

Taller ELRC en España
enero 2018

Iniciativas de servicios públicos digitales y datos abiertos en España

David Pérez Fernández

SESIAD – Plan de Impulso de las Tecnologías del Lenguaje
Ministerio de Energía, Turismo y Agenda Digital



Datos Abiertos en España








- En 2013 la UE publica la reforma de la Directiva 2003/98/CE relativa a la **reutilización de la información del sector público (RISP)**.
- España a través de la transposición de la Directiva en 2015, hace suya la iniciativa y le da apoyo.
- La iniciativa **Aporta** (RED.es) nace ya en 2009 con el objetivo de crear las condiciones para el desarrollo del mercado de la reutilización de la información del sector público, y para dar apoyo a la administración para la publicación de la información de acceso no restringido que recoge.

Catálogo de datos

Categoría	
	Sector público (3598)
	Sociedad y bienestar (2687)
	Economía (2477)
	Demografía (2143)
	Medio ambiente (1537)
	Cultura y ocio (1124)
	Educación (1099)
	Hacienda (1085)
	Empleo (1003)
	Urbanismo e infraestructuras (750)

	Turismo (743)
	Salud (720)
	Transporte (708)
	Ciencia y tecnología (658)
	Medio Rural (498)
	Seguridad (492)
	Vivienda (463)
	Energía (449)
	Comercio (429)
	Legislación y justicia (384)
	Industria (237)
	Deporte (202)

<http://datos.gob.es> (15 enero 2018)

	<p>Spain - Clic&Fish</p> <p>Sector: Agriculture, Fisheries, Forestry & Foods Nature: Service Description: This service shows live data of Spain's river conditions that...</p> <p>07/11/2017</p>	Spain
	<p>Spain - Conoces Aragón?</p> <p>Sector: Education, Culture & Sport Nature: Application Description: This app is a quiz on Aragón's history and culture</p> <p>07/11/2017</p>	Spain
	<p>Spain - iMar</p> <p>Sector: Environment Nature: Application Description: This app provides information on real-time sea conditions</p> <p>07/11/2017</p>	Spain
	<p>Spain - Subastas Agencia Tributaria</p> <p>Sector: Justice, Legal System & Public Safety Nature: Application Description: This app allows citizens to inform themselves on and bid for...</p> <p>07/11/2017</p>	Spain
	<p>Spain - Puntos de Acceso a Internet</p> <p>Sector: Regions & Cities Nature: Application Description: This application offers citizens information about free Internet...</p>	Spain

Ejemplos de re-utilización de
Datos Abiertos
(fuente:
<https://www.europeandataportal.eu/es/using-data/tell-us-your-story>)

Administración electrónica en España



Administración electrónica en España

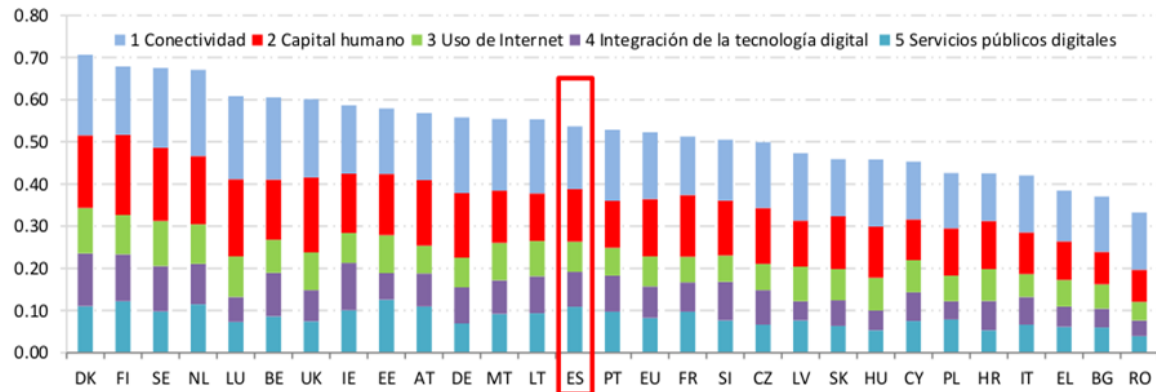


- Posición de España en el Mercado Único Digital

1 Conectividad	Banda ancha fija, banda ancha móvil, velocidades y precios de la banda ancha
2 Capital humano	Uso de internet, competencias digitales básicas y avanzadas
3 Uso de internet	Uso por parte de los ciudadanos de los contenidos, las comunicaciones y las transacciones en línea
4 Integración de la tecnología digital	Digitalización de las empresas y comercio electrónico
5 Servicios públicos digitales	Administración electrónica

	España		Grupo	UE
	puesto	resultado	resultado	resultado
DESI 2017	14	0,54	0,54	0,52
DESI 2016 ²	15	0,51	0,51	0,49

Clasificación del Índice de la Economía y la Sociedad Digitales (DESI) de 2017



Digital Economy and Society Index
<https://ec.europa.eu/digital-single-market/en/desi>

- Servicios públicos digitales en España

5 Servicios públicos digitales	España		Grupo	UE
	puesto	resultado	resultado	resultado
DESI 2017	6	0,72	0,59	0,55
DESI 2016	5	0,71	0,56	0,51

	España				UE
	DESI 2017		DESI 2016		DESI 2017
	valor	puesto	valor	puesto	valor
5a1 Usuarios de la administración electrónica <small>% de usuarios de internet (año anterior)</small>	40 %	11	38 %	11	34 %
	2016	↑	2015		2016
5a2 Formularios precumplimentados <small>Resultado (0 a 100)</small>	67	10	68	9	49
	2016	↓	2015		2016
5A3 Compleción de los servicios en línea <small>Resultado (0 a 100)</small>	89	11	91	7	82
	2016	↓	2015		2016
5a4 Datos abiertos²⁴ <small>% de la puntuación máxima</small>	91 %	1	82 %	1	59 %
	2016	↑	2015		2016

Estrategia TIC 2015-2020 de España

- Garantizar que *todos los ciudadanos* dispongan de acceso electrónico a la Administración Pública ("digital por defecto").
- Se establece un *marco estratégico* global como continuación de la Agenda Digital
- Incluye la adopción de un programa de digitalización para construir una *infraestructura informática interoperable*.

