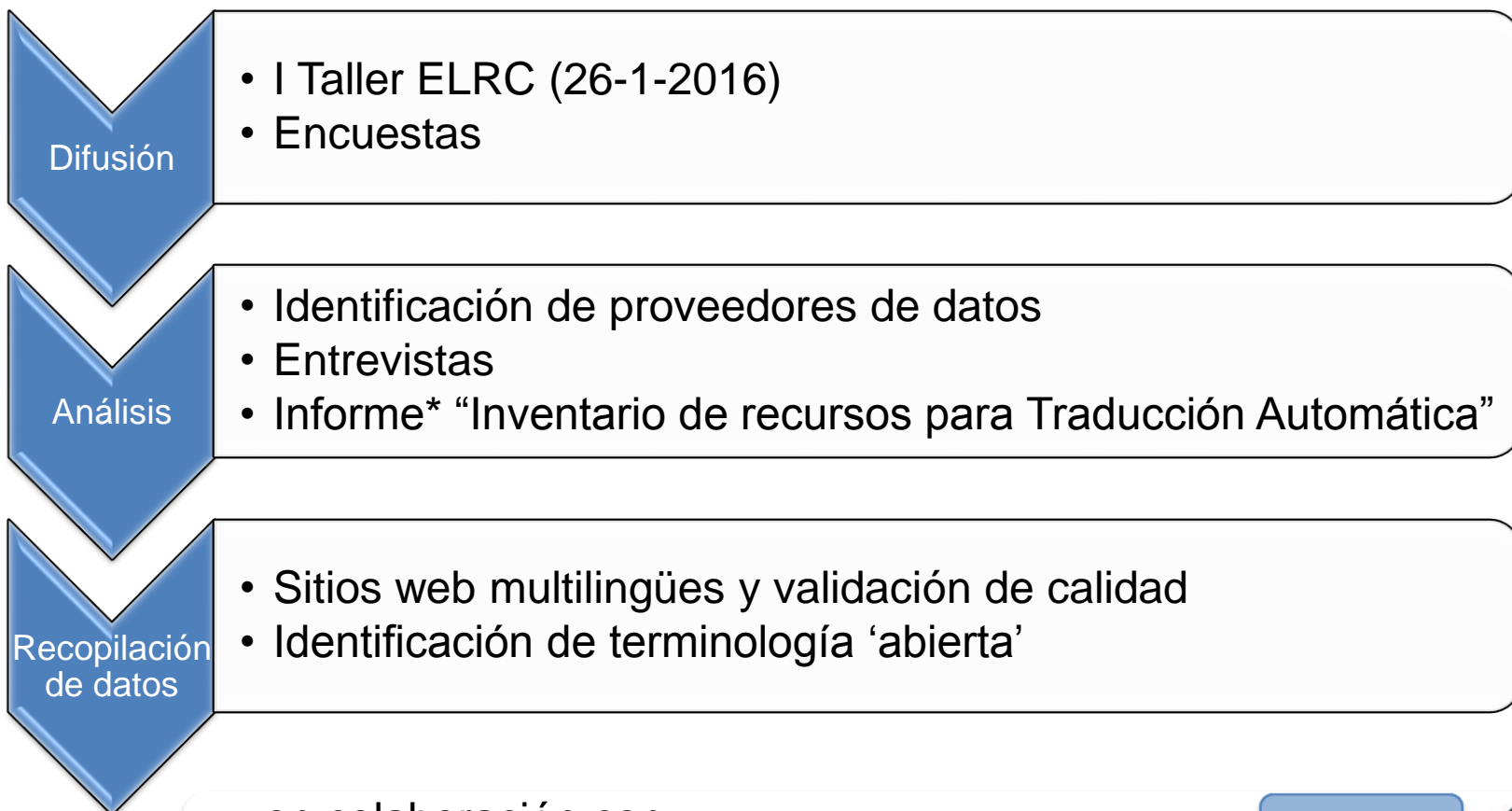


Segundo taller de Coordinación Europea de Recursos de Idiomas (ELRC)

ELRC en España

Núria Bel

Universidad Pompeu Fabra



- en colaboración con
- <http://www.agendadigital.gob.es/tecnologias-lenguaje/actuaciones/Paginas/estudios-realizados.aspx>

(ReTeLe)



[Home](#) [Browse Resources](#) [Help](#) [About](#) [Register](#) [Login](#)



Memorias de traducción

Documentos y sus traducciones

Contenidos Web multilingües

Terminologías

Type in your keywords, please...



