

# ESTACIÓN DE TRANSMISIÓN NACIONAL

## ELRI

### GUÍA DEL USUARIO

---

#### *Resumen ejecutivo*

---

Este documento proporciona directrices para los usuarios de las Estaciones Nacionales de Datos desarrolladas en el marco del proyecto ELRI (Infraestructura Europea de Recursos Lingüísticos).<sup>1</sup>

Nivel de disseminación	Público
Versión	V2.0
Fecha	2019/03/06
Autór(es)	Thierry Etchegoyhen Maite Melero Consortio ELRI



**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.*

---

<sup>1</sup> [www.elri-project.eu](http://www.elri-project.eu)

## Índice

1	Introducción .....	3
2	Unirse a una estación nacional de transmisión ELRI .....	4
2.1	Inscripción .....	4
2.2	Autorización .....	5
3	Acceso a los servicios de ELRI .....	6
4	Uso de los servicios de ELRI .....	7
4.1	¿Qué es un recurso lingüístico? .....	7
4.2	¿Qué ficheros se pueden descargar? .....	8
4.3	¿Qué ficheros se pueden subir? .....	9
4.4	¿Quién comparte con quién? .....	9
5	Subida y descarga de recursos .....	10
5.1	Contribución de recursos lingüísticos .....	10
5.2	Descarga de recursos lingüísticos .....	15
6	Contactar con el equipo de la estación de datos ELRI .....	17

## 1 Introducción

Las estaciones de transmisión nacionales ELRI son aplicaciones web diseñadas para ofrecer servicios que facilitan la preparación y el intercambio de recursos lingüísticos a diferentes niveles: con todas las organizaciones públicas registradas a nivel nacional, entre grupos específicos de las organizaciones registradas, con la Comisión Europea, o como Datos Abiertos.

La red ELRI está restringida a los organismos públicos de los estados miembros. En otras palabras, sólo el personal de una entidad pública puede registrarse como usuario del sistema en su estado miembro. Por consiguiente, los recursos sólo se comparten entre organizaciones públicas registradas. Los recursos cuyos propietarios estén dispuestos a compartirlos con un público más amplio pueden optar por compartirlos con grupos específicos más extendidos y estos recursos se subirán a repositorios dedicados, como el ELRC-SHARE o los portales europeos de datos abiertos.

Cada estación de transmisión nacional se aloja en servidores de un estado miembro<sup>2</sup> y la mantiene un equipo especializado de ese país. Este equipo supervisa la gestión de usuarios y el manejo adecuado de los recursos, a la vez que actúa como interlocutor de cara a los usuarios de los servicios de ELRI en dicho estado miembro.

En este documento, los usuarios pueden encontrar información sobre cómo conectarse y utilizar la estación nacional de datos de su país.

La pantalla principal de la estación ELRI en España se muestra en la Figura 1.

---

<sup>2</sup> Actualmente hay estaciones desplegadas en Francia, Irlanda, Portugal y España.



Figura 1 Pantalla principal de la estación de transmisión ELRI en España

## 2 Unirse a una estación nacional de transmisión ELRI

Únicamente los usuarios registrados y autorizados de organizaciones públicas del Estado miembro en el que está desplegada la estación ELRI tienen acceso a dicha estación. Por lo tanto, la adhesión a una estación ELRI requiere de dos etapas:

1. Los usuarios que deseen unirse a la red pueden solicitar la creación de una cuenta a través de la página de registro de la estación nacional de datos.
2. Los miembros del equipo de soporte de la estación de datos conceden acceso a los usuarios pertenecientes a una organización pública.

Las secciones 2.1 y 2.2 proporcionan más detalles sobre cada uno de estos dos pasos.

### 2.1 Inscripción

Los nuevos usuarios pueden inscribirse en línea a través de la página de **Registro** que se muestra en la Figura 2.

