



ELRI NATIONAL RELAY STATION

ELRI

USER GUIDELINES

This document provides guidelines for users of the National Relay Stations developed within the European Language Resource Infrastructure (ELRI) project.¹

Dissemination Level	Público
Version	V2.1
Date	2019/03/07
Author(s)	Thierry Etchegoyhen ELRI Consortium



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.

¹ www.elri-project.eu



Contents

1	Introduction	2
2	Joining the National Relay Station (NRS)	3
2.1	User Registration.....	3
2.2	User Authorisation	4
3	Accessing ELRI services	4
4	Using ELRI services.....	5
4.1	What is a language resource?	5
4.2	What can be downloaded?	6
4.3	What can be uploaded?	7
4.4	Who is sharing with whom?	7
5	Walkthrough for Users.....	8
5.1	Contributing language resources	8
5.2	Downloading language resources	12
6	Contacting the NRS team	15

1 Introduction

ELRI National Relay Stations (NRS) are web applications designed to offer services that facilitate the preparation and sharing of language resources (LRs) at different levels: with all registered public organisations at the national level, between specific groups of registered organisations, with the European Commission or as Open Data.

The network is restricted to public organisations of the participating European Union Member States². This means that for the Irish NRS, only staff from an Irish public body can register as users. It also means that LRs are only shared among registered public organisations. Resources whose owners are willing to share with a larger audience (outside Ireland) can opt to share them with dedicated larger groups and these resources will be placed in dedicated repositories such as the ELRC-SHARE or the European Open Data portals.

The Irish NRS is deployed and maintained by a dedicated team based in Dublin City University (DCU). This team oversees user management and the proper handling of resources, while also acting as an interlocutor for users of the ELRI services and for the relevant languages.

In this document, users can find information about joining and using the National Relay Station in Ireland.

The front page of the NRS in Ireland is shown in Figure 1.

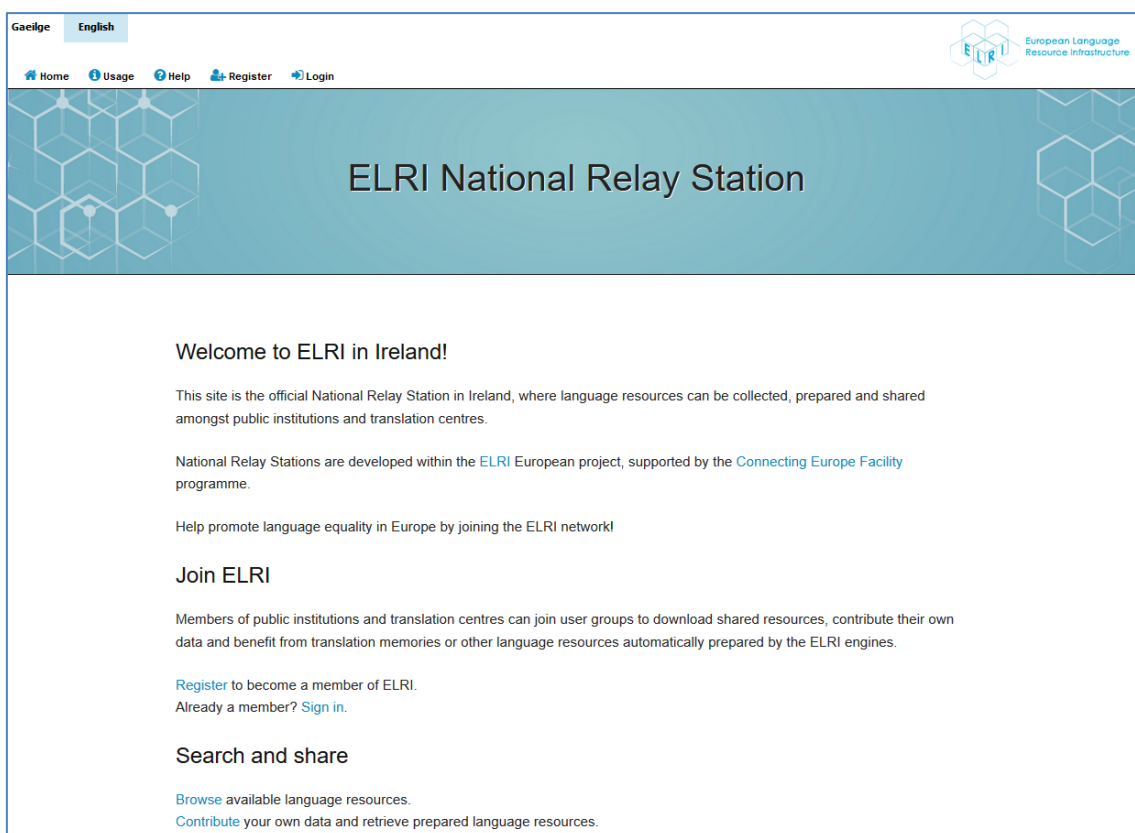


Figure 1 National Relay Station front page

² Currently: France, Ireland, Portugal and Spain.