Consultable en
elrc-share.eu

Welcome to the ELRC-SHARE repository!



Attribution details: See COPYRIGHT file which contains Source owners

This is a parallel corpus of bilingual texts crawled from multilingual websites, which contains 21,007 TUs.

Period of crawling : 15/11/2016 - 23/01/2017

A strict validation process has been followed, which resulted in discarding:

- TUs from crawled websites that do not comply to the PSI directive,

... [Read More](#)

[← Back](#)

Distribution

Availability: Available

Licences

Open Under-PSI

Distribution Details

Attribution Details: See COPYRIGHT file which contains Source owners

Contact Persons

[Valérie Mapelli](#) 

[Núria Bel](#) 

text

Bilingual text corpus

Languages

Spanish; Castilian (es) (415,720 Tokens)

English (en) (382,085 Tokens)

Linguality

Linguality type: Bilingual

Multi-linguality type: Parallel

Text Format

TMX

Size

- Unidades de Traducción

- Segmentos alineados (no tienen por qué ser frases enteras)

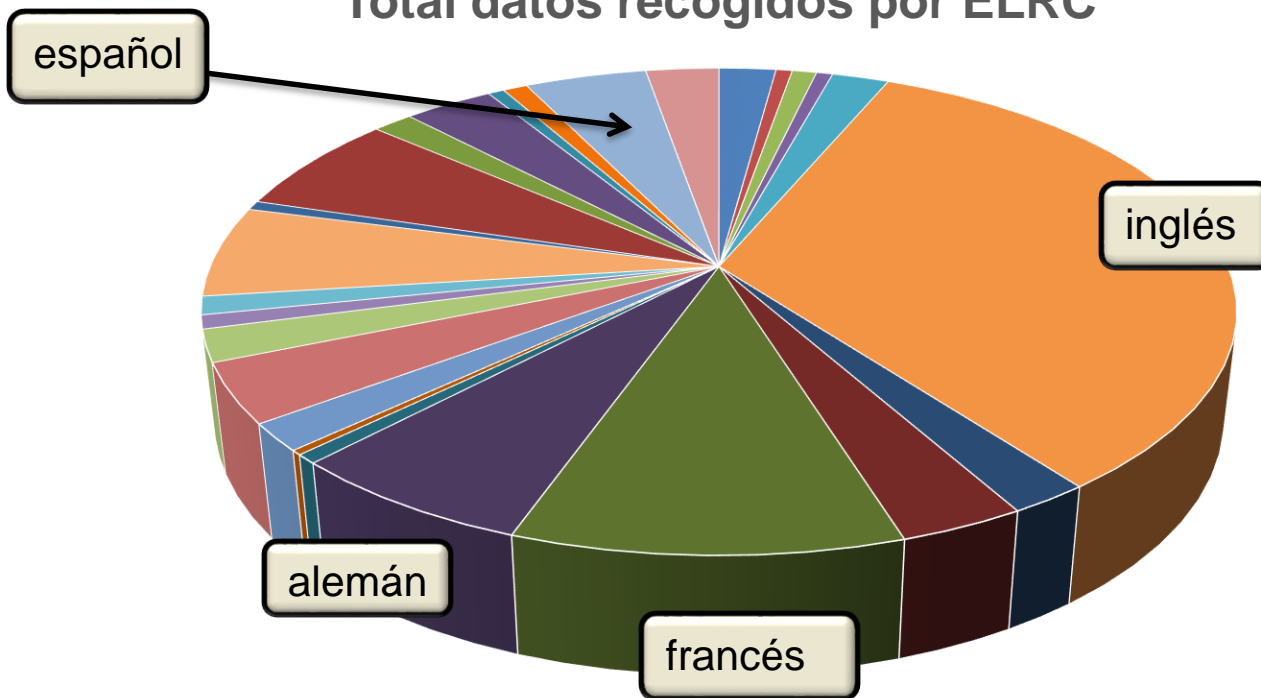
[97281f0f8a907ac37c2be909305d1037; 6.5632634; #psi_Yes]

So, we must consider very carefully whether or not to prosecute.