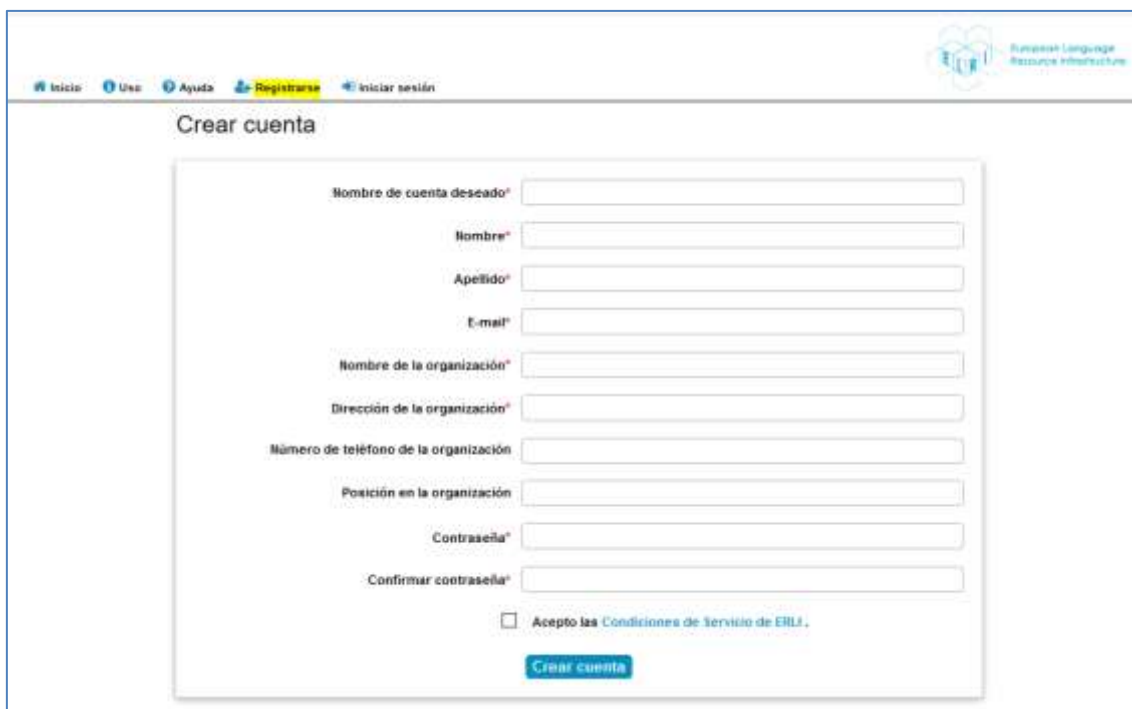


Figura 2 Página de registro

Después de proporcionar la información obligatoria solicitada<sup>3</sup>, los usuarios deben aceptar las Condiciones de Servicio de ELRI, marcando la casilla correspondiente antes de hacer clic en el botón *Crear Cuenta*.

Una vez enviada la solicitud de registro, ésta será evaluada y eventualmente validada como se describe en la Sección 2.2.

## 2.2 Autorización

Las solicitudes de registro llegan al equipo ELRI responsable en España. El personal de ELRI para España revisará la solicitud y, tras verificar que el usuario pertenece a una institución pública española, validará la cuenta solicitada.

Una vez validado, el usuario recibirá una notificación por correo electrónico, tal y como se muestra en la Figura 2, y podrá acceder a los servicios ELRI ofrecidos por la estación nacional de datos.

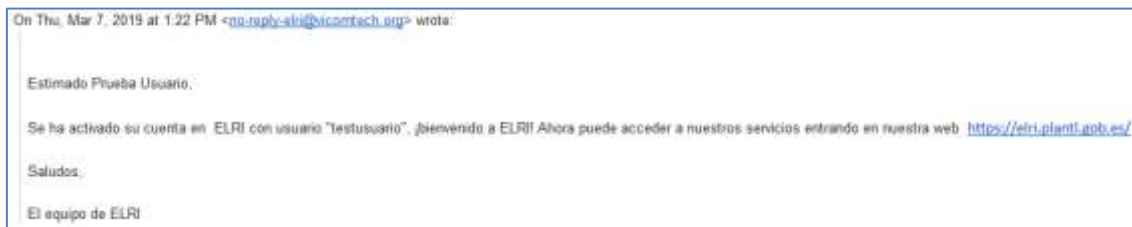


Figura 2 Email de notificación de validación de inscripción

<sup>3</sup> Indicado con un asterisco rojo.

### 3 Acceso a los servicios de ELRI

Sólo los usuarios registrados y autorizados tienen acceso a los servicios prestados por ELRI, es decir, reaprovechar sus textos multilingües para construir recursos lingüísticos útiles para la traducción, buscar recursos de interés aportados por otros usuarios y compartir estos recursos con otras organizaciones o con la comunidad. Por consiguiente, antes de acceder a dichos servicios, los usuarios deben iniciar sesión en el sistema a través de la página de **inicio de sesión** que se muestra en la Figure 3.

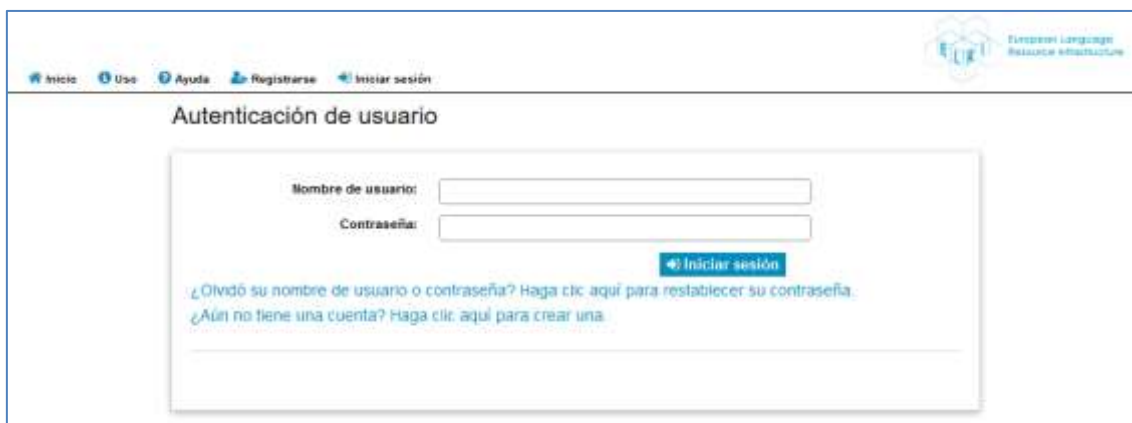


Figure 3 Pantalla de verificación de usuario

Una vez que se ha iniciado sesión, los usuarios tienen acceso a las opciones del menú que se muestran en la Figura 5.