Plan de transformación Digital de la AGE y sus Organismos Públicos

<https://administracionelectronica.gob.es/dam/jcr:898162f1-2682-483e-9e43-50f2d3a08eff/20151002-Plan-transformacion-digital-age-oopp.pdf>

Servicios públicos digitales multilingües



- Los Servicios públicos digitales multilingües facilitan el acceso de los ciudadanos a la Administración electrónica con independencia de su lengua de origen.
 - La *realidad multilingüe de España* debe poder reflejarse en los servicios públicos digitales.
 - El multilingüismo es la clave para facilitar los servicios públicos digitales transfronterizos en un *Mercado Único Digital europeo*.
- Las tecnologías del lenguaje, particularmente la traducción automática, facilitan la incorporación del multilingüismo a los SPDs.

Tecnologías del Lenguaje y Servicios Digitales



El Plan de Impulso de las Tecnologías del Lenguaje

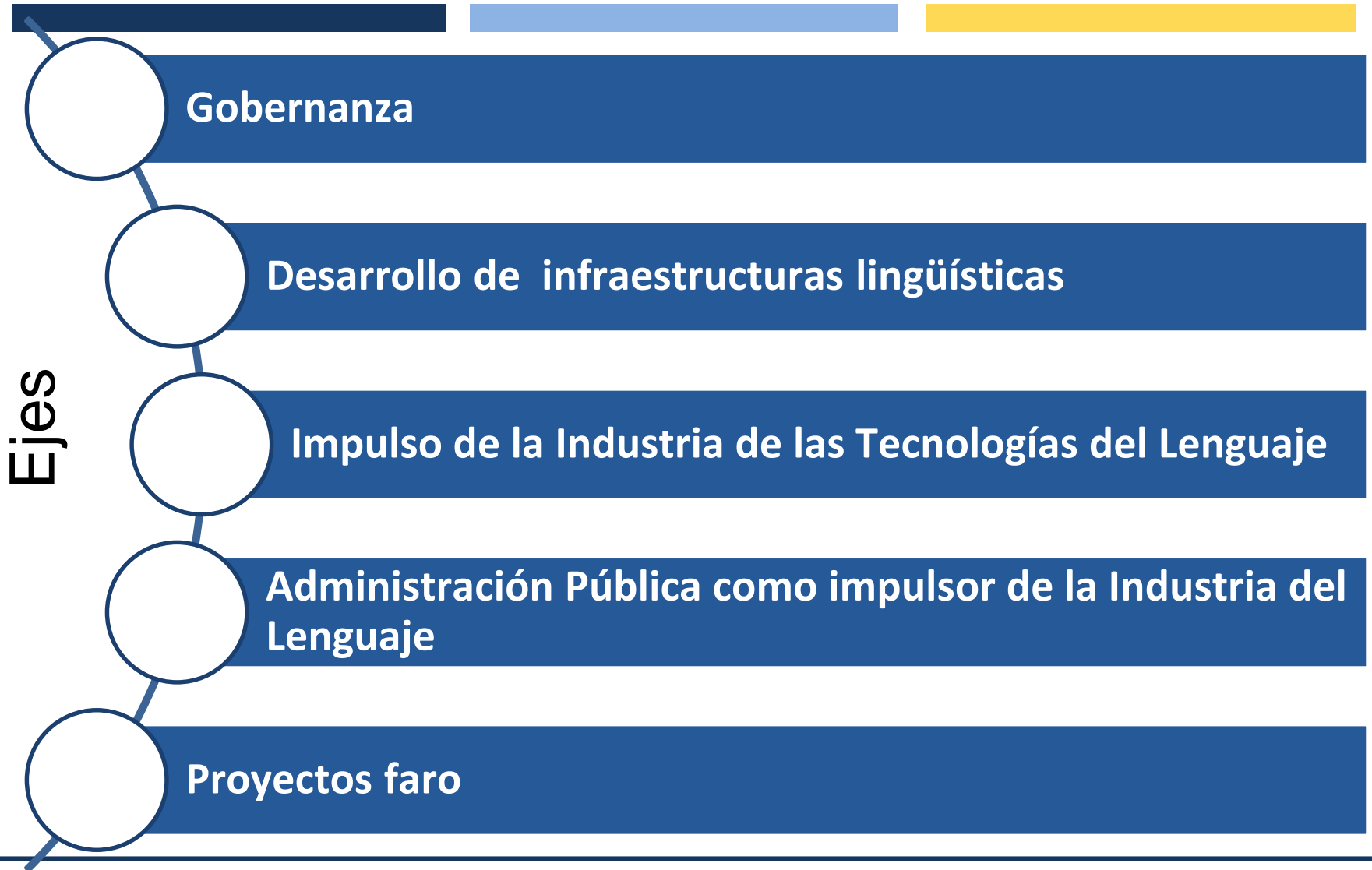


El **Plan de Impulso de las Tecnologías del Lenguaje** (Plan TL) tiene como objetivo fomentar la industria de las Tecnologías del Lenguaje y el desarrollo de las infraestructuras lingüísticas, contando con la Administración Pública como impulsora de la demanda.

