

Plan de Impulso de las Tecnologías del Lenguaje

20 de Octubre de 2015

Industria
Energía
Turismo
Telecomunicaciones
Sociedad de la Información

1. Introducción

2. Plan de Impulso de las Tecnologías del Lenguaje

- Eje 1: Infraestructuras Lingüísticas
- Eje 2: Impulso de la Industria
- Eje 3: La Administración
- Eje 4: Proyectos Faro

3. Mapa del Plan

OPORTUNIDADES

- Alto potencial de **internacionalización** del español y de cooperación con **Iberoamérica**.
- Nuevos **servicios públicos** de la AGE para ciudadanos y empresas en sectores estratégicos (sanidad, turismo, seguridad, etc.).
- **Mercado en fuerte crecimiento** asociado a la innovación y a el desarrollo.





FORTALEZAS

- **Buena gobernanza** del idioma español (RAE, Asociación de Academias de la Lengua Española en Iberoamérica).
 - **Alto nivel investigador** en PLN, con una adecuada coordinación (SEPLN).
 - **Recursos lingüísticos ya disponibles en la AGE** como fuente principal de datos para la industria y la investigación.
- 



DEBILIDADES

- **PYMES no alcanzan la capacidad industrial** suficiente para:
 - Competir en el mercado **internacional**.
 - Completar la **cadena de valor** en España.
- **Dificultad de transferencia de conocimiento** del sector investigador a la industria debido al carácter transversal y multidisciplinar del procesamiento de LN.



AMENAZAS

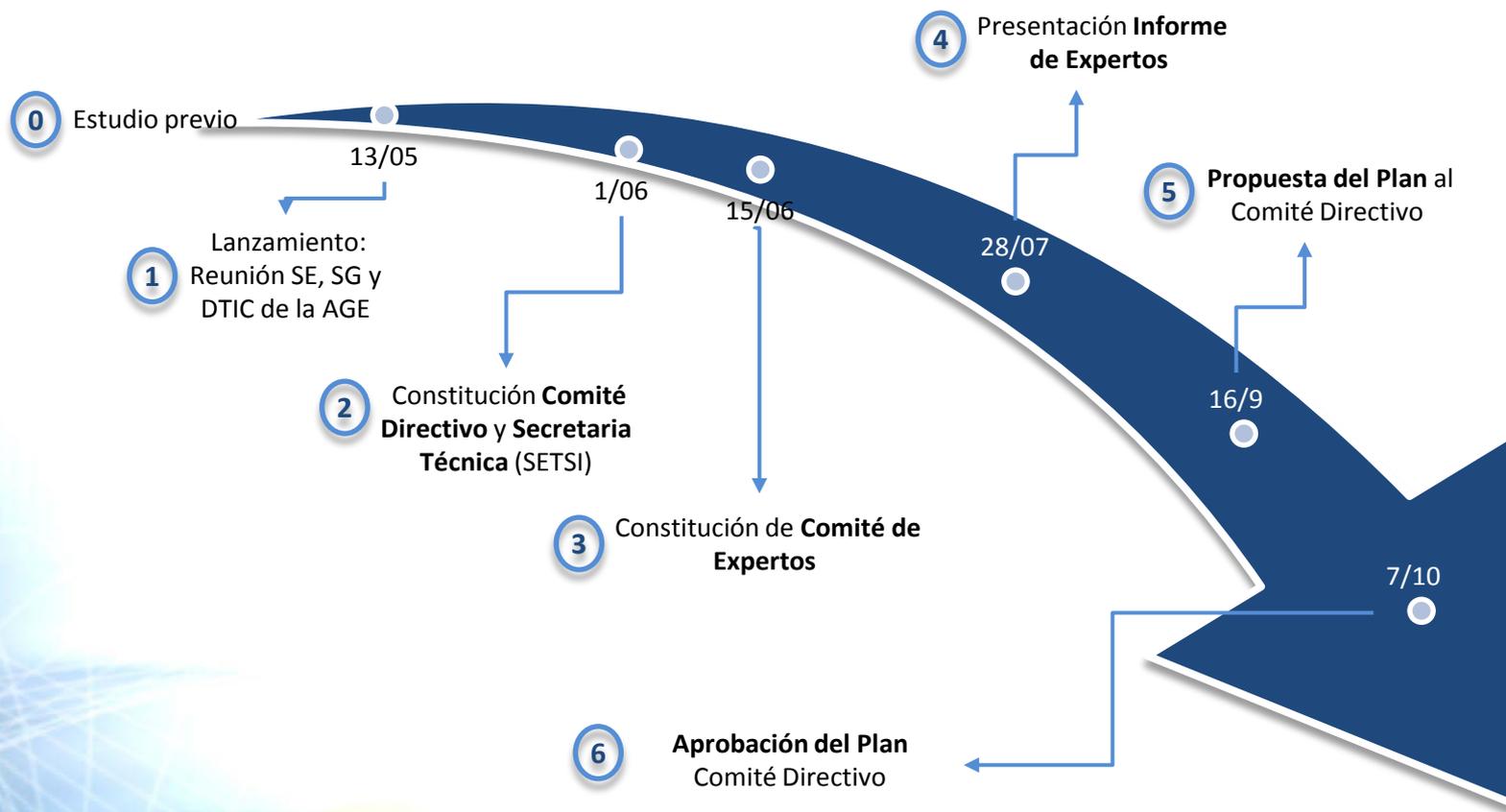
- **Pérdida de competitividad** económica e industrial de España e Iberoamérica frente a EEUU.
- **Subdesarrollo digital** del español.
- **Extinción digital** de las lenguas **cooficiales**.
- **Fuga de investigadores y profesionales** y deterioro del sector investigador español.