Por lo tanto, debemos examinar con mucha cautela si es necesario emprender o no acciones judiciales.

- Palabras (Tokens) x cada lengua
- Términos x cada lengua
 - Un término puede tener más de una palabra

Total datos recogidos por ELRC

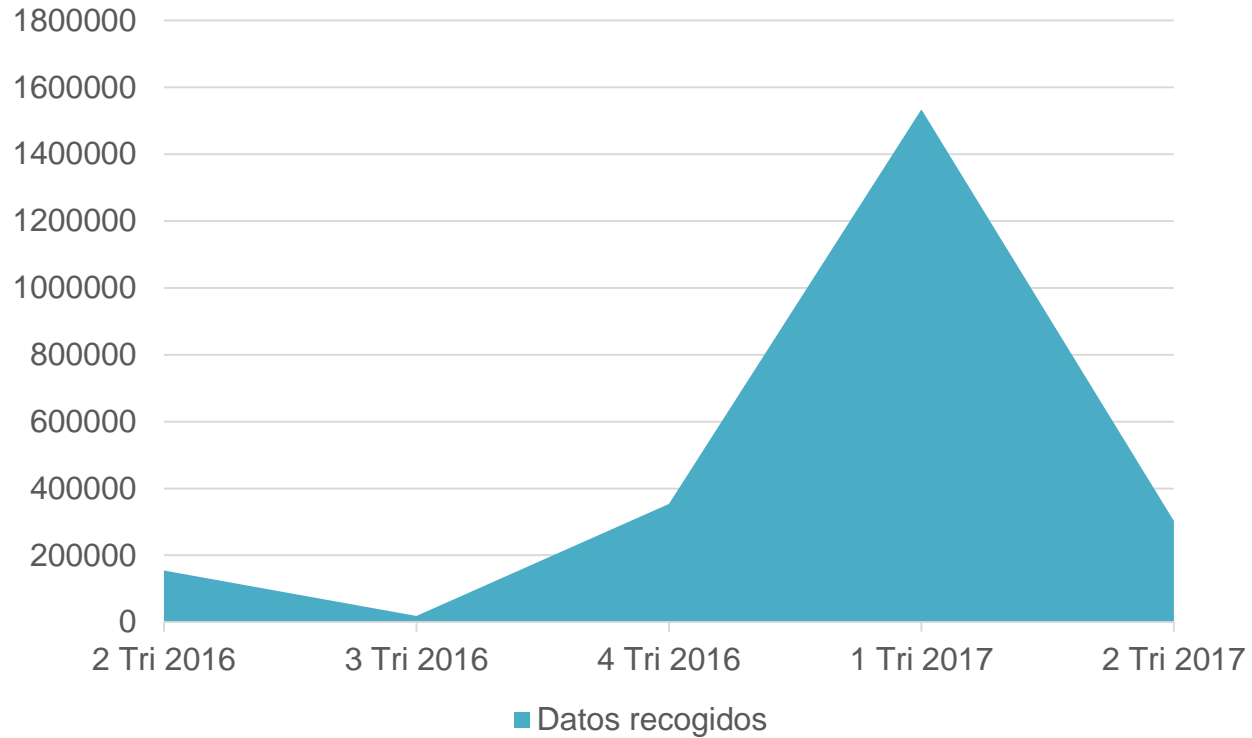


- | | | | | | |
|-------------|------------|--------------|--------------|-----------------|----------------|
| ■ Bulgarian | ■ Croatian | ■ Czech | ■ Danish | ■ Dutch/Flemish | ■ English |
| ■ Estonian | ■ Finnish | ■ French | ■ German | ■ Hungarian | ■ Icelandic |
| ■ Irish | ■ Italian | ■ Latvian | ■ Lithuanian | ■ Maltese | ■ Modern Greek |
| ■ Norwegian | ■ Polish | ■ Portuguese | ■ Romanian | ■ Slovak | ■ Slovenian |
| ■ Spanish | ■ Swedish | | | | |

