



EUROPEAN LANGUAGE RESOURCE INFRASTRUCTURE



Co-financed by the European Union
Connecting Europe Facility

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

Proyecto

- ▶ Co-financiado por la Unión Europea - Connecting Europe Facility Programme (CEF)
- ▶ Del 02/10/2017 al 30/09/2019
- ▶ Desarrollo de una infraestructura para facilitar la recopilación, preparación y compartición de recursos lingüísticos que ayuden a mejorar los servicios de traducción europeos
- ▶ 7 socios, 4 Estados Miembros: Francia, Irlanda, Portugal y España



Participantes

- ▶ Vicomtech, España:
 - ▶ Coordinador del proyecto
 - ▶ Centro de investigación aplicada con amplia experiencia en tecnologías del lenguaje y transferencia de tecnología
- ▶ Agencia para a Modernizaçao Administrativa (AMA), Portugal:
 - ▶ Instituto público responsable de la modernización administrativa y de la transformación digital
- ▶ Dublin City University (DCU), Irlanda:
 - ▶ Entre los principales centros europeos en tecnologías del lenguaje
- ▶ Evaluations and Language resources Distribution Agency (ELDA), Francia:
 - ▶ Líderes europeos en recopilación, preparación y distribución de recursos lingüísticos
- ▶ Universidade de Lisboa, Faculdade de Ciências (FCUL), Portugal:
 - ▶ Expertos en tecnologías del lenguaje y infraestructuras de compartición de recursos lingüísticos
- ▶ Linkare, Portugal:
 - ▶ Expertos en desarrollo software development y componentes CEF
- ▶ Secretaría de Estado para el Avance Digital (SEAD), España:
 - ▶ Institución pública responsable del Plan de Tecnologías de las Lenguas



¿Por qué ELRI?

- ▶ Compartir datos no es fácil:
 - ▶ ¿Puedo compartir mis datos?
 - ▶ ¿Con quién?
 - ▶ Requiere tiempo y esfuerzo: ¿Por qué lo haría?
- ▶ No es fácil preparar y distribuir datos:
 - ▶ Requiere conocimiento en tecnologías del lenguaje
 - ▶ Requiere conocimiento en temas administrativos y legales
 - ▶ Requiere conocimiento en desarrollo de infraestructuras software
 - ▶ Requiere tiempo para preparar recursos lingüísticos multilingües
- ▶ Pero...
 - ▶ Se necesitan recursos lingüísticos de los Estados Miembros para dar soporte a una Europa realmente multilingüe
 - ▶ Se puede desarrollar un entorno flexible que beneficie a todos los actores a partir del conocimiento actual



Objetivos

- ▶ Complementar otras iniciativas de la UE como ELRC
- ▶ Desarrollar nuevos medios para recopilar, preparar y compartir recursos lingüísticos
- ▶ Mejorar los servicios de traducción humana y automática a nivel nacional y europeo
- ▶ Proveer beneficios a corto y largo plazo para los actores:
 - ▶ Instituciones públicas y centros de traducción públicos de los Estados Miembros
 - ▶ Instituciones europeas e infraestructura de servicios digitales
 - ▶ Usuarios y desarrolladores de servicios de traducción europeos
- ▶ Tener en cuenta dificultades:
 - ▶ Prácticas
 - ▶ Técnicas
 - ▶ Legales y administrativas



Enfoque

- ▶ Desarrollo de una infraestructura para simplificar la recopilación de recursos en los distintos estados miembros:
 - ▶ Estaciones nacionales (National Relay Stations: NRS) para Estados Miembros
 - ▶ Permitir la recopilación, preparación y gestión local de recursos
 - ▶ Start local, build global!
- ▶ Integrar tecnología de procesamiento de lenguaje para:
 - ▶ Automatizar la creación de memorias de traducción a partir de datos de usuarios y la limpieza o formateo de memorias
 - ▶ Automatizar la preparación de recursos lingüísticos adecuados
- ▶ Involucrar a los centros de traducción públicos y a las instituciones públicas de los Estados Miembros
- ▶ Proveer medios flexibles para compartir datos:
 - ▶ A diferentes niveles: nacional, europeo o Open Data
 - ▶ Compartir en función de los recursos considerados
 - ▶ Soporte y validación antes de compartir recursos



Dificultades y respuestas

- ▶ No es fácil compartir datos:
 - ▶ ¿Puedo? *Proveemos medios dedicados en su Estado Miembro como soporte a la decisión*
 - ▶ ¿Con quién? *Las estaciones nacionales proveen medios flexibles para compartir según las necesidades*
 - ▶ ¿Por qué lo haría? *Puede beneficiar de recursos preparados automáticamente a partir de sus datos para ayudar a optimizar sus propios procesos de traducción*
- ▶ No es fácil preparar y distribuir datos:
 - ▶ *Automatizamos la creación de recursos lingüísticos*
 - ▶ *Proveemos una infraestructura accesible a miembros de organismos y centros de traducción públicos*
 - ▶ *La distribución se gestiona en función de las condiciones determinadas por los usuarios*