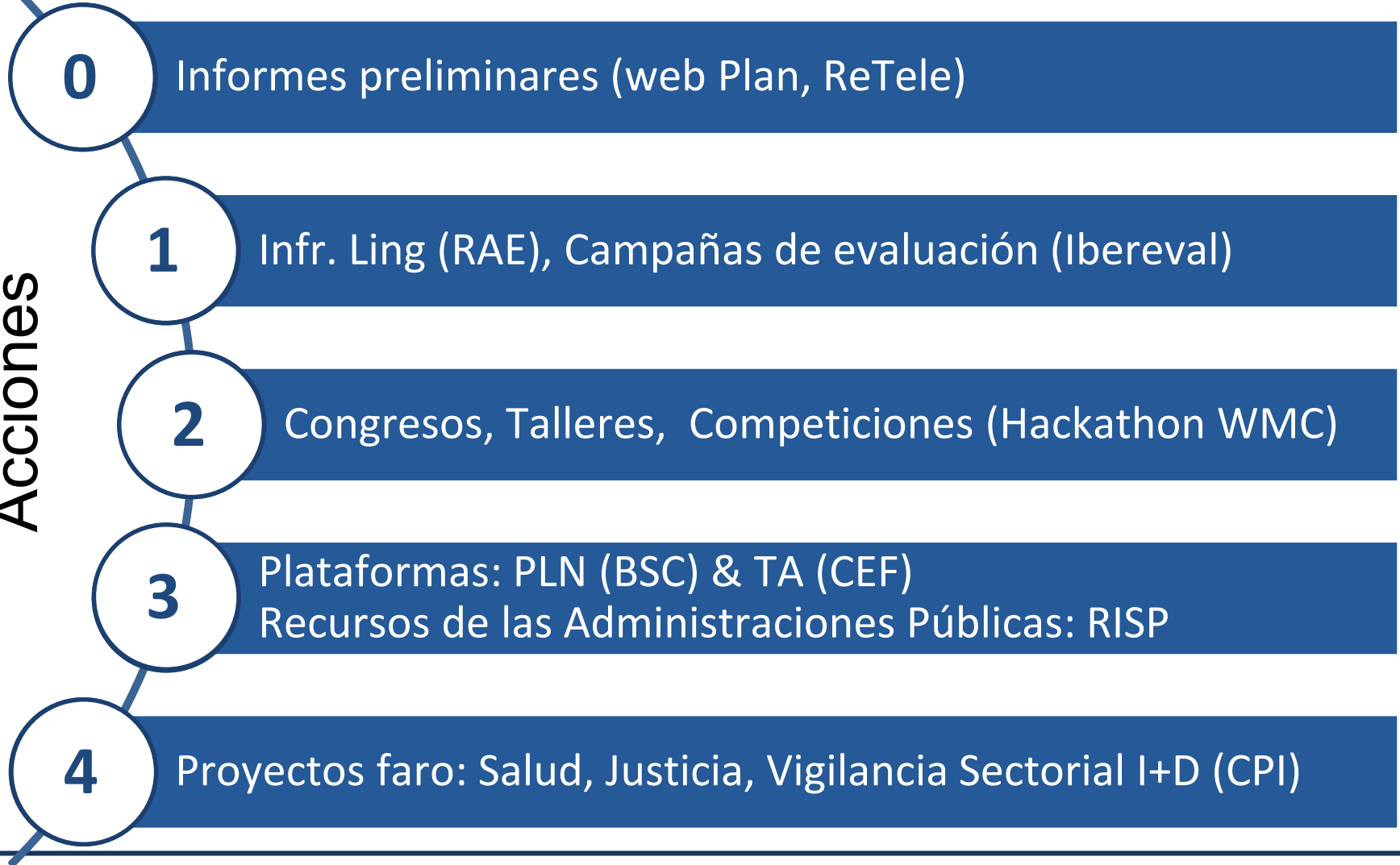
Plan TL
Plan de Impulso de las
Tecnologías del Lenguaje