2 Joining the National Relay Station (NRS)

The NRS is accessible only to registered and authorised users from public organisations. Joining the NRS is thus a two-step process:

1. Users willing to join the network can request the creation of an account via the NRS registration page.
2. Access is granted by members of the dedicated NRS team to users who work in a public organisation.

Sections 2.1 and 2.2 below provide more details on each of these two steps.

2.1 User Registration

New users can register online via the **Register** page shown in Figure 2.

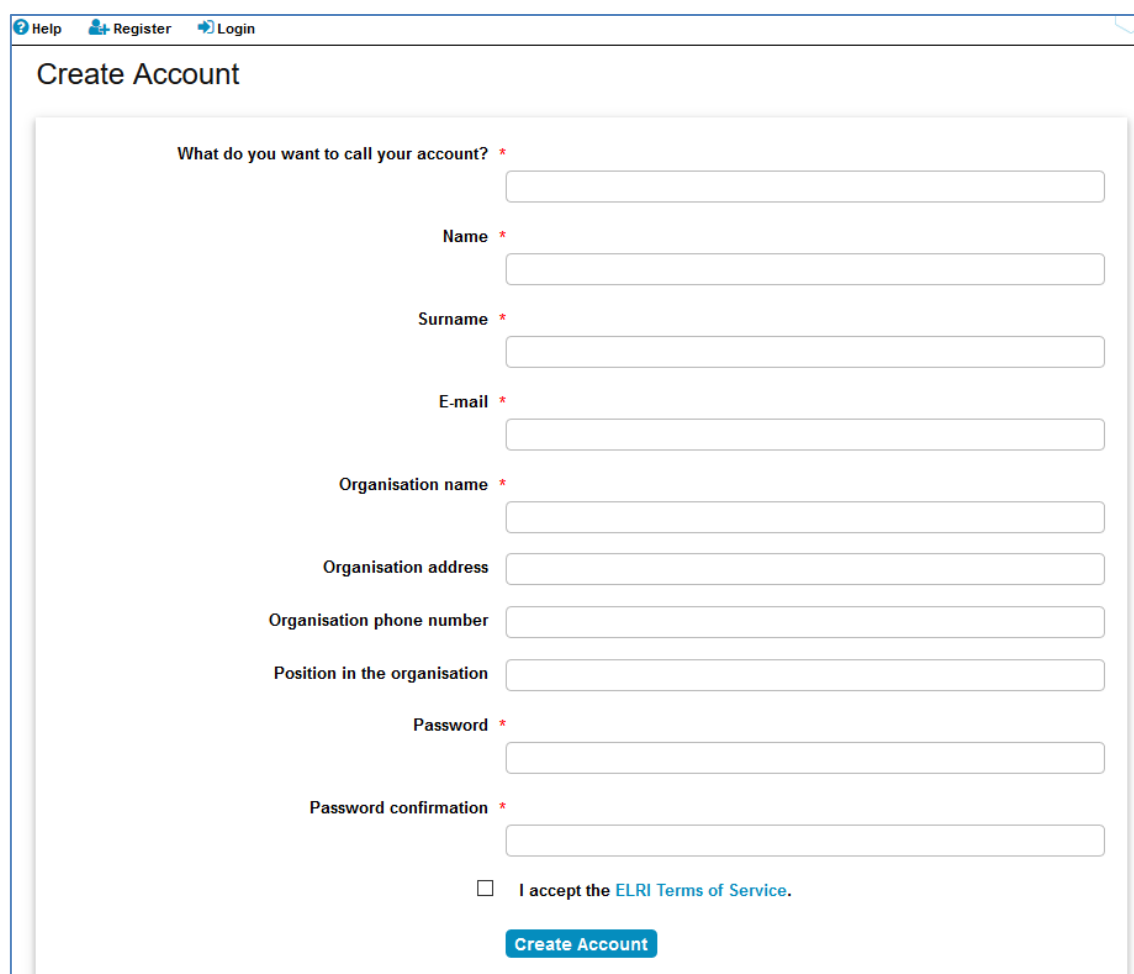


Figure 2 Registration page

After providing the requested mandatory information³, users must accept the ELRI Terms of Service by ticking the relevant checkbox prior to clicking on the *Create Account* button.