Beneficios

- ▶ Para organismos y centros de traducción públicos, como para servicios de traducción institucionales europeos
- ▶ Corto plazo:
 - ▶ Obtener memorias de traducción limpias a partir de conjuntos de documentos o memorias de traducción existentes
 - ▶ Obtener medios flexibles para compartir datos: grupos de organismos nacionales, repositorio de la Comisión Europea o Open Data
 - ▶ Obtener acceso a recursos compartidos por otras instituciones en su estado miembro o en Europa
- ▶ Medio y largo plazo:
 - ▶ Beneficiarse de servicios eTranslation adaptados para sus idiomas
 - ▶ Servicios más eficientes en una Europa multilingüe



Experiencia en RL

- ▶ ELRC National Anchor Points (NAP):
 - ▶ Expertos que dan soporte a la recopilación de recursos europeos en la Unión Europea
 - ▶ 2 por Estado Miembro
 - ▶ NAP Servicios Públicos: representantes de instituciones públicas
 - ▶ NAP Tecnología: representantes en tecnologías del lenguaje
- ▶ 4 National Anchor Points en ELRI:
 - ▶ Paulo Vale, Portugal (Public NAP)
 - ▶ António Branco, Portugal (Technology NAP)
 - ▶ Andy Way, Irlanda (Technology NAP)
 - ▶ David Pérez Fernández, España (Public NAP)
- ▶ Instituciones líderes en europea en tratamiento de RL



Experiencia tecnológica

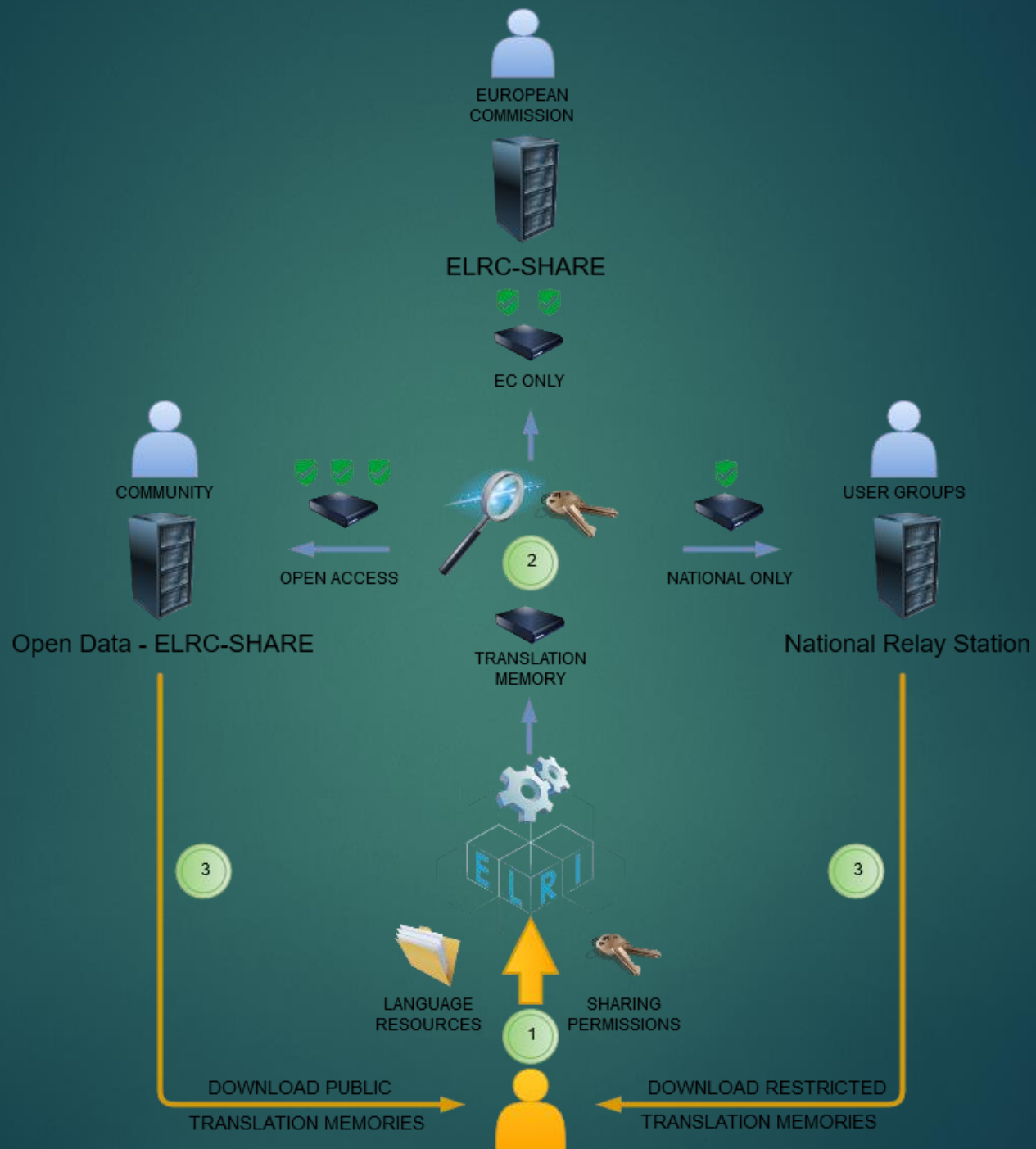
- ▶ Herramientas de procesamiento robustas:
 - ▶ Crear recursos estructurados como memorias de traducción a partir de datos dispersos en múltiples idiomas
 - ▶ Limpiar recursos existentes como memorias de traducción
 - ▶ Herramientas integradas en cadenas de procesamiento
- ▶ Experiencia en tratamiento de datos reales:
 - ▶ De instituciones públicas
 - ▶ En idiomas y dominios variados
- ▶ Desarrolladores de infraestructuras para recursos lingüísticos
- ▶ Aspectos administrativos y legales en compartición de recursos



