

Segundo taller del proyecto ELRC en Madrid

¿Cómo identificar y gestionar los datos públicos? Preguntas y respuestas

Elena Montiel-Ponsoda (moderadora)

ReTeLe – Red de recursos para tecnologías de la lengua

Ontology Engineering Group-Universidad Politécnica de Madrid



El **DMP** o **plan de gestión de datos** proporciona guías y estándares para el tratamiento de los datos durante su producción y después, durante su uso. Cubre todo el **ciclo de vida** de los datos y define una política de datos que permita gestionarlos de forma eficiente y asegurar su sostenibilidad.



- **Anticipar posibles cuestiones legales**
 - Asegurar la propiedad intelectual de los datos
 - **Acordar con todas las partes involucradas la propiedad los datos (p.ej. en relación con proveedores de servicios lingüísticos: asegurarse de tener todos los derechos)**
 - Asegurarse de la propiedad de toda la documentación producida (p.ej. Memorias de traducción)
 - Comprobar los derechos de privacidad de antemano y planificar tareas de anonimización si es necesario
- **Definir un plan de gestión**
 - Que tiene que responder al objetivo principal (escritura de documentos, traducción, etc.)
- **Establecer un plan para la reutilización** (de la documentación a los recursos)
 - Solicitar datos en formatos reusables (no sólo PDFs sino también TMX/XML/)
 - Asegurar el almacenamiento adecuado de los datos (no CDs?)
- **Prever la publicación y compartición de los datos** como parte del plan de reutilización de la información del Sector Público

Preguntas sobre la identificación y gestión de los datos



Su opinión y experiencia son importantes para ELRC. Por favor, ¡compártalas con nosotros!

Estimado participante, |

Queremos agradecerle su participación en este Segundo Taller ELRC en España. Sus conocimientos y experiencia son fundamentales para ELRC y para el éxito de la adaptación del sistema CEF e ~~Traducción~~ a las necesidades de los servicios de la administración pública en España.

Por este motivo, le estaríamos muy agradecidos si pudiera contestar a las siguientes preguntas. Se trata de unas preguntas a las que puede contestar rápidamente a partir de sus conocimientos y experiencia. Sus respuestas nos ayudarán a obtener información general sobre los procedimientos y procesos en diferentes servicios y ministerios.

¿Dispone su organización de recursos/traducciones que puedan ser potencialmente interesantes para ELRC?
En caso afirmativo, por favor denos algunos detalles (lenguas, dominios, etc.) sobre qué tipo de recursos son.

¿Usa su organización algún plan de gestión de datos, es decir guías y estándares para hacer que los datos creados por su organización puedan ser encontrados por terceros y para que los datos sean accesibles, interoperables y reutilizables?

¿Utiliza su organización el sistema MT@EC o cualquier otra tecnología de traducción asistida por ordenador (por ejemplo, memorias de traducción) o herramientas lingüísticas (por ejemplo, correctores ortográficos y gramaticales)? En caso afirmativo, por favor especifique qué tipo de herramientas.

1. ¿Dispone su organización de recursos/traducciones que puedan ser potencialmente interesantes para ELRC?
 - En caso afirmativo, por favor denos algunos detalles (lenguas, dominios, etc.) sobre qué tipo de recursos son.

Los recursos se han clasificado en tres grandes bloques: corpus, gramática o modelo, y recurso léxico-conceptual.

a) Corpus: conjunto de documentos o textos en una o varias lenguas. P. ej.:

- Conjunto de documentos de la administración oficial (por ejemplo, decisiones de ministerio, actos legales, etc.).
- Conjunto de documentos de un periódico, revista, boletín, blog de artículos, etc.
- Corpus paralelo (por ejemplo, documentos de traducción con el documento original, sin procesar).
- Memorias de traducción (por ejemplo, documentos de traducción alineados por segmentos con sus documentos originales).

b) Recursos léxico-conceptuales:

- Recursos léxicos y terminológicos, glosarios, etc., incluyendo alguna otra información (definiciones, ejemplos, traducciones equivalentes, información lingüística...) o sin ella.



2. ¿Usa su organización algún **plan de gestión de datos**, es decir guías y estándares para hacer que los datos creados por su organización puedan ser encontrados por terceros y para que los datos sean **accesibles, interoperables y reutilizables**?

3. ¿Utiliza su organización el sistema **MT@EC** o cualquier otra **tecnología de traducción asistida por ordenador** (p.ej., memorias de traducción) o herramientas lingüísticas (p.ej., correctores ortográficos y gramaticales)?
- En caso afirmativo, por favor especifique de qué tipo de herramientas se trata.

4. ¿Sabe de algún servicio o servicios en España con necesidad de gestionar el multilingüismo que pudiera beneficiarse del sistema CEF *eTranslation*?

-
5. ¿Necesita disponer de asistencia remota o *in situ* para seleccionar, preparar y procesar recursos lingüísticos?

6. ¿Cree que hay otras personas o instituciones que deberían ser informadas e invitadas a participar en ELRC?
- Por favor, denos su nombre y dirección.

¿Otras preguntas o dudas?

1. ¿Cómo puedo saber si puedo donar mi corpus a ELRC?
2. ¿Si a partir de un corpus he creado un recurso terminológico y lo he combinado con otro diccionario que tenía, puedo compartirlo?
3. ¿Si externalizo la traducción de un texto, a quién pertenece la traducción, a mi institución o a la agencia de traducción?
4. ¿Qué puedo hacer con un recurso que contiene datos personales? ¿Tendría que anonimizar un corpus de licitaciones, p. ej.?

5. Mi organización tiene datos, pero no tengo recursos para identificarlos y crear un plan de gestión de datos, ¿puede ELRC ayudarme?
6. Tengo un corpus pero está en pdf, ¿qué puedo hacer?
7. La mayoría de mis recursos son bases de datos numéricas con algo de texto, ¿serían útiles para ELRC?
8. ¿Si ya cedí mis recursos a CLARIN, tengo que volver a cederlos a ELRC?

Segundo taller del proyecto ELRC en Madrid

¿Cómo identificar y gestionar los datos públicos? Preguntas y respuestas

Elena Montiel-Ponsoda (moderadora)

ReTeLe – Red de recursos para tecnologías de la lengua

Ontology Engineering Group-Universidad Politécnica de Madrid

