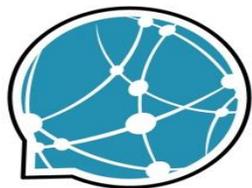


Plan TL

Plan de Impulso de las
Tecnologías del Lenguaje



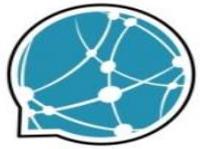
GOBIERNO
DE ESPAÑA

MINISTERIO
DE ECONOMÍA
Y EMPRESA

PLAN DE IMPULSO DE LAS TECNOLOGÍAS DEL LENGUAJE

Situación actual y próximas acciones

1. Plan de Impulso de Tecnologías del Lenguaje
2. Eje 0. Gobernanza
3. Eje 1. Desarrollo de infraestructuras lingüísticas
4. Eje 2. Impulso de la industria de las tecnologías del lenguaje
5. Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje
6. Eje 4. Proyectos Faro



- 1. Plan de Impulso de Tecnologías del Lenguaje**
2. Eje 0. Gobernanza
3. Eje 1. Desarrollo de infraestructuras lingüísticas
4. Eje 2. Impulso de la industria de las tecnologías del lenguaje
5. Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje
6. Eje 4. Proyectos Faro

- Rama de **Inteligencia Artificial**.
- **Desarrollo en España del sector** del procesamiento del lenguaje natural, de la traducción automática o asistida por ordenador y de los sistemas conversacionales.
 ➔ **Emprendimiento**.
- Aprovechar estas novedosas capacidades para mejorar el **servicio público**.

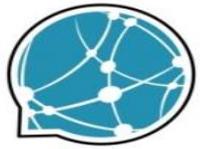
 www.PlanTL.gob.es

 plantechnologiaslenguaje@mineco.es

 Plan TL - Tecnologías Lenguaje

El Plan se articula en los siguientes cinco ejes:

- Eje 0: *Gobernanza.*
- Eje 1: *Desarrollo de infraestructuras lingüísticas.*
- Eje 2: *Impulso de la industria de las tecnologías del lenguaje.*
- Eje 3: *La Administración Pública como impulsora de la Industria del Lenguaje.*
- Eje 4: *Proyectos Faro.*



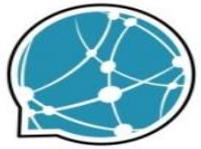
1. Plan de Impulso de Tecnologías del Lenguaje
2. **Eje 0. Gobernanza**
3. Eje 1. Desarrollo de infraestructuras lingüísticas
4. Eje 2. Impulso de la industria de las tecnologías del lenguaje
5. Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje
6. Eje 4. Proyectos Faro

-Coordinación,

-Colaboración,

-Intercambio de ideas, documentación e información
entre los diferentes agentes

- Estudio de sistemas conversacionales
- Gobernanza de las Infraestructuras Lingüísticas



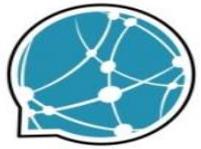
1. Plan de Impulso de Tecnologías del Lenguaje
2. Eje 0. Gobernanza
- 3. Eje 1. Desarrollo de infraestructuras lingüísticas**
4. Eje 2. Impulso de la industria de las tecnologías del lenguaje
5. Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje
6. Eje 4. Proyectos Faro

Eje 1. Desarrollo infraestructuras lingüísticas

- Definición de II.LL.
- Objetivos:
 - Impulsar la industria.
 - Frontera de la innovación.
 - Reducir la distancia respecto del inglés.
 - Infraestructura pública: mejorar procesos y garantizar acceso e interoperabilidad.

Eje 1. Desarrollo infraestructuras lingüísticas

- Trabajos en infraestructuras
- Informes
- Infraestructuras desarrolladas: corpus
- Campañas de evaluación
- Retos: IEC, EFE, otros grupos editoriales, I.Cervantes



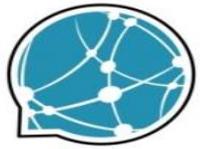
1. Plan de Impulso de Tecnologías del Lenguaje
2. Eje 0. Gobernanza
3. Eje 1. Desarrollo de infraestructuras lingüísticas
4. **Eje 2. Impulso de la industria de las tecnologías del lenguaje**
5. Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje
6. Eje 4. Proyectos Faro

Eje 2. Impulso de la industria de TL

- Visibilidad
- Formación, transferencia del conocimiento
- Internacionalización.