Ritmo de recopilación ELRC 2016 - 2017

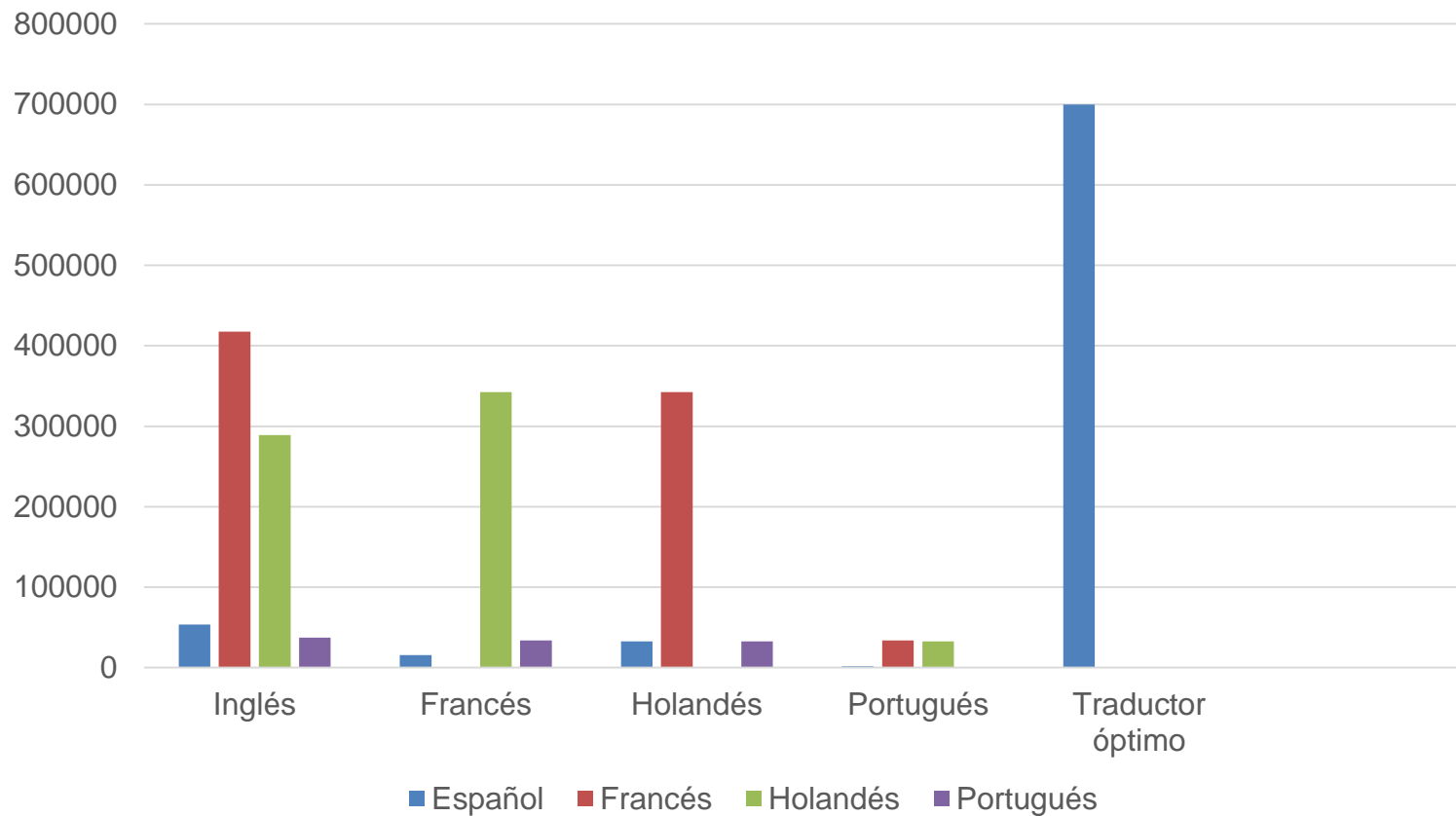


Totales datos bi-multilingües recogidos





Datos recogidos: ejemplos por idiomas - UT





Datos textuales infravalorados

- Archivo poco sistemático
- No hay conjuntos de datos
- Formatos finales de difícil recuperación



