

# META-FORUM 2016

## Towards a Shared Programme between European Countries and the European Union?

July 5th

Industria  
Energía  
Turismo  
Telecomunicaciones  
Sociedad de la Información

# The Business Plan of a National Language Technologies Plan

# Flagship projects: Public Administration as a driver of Language Industry (I)



How much money can we save if ...?:

- **Using LT for innovation monitoring: processing patent data, R&D projects, papers, ...**  
Improve aid innovation evaluation and monitor process, duplicate project detection, better innovation evaluation ...)  
First 2 R&D organizations quantify it on 6M€ without including HR cost reduction.

Patent evaluation, OEPM has 200 people working in Spain, 3K in European Patent Office.  
This year we will reduce by 15% the time dedicated to the recovery of similar patents.  
Better patent references and patent awards and HR cost reduction.

In both cases use IR techniques based on gaussian models, word embeddings and deep learning models.

- **Legal information processing.** Can we reduce time used by judges (5K in Spain) seeking similar sentences (6M sentences just in Spanish High Courts) and legal grounds (100K docs just EU)?

One of the main problems in the Spanish judicial proceedings (currently 2,5M open proceedings) is the classification of documents provide. Can we improve it using LT?

Can we do the same with parliamentary initiatives?

# Flagship projects: Public Administration as a driver of Language Industry (II)



- **Electronic Medical Record** processing. 3K people in Spain are tagging EMR with CIE-9, CIE-10 tags to measure hospital activity. We start this year a project to optimize this task.
- **Drug data sheets analysis** using LT. Also using EMR to discover drug incompatibilities.
- **Phenotyping and genomics**. Massive correlation between personal phenotype and genomic information.
- **Digitized heritage** LT processing. Spanish National Library project to process current web crawling and digitalized heritage archives LT.
- **Advanced citizen assistance**. Multichannel & multilanguage QA problem based on domain specific knowledge base.
- **Other** application fields: Education, Tourism, Defence ...

# Language Technologies Plan: Action areas



---

1. Language infrastructures development.

---

2. Boost of the language technologies industry

Improvement of the **visibility** and **knowledge transference** of the sector (from academy to industry).

Support for **internationalization** and **commercialization** of the sector.

---

3. Public Administration as a driver of Language Industry

Platforms for **natural language processing** and **automatic translation** in the public administrations.

Linguistic resources of the public administrations and **reuse policy of public sector information** (open ling data on RISP).