Eje 2. Impulso de la industria de TL

- Estudio sobre una estrategia formativa preliminar.
- **Dos Hackathon de Tecnologías del Lenguaje**, varios **infodays** en Madrid, Murcia y Sevilla, con objeto de dar visibilidad a las Tecnologías del Lenguaje
- Estudio de caracterización del sector de Tecnologías del Lenguaje en España, México y Argentina y en la Unión Europea (proyectos europeos, planes nacionales de otros países UE...).
- Retos: INAP



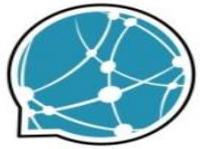
1. Plan de Impulso de Tecnologías del Lenguaje
2. Eje 0. Gobernanza
3. Eje 1. Desarrollo de infraestructuras lingüísticas
4. Eje 2. Impulso de la industria de las tecnologías del lenguaje
5. **Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje**
6. Eje 4. Proyectos Faro

Eje 3. La Admón. Pública como impulsora de la Industria del Lenguaje

- Plataformas comunes.
- Recursos lingüísticos de las Administraciones Públicas y Política de Reutilización de la Información del Sector Público.

Eje 3. La Admón. Pública como impulsora de la Industria del Lenguaje

- Plataformas: BSC-CNS
- Proyectos europeos.
- Informe Datos abiertos.
- Retos: otras plataformas, SGAD



1. Plan de Impulso de Tecnologías del Lenguaje
2. Eje 0. Gobernanza
3. Eje 1. Desarrollo de infraestructuras lingüísticas
4. Eje 2. Impulso de la industria de las tecnologías del lenguaje
5. Eje 3. La Administración Pública como impulsora de la Industria del Lenguaje
- 6. Eje 4. Proyectos Faro**

Eje 4. Proyectos Faro

- Poner en valor desde el primer momento las demás medidas del Plan.
- Servicios públicos concretos de alto impacto social.
- Productos y servicios acabados.
- Demostrar las capacidades y beneficios de las tecnologías del lenguaje.

Eje 4. Proyectos Faro

- Sanidad
- Justicia
- Cultura
- Inteligencia Competitiva
- Otros



Resultados esperados (Output tangibles) ¿Qué se pretende conseguir?

- **Recursos para la industria:**
Corpus clínicos anotados y anonimizados
Guías clínicas
Protocolos
- **Componentes SW para procesar HCE.**
- **Herramientas avanzadas para médicos.**
- **Plataformas PLN y TA.**
- **Plataforma Campañas Evaluación:**
Círculo virtuoso de mejora de la técnica asociando competición y licitación

Beneficiarios por sectores ¿A quién servirían estos resultados?



- **INDUSTRIA TL:**
Recursos
Participación en campañas de evaluación
- **MÉDICOS:**
Herramientas avanzadas de asistencia a la decisión
- **PACIENTES:**
Apoyo a pacientes crónicos
Mejor atención al paciente
Adaptación a un lenguaje cercano
- **GERENTES SANIDAD:**
Apoyo en diseño de políticas públicas
Gestión más eficaz de ayudas I+D+i



¿Cómo se implementan?



Fase 0: Convenio

Fase 1: Licitaciones y prototipos iniciales

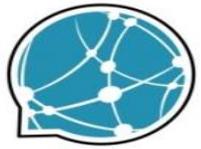
Fase 2 Despliegue

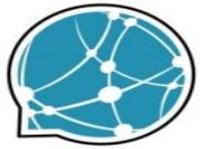
Fase 3 Internacionalización

CASOS DE USO

- **SAS-FISEVI: Procesar HCE (ictus, cancer, pruebas diagnósticas y tratamiento)**
- **Hospital 12 Oct: Terminología, Estandarización y Formación personal de hospitales**
- **AEMPS (BIFAP y FTM): Proc. HCE (AP y Pediatría) y Base Conocimiento Farmacológica**
- **ISCIII y RANM: Clasificadores automáticos DESC y Apertura corpus**
- **FCRi: Traducción automática a lenguas cooficiales en dominio de Salud y comparación de I+D+i territorial (CCAA, España, Europa)**
- **Hospital Clínic: Etiquetado CIE-10 y base conocimiento farmacológica**
- **Hospital Son Espases: Procesar HCE (infecciones hospitalarias)**
- **Aguas: Procesar HCE (informes alta ictus)**
- **MSSSI-FID: Colaboración en evaluación proyectos FID Salud- TL**
- **IACS: Procesamiento guías clínicas (extracción procedimientos)**
- **Canarias: Sistema conversacional de asistencia a atención primaria multilingüe**
- **Sanitas: Procesamiento HCE**

SANIDAD





Resultados esperados (Output tangibles) ¿Qué se pretende conseguir?