Incertidumbre legal

- Privacidad y confidencialidad
- Falta de familiaridad con directiva ISP



Falta de protocolos y recursos

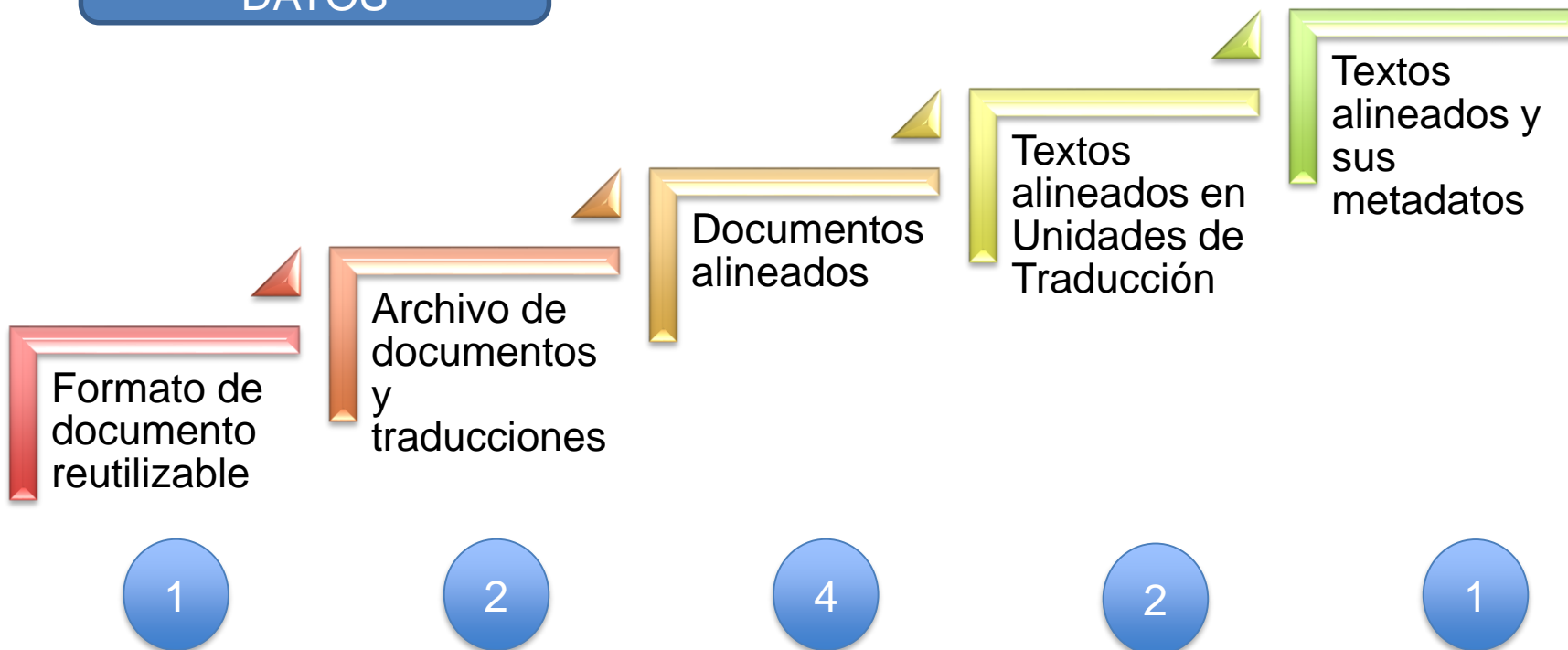
- Cadena de autorización
- Trabajo manual individualizado

Resultados en España

- Terminología: conjuntos de datos abiertos
- Contenidos web multilingües validados
- Sensibilización de diferentes actores
- Creación del Modelo de Madurez de las organizaciones como proveedores de recursos
- Contribución a Datos de traducción en el Plan de Impulso de las Tecnologías del lenguaje



