ELRC-SHARE Repository Walkthrough for Contributors

Author(s):   Miltos Deligiannis (ILSP)
                        Stelios Piperidis (ILSP)
                        Maria Giagkou (ILSP)
Dissemination Level: Public
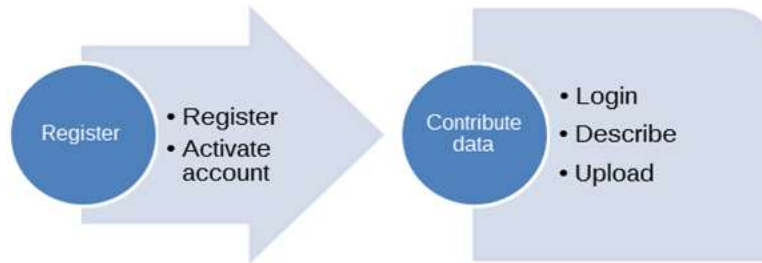**Version No.:**   V4.0
Date: 2019-03-04

Contents

1 Introduction

This document is a simple step-by-step guide to contributing language resources for the purposes of CEF eTranslation through the ELRC-SHARE repository.
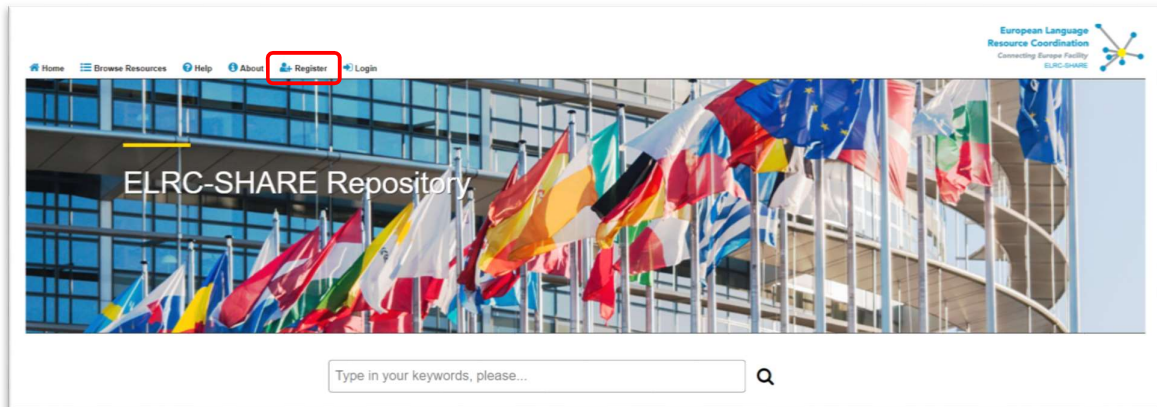
This is done in a simple two-step procedure:

- user registration or, if the user is already registered, login
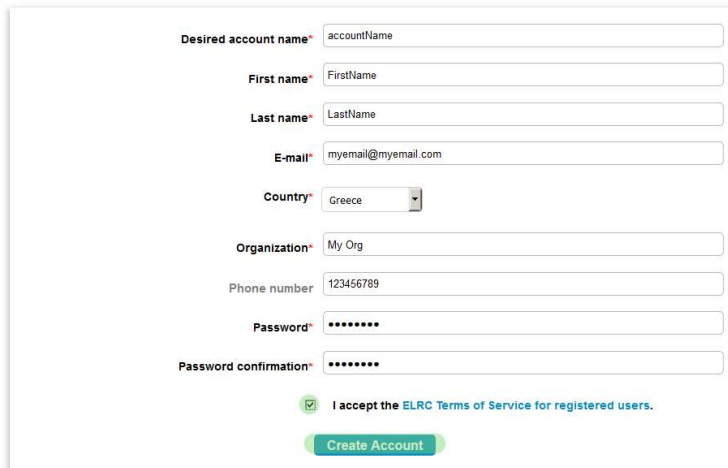- description and uploading of the data

2 Registration and login

Only registered users can contribute resources.

To register, please go to the ELRC-SHARE repository (https://elrc-share.eu) and click on **Register** at the top menu.


On the registration page, fill in all the required information.

Click on the **ELRC Terms of Service for registered users** link to read the terms of use. If you accept the terms of use, check the relevant checkbox.

Click on the **Create Account** button.