After the registration request is sent, it will be evaluated by the dedicated NRS team in DCU and authorised as described in Section 2.2.

³ Indicated with a red asterisk.

2.2 User Authorisation

Registration requests are received by the aforementioned team in DCU. Members of this team will review the request and, having verified that the user belongs to a public institution in Ireland, will authorise the requested account.

Upon authorisation, the user will receive an email notification, as shown in Figure 3, and will be able to access the ELRI services offered by the NRS.

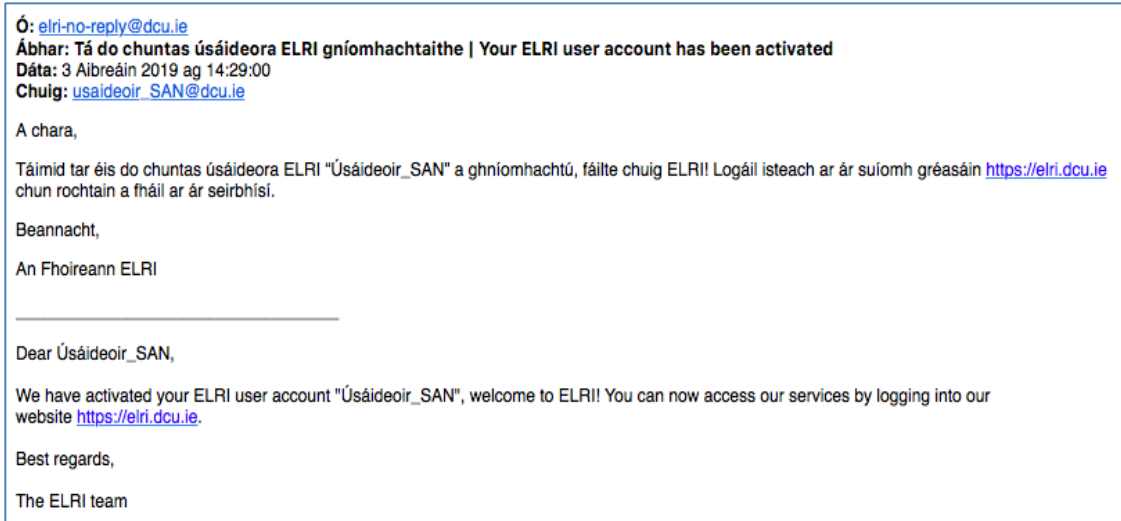


Figure 3 Registration validation notification email

3 Accessing ELRI services

Only registered and authorised users have access to the services provided by the European Language Resource Infrastructure (ELRI), namely browsing, downloading and contributing language resources. Prior to accessing these services, users must therefore first log into the system via the **Login** page shown in Figure 4 with the username and password that they chose during registration.

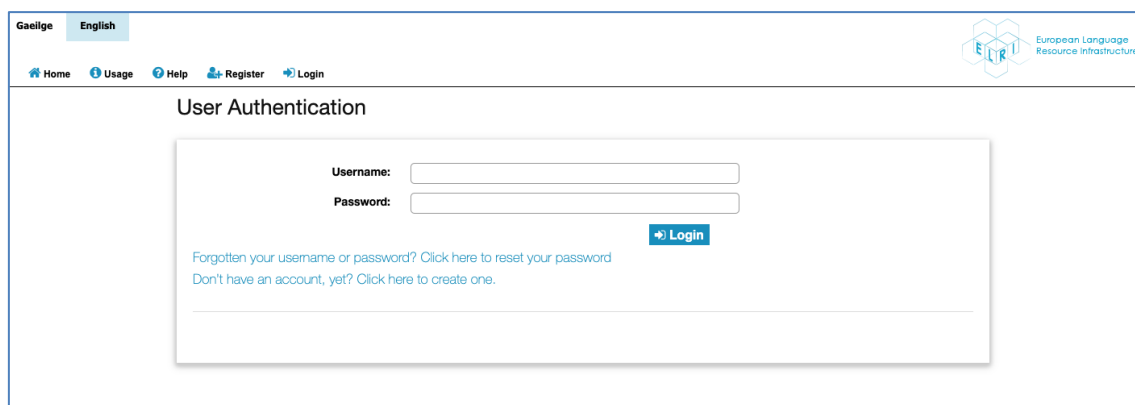


Figure 4 User authentication page

Once logged in, users have access to the menu options shown in Figure 5.