Figura 5 Opciones de menú para usuarios registrados

Desde este menú, los usuarios tienen acceso a los siguientes servicios:

- Inicio: enlace a la página de inicio.
- Navegar: enlace a la página donde los usuarios pueden buscar y descargar recursos lingüísticos (Ver Sección 4).
- Contribuir: enlace a la página donde los usuarios pueden contribuir recursos lingüísticos (Ver Sección 5).
- Uso: enlace a las instrucciones de uso del sistema.
- Condiciones del Servicio: enlace a las Condiciones del Servicio de ELRI.
- Ayuda: enlace a una página de ayuda con información de contacto y documentación de soporte.
- Nombre de usuario<sup>4</sup>: enlace a una página con información del perfil del usuario que está conectado en ese momento.
- Salir: desconecta al usuario conectado del sistema y cierra la sesión.

<sup>4</sup> jdoe en el ejemplo.

## 4 Uso de los servicios de ELRI

A través de sus estaciones de transmisión de datos, ELRI proporciona las herramientas básicas necesarias para preparar y compartir recursos lingüísticos. A continuación, explicamos qué son los recursos lingüísticos, cómo pueden contribuir los usuarios y cómo funciona el uso compartido en una estación ELRI.

### 4.1 ¿Qué es un recurso lingüístico?

El término "*recurso lingüístico*" engloba todos los datos que son relevantes para la prestación de servicios lingüísticos en un idioma o grupo de idiomas determinado. Esto incluye los siguientes tipos principales de recursos:

- Corpus: es el tipo principal de recurso lingüístico. Es, fundamentalmente, un conjunto preparado de documentos con texto en uno o más idiomas. Un corpus puede ser *monolingüe*, es decir, que contiene texto en un solo idioma, normalmente en un solo archivo, o *bilingüe*, es decir, que contiene textos traducidos, en dos idiomas, normalmente en archivos separados. Se pueden crear corpus a partir de colecciones de documentos en diferentes formatos (por ejemplo, MS Word o PDF), extrayendo y procesando el contenido de los documentos.
- Memoria de traducción: uno de los recursos más útiles, tanto para los traductores humanos como para los servicios de traducción automática, es la memoria de traducción. Se trata de un documento estructurado, con la extensión estándar de archivo TMX, en el que las traducciones entre idiomas se emparejan por unidades de traducción (por lo general, una oración fuente u original emparejada con su equivalente traducido en el idioma de destino). Un ejemplo del contenido de TMX se muestra en la Figura 4.
- Archivos de intercambio de terminología: las correspondencias terminológicas en diferentes idiomas también pueden almacenarse de forma estructurada, normalmente en archivos TBX, que son similares a los archivos TMX descritos anteriormente pero que contienen términos equivalentes pareados en lugar de frases pareadas.

Existen otros tipos de recursos lingüísticos, como las descripciones gramaticales de un idioma o los tesauros. Sin embargo, a efectos prácticos, los tres tipos de recursos descritos anteriormente son los más valiosos a la hora de mejorar los servicios de traducción.

Las memorias de traducción y las bases de datos terminológicas, en particular, son recursos muy importantes para los servicios de traducción: los traductores humanos dependen de ellos para mejorar su productividad, ya que pueden recuperar y reutilizar el material previamente traducido, mientras que los sistemas modernos de traducción automática aprenden a traducir a partir de las traducciones humanas almacenadas en corpus y memorias de traducción bilingües.

Proporcionar recursos lingüísticos de calidad es, por tanto, una parte crucial en la construcción de una Europa verdaderamente multilingüe. En esta línea, ELRI aspira a proporcionar servicios que faciliten la creación y el intercambio de estos recursos.

```
[...]
<tu tuid="1" creationdate="2018-09-07T12:01:58Z" creationid="ELRI">
  <tuv xml:lang="en">
    <seg>This is an example.</seg>
  </tuv>
  <tuv xml:lang="es">
    <seg>Esto es un ejemplo.</seg>
  </tuv>
</tu>
<tu tuid="2" creationdate="2018-09-07T12:01:58Z" creationid="ELRI">
  <tuv xml:lang="en">
    <seg>A translation memory contains translation units.</seg>
  </tuv>
  <tuv xml:lang="es">
    <seg>Una memoria de traducción contiene unidades de traducción.</seg>
  </tuv>
</tu>
[...]
```

Figura 4 Ejemplo de contenido TMX

## 4.2 ¿Qué ficheros se pueden descargar?

Los recursos que se comparten vienen en un formato preparado, creado automáticamente por los motores ELRI a partir de los datos cargados por los usuarios. En otras palabras, los datos originales aportados por los usuarios (por ejemplo, la recopilación de documentos en varios idiomas) no se comparten tal cual: los datos originales se someten a una serie de fases de procesamiento que conducen a la creación de recursos lingüísticos estructurados, que luego se comparten, de acuerdo con los permisos especificados por los usuarios que subieron los datos originales.