Conclusiones:

- El sector de las tecnologías del lenguaje representa un **sector emergente transversal** vinculado a la innovación con capacidad para promover el crecimiento, la competitividad y el empleo de calidad.
- Su **desarrollo es imparable**, pero si no aprovechamos la oportunidad, serán otros los que ocupen este espacio.
- España tiene los medios, pero es necesario **impulsar y coordinar las actuaciones** desde la Administración General del Estado, en coordinación con las Comunidades Autónomas y las Unión Europea, y en colaboración con Iberoamérica, para aprovechar la oportunidad.

Introducción

- Calendario de elaboración del Plan:



Plan de Impulso de las Tecnologías del Lenguaje.



- Ejes de actuación:

1. Desarrollo de infraestructuras lingüísticas.

2. Impulso de la industrias de las Tecnologías del Lenguaje.

Mejora de la **visibilidad** y de la **transferencia** del sector.

Apoyo a la **internacionalización** y comercialización del sector.

3. La Administración como impulsor de la Industria del Lenguaje.

Plataformas de procesamiento de lenguaje natural y traducción automática en las Administraciones Públicas.

Recursos lingüísticos de las administraciones públicas y política de **reutilización** de la información del sector público.

- Proyectos Faro.

Eje 1: Desarrollo de infraestructuras lingüísticas

- Objetivos:
 - **Impulsar la industria PLN en español y lenguas cooficiales** poniendo a su disposición infraestructuras lingüísticas de propósito general.
 - **Reducir la distancia** que nos separa del **inglés**.
 - **Llevar la industria a la frontera de la innovación** facilitando el uso abierto de infraestructuras lingüísticas básicas.



Presupuesto: **30 M€**.



Eje 1: Desarrollo de infraestructuras lingüísticas



Medidas:

- Seleccionar normas técnicas de **interoperabilidad**, políticas de **licencias** y mecanismos de **protección de datos personales** adecuados para la generación de recursos lingüísticos.
- Adquirir o desarrollar **herramientas comunes** para la generación y evaluación de infraestructuras lingüísticas.
- Elaborar y ejecutar un plan de desarrollo de infraestructuras lingüísticas. Realizar un **inventario** de infraestructuras lingüísticas actualmente disponibles. Evaluar la evolución de la cantidad, calidad y disponibilidad de infraestructuras lingüísticas.
- Facilitar el **acceso público** a las infraestructuras lingüísticas existentes.

Eje 2: Impulso de las industrias de las Tecnologías del Lenguaje

Línea 1: Mejora de la **visibilidad** y la **transferencia** del sector.

Objetivos:

- Mejorar la visibilidad y capacidad de internacionalización del sector del NLP y TA.
- Trasladar la excelencia investigadora española a la industria.



Garantizar la disponibilidad de doctores especialistas y formar nuevos investigadores y desarrolladores en estas áreas.

Presupuesto: **1 M€.**



Eje 2: Impulso de las industrias de las Tecnologías del Lenguaje



Medidas:

- Coordinación con otras Administraciones y con **Iberoamérica**.
- Mejorar la **formación**: Inclusión de asignaturas específicas sobre tecnologías del lenguaje en el currículum académico, impulso de la creación de contenidos formativos en línea [MooCs], hackathones, apoyo a programas de doctorado industrial, másters y becas especializadas.
- Elevar la **visibilidad**: Jornadas de formación básica para PyMES y profesionales; Congresos, foros y participación en ferias nacionales e internacionales. Impulso a centros demostradores en línea (SaaS).