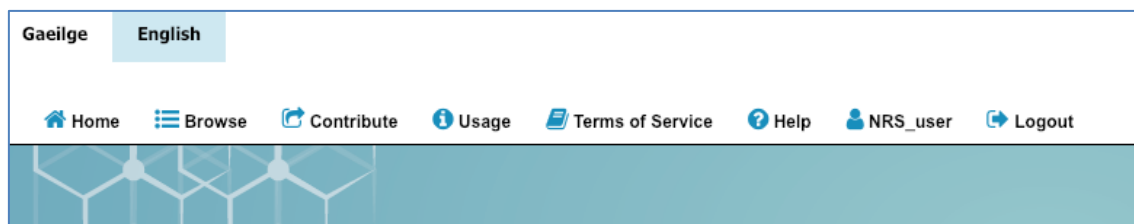


Figure 5 Menu options for logged users

From this menu, users have access to the following services:

- Home: link to the home page.
- Browse: link to the page where users can search and download LR (See Section 4).
- Contribute: link to the page where users can contribute LR (See Section 5).
- Usage: link to instructions on how to use the system.
- Terms of Service: link to the ELRI Terms of Service.
- Help: link to a help page with contact information and support documentation.
- Username⁴: link to a page with profile information for the currently logged-in user.
- Logout: disconnects the logged-in user from the system and closes the session.

4 Using ELRI services

Via the NRS, ELRI essentially provides the means to prepare and share language resources. In the following sections, we explain what language resources are, how users can contribute them, and how sharing works in the NRS.

4.1 What is a language resource?

The term *language resource* encompasses all data that is relevant to provide linguistic services in a given language or group of languages. This includes the following main types of resources:

- Corpus: this is the main type of LR and is essentially a prepared set of documents with text in one or more language(s). A corpus can be *monolingual*, i.e. containing text in one language only in a single file, or *bilingual*, i.e. containing translated texts in two languages and in separate files. Corpora can be created from collections of documents in different formats (e.g., MS Word or RTF), by extracting and processing the content of the documents.
- Translation memory: one of the most useful resources for both human translators and machine translation services is the translation memory. The translation memory is a structured document with the standard .tmx file extension, where translations between languages are paired in terms of translation units (usually sentences aligned in pairs, i.e. a source/original and its translated equivalent in the target language). An example of .tmx content is provided in Figure 6.
- Terminology exchange files: terms which correspond to each other in different languages can also be stored in a structured way, typically in .tbx files that are similar to the .tmx files described above but contain paired equivalent terms instead of paired sentences.

⁴ NRS_user in the example.

Other types of language resources exist, such as grammatical descriptions of a language or thesauri. However, for practical purposes, the three types of LRs described above are the most valuable to improve translation services.

Translation memories and terminology files, in particular, are extremely important resources for translation services: human translators rely on them to improve their productivity, by being able to retrieve and reuse previously translated material, whereas modern machine translation systems learn to translate from the human translations stored in bilingual corpora and translation memories.

Providing quality language resources is thus a crucial part of building a truly multilingual Europe and ELRI aims to provide services that facilitate the creation and sharing of these resources.

```
[...]
<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>
[...]
```

Figure 6 Example of TMX content

4.2 What can be downloaded?

The resources that are shared come in a prepared format, automatically created by the ELRI engines from the data uploaded by users. In other words, the original data contributed by users (e.g., collection of documents in multiple languages) is not shared “as is”. Instead, the original data undergoes a series of processing steps that lead to the creation of structured language resources, which are then shared according to the permissions specified by the users who uploaded the original data.

These structured LRs come in the following formats:

- **Translation memories:** files in standard .tmx format, where translations are paired at the sentence level, following an alignment process (as shown in Figure 6).
- **Monolingual files:** when a document in a given language does not have a translation counterpart in another language, the ELRI engines still produce a processed file from

the original file, with the file content split into separate sentences, one sentence per line.

- **Terminology database files:** files in standard .tbx format, where equivalent terms in different languages are paired.

All downloadable resources are available in a compressed file (*archive.zip*), which contains one or more files in the format(s) specified above.

The steps related to uploading resources are described in detail in Section 5.2.