A notification message appears that acknowledges receipt of the request and informs you that further instructions have been sent to your email address.
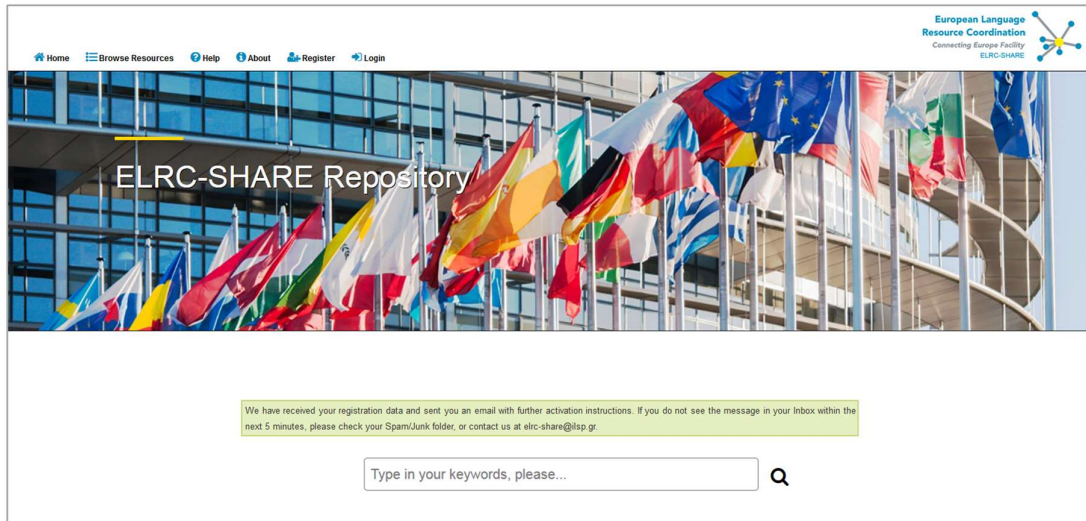
Check your email account and click on the activation link indicated in the email message. Your account will be activated and you will be directed to the **Data Contribution Form** (see section 3).

Once registered, you can log in anytime and contribute data; simply click on **Login** at the top menu and fill in your username and password, as set during registration.
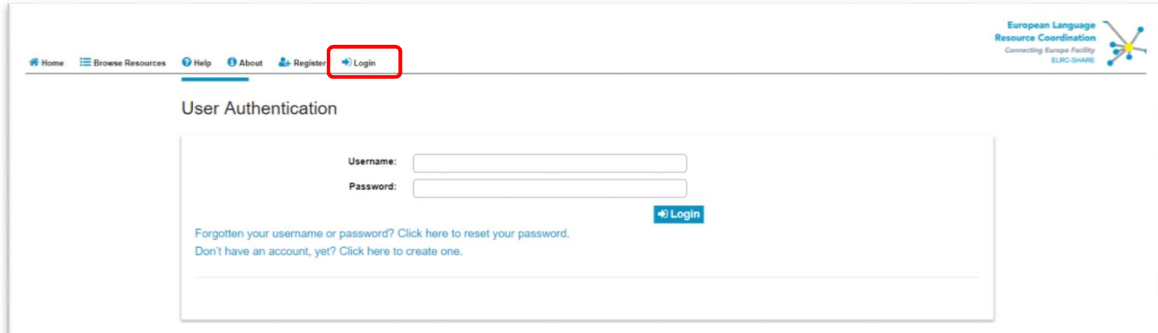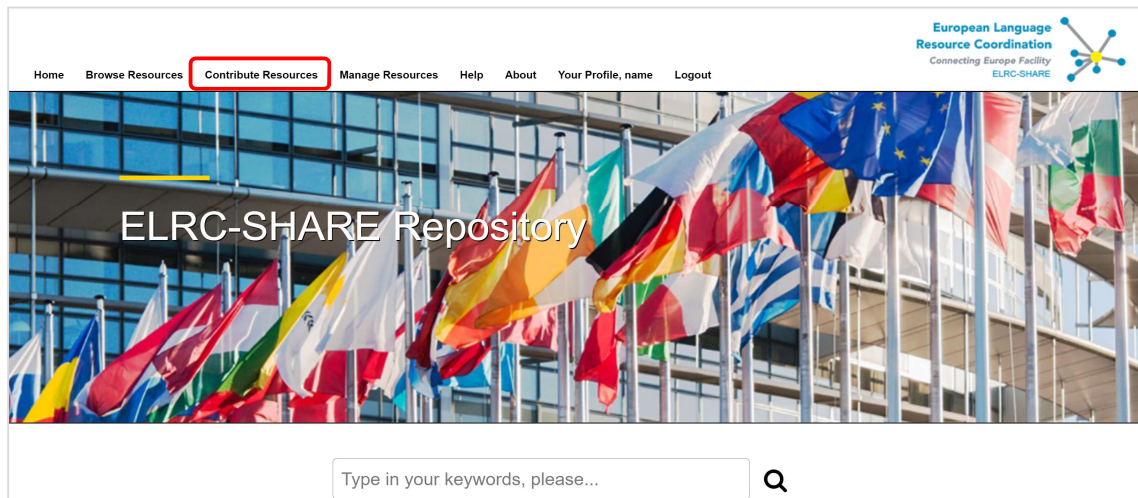
3 Contribution of Resources