- Recursos para la industria
Corpus de textos legales
Terminologías
- Componentes SW
Anonimizadores
Transcripción de vistas judiciales
Detección de eventos
Enlazar textos con leyes
Análisis de conjunto

Beneficiarios por sectores ¿A quién servirían estos resultados?



- El sector de justicia
Jueces
Gestores
- Ciudadanos
- Industria TL
- Industria de servicios legales informatizados
Capacitación en TL y Empleo
- Impacto
 - Visibilidad emprendedores y PyMEs
 - Demanda de TL por AAPP
 - Ahorro y eficiencia AAPP

JUSTICIA

Convenio



¿Cómo se implementan?



Fase 0: Convenio

Fase 1: Licitaciones y
prototipos iniciales

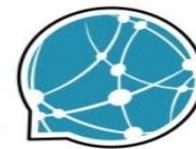
Fase 2 Despliegue

Fase 3 Internacionalización

CASOS DE USO

- CENDOJ:
Sentencias anonimizadas
Herramientas para jueces
Detección de eventos
- JUSTICIA:
Procesamiento los expedientes judiciales
Transcripción de vistas judiciales

Corpus
Guías de anonimización
Herramientas



Resultados esperados (Output tangibles) ¿Qué se pretende conseguir?

Beneficiarios por sectores ¿A quién servirían estos resultados?



Recursos para la industria:

- Corpus textuales (RAE-BNE)
- Corpus audiovisuales
- Transcripción voz a texto (RTVE)

Componentes SW

- Anotadores
- Transcriptors de voz-a texto

Plataformas

- Plataforma PLN
- Plataforma Campañas Evaluación

- Sector TL (PYMES e Academia)
- Sociedad: Accesibilidad
- Humanidades digitales

Impacto

- Capacitación en TL y Empleo
- Visibilidad emprendedores y PyMEs
- Demanda de TL por AAPP
- Ahorro y eficiencia AAPP

CULTURA
BNE RTVE RAE

Convenio



¿Cómo se implementan?



Fase 0: Convenio

Fase 1: Licitaciones y
prototipos iniciales

Fase 2 Despliegue

Fase 3 Internacionalización

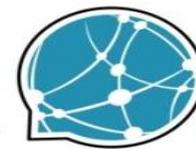
CASOS DE USO

- Corpus
- Herramientas

- RAE:
Recopilación de un corpus de noticias de 300 millones de palabras.
Anotación del corpus.

- BNE:
Asistencia a la mejora de los resultados OCR
Asistencia a la catalogación
Grafos de conocimiento enlazado
Mejorar accesibilidad al patrimonio cultural

- RTVE:
Corpus audiovisual
Entrenar transcriptors de voz



Resultados esperados (Output tangibles)
¿Qué se pretende conseguir?

Beneficiarios por sectores
¿A quién servirían estos resultados?



- Componentes SW
- Corpus
- Herramientas avanzadas para apoyo en el diseño de políticas públicas.
- Herramientas avanzadas para detección de fraude
- Campañas Evaluación: Círculo virtuoso de mejora de la técnica asociando competición y licitación

- Responsables de políticas públicas
- Gestor de ayudas
- Evaluadores de ayudas (i.e. detección de fraude)
- Investigador/Desarrollador
- Empresas/Org. Públicos Investigación