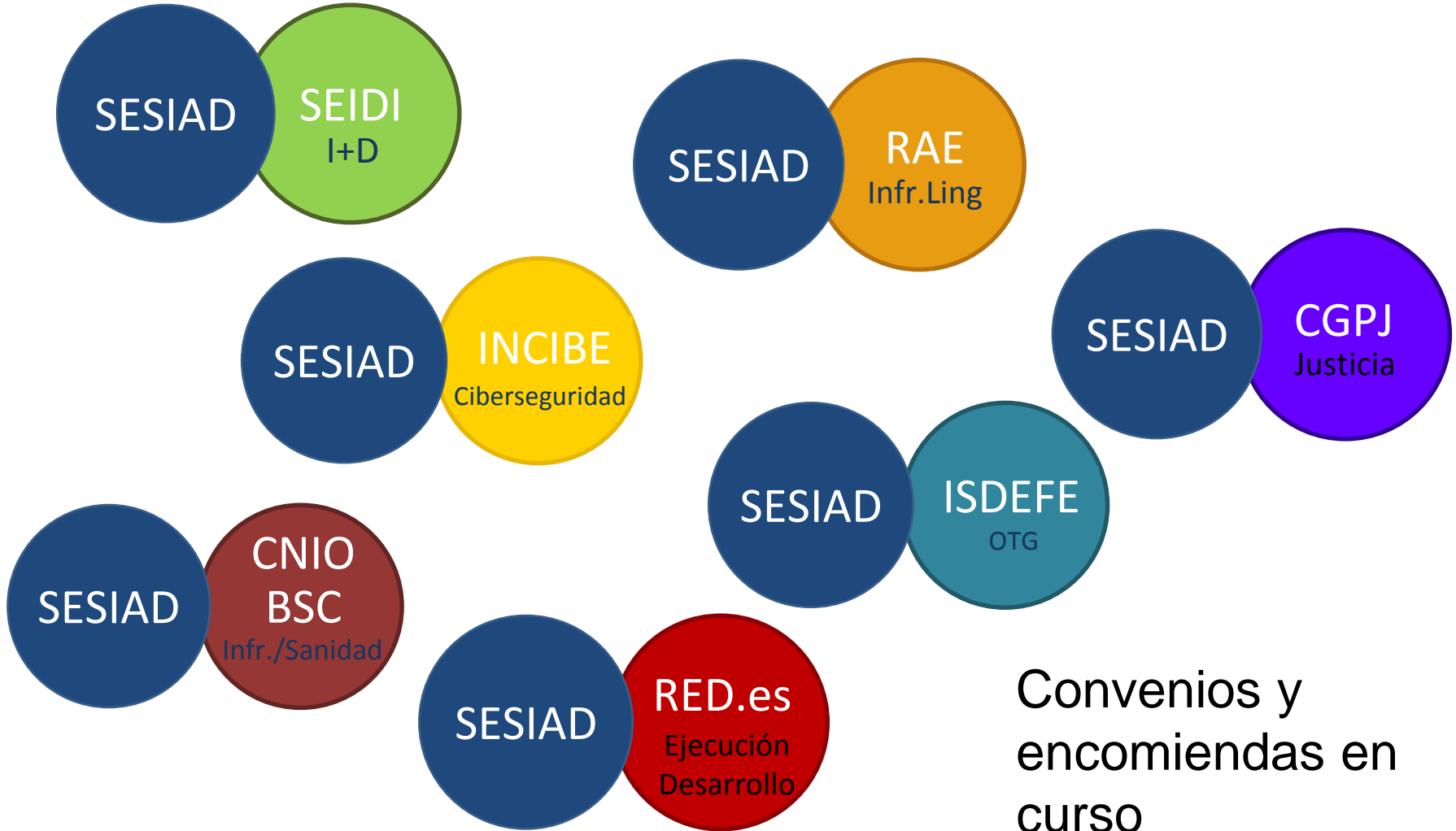


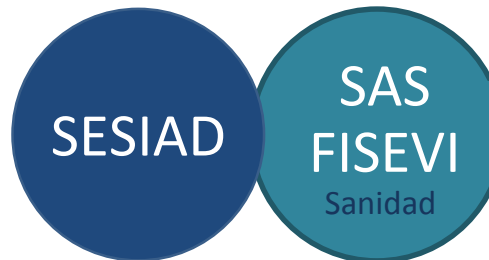
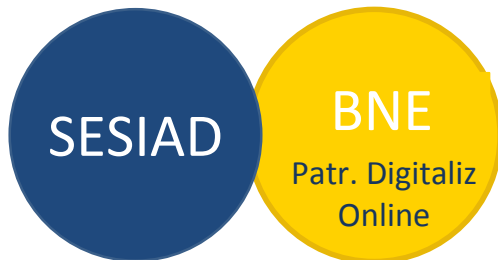
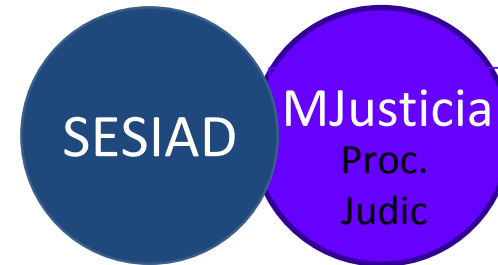
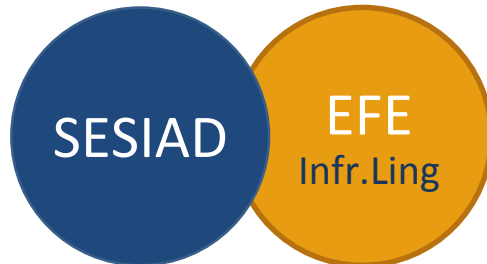
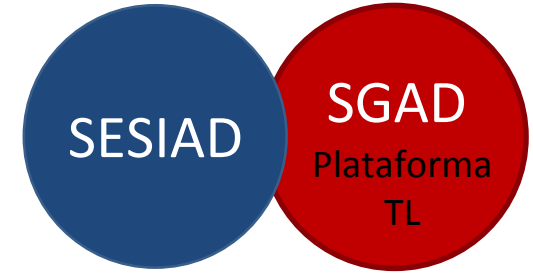
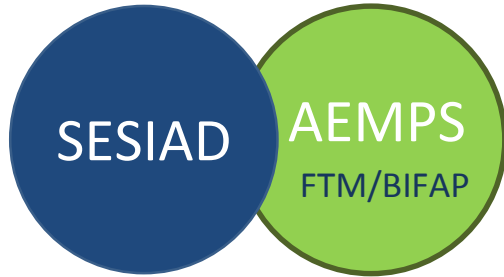
<http://www.plantl.es>