4.3 What can be uploaded?

As previously indicated, the NRS integrates the ELRI data processing engines. The function of these engines is to create structured language resources from raw data provided by users in different languages and in a large number of formats, including MS Word doc(x), ODT or RTF.⁵

Users can thus simply upload different documents in one or more languages and let the ELRI engines prepare structured resources in the background. After the resources have been prepared and validated by DCU, they will then be published on the NRS and become visible according to the sharing restrictions specified by the user who contributed the resources.

The main scenarios regarding data uploading are the following:

- *Upload a collection of documents that contain the same content translated to or from Irish to or from another language (e.g. Irish/English, Irish/French, Irish/Polish etc.)* The ELRI engines will then produce translation memories (.tmx files) from the translated content by automatically pairing sentences from the original documents.
- *Upload existing translation memories.* In practice, translation memories can contain errors, such as translation units in the wrong language or misaligned sentences. In that case, the ELRI engines provide the functionality to clean up existing translation memories automatically.
- *Upload terminology database files or other resources.* Users can also provide files that do not require automatic processing. These files will also be examined and validated by the ELRI team in DCU prior to publishing and sharing the resources.

The steps required to upload language resources are described in detail in Section 5.1.

4.4 Who is sharing with whom?

ELRI aims to provide flexible means to share data, acknowledging that sharing restrictions may need to vary for specific resources.

Sharing via the NRS is done via **groups**, where users can browse and download only those resources that are shared with a group that they belong to. There are three different groups to which users of the NRS can belong to by default:

- **National Organisations:** This group includes all registered users of the NRS and resources shared with this group will thus be accessible to all registered users of the NRS.
- **National Organisations & European Commission:** This group includes all registered users of the NRS and the European Commission, who may then utilise the shared resources to improve the automatic translation services provided freely to public administrations of the EU Member States. Resources shared with this group will thus

⁵ The complete list of accepted file formats is provided in the *Contribute* page of the NRS.



be accessible to all registered users of the NRS and will also be shared with the European Commission by being placed on the ELRC-SHARE repository.⁶

- **OpenData:** This group includes all registered users of the NRS and all users of the free Open Data portal of the European Union.⁷ Resources shared with this group will thus be placed in the Open Data repository in addition to the NRS.

These default groups are always available to data contributors and aim to cover the most frequent cases of resource sharing. If different sharing needs arise for specific resources, users may request the *ad hoc* creation of specific groups by contacting the dedicated ELRI team in DCU.

As a simple example, users from a given public organisation (e.g. *OrganisationX*) may need to restrict the sharing of a resource to members of *OrganisationX* and the European Commission alone, to specifically help improve the machine translation services of the European Union, without further sharing because of specific distribution restrictions.

This may be achieved by requesting the creation of a specific group joining the two entities, e.g. *OrganisationX & EuropeanCommission*. Users from *OrganisationX* will then be able to contribute their resource and select this specific group to match their desired level of sharing.

5 Walkthrough for Users

The two sections below provide a step-by-step guide for users willing to access or contribute language resources. To access the services, users must first log into the system, as described in Section 3.

5.1 Contributing language resources

To contribute their own resources, users should click on the **Contribute** option in the menu bar, which opens the page where information on the resource can be provided. The top of the page is shown in Figure 7.

⁶ <https://elrc-share.eu/>

⁷ <https://data.europa.eu/euodp/en/home>

Uploading your data

Please provide the required information below for the data you wish to upload. Once uploaded, the ELRI engines will process your data with the goal of providing you, and users of the group of your choice, with a fully prepared language resource.

Once your resource has been fully processed, you will be notified via email and will be able to download it from this repository.

Resource Title*

Press Release from Department of XXXX May 2018

The name by which the resource is already known or by which you would like it to be known; e.g. "The GSRT bilingual corpus of Greek-English bulletins"

Resource short description

An Irish and English version of a press release sent 12/5/18 regarding the launch of a new initiative

[Optional] A short description, including any information considered useful about the resource, e.g. whether it's a dataset (collection of documents) or a lexicon, glossary, terminological resource, etc., its size, language(s), classification information (e.g. health reports, news bulletins, lexicon of sports terminology etc.)

Figure 7 Basic resource information

Under "Resource Title," please enter a user-defined name for the resource and under "Resource short description" provide a description for the resource; for example,

"An Irish and English version of a press release sent 12/5/18 regarding the launch of a new initiative."

Further down the page, the user must specify the language(s) of the resource (this action is mandatory, as shown by the red asterisk). In this simple example, the files contain texts in Irish and English, so these two languages need to be selected, as shown in Figure 8.