Estos recursos lingüísticos estructurados vienen en los siguientes formatos:

- Memorias de traducción: archivos en formato TMX estándar, donde las traducciones se emparejan a nivel de oración, siguiendo un proceso de alineación original-traducción.
- Ficheros monolingües: cuando un documento en un idioma determinado no tiene una contrapartida de traducción en otro idioma, los motores ELRI procesan igualmente el fichero original, dividiendo el contenido del fichero en frases separadas, una por línea, dejándolo preparado para posteriores procesamientos.
- Archivos de intercambio de terminología: archivos en formato TBX estándar, donde se emparejan términos equivalentes en diferentes idiomas.

Todos los recursos descargables vienen en un archivo comprimido (*archive.zip*) que contiene uno o más archivos en el formato(s) especificado anteriormente.

Los pasos relacionados con la descarga de recursos se describen en detalle en la sección 5.1.



### 4.3 ¿Qué ficheros se pueden subir?

Como se ha indicado anteriormente, las estaciones ELRI integran motores de procesamiento de datos. La función de estos motores es crear recursos de lenguaje estructurados a partir de datos en bruto, proporcionados por los usuarios, en diferentes idiomas y en un gran número de formatos, incluyendo MS Word doc(x), xls(x), rtf, odt, txt, pdf, tmx, tbx, sdtm, xml y archivos comprimidos tipo zip.

De esta manera, los usuarios pueden simplemente cargar diferentes documentos en uno o más idiomas y dejar que los motores ELRI preparen recursos estructurados en segundo plano. Una vez que los recursos hayan sido preparados y validados por el equipo de ELRI en España, se publicarán en la aplicación web y serán visibles de acuerdo con las restricciones de uso compartido especificadas por el usuario que aportó dichos recursos.

Los principales escenarios relacionados con la carga de datos son los siguientes:

- *Subir una colección de documentos con el mismo contenido traducido, en dos o más idiomas diferentes.* Los motores ELRI producirán memorias de traducción a partir del contenido traducido, al emparejar automáticamente los documentos, en primer término, y, a continuación, las frases dentro de cada documento.
- *Subir memorias de traducción preexistentes.* En la práctica, las memorias de traducción pueden contener errores, como por ejemplo segmentos en un idioma incorrecto o frases mal alineadas. En ese caso, los motores ELRI proporcionan la funcionalidad de limpiar automáticamente las memorias de traducción ya existentes.
- *Subir archivos de terminología u otros recursos.* Los usuarios también pueden aportar archivos que no requieren procesamiento automático. Estos archivos también serán examinados y validados por el equipo de ELRI antes de publicarlos para ser compartidos.

Los pasos necesarios para subir los recursos lingüísticos se describen en detalle en la Sección 5.1.

### 4.4 ¿Quién comparte con quién?

ELRI tiene como objetivo facilitar los medios para compartir datos, de manera flexible, puesto que las restricciones de acceso pueden variar dependiendo de cada recurso.

La compartición a través de una estación ELRI se realiza a través de **grupos**. Los usuarios pueden buscar y descargar sólo aquellos recursos que se comparten en los grupos a los que pertenecen. Hay tres grupos principales a los que cualquier usuario de una estación ELRI pertenece por defecto:

- **Organizaciones nacionales:** Este grupo incluye todos los organismos públicos de España, y por tanto, a todos los usuarios registrados en la estación nacional ELRI. Los recursos compartidos con este grupo serán accesibles a todos los usuarios registrados en la estación española.
- **Organizaciones nacionales + Comisión Europea:** Este grupo incluye a los organismos públicos españoles y a la Comisión Europea, que puede utilizar los recursos lingüísticos para mejorar los servicios de traducción automática que presta gratuitamente a las administraciones públicas de los Estados miembros. Los recursos compartidos con este

grupo serán accesibles a todos los usuarios registrados de la estación ELRI y también a los servicios de la Comisión Europea, a través del repositorio europeo ELRC-SHARE.<sup>5</sup>

- **Open Data:** Este grupo incluye a todos los usuarios registrados de la estación ELRI y a todos los usuarios del portal gratuito de datos abiertos de la Unión Europea.<sup>6</sup> Los recursos compartidos con este grupo se colocarán en el repositorio de Open Data además de la estación ELRI.

Estos grupos por defecto están siempre disponibles para los proveedores de datos y su objetivo es cubrir los casos más frecuentes de intercambio de recursos. En caso de que surjan diferentes necesidades de compartición de recursos específicos, los usuarios podrán solicitar la creación *ad hoc* de grupos específicos poniéndose en contacto con el equipo especializado de ELRI en España.