Proceso

- ▶ Registrarse en la NRS de su Estado Miembro:
 - ▶ Registro por la página web
 - ▶ El equipo ELRI autorizará a los miembros de instituciones y centros de traducción públicos
- ▶ Subir sus datos:
 - ▶ Documentos multilingües, memorias, etc.
 - ▶ Especificar con qué grupo(s) compartir los recursos
 - ▶ Los datos se preparan automáticamente y se validan manualmente por el equipo ELRI
- ▶ Descargar recursos:
 - ▶ Sus propios recursos preparados automáticamente
 - ▶ Otros recursos compartidos por otros organismos





Compartir en Europa

- ▶ National Relay Stations
- ▶ Datos compartidos más allá del Estado Miembro:
 - ▶ Beneficia a más entidades
 - ▶ Esencial para mejorar servicios europeos de traducción, p.ej., eTranslation.
- ▶ Recursos preparados mediante ELRI pueden compartirse:
 - ▶ A nivel europeo a través del repositorio ELRC-SHARE:
 - ▶ Sólo con la Comisión Europea para mejorar los servicios eTranslation disponibles para las administraciones públicas
 - ▶ Con todos los usuarios de ELRC-SHARE (p.ej., Open Under PSI).
 - ▶ Con la comunidad en repositorios Open Data:
 - ▶ Algunos recursos se pueden compartir sin restricciones
 - ▶ Beneficia a la comunidad investigadora en tecnologías del lenguaje
 - ▶ ELRI subirá los datos especificados como abiertos a estos repositorios



Etapas

- ▶ 2018:
 - ▶ Diseño y desarrollo de la infraestructura
 - ▶ Integración y despliegues en Estados Miembros
 - ▶ Diseminación de la iniciativa
- ▶ 2019:
 - ▶ Talleres de información y uso:
 - ▶ En los distintos Estados Miembros del proyecto
 - ▶ Primavera 2019
 - ▶ Extensión a más participantes
 - ▶ Recopilación activa de recursos



FAQ

- ▶ ¿Puedo compartir mis datos con grupos reducidos?
 - ▶ *Sí. Aunque el objetivo sea compartir con la mayor cantidad de usuarios posibles, podemos crear grupos específicos*
- ▶ ¿Está abierta la estación nacional en España?
 - ▶ *Lista para apertura pública*
- ▶ ¿Se mantendrán las estaciones a final de proyecto?
 - ▶ *Sí. Despliegues y mantenimiento por organismos públicos*
- ▶ ¿Tratáis la anonimización de datos?
 - ▶ *Recomendamos contactar a ELRC para un procesamiento específico en este caso*
- ▶ Tengo otras preguntas...
 - ▶ *Sesión de preguntas en breve :)*
 - ▶ *¡Sesiones prácticas e interactivas en este taller ELRI!*



Información

- ▶ Sitio web del proyecto:
 - ▶ www.elri-project.eu
 - ▶ Información sobre el proyecto
 - ▶ Eventos, enlaces útiles y más
- ▶ Twitter:
 - ▶ @ElriProject
 - ▶ Últimas noticias



Contact

- ▶ Coordinación:
 - ▶ elri-coordination@vicomtech.org
- ▶ Contactos en España:
 - ▶ Vicomtech:
 - ▶ Thierry Etchegoyhen
 - ▶ SEAD:
 - ▶ David Pérez Fernández
 - ▶ Maite Melero
- ▶ Lista de contactos:
 - ▶ <http://www.elri-project.eu/contact-us.html>



¡Muchas gracias!