---

4. Flagship projects

Health

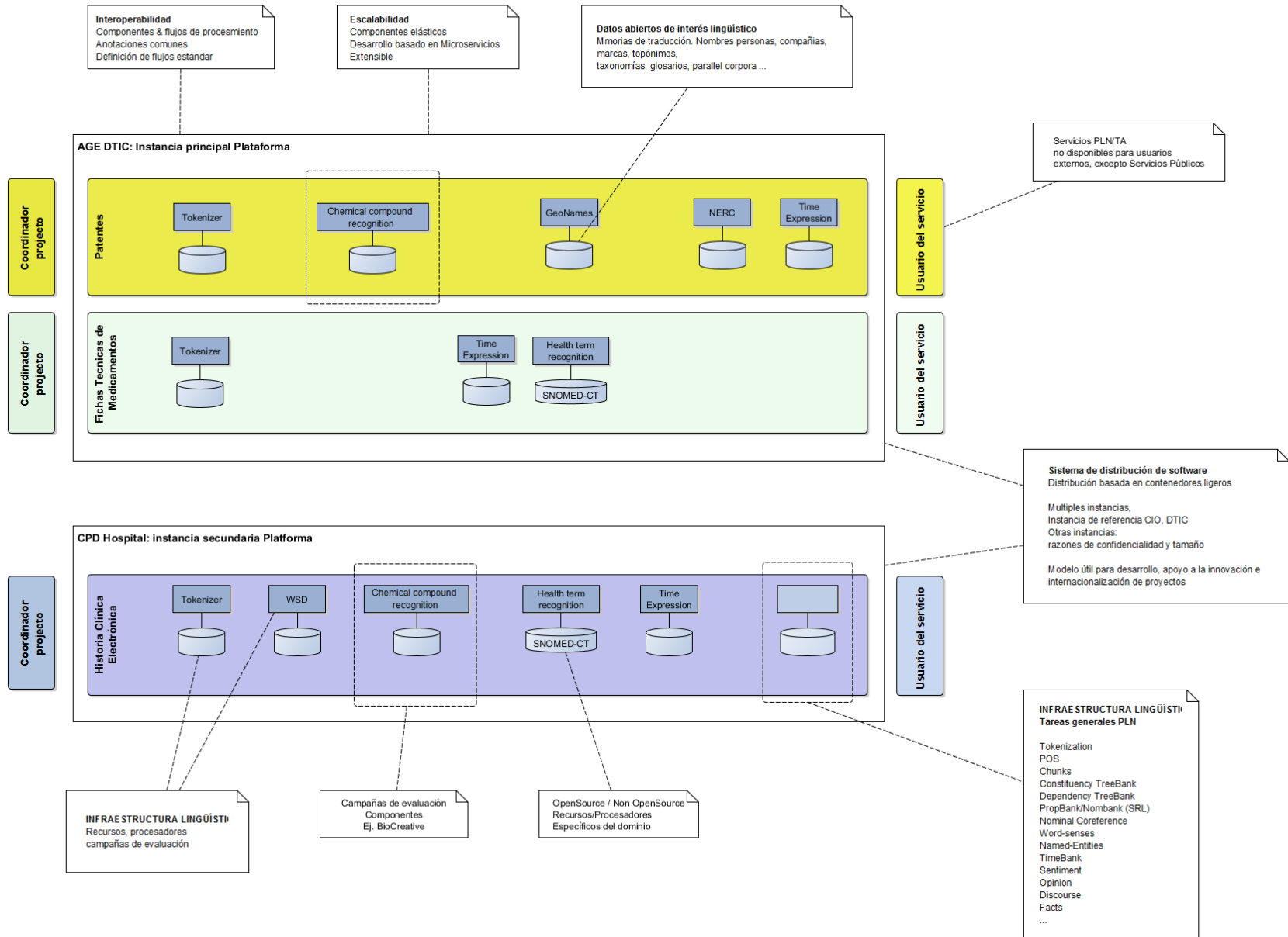
Justice

Education

Tourism, Sectorial Monitoring, Digitised Heritage, etc.

---

# 3.1: NLP Scalable Platform



# WP 2016-17: main actions

## Flagship projects:

- Health 1: Electronic Medical Record processing (EMR)
- Health 2: Drug data sheets processing (FTM)
- Health 3: Phenotyping and genomics
- Justice: legal information processing
- Touristic intelligence
- Sectorial monitoring for innovation
- Digitized and online heritage
- Advanced attention to the citizen
- Education

## Cross projects

- Language infrastructure
- Natural language processing platform for public administrations
- Automatic translation platform for public administrations
- Open data of language data