Por ejemplo, los usuarios de una determinada organización pública (denominada *OrganizaciónX*) pueden tener que restringir el uso compartido de un recurso a los miembros de la propia *OrganizaciónX* y los de la Comisión Europea, para ayudar específicamente a mejorar los servicios de traducción automática de la Unión Europea, sin necesidad de compartirlo con otras organizaciones, debido a restricciones específicas de distribución.

Esto puede lograrse solicitando la creación de un grupo específico que reúna las dos entidades, por ejemplo, *OrganizaciónX + Comisión Europea*. Los usuarios de *OrganizaciónX* podrán entonces contribuir con sus recursos a la infraestructura europea, y seleccionar este grupo específico para que coincida con el nivel de uso compartido que desean.

## 5 Subida y descarga de recursos

Las dos secciones siguientes ofrecen una guía paso a paso para los usuarios que deseen acceder a los recursos lingüísticos o aportar nuevos recursos. Para acceder a los servicios, los usuarios deben primero iniciar sesión en el sistema, como se describe en la Sección 3.

### 5.1 Contribución de recursos lingüísticos

Para contribuir con sus propios recursos, los usuarios deben hacer clic en la opción **Contribuir** de la barra de menú, que abre la página donde se puede proporcionar información sobre el recurso. La parte superior de la página se muestra en la Figura 7.

---

<sup>5</sup> <https://elrc-share.eu/>

<sup>6</sup> <https://data.europa.eu/euodp/en/home>

Contribuir
Uso
Condiciones de servicio
Ayuda
jdoe
Cerrar sesión

## Subir datos

Por favor, proporcione la información requerida a continuación para los datos que desea subir. Una vez subidos, los motores de ELRI procesarán sus datos con el objetivo de proporcionarle a usted, y a los usuarios del grupo de su elección, un recurso lingüístico totalmente preparado.

Una vez que su recurso haya sido completamente procesado, se le notificará por correo electrónico y podrá bajárselo desde este repositorio.

**Título del recurso\***

El nombre por el que el recurso es conocido o por el que desea que se conozca; p.e. "El corpus bilingüe GSRT de noticias griego-inglés"

**Descripción breve del recurso**

[Opcional] Una breve descripción, incluyendo cualquier información que se considere útil sobre el recurso, por ejemplo, si se trata de un corpus (colección de documentos) o de un léxico, glosario, recurso terminológico, etc., su tamaño, idioma(s), clasificación (por ejemplo, informes clínicos, boletines de noticias, glosario de terminología deportiva, etc.)

Figura 7 Información básica de los recursos

En esta parte de la página, el único campo obligatorio (señalado por el asterisco rojo) es un nombre para el recurso (definido por el usuario). Sin embargo, proporcionar una breve descripción del recurso resulta útil para los revisores del equipo ELRI, que examinarán el recurso cargado con el fin de validarlo.

Más adelante en la página (Figura 8), el usuario debe especificar el idioma(s) del recurso (esta acción también es obligatoria, como muestra el asterisco rojo). En este sencillo ejemplo, los archivos contienen textos en inglés y francés y es necesario seleccionar estos dos idiomas. Además, el usuario tiene la opción de seleccionar también el(los) dominio(s) relevante(s) de la lista de opciones proporcionada bajo el título correspondiente ("Otro / Desconocido").

[Contribuir](#)
[Uso](#)
[Condiciones de servicio](#)
[Ayuda](#)
[jdoe](#)
[Cerrar sesión](#)

**Idioma(s)\***

Inglés  
Francés  
Alemán  
Italiano  
Irlandés  
Español  
Portugués

Idioma(s) del recurso (para recursos multilingües, mantenga pulsada la tecla CTRL para seleccionar varios idiomas)

**Dominio(s)**

Otro/desconocido  
eCulture  
Intercambio electrónico de información de la Seguridad Social  
eHealth  
eJustice  
eProcurement  
Datos públicos abiertos  
Internet más seguro  
Sistema de interconexión de registros mercantiles  
Ciberseguridad

[Opcional] Dominio se refiere a la información que pertenece a un campo de actividad específico (por ejemplo, cultura o salud) y está relacionada con uno de los bloques de construcción de los servicios digitales europeos indicados en la lista. Los recursos digitales que no pertenecen claramente a uno de los dominios proporcionados se pueden dejar como Otros/desconocidos.

Figura 8 Idiomas y dominio

Finalmente, la parte inferior de la página permite al usuario especificar las condiciones de licencia asociadas al recurso, cargar los archivos y seleccionar el grupo o grupos con los que compartir el recurso, como se muestra en la Figura 9. Se trata de acciones obligatorias, como indican los asteriscos rojos.