Eje 2: Impulso de las industrias de las Tecnologías del Lenguaje

- **Línea 2:** Apoyo a la **internacionalización** y comercialización.

Objetivos:

- Mejorar la internacionalización, especialmente en el mercado iberoamericano y norteamericano.
- Aprovechar mejor herramientas de **financiación** existentes.
- Papel tractor de las **grandes empresas** españolas asentadas en el mercado iberoamericano.



Presupuesto: **1 M€.**



Eje 2: Impulso de las industrias de las Tecnologías del Lenguaje

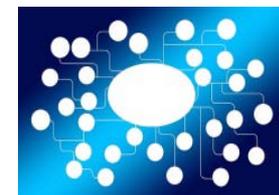


Medidas:

- Cooperación con **Iberoamérica**.
- Coordinación con AECID, Red de Consejeros Científicos en el Exterior, Acción Estratégica de Economía y Sociedad Digital, Invest in Spain ...
- Ferias, congresos o eventos en los que, con la colaboración del ICEX, pueda darse difusión a los productos y proyectos realizados por las empresas españolas del sector.

Eje 3: La Administración como impulsor de la Industria del Lenguaje

- **Línea 1: Plataformas comunes** de procesamiento de lenguaje natural y traducción automática en las Administraciones Públicas.



Objetivos:

Presupuesto: **4 M€.**

- Promover servicios avanzados para el ciudadano.
- Mejorar el funcionamiento de la Administración.
- CORA: reutilizar, simplificar y alcanzar economías de escala.
- Mejorar la accesibilidad a personas con necesidades especiales.



Eje 3: La Administración como impulsor de la Industria del Lenguaje



Diseño y creación de una **plataforma común de procesamiento del lenguaje natural y de traducción automática para las Administraciones Públicas** con los siguientes requisitos esenciales:

- **Facilitar la puesta en marcha de servicios** avanzados basados en procesamiento de lenguaje natural y traducción automática en la Administración General del Estado y las Comunidades Autónomas.
- Desarrollar una infraestructura **escalable** basada en componentes para el procesamiento paralelo de grandes corpus documentales.
- Mantener las garantías de **confidencialidad** propias de los servicios públicos.
- Agregar distintos componentes y recursos lingüísticos al flujo de procesamiento lingüístico con diferentes modelos de licenciamiento y modos de ejecución del procesamiento.

Eje 3: La Administración como impulsor de la Industria del Lenguaje



- **Herramientas comunes** de anonimización, editores, post-edición de traducciones automáticas, etc.
- Esta plataforma **pondrá a disposición infraestructuras lingüísticas** de propósito general (Eje 1) y específicos de dominio (proyectos verticales, Eje 4).
- Será un lugar para la explotación y estandarización de los recursos lingüísticos generados al amparo de la política de Reutilización de la Información del Sector Público (**RISP**).
- Permitirá distintos modelos de ejecución y distribución (embebido, cluster local, remoto y ejecución en centros de supercomputación).
- Coordinación otros programas/proyectos europeos (CEF.AT, ...).

Eje 3: La Administración como impulsor de la Industria del Lenguaje

- **Línea 2:** Recursos lingüísticos de las Administraciones Públicas y política de reutilización de la información del sector público.



Objetivos:

- Aprovechar el enorme potencial de los **datos abiertos de interés lingüístico** como recursos abiertos de la industria del lenguaje.
- **Estandarizar**, modelo de **licencias** y **protección de datos personales**.
- Desarrollo de **herramientas comunes**.

Presupuesto: **2 M€**.



Proyectos Faro



Objetivos:

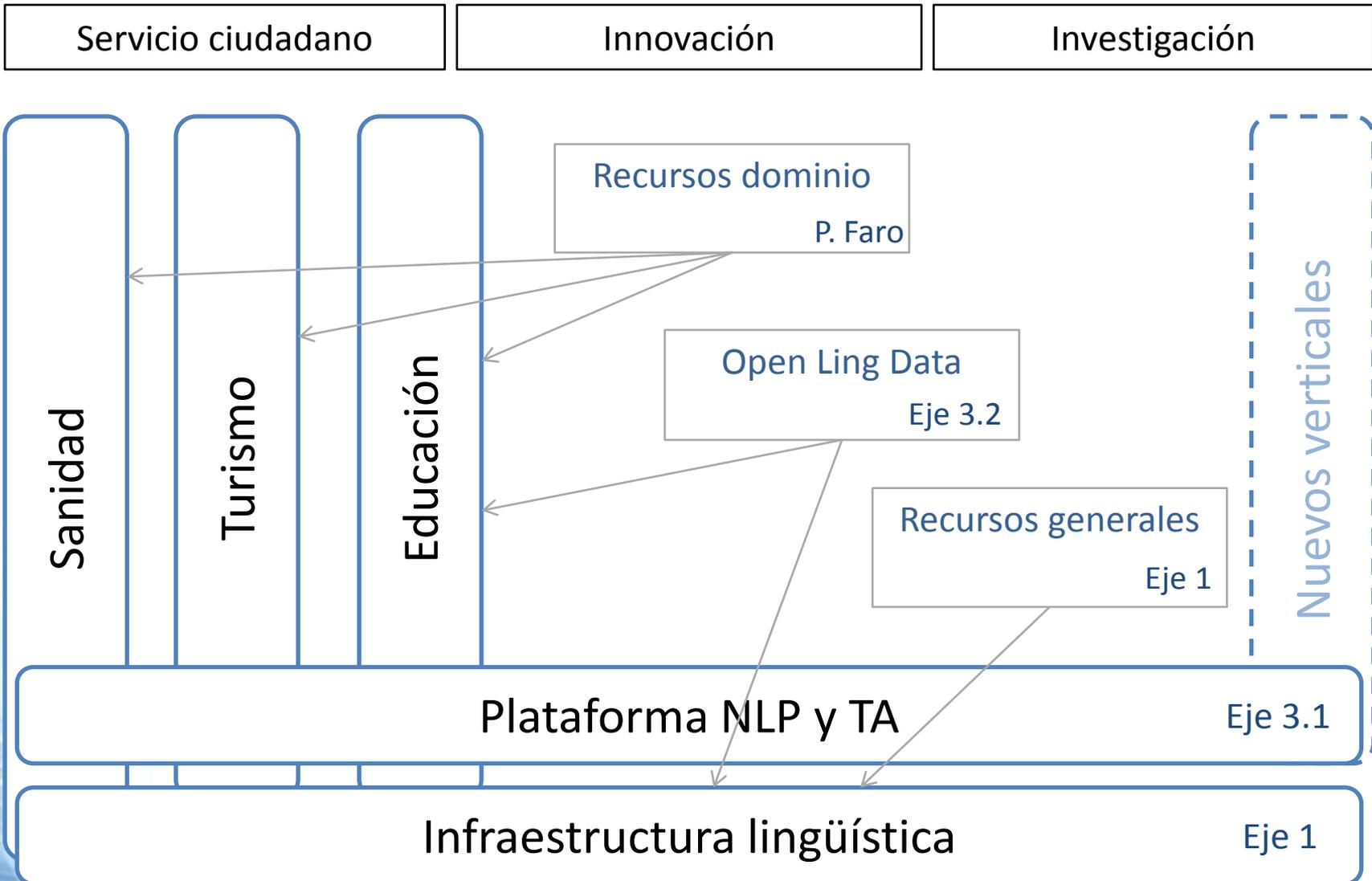
- Ofrecer **nuevos servicios** públicos o mejorar la capacidad y calidad de servicios públicos existentes mediante la aplicación de las tecnologías del procesamiento del lenguaje natural.
- Servir de **demostración** para otros sectores, dentro y fuera de España, de las capacidades y beneficios de las tecnologías del lenguaje.
- Generar **infraestructuras reutilizables** para otros proyectos del mismo u otro dominio.
- **Aplicación** inmediata de medidas horizontales del Plan; uso de **infraestructuras lingüísticas** generales y **plataformas** comunes.

Presupuesto: **49 M€.**

Criterios de búsqueda de los **Proyectos Faro**:

- **Compromiso** de los órganos competentes. Asegurar el liderazgo de los que conocen bien el problema y tienen competencias para resolverlo.
- **Concreción**. Responder a problemas ya identificados que justifiquen la idoneidad y el momento de la puesta en marcha del proyecto.
- Elevado **impacto** económico y social.
- Generación de **recursos reutilizables**.
- **Sinergias** con las demás medidas del Plan de Impulso de las Tecnologías del Lenguaje.
- Especial atención a la adquisición de **experiencia para futuros proyectos**.

Mapa del Plan



Gracias.