## Other actions

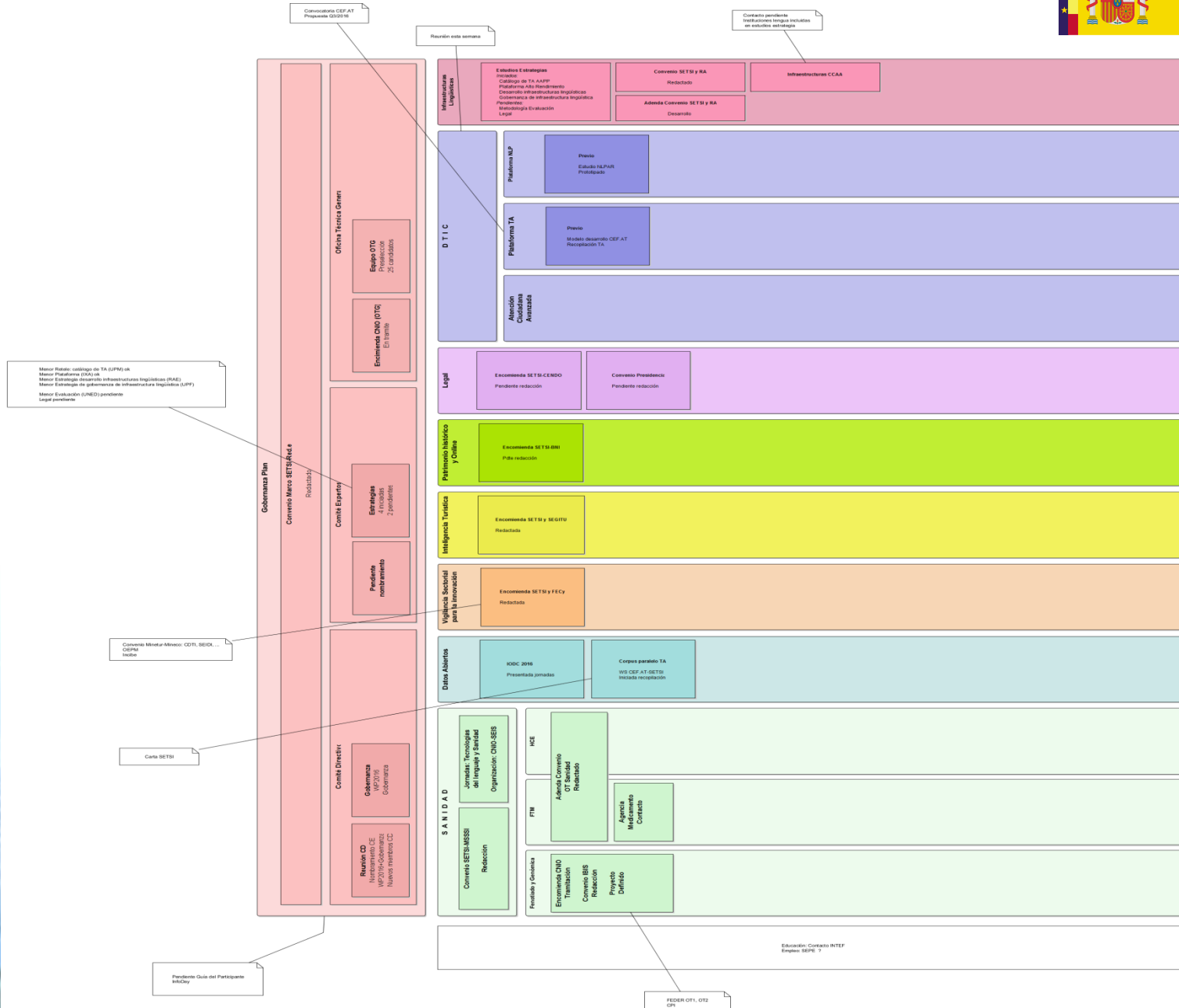
- Studies and strategies
- Internationalization
- Training

# WP 2016-17: main projects (II)



GOBIERNO DE ESPAÑA

MINISTERIO DE INDUSTRIA, ENERGÍA Y TURISMO





**EU LT Collaboration**

**=**

**EU LT Plan?**

# Cooperation areas



## Cross areas

- Language infrastructure (LT resources, evaluation campaigns, general purpose processors):
  - Technical infrastructures
  - Open language data
  - Standards, License & sustainability models
- NLP & TA platform
  - Design and implementation
  - Relation to other Public Administration infrastructures (CEF.AT, Open Data DSI ...)

## Flagship projects

- Results, learned lessons, ideas, new use cases, ...
- Domain specific language infrastructures
- European citizen services

# Thank you