Contribuir
Uso
Condiciones de servicio
Ayuda
jdoe
Cerrar sesión

**Condiciones de licencia\***  
Creative Commons Zero 1.0 (CC0-1.0)

Las condiciones de licencia para el recurso. [Más información sobre las licencias disponibles](#)

**Subir recurso\***  
Examinar... 2 archivos seleccionados.  
Por favor, cargue un solo archivo .zip o cualquier cantidad de archivos del siguiente tipo: .doc(x), .odt, .rtf, .pdf, .tmx, .sdltn, .xml, .tbx, .xls(x) o .txt hasta 100MB.  
En caso de que los datos que desea subir sean mayores de 100MB, por favor póngase en contacto con nosotros para proporcionar un enlace SFTP de carga directa o contemple la posibilidad de subir archivos más pequeños.

**Grupo(s)\***  
Organizaciones Nacionales  
Open Data  
Organizaciones Nacionales + Comision Europea

Idioma(s) del recurso (para recursos multilingües, mantenga pulsada la tecla CTRL para seleccionar varios idiomas)

☒ Estoy de acuerdo con los [Términos de servicio de subida de recursos lingüísticos para los titulares de datos](#)

Enviar
Reiniciar

Figura 9 Licencia, carga de archivos y selección de grupos

Después de estar de acuerdo con las Condiciones del Servicio relacionadas con la carga del recurso y hacer clic en el botón **Enviar**, los archivos se cargan. Si la carga se realiza correctamente, aparecerá el mensaje que se muestra en la Figura 10.

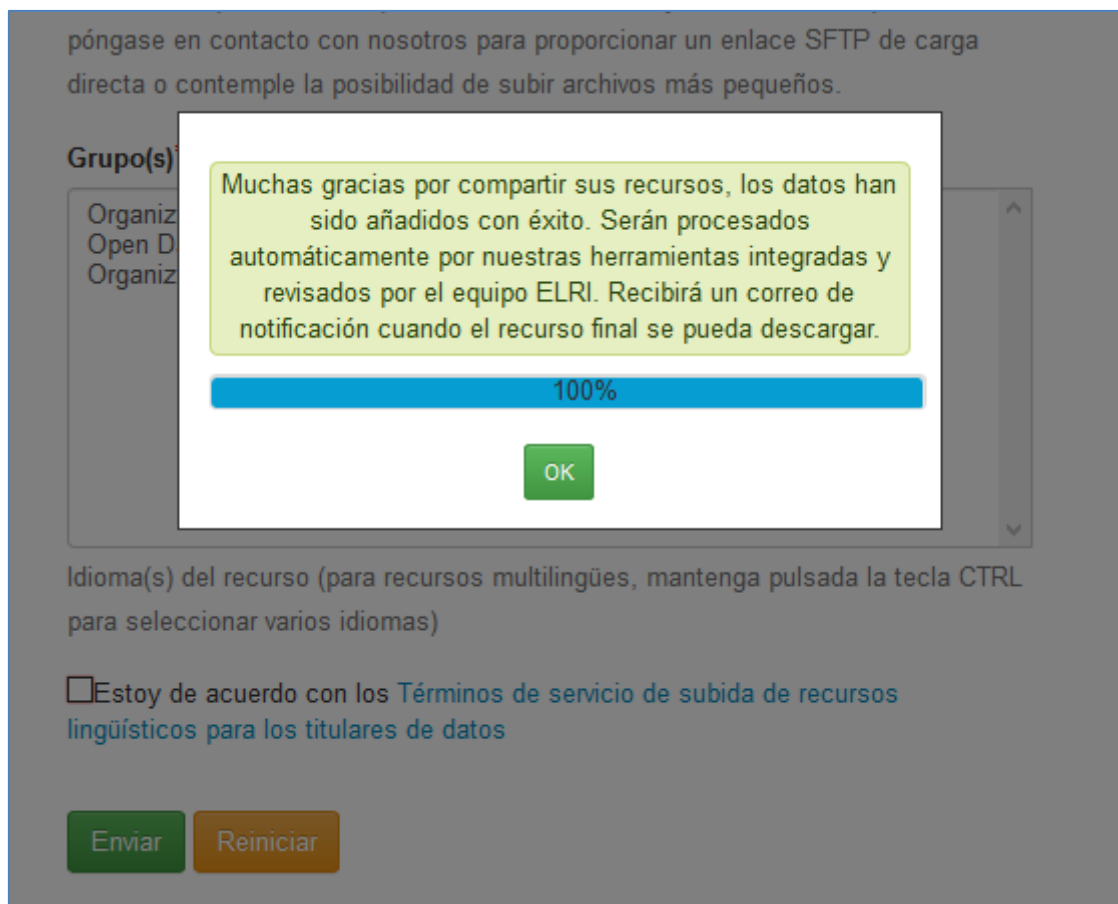


Figura 10 Contribución exitosa de recursos

A partir de este momento, los archivos cargados serán procesados por los motores de ELRI y los resultados junto con la información proporcionada serán revisados por el equipo de ELRI a cargo de la estación de datos. Este equipo se asegura de que los resultados del procesamiento y la información legal sean válidos. Después de esta validación, se publicará el recurso, en las condiciones establecidas por el usuario que lo ha aportado, y se notificará por correo electrónico a dicho usuario (Figura 11).