In addition, the user has the option of also selecting the relevant domain(s) from the list of options provided under the relevant heading ("Other / Unknown").

Language(s)*

English

French

German

Italian

Irish

Spanish

Portuguese

The language(s) of the resource; for resources with multiple languages, hold down CTRL key to select multiple values

Domain(s)

Other / Unknown

eCulture

Electronic Exchange of Social Security Information

eHealth

eJustice

eProcurement

Public Open Data

Safer Internet

Business Registers Interconnection System

Cybersecurity

[Optional] Domain refers to information that belongs to a specific field of activity (e.g., culture or health) and is related to one of the building blocks of the European digital services indicated in the provided list. Digital resources that do not clearly belong to one of the provided domains can be left as Other/Unknown.

Figure 8 Languages and domain

Finally, the bottom of the page allows the user to specify the license conditions associated with the resource, upload the files and select the group(s) with which to share the resource, as shown in Figure 9. These are mandatory actions, as indicated by the red asterisks.

Licensing conditions* Creative Commons - Attribution 4.0 (CC-BY-4.0)

The licensing conditions for the resource. [More information about the available licences.](#)

Upload Resource*

Choose Files 2 files

Please upload a single **.zip** file or any number of files of the following type: **.doc(x), .odt, .rtf, .pdf, .tmx, .sdlm, .xml, .tbx, .xls(x), or .txt** file up to 100MB.

If the data you wish to upload is larger than 100MB, please contact us to provide an SFTP link for direct download, or else please consider uploading smaller files

Group(s)*

Open Data/Sonrai Oscailte
National Organisations & European Commission/Eagriochtaí Naisiunta & Coimisiún
 National Organisations/Eagriochtaí Naisiunta

The group(s) that the language resource can be shared with (hold down the CTRL key – or CMD key for Apple users – to select multiple groups).

☒ I agree with the [Language Resource upload Terms of Service for Data Holders](#)

Submit Reset

Figure 9 License, file upload and group selection

After agreeing with the Terms of Service related to uploading the resource and clicking on the **Submit** button, the files are uploaded. If the upload is successful, the message shown in Figure 10 will appear.

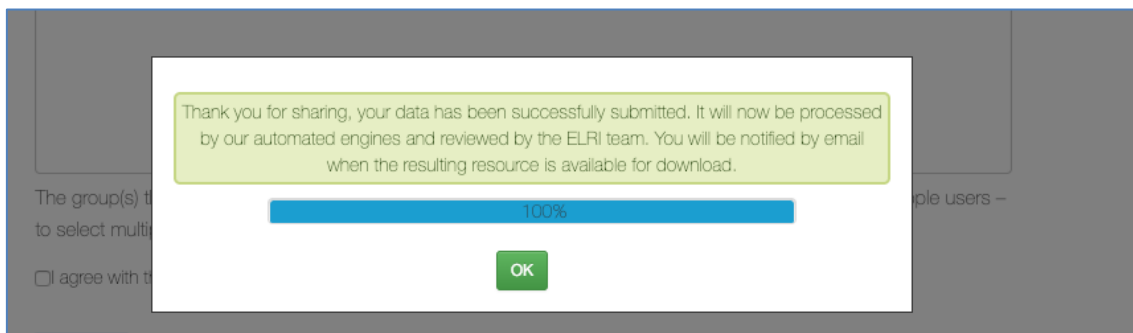


Figure 10 Successful resource contribution



From this point on, the uploaded files will be processed by the ELRI engines and the results along with the provided information will be reviewed by the ELRI team in DCU. Once the resource is validated by DCU (who make sure that the processing results and legal information are valid) the resource will be published. The user who contributed the resource will be notified by email of its publication, as shown in Figure 11.

Ó: elri-no-reply@dcu.ie
Ábhar: Acmhainn foilsithe | Published Resource
Dáta: 4 Aibreán 2019 ag 10:09:00
Chuig: usaideoir_SAN@dcu.ie

A chara,

Tá d'acmhainn Preasraiteas on Roinn ----- Meitheamh 2018 foilsithe ar an Stáisiún Athsheachadta Náisiúnta ELRI. Is féidir leat logáil isteach anois ar ár suíomh gréasáin <https://elri.dcu.ie> agus d'acmhainn ullmhaithe a íoslódáil ón leathanach Brabhsála.

Go raibh maith agat as na seirbhísí ELRI a úsáid. Ná biodh leisce ort teagmháil a dhéanamh linn má tá fiosruithe ar bith agat.