INTELIGENCIA COMPETITIVA

Convenio

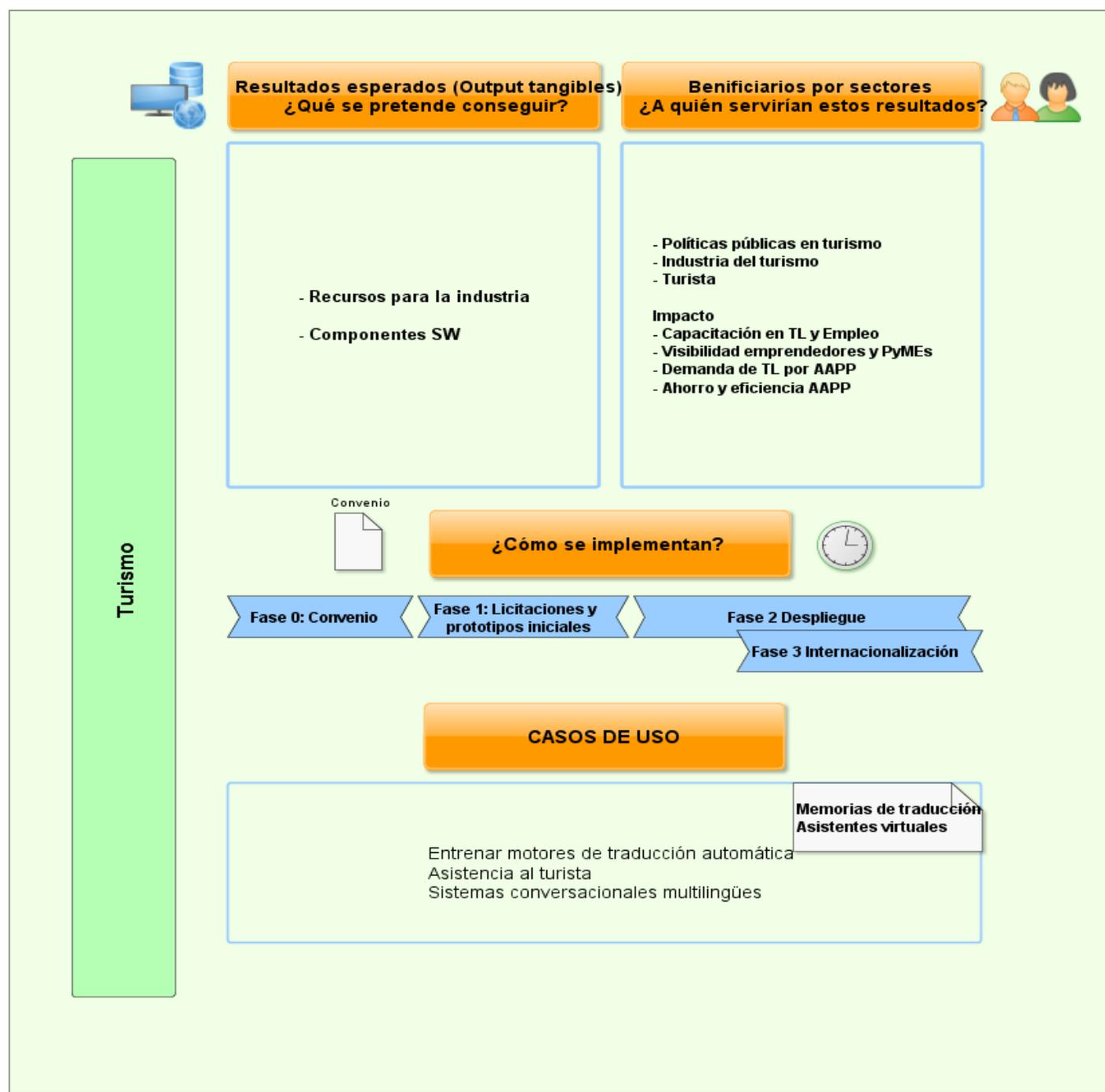
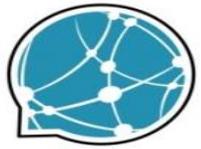


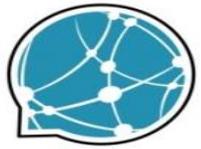
¿Cómo se implementan?



CASOS DE USO

CASO DE USO	Perfil de usuario				
	Decisor	Gestor Ayudas	Evaluador	Investigador	Empresa
DISEÑO POLÍTICAS PÚBLICAS (seguimiento, prospectiva, planificación)	X	X	X	X	X
Herramientas de soporte a EVALUACIÓN (evaluadores, similitud de doc., clasificación taxonomías, estimación innovación)		X	X	X	X
SISTEMA DE ALARMAS (plagio, patrones de fraude)		X	X		X
SISTEMA DE RECOMENDACIÓN (evaluadores, licitación, formación, cruce, conocimiento implícito)	X	X		X	X





Resultados esperados (Output tangibles) ¿Qué se pretende conseguir?

- Recursos lingüísticos para la industria
- Componentes SW
- Perfilado de oferta y demanda de empleo
- Análisis de la relación entre oferta y demanda de empleo
- Sistema de sugerencias para oferentes y demandantes
- Análisis de la relación entre oferta formativa y oferta de empleo

Beneficiarios por sectores ¿A quién servirían estos resultados?



- Desempleados y demandantes de empleo
- Política pública de empleo
- Formación para el empleo
- Oferentes de empleo

EMPLEO
SEPE

Convenio



¿Cómo se implementan?



