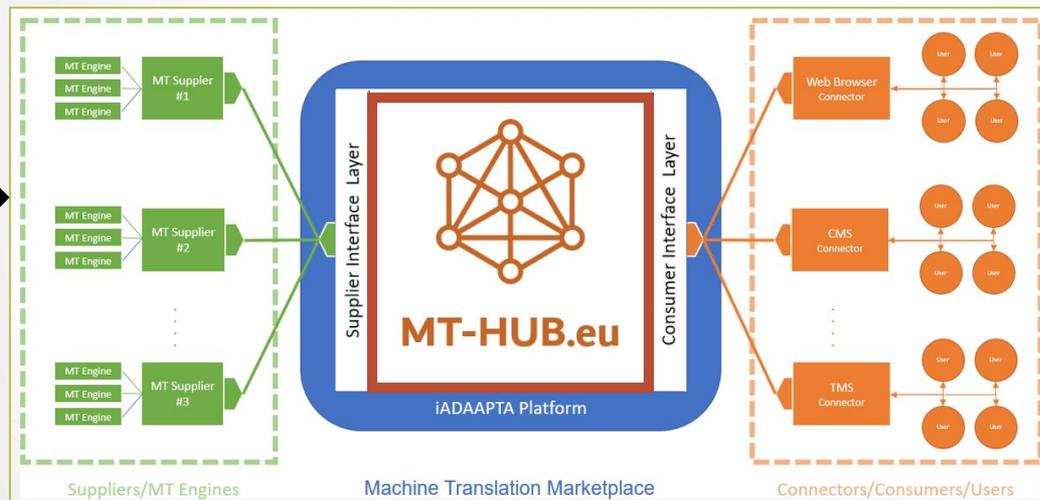
¿Cómo?



- Impulsando la **innovación** en las Tecnologías del Lenguaje y de la Traducción.
- Promoviendo la **recolección de recursos lingüísticos**.
- Proporcionando **soluciones tecnológicas** a la Administración Pública (CE y estados miembros).
- Fomentando la colaboración entre la **industria** y el sector público.

ELRC: Infraestructura de Intercambio de Datos Lingüísticos en Europa





Estación Nacional de Transmisión de datos de ELRI en España

¡Bienvenidos a ELRI en España!

Este sitio web es la Estación Nacional de Transmisión de datos de ELRI en España, donde las instituciones de traducción públicas pueden recopilar, preparar y compartir recursos lingüísticos.

Las estaciones nacionales de transmisión se desarrollan dentro del proyecto europeo ELRI, financiado por la European Facility.

¡Contribuya a promover la igualdad lingüística en Europa uniéndose a la red ELRI!

Unirse a ELRI

Los miembros de las instituciones públicas y centros de traducción pública pueden unirse a un grupo de recursos compartidos, aportar sus propios datos y beneficiarse de memorias de traducción u otros recursos. [Proporcionar información acerca de los miembros de ELRI](#)

- Storage
- TU cleanup
- TMX generation
- Sentence alignment
- Document alignment
- Preprocessing
- Sentence split
- Language identification
- Text extraction
- Multilingual documents (odt, doc(x), rtf, pdf, txt)



vicomtech
your R&D partner for smart digital solutions

everis
an NTT DATA Company

prompsit

pangeanic

CIKLOPEA

KantanMT.com

Linkare
FACING INNOVATION TOGETHER

Iconic
Translation Machines

Unbabel

TILDE

DCU
Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

F C **Ciências
ULisboa**

Limsi

**el
Da**



ama | AGÊNCIA PARA A
MODERNIZAÇÃO
ADMINISTRATIVA
PRESIDÊNCIA DO CONSELHO DE MINISTROS

Jornada sobre Tecnologías de la Traducción para la Administración Pública

9:45-10:15 La traducción como Big Data europea: estudio de los costes de los contratos públicos de traducción en los países de la UE - *Manuel Herranz - Pangeanic*

10:15-10:30 La magia de las palabras: el nuevo traductor neuronal entre euskera y castellano - *Xabier Arrieta Goiri - Director de Informática y Telecomunicaciones del Gobierno Vasco*

10:30-11:00 Pausa café

11:00-11:45 Demostración en vivo de NEC-TM, asistencia a la traducción para Administraciones Públicas - *Amando Estela - Pangeanic*

11:45-12:45 Traducción Automática para las Administraciones Públicas: plataforma MT-HUB y proyecto NTEU de motores neuronales - *Manuel Herranz y Mercedes García - Pangeanic*

12:45-13:05 Hugo.lv - AI powered language services for Latvian public sector - *Artūrs Vasilevskis - Tilde*

13:05-13:20 El Plan de Impulso a las Tecnologías del Lenguaje en la estrategia de España para la Inteligencia Artificial - *David Pérez Fernández - Gabinete del Secretario de Estado para Digitalización e Inteligencia Artificial*

13:20-13:45 Sesión de discusión, ruegos y preguntas

13:45-15:00 Clausura de la sesión y vino español