Beannacht,

An fhoireann ELRI

Dear Úsáideoir SAN,

Your resource Preasraiteas on Roinn ----- Meitheamh 2018 has been published on the ELRI National Relay Station. You can now log into our website <https://elri.dcu.ie> and download your prepared resource via the Browse page.

Thank you for using the ELRI services, do not hesitate to contact us for any enquiries.

Best regards,

The ELRI team

Figure 11 Notification email for resource publication

Once published, the resource will be accessible from the catalogue, both to the user who uploaded it and to all members of the selected group(s), as described in Section 5.2.

5.2 Downloading language resources

To access language resources, users should click on the **Browse** menu option from the menu bar, which opens the resource catalogue page, as shown in Figure 12.

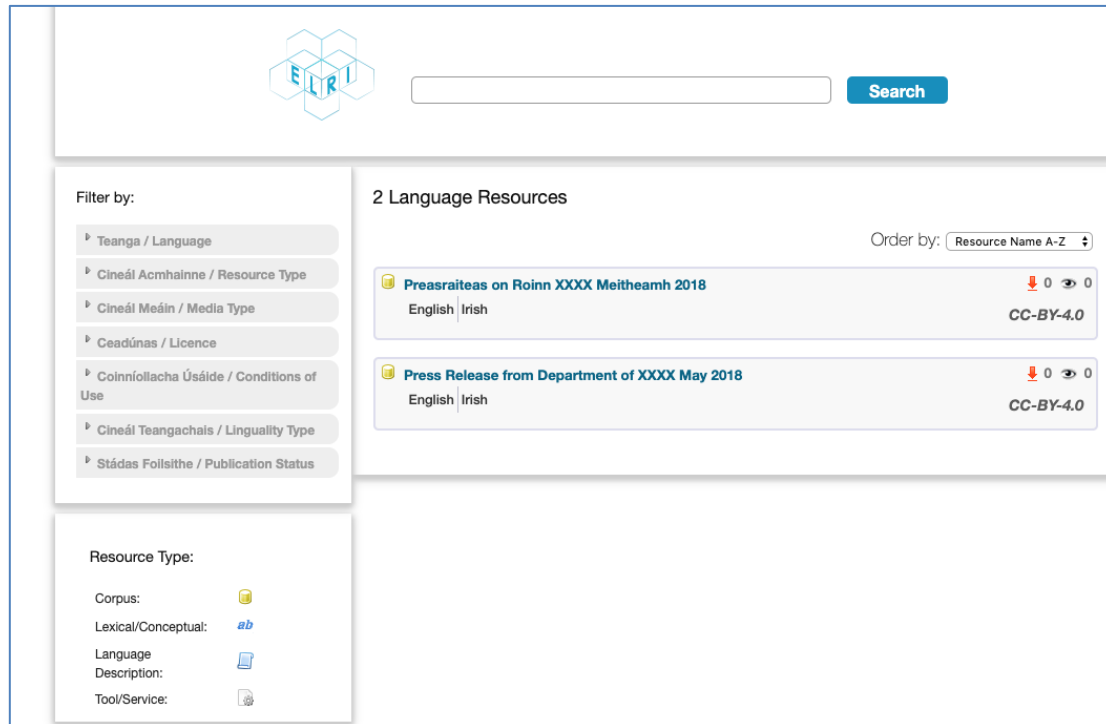


Figure 12 Resource catalogue page

From this page, users can search for specific resources via the search box at the top of the page, view the available resources page by page, or filter the resources via the options in the dedicated panel on the left of the page.

To access a specific resource, users just need to simply click on the resource name, which opens a new page with specific information related to the resource, as shown in the example in Figure 13.

Press Release from Department of XXXX May 2018 📄

An Irish and English version of a press release sent 12/5/18 regarding the launch of a new initiative.

[← Back](#)[Download](#)

Distribution

Availability: Under Review

Licences

CC-BY-4.0

Conditions: Attribution

Distribution Details

Contact Person

NRS User 📧

text 📄

Bilingual text corpus

Languages

Irish (ga)

English (en)

Linguality

Linguality type: Bilingual

Metadata

Created: 04/04/2019

Last Updated: 04/04/2019

Metadata Language: English (en)

Figure 13 Example resource information

To download the resource, users must click on the **Download** button. This opens the page shown in Figure 14, where the user must accept the license terms associated with this specific resource.