Datasets / Language Resources can be contributed to the ELRC-SHARE repository through the following two alternative channels:
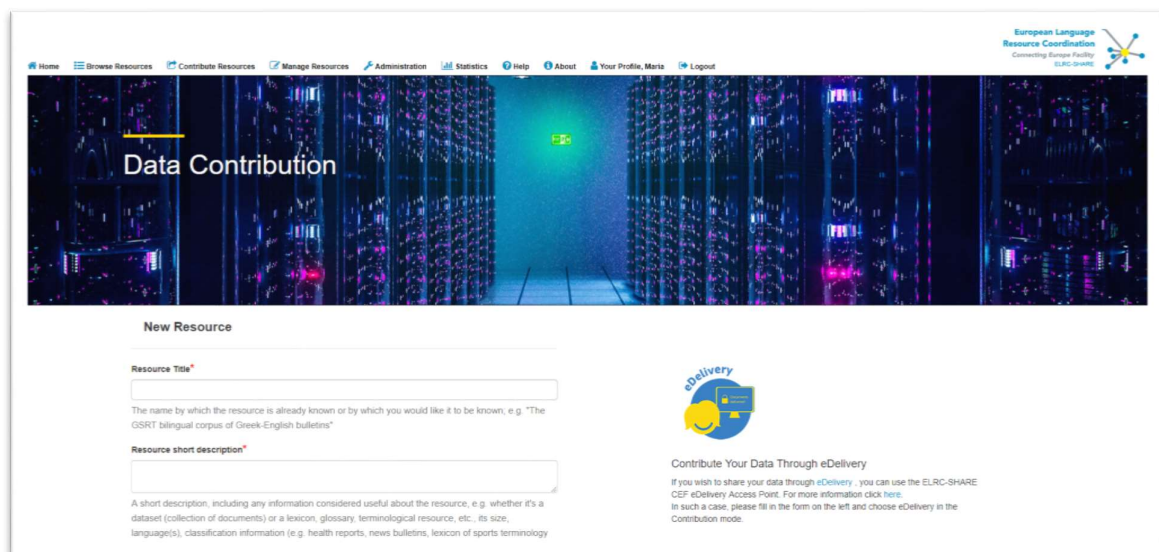
a. the online **Data Contribution Form** (recommended for creating a single language resource at a time) or
b. the **ELRC-SHARE-client API** (recommended for batch creating multiple resources or for batch updating the metadata of your previously created resources).

3.1 Online Data Contribution Form

Once logged in, click on **Contribute Resources** from the top menu.

The following web form appears:

Fill in the fields as appropriate. All fields marked with a red asterisk (*) are mandatory.