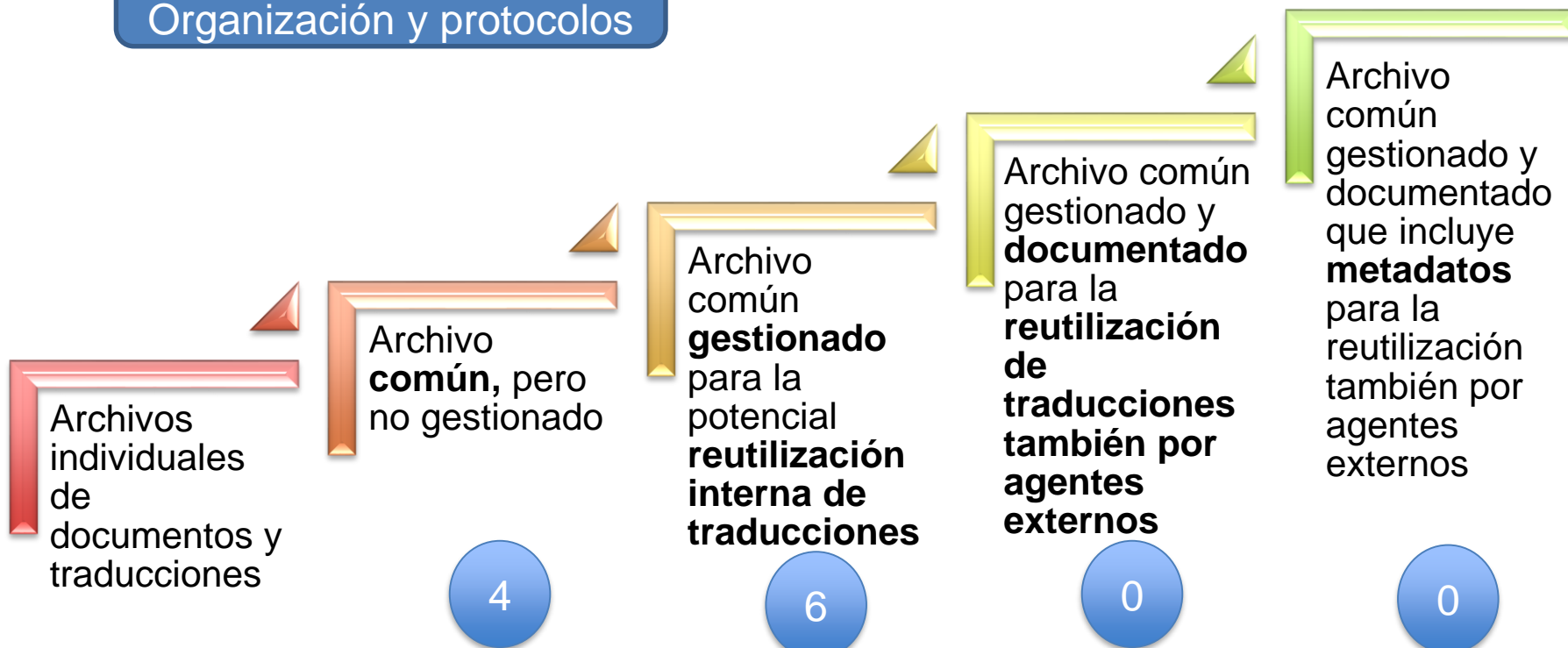
Modelo de Madurez DATOS



Casos del “Inventario de recursos ...”



Modelo de Madurez Organización y protocolos



Casos del “Inventario de recursos ...”

¿Qué fue bien en ELRC?

Sensibilizar y recoger datos fue más fácil cuando los traductores usaban herramientas de traducción asistida

Hubo apoyo institucional, en muchos casos de lenguas con pocos hablantes

Confirmación legal de que la mayoría de los documentos de la AP pueden ser datos abiertos

¿Qué fue mal?

Resistencia a modificar la gestión de datos documentales, a enriquecer los datos con metadatos, ...

Poco interés en la traducción automática



Para el futuro ...

Mejorar el conocimiento del valor de los conjuntos de datos textuales

Mejorar protocolos y métodos de archivo incluyendo información sobre confidencialidad, licencias, etc.

Recuperación de los datos de los contratos de traducción externalizados

Encontrar métodos óptimos de recuperación de los datos de traducción internos

¡Gracias por su atención!

Email: info@lr-coordination.eu

Website: www.lr-coordination.eu