Fase 0: Convenio

Fase 1: Licitaciones y
prototipos iniciales

Fase 2 Despliegue

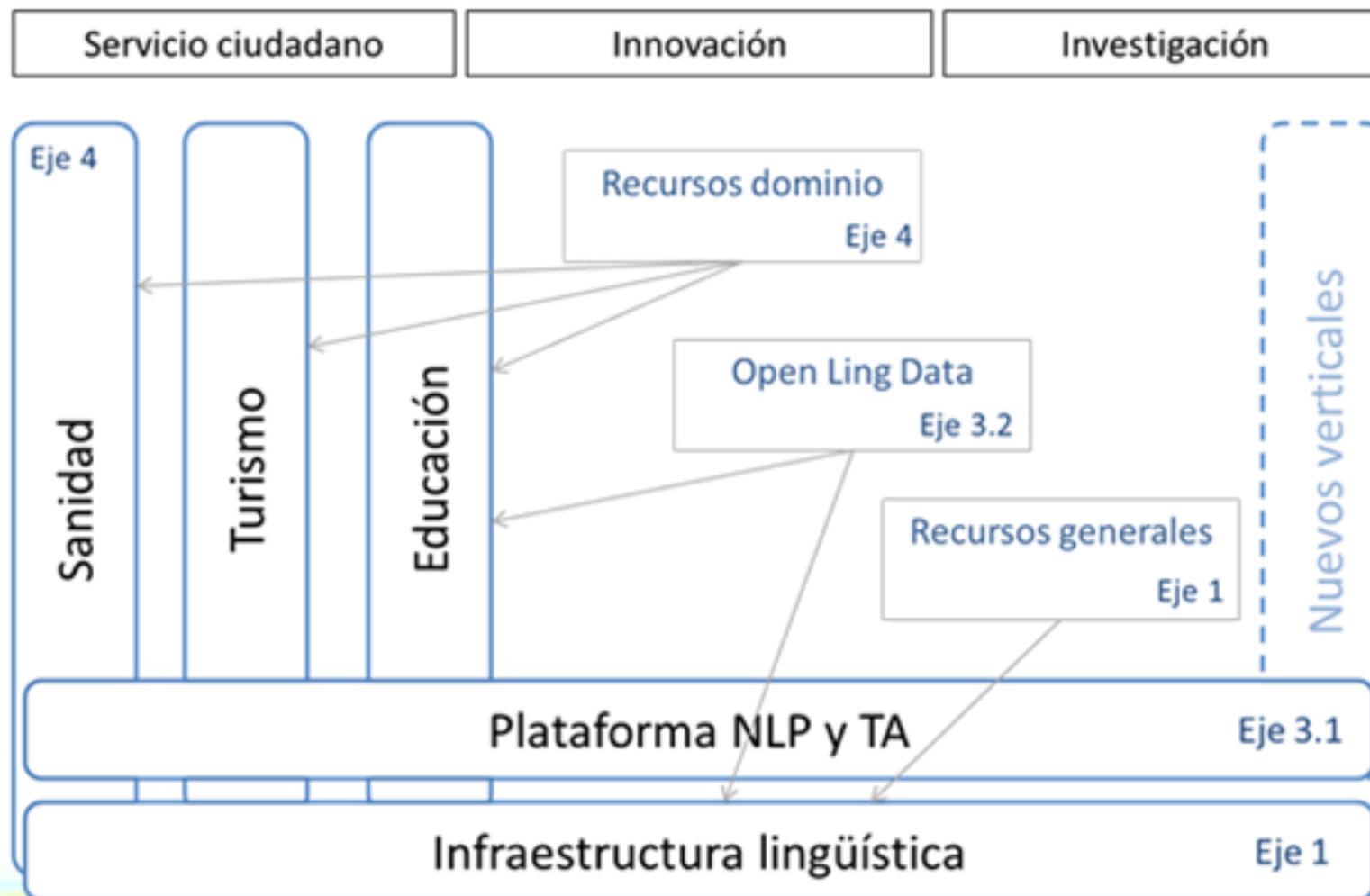
Fase 3 Internacionalización

CASOS DE USO

SEPE (Servicio Público de Empleo Estatal):
Emplear tecnologías del lenguaje sobre los corpus de ofertas de empleo, demandas de empleo y oferta formativa para demandantes de empleo para:

- Perfilar demandantes y oferentes de empleo.
- Analizar la relación entre oferta y demanda de empleo.
- Crear un sistema de sugerencias para demandantes y oferentes.
- Analizar la relación entre demanda de empleo y oferta formativa.

- Oferta de empleo
- Demanda de empleo
- Oferta formativa





¡ ¡ Gracias!!

 www.PlanTL.gob.es

 plantecnologiaslenguaje@mineco.es

 Plan TL - Tecnologías Lenguaje