Figura 11 Correo electrónico de notificación para la publicación de recursos

Una vez publicado, el recurso será accesible desde el catálogo, tanto para el usuario que lo subió como para todos los miembros de los grupos seleccionados, tal y como se describe en la Sección 5.2.

## 5.2 Descarga de recursos lingüísticos

Para acceder a los recursos lingüísticos, los usuarios deben hacer clic en la opción **Examinar** de la barra de menús, que abre la página del catálogo de recursos, tal y como se muestra en la Figura 12.

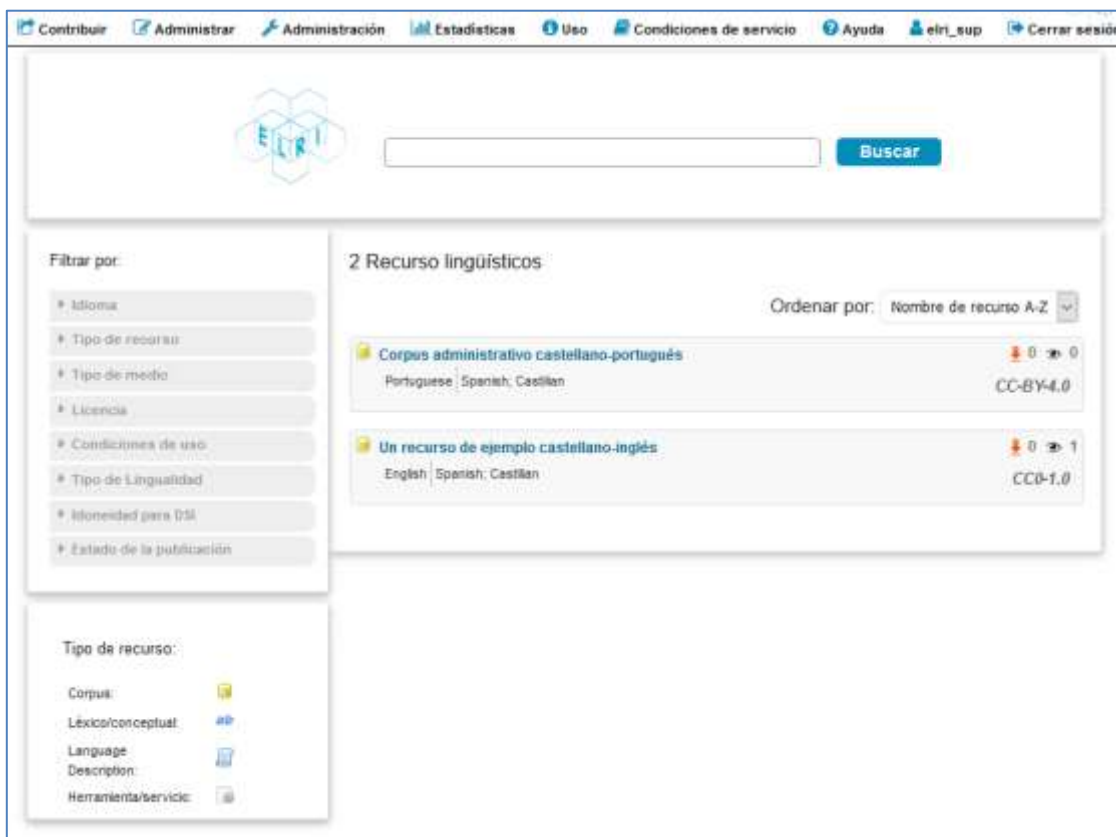


Figura 12 Página del catálogo de recursos

Desde esta página, los usuarios pueden buscar recursos específicos a través del cuadro de búsqueda en la parte superior de la página, ver los recursos disponibles página a página, o filtrar los recursos a través de las opciones del panel que aparece en la parte izquierda de la página.

Para acceder a un recurso específico, los usuarios simplemente deben hacer clic en el nombre del recurso. Esta acción abre una nueva página con información específica relacionada con el recurso, como se muestra en el ejemplo de la 13.