- **Resource title**: 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:** 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.)
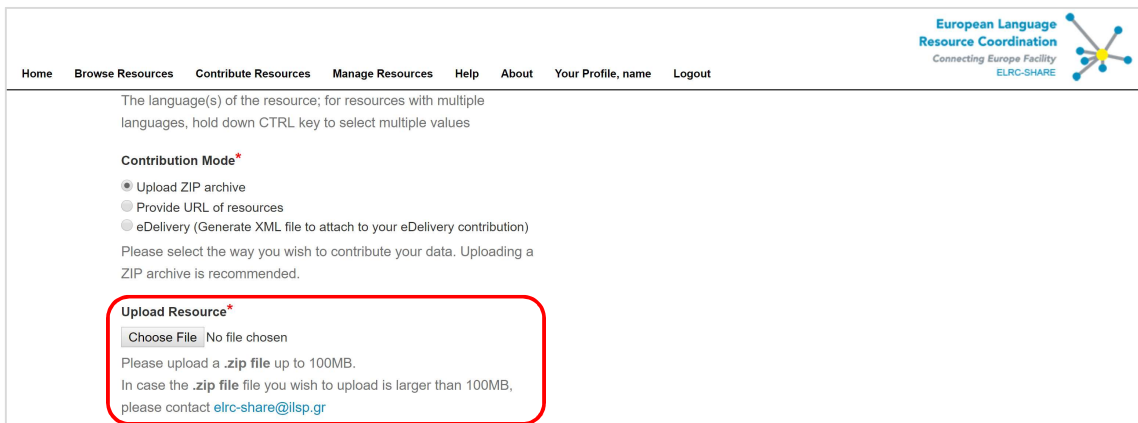- **Language(s):** The language(s) of the resource (to select multiple languages, click and press the CTRL button)
- **Licence:** If known or possible, select one of the standardized licences to determine the conditions/terms under which you are sharing this resource (cf. here). Please note that ELRC may check and validate legal conditions with you.
- **Data owner and other information**: The name of the Data owner / Intellectual Property Rights holder. It can be one or more person(s) and/or organization(s). Write here any other information concerning the terms/conditions under which you are sharing the resource, especially if you were not able to indicate a **Licence** in the field above.
- **Contribution mode:** Select the respective radio button, to choose one of the three alternative available options:
  1. Upload zip archive (recommended), or
  2. Provide a URL, or
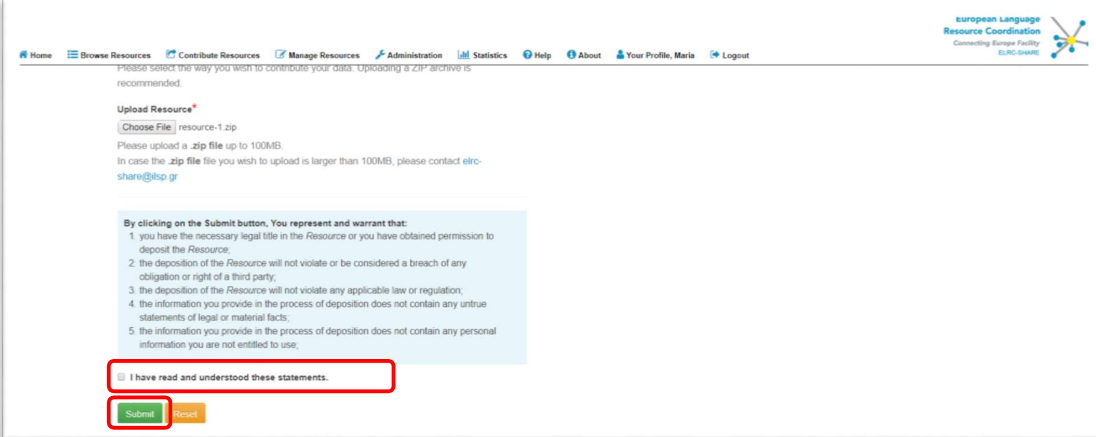 3. eDelivery

3.1.1   Upload a .zip file

When the "Upload zip archive" option is selected as the preferred contribution mode, the **Upload Resource** section appears on screen.

Click on **Browse / Choose file** (naming depends on your browser). In the appearing window, browse your local folders for the intended zipped archive (**.zip**) and select it.

Read carefully the statements that follow and confirm that you have read and understood them by checking the respective checkbox. Then click on the **Submit** button to finalise the contribution of the resource.

Important note: Only zipped files up to 100 MB are allowed for upload through the online contribution form; if a different file format is selected or if the .zip file you are trying to upload exceeds 100 MB, a warning message appears and uploading fails.

3.1.2   Provide a URL

Use this option in case the zipped file you would like to contribute exceeds 100 MB in size. When the "Provide URL of resources" option is selected as the preferred contribution mode, the **Resource URL** section appears on screen.

Type in the URL or FTP address where your dataset is stored. Please note that the URL or FTP link you indicate in this field should be a directory listing of the files you intend to share and should be freely accessible. In case credentials are required to access the directory listing, please make sure you contact us at elrc-share@ilsp.gr to provide us with the required details for accessing your files.

Read carefully the statements that follow and confirm that you have read and understood them by checking the respective checkbox. Then click on the **Submit** button to finalise the contribution of the resource.

3.1.3   eDelivery

Use this option in case your organization has deployed and uses a CEF eDelivery Access Point for secure transfer of datasets up to 400 MB.