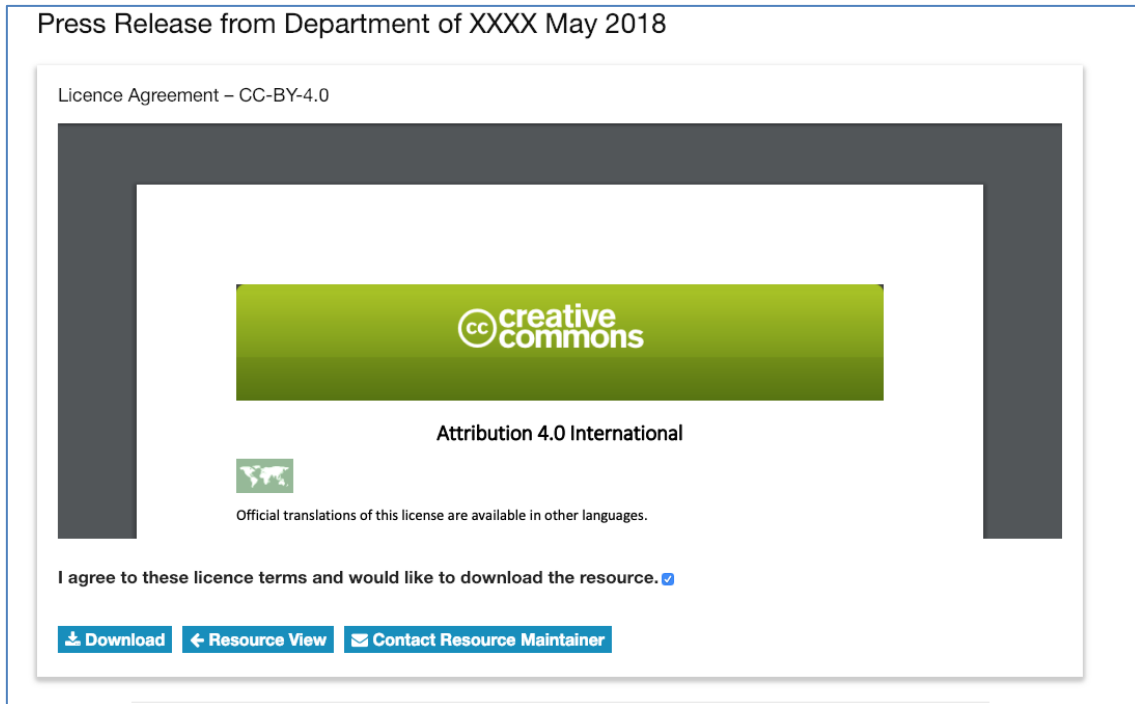


Figure 14 License page

Once the license is accepted by ticking the checkbox, the download can take place and a compressed file named *archive.zip* will be downloaded. This file contains the processed resource, e.g. a translation memory in .tmx format in the example shown in Figure 15.

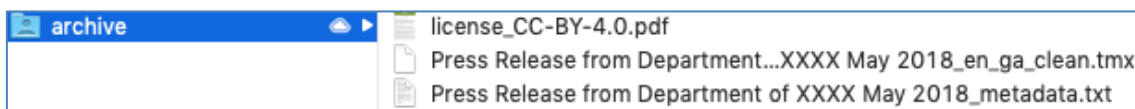


Figure 15 Downloaded resource

6 Contacting the NRS team

For any enquiries or requests, the ELRI team can be contacted via the **Help** page accessible from the menu bar and sending an email to the address provided, as shown in Figure 16.

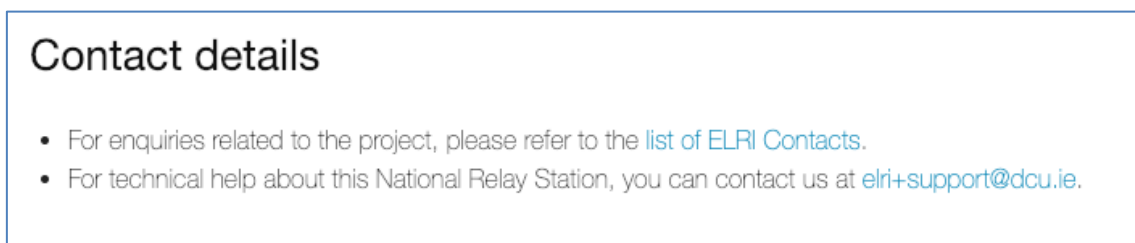


Figure 16 Contact information