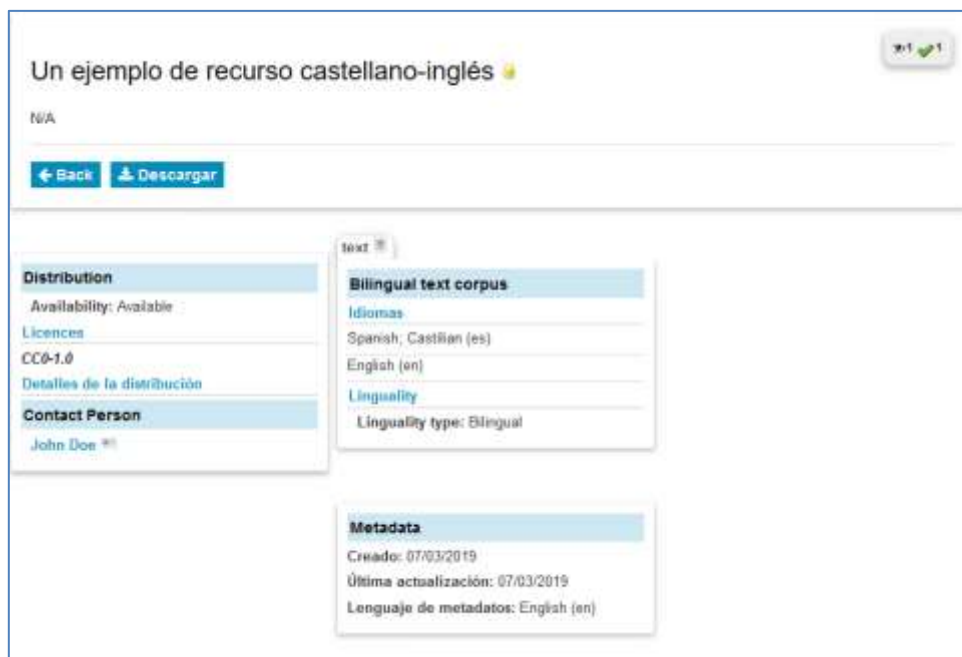


Figura 13 Ejemplo de información de recursos

Para descargar el recurso, los usuarios deben hacer clic en el botón **Descargar**. Se abre la página mostrada en la Figura 14, donde el usuario debe aceptar los términos de licencia asociados a este recurso específico.

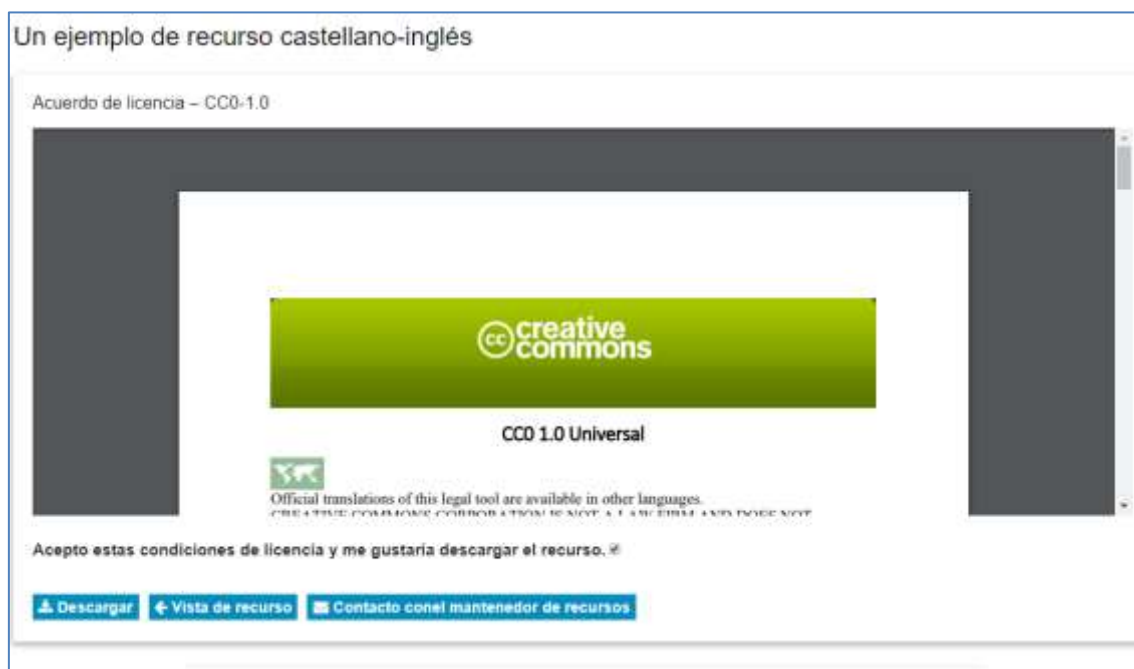





Figura 14 Página de licencia

Una vez aceptada la licencia, se puede realizar la descarga y se descargará un archivo comprimido llamado *archive.zip*. Este archivo contiene el recurso procesado, por ejemplo, una memoria de traducción en formato TMX, en el ejemplo mostrado en la Figura 15.



 license_CC0-1.0.pdf	Adobe Acrobat Document	165 KB
 Un ejemplo de recurso castellano-inglés_en_es_clean.tmx	TMX File	2 KB
 Un ejemplo de recurso castellano-inglés_metadata.txt	TXT File	1 KB

*Figura 15 Recurso descargado*

## 6 Contactar con el equipo de la estación de datos ELRI

Para cualquier consulta o solicitud, el equipo de ELRI puede ser contactado a través de la página de **Ayuda** accesible desde la barra de menú y enviando un correo electrónico a la dirección proporcionada, como se muestra en la Figura 16.

### Detalles de contacto

- Para consultas relacionadas con el proyecto, consulte la [lista de contactos ELRI](#).
- Para obtener ayuda técnica sobre este repositorio nacional, puede ponerse en contacto con nosotros en [elri-nrs-support@vicomtech.org](mailto:elri-nrs-support@vicomtech.org).

*Figura 16 Información de contacto*