Acciones

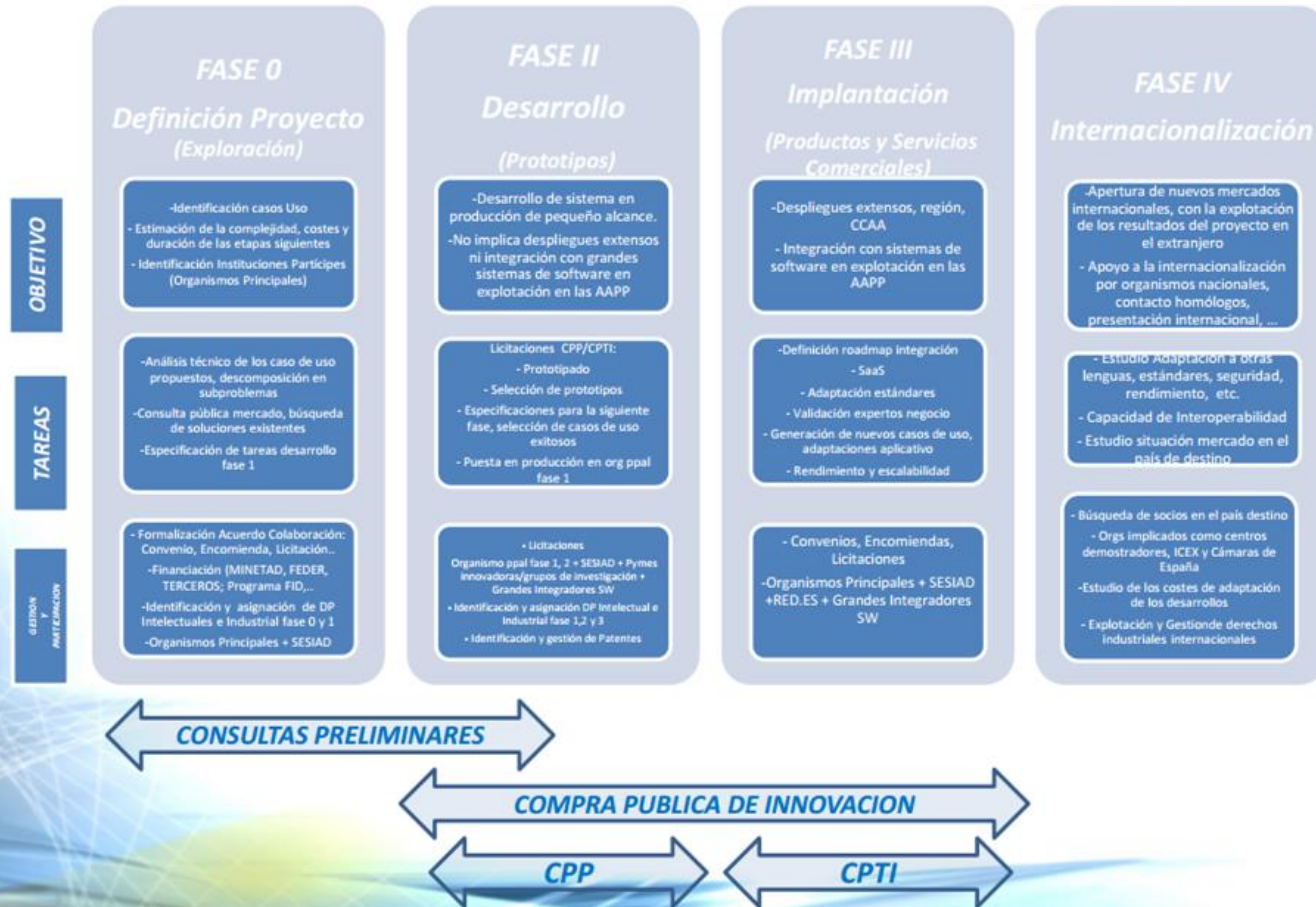






Convenios y
encomiendas en
desarrollo

CPI en el PLN



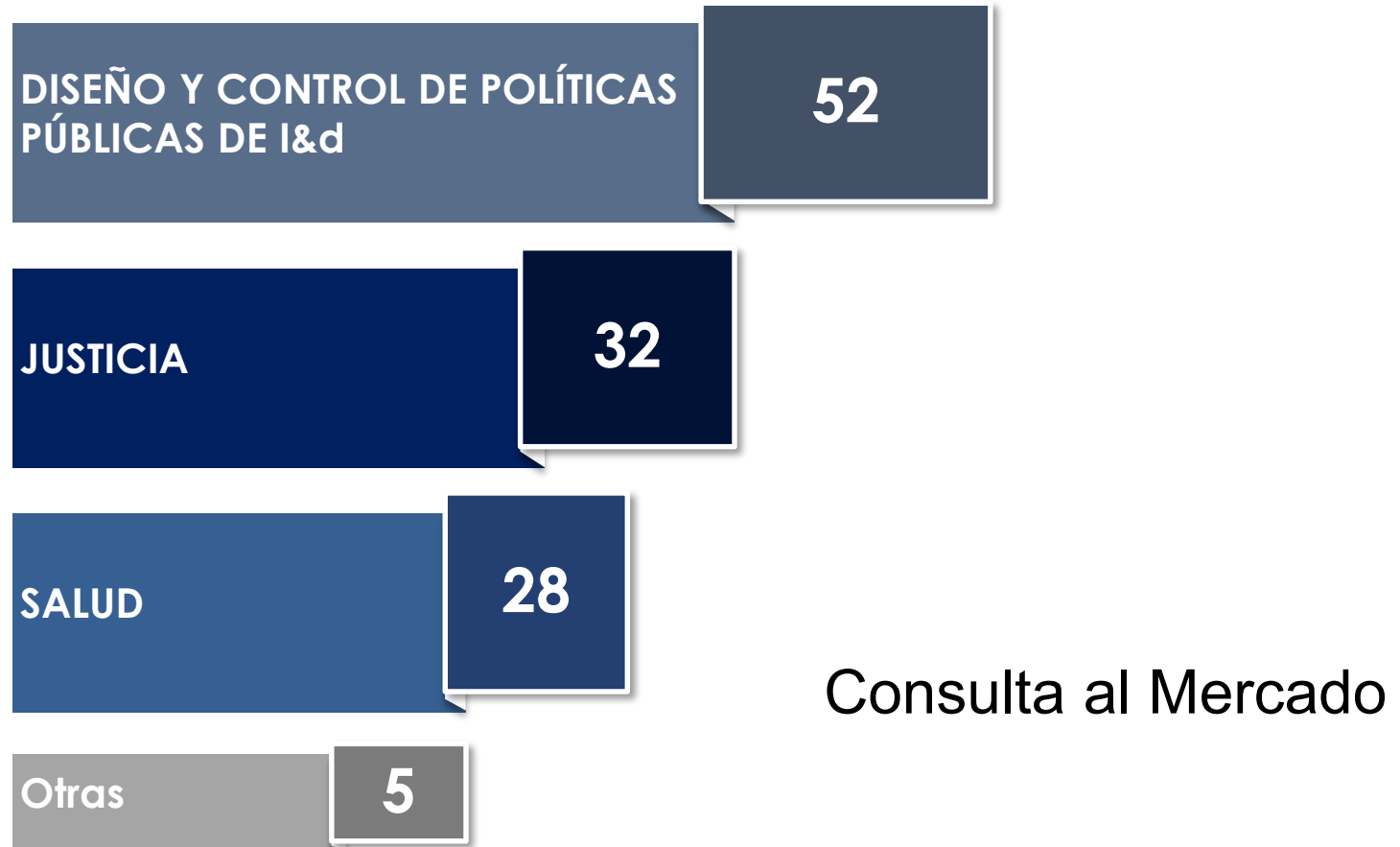
Compra Pública Innovadora (Instrumento de financiación)

- Fomentar la innovación
- Validar nuevas tecnologías
- Efecto demostrador incentivador de demanda
- Mejorar los servicios públicos a los ciudadanos

Consulta Pública (2017)

