European Language Resource Coordination (ELRC) is a service contract operating under the EU's Connecting Europe Facility SMART 2014/1074 and SMART 2015/1091 programmes.



ELRC-SHARE Repository Walkthrough for Editors

Author(s): Miltos Deligiannis (ILSP)

> Penny Labropoulou (ILSP) Stelios Piperidis (ILSP)

Maria Giagkou (ILSP)

Dissemination Level: Public Version No.: V4.0

Date: 2019-03-04





Contents

<u>1</u> <u>Introduction</u>	<u> </u>
1.1 What is an ELRC-SHARE Editor	3
1.2 Becoming an ELRC-SHARE editor	3
2 Basic concepts and background	6
2.1 Types of resources on ELRC-SHARE	6
2.2 Publication status of resources	6
2.3 The ELRC-SHARE metadata schema	6
3 The Editor user interface	7
3.1 Editor menu	7
3.2 Managing form	8
3.3 Editing form	9
4 Adding new resources (metadata re	cords) 10
4.1 Create a new resource from scratch	10
4.2 Upload a resource description (in .xn	nl) 11
5 Describing resources / editing meta-	data records 12
5.1 Selecting a resource for editing	12
5.2 Editing of metadata records from co	ntributors 13
5.2.1 Administrative information	14
5.2.2 Resource type-specific information	n 21
5.3 Uploading a dataset	23
5.4 Downloading a resource	25
6 Adding legal documentation to lang	juage resources 27
7 Adding validation report to languag	e resources 28
8 Publishing resource descriptions (m	etadata records) 31
9 Working on resources from crawling	32



1 Introduction

The current document guides editors of the ELRC-SHARE repository through the processes of creating, managing, editing and publishing resource descriptions (metadata records) through the ELRC-SHARE metadata editing (documentation) environment.

1.1 What is an ELRC-SHARE Editor

An ELRC-SHARE editor is a repository registered user that has elevated permissions to create, describe and manage metadata records.

Metadata records and the corresponding uploaded datasets may be added to the repository in one of the following modes:

- (a) contributed by any registered user through the online Data Contribution Form,
- (b) created by an Editor through the ELRC-SHARE-client API
- (c) created by an Editor directly through the ELRC-SHARE metadata editing environment

An Editor can manage the metadata records of resources he/she has created in any of the above modes (a-c). All resources created by an Editor are considered as "**owned**" by the Editor and they are made available to him/her for further manipulation through the ELRC-SHARE documentation environment¹.

Note that the use of "editor status" herein actually describes three sub-groups of users with editor rights in ELRC-SHARE: Global editors (ELRC and CEF-funded projects consortium members), ELRC NAPs, and Legal reviewers. The operational functions on ELRC-SHARE available to each editor sub-group are presented in the following table.

	Edit / update		Ingest		Publish		Delete		Download	
	Own	All	Own	All	Own	All	Own	All	Own	All
	LRs	LRs	LRs	LRs	LRs	LRs	LRs	LRs	LRs	LRs
Global Editors										
Legal			_							
Reviewers		-	_		-				-	
NAPs	•									

1.2 Becoming an ELRC-SHARE editor

Only ELRC consortium members, CEF-funded projects (generic services) and National Anchor Points (NAPs) can request an elevated Editor status.

_

¹ According to the ELRC project operations, Editors that are members of the ELRC project consortium are also considered "owners" of resources contributed through the online Data Contribution Form by specific countries