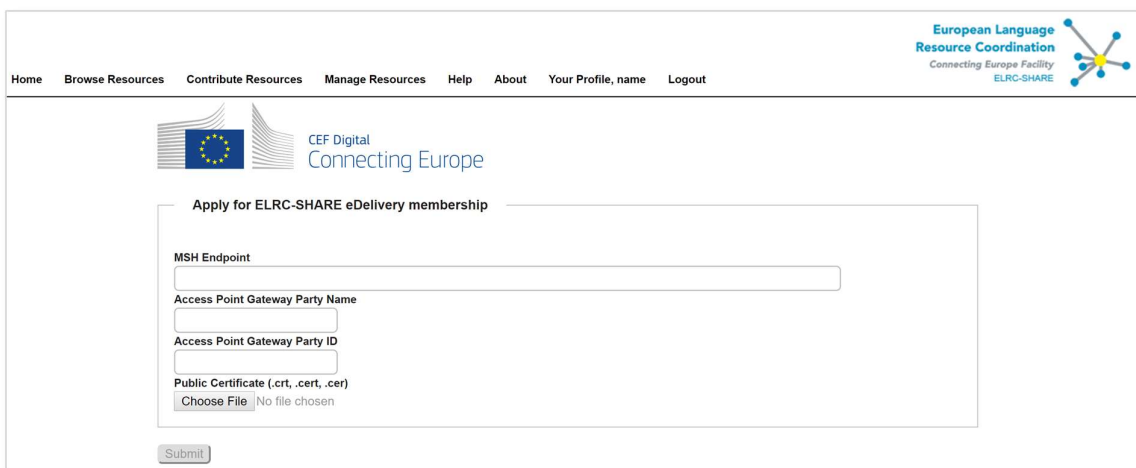When the "eDelivery" option is selected as the preferred contribution mode, simply read carefully the statements that follow and confirm that you have read and understood them by checking the respective checkbox. Then click on the **Submit** button. An .xml file containing the information you filled in the preceding fields of the online Contribution Form is generated and downloaded on your computer. Attach this .xml file to the message you will be transferring to ELRC-SHARE through your eDelivery implementation.

To join the ELRC-SHARE eDelivery Network of trusted parties, fill in the online application available here https://elrc-share.eu/accounts/edelivery_membership_application/.

The following information should be provided through the form:

MSH Endpoint	The endpoint of your Access Point
Gateway Party Name	The name of your Access Point's Gateway
Gateway Party ID	The identifier of your Access Point's Gateway
Public Certificate	Your Access Point's X.509 public certificate (.crt, .cert, .cer)

Once your application to join the ELRC-SHARE eDelivery trusted network is approved, the required ELRC-SHARE Access Point information will be emailed to you. Include this information in your Access Point configuration and complete the process of trust establishment. From this point onwards, you can send your data to the ELRC-SHARE Access Point using your implementation specific software.

The ELRC-SHARE Access Point accepts messages containing two files as attachments:

1. an xml description of your dataset*
2. the actual dataset in zip format, up to 400MB

*The xml file is automatically generated and downloaded on your computer, once you fill in the required fields on the contribution form and select the eDelivery contribution mode.

3.2 ELRC-SHARE-client API

The ELRC-SHARE-client is Python API with an interactive command-line tool for retrieving, creating and updating ELRC-SHARE resources. Its use is recommended for batch creating multiple resources or for batch updating the metadata of your previously created resources and it requires at least some basic programming skills in Python. As such, it is mainly addressed to the CEF-funded projects (generic services) that plan to deliver their datasets through the ELRC-SHARE repository.

The ELRC-SHARE-client package can be downloaded from https://gitlab.com/ilsp-nlpli-elrc-share/elrc-share-client. Documentation and help for the API is also provided through the above link.

4 Data contribution by CEF-funded projects (generic services projects)

In case you need to contribute language resources through ELRC-SHARE on behalf of a CEF-funded project (CEF generic service), please follow these steps:

1. Register to ELRC-SHARE as contributor (see section 2).
2. Contact the ELRC-SHARE repository administrators at elrc-share@ilsp.gr to request **Editor** access rights. State the username you created during registration, your affiliation and role in the context of the CEF-funded project you represent.
3. Contribute your dataset(s) by selecting any of the options described above, i.e. through the online Contribution Form (section 3.1) or by using the ELRC-SHARE-client API (section 3.2). Alternatively, once you are granted Editor access rights, you can contribute esources directly through the ELRC-SHARE documentation environment (see the ELRC-SHARE Walkthrough for Editors).

> ⚠️ **Important note:**
>
> **The responsibility to upload, appropriately describe the datasets with metadata, ingest and publish metadata records that can be browsed through the ELRC-SHARE public inventory lies with the CEF-funded project itself.**
>
> **ELRC will by no means edit or otherwise manipulate an external project's datasets and/or metadata hosted in ELRC-SHARE.**

5 Viewing and accessing your contributed data

Once a resource is contributed to ELRC-SHARE through the online Contribution Form, it has to be imported to the repository database. This process is performed by the ELRC-SHARE technical team, usually within the working day of the contribution. In case the contribution is made during non-working hours, it may be possible that the contributed resource will be imported during the next working day.

In any case, please note that a contributed resource will be available for browsing and viewing in the public section of the repository (https://www.elrc-share.eu/repository/search/), only after it is fully documented.