- Consulta preliminar al mercado
- Expresión de interés del sector público

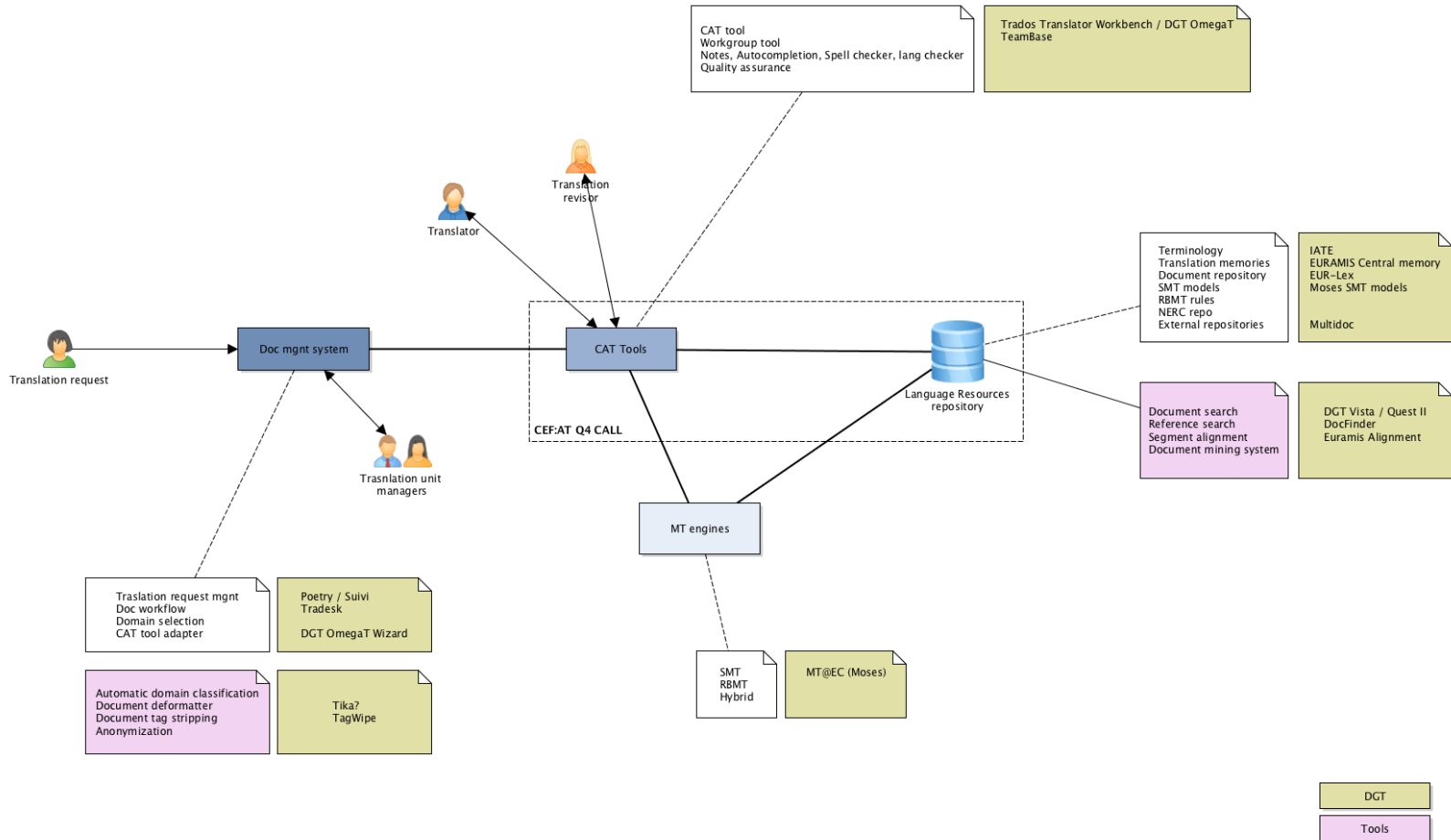
Plan de Impulso de las Tecnologías del Lenguaje



- Participación de SESIAD / PlanTL en proyectos CEF (Mecanismo “Conectar Europa”):
 - **iADAATPA**: Traducción automática inteligente, adaptada a dominio, para las Administraciones Públicas
 - Desarrollo de motores de Traducción Automática neuronal para su integración en eTranslation
 - **ELRI**: Infraestructura de Recursos Lingüísticos Europeos
 - Creación de nodos nacionales conectados al repositorio ELRC
- Proyectos TA: Red.es-SESIAD

- Participación de SESIAD / PlanTL en proyectos CEF (Mecanismo “Conectar Europa”):
 - [iADAATPA](#): Traducción automática inteligente, adaptada a dominio, para las Administraciones Públicas
 - Desarrollo de motores de Traducción Automática neuronal para su integración en eTranslation
 - [ELRI](#): Infraestructura de Recursos Lingüísticos Europeos
 - Creación de nodos nacionales conectados al repositorio ELRC
- Proyectos TA: Red.es-SESIAD

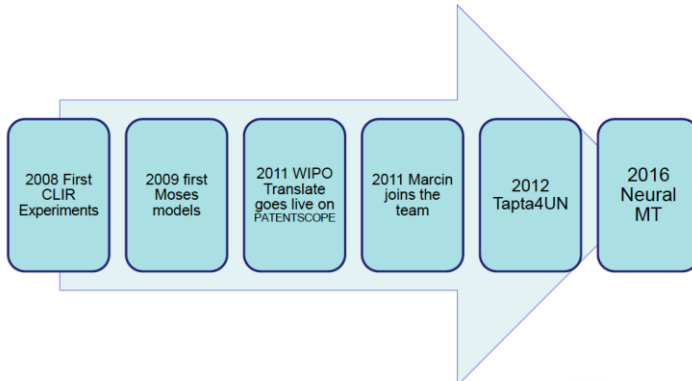
Computer Aided Translation Environment



- PCT : translation of patent titles/abstracts and international search reports
 - 10 languages: ar, de, en, es, fr, ja, ko, pt, ru, zh
 - 51 internal translators + ~20 agencies
 - ~ 355'000 words translated per day
- Madrid : translation of goods and services descriptions
 - 3 languages: en, fr, es
 - 7 internal translators, + 4 agencies
 - ~ 80'000 words translated per day
- Conferences : translation of documents, reports for assemblies member states...
 - 6 languages: ar, en, es, fr, ru, zh
 - 40 internal translators, ~ 150 external translators
 - ~ 42'100 words translated per day



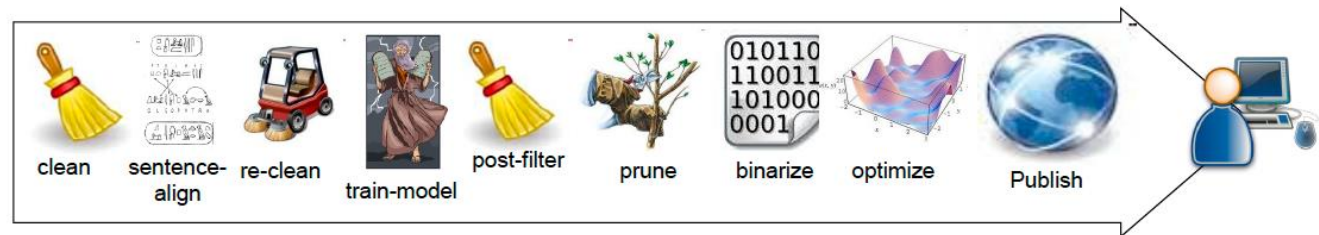
From: *Practical use of machine translation in international organizations*
Bruno Pouliquen. World Intellectual Property Organization (WIPO)



Tapta: from documents to MT



Fully automatic (preparation/training/publishing etc.)



Un ejemplo: WIPO



	Segments	Words
PT:	7M	209M
JA:	112M	4'000M
ZH:	62M	1'872M
FR:	18M	488M
DE:	40M	1'455M
RU:	13M	386M
KO:	44M	1'481M

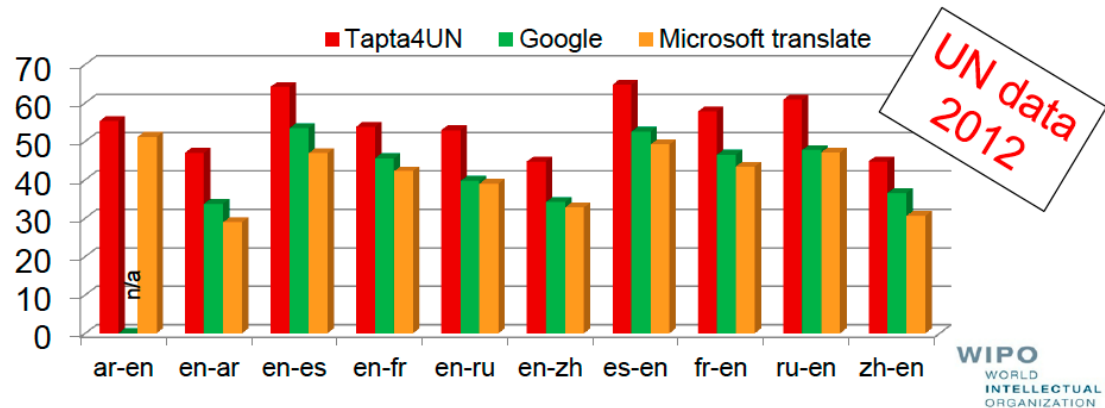
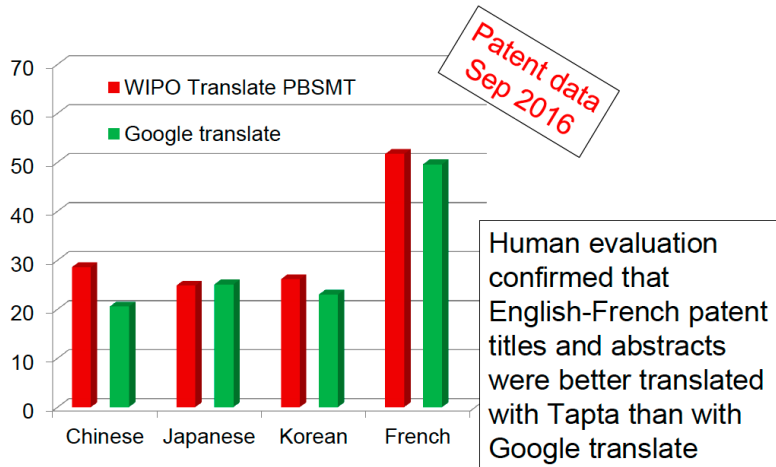
Patent data 2016

Inst. date	Organization	Words/day	Languages	Corpus size (Mw)
Feb 2011	PATENTSCOPE	210'000	de,en,es,fr,ja,ko,pt,ru,zh (ar)	4'000 en-zh
Sep 2012	International Telecommunication Union	400	en,fr,zh (es, ru, ar)	84 en-fr
Oct 2012	United Nations	210'000	ar,en,es,fr,ru,zh (de)	400 en-es
Feb	MADRID The International Trademark System	24'000	en,fr,es	22 en-fr
	WIPO PCT The International Patent System	113'000	de,en,es,fr,zh (ja,ko,pt,ru)	790 en-fr
	INTERNATIONAL MARITIME ORGANIZATION	2'000	en,fr,es, ar,ru (zh)	54 en-fr
	issa	prototype (not used)	de,en,es,fr (ar,ru,zh)	0.7 en-fr

Date	Organization	Words/day	Languages	Corpus size Mw
Aug 2014	International Labour Organization	prototype	en,es,fr, (ar,de,ru,zh)	34 en-fr
Nov 2014	WORLD TRADE ORGANIZATION	Prototype, almost in production	en,es,fr	400 en-fr
Jun 2015	Language Division	production	ar,en,fr,es,ru,zh (de)	55 en-fr
Jun 2015	Food and Agriculture Organization of the United Nations	production	ar,en,fr,es (ru,zh)	41 en-fr
Oct 2015	International Patent Classification	production	en,fr,pt,sk,rs,pl	0.7 en-fr
Fev 2016	EAPO	Production, not used	En,ru	
Sep 2016	The Global Fund To Fight AIDS, Tuberculosis and Malaria	Early production	en, fr, es, ru	n/a

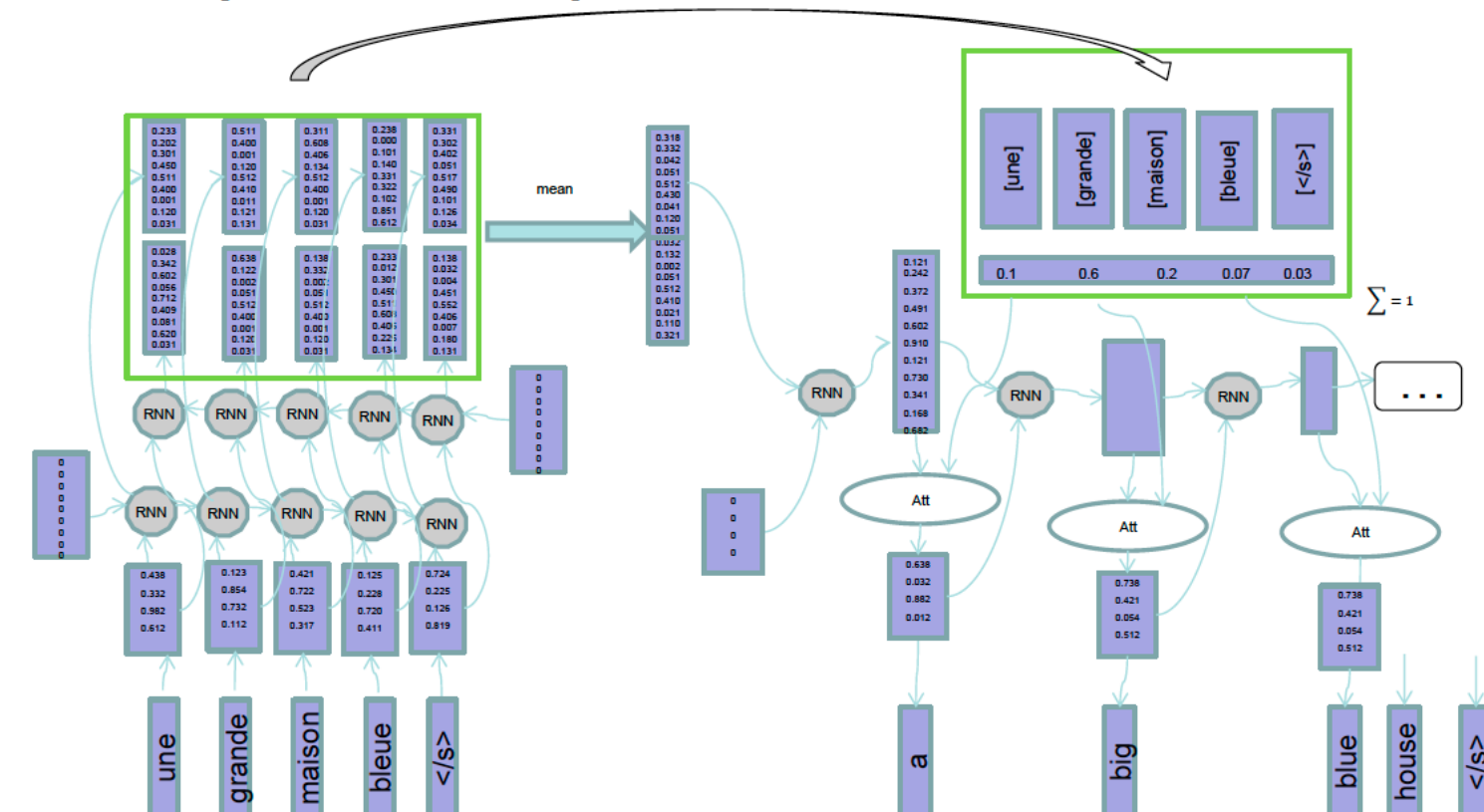
- 20 servers (running 24/7)
- 9 international organizations (production + prototypes)
- 300 cores
- 500'000 words translated per day

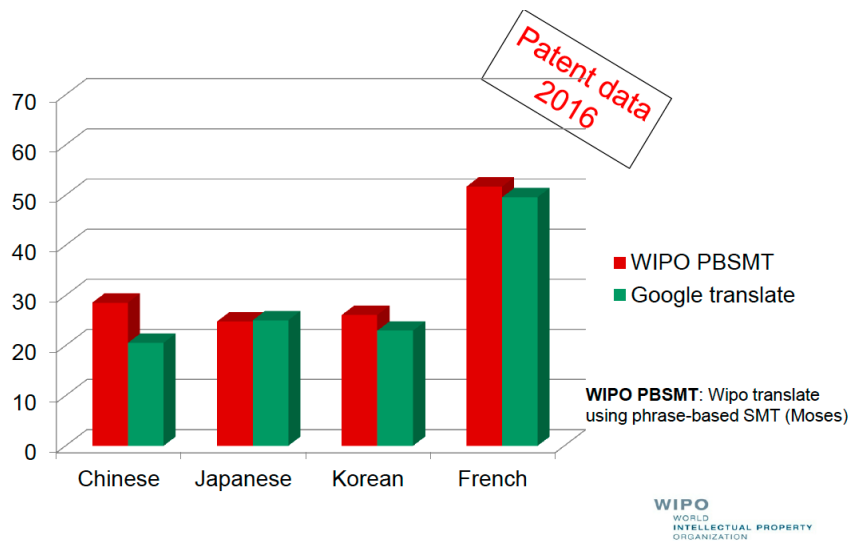
Un ejemplo: WIPO



Neural Machine Translation

Attentional encoder-decoder [Bahdanau et al., 2014], implemented in Nematus [Sennrich et al., 2016]





- Corpus paralelo sostenible: Coppa v2
- Servicio de mayor calidad
 - Adaptado a dominio
 - Terminología específica
- Control de confidencialidad

¿Preguntas?

Gracias por su atención

David Pérez Fernández

dperezf@minetad.es

www.plantl.es

Email: info@lr-coordination.eu

Website: www.lr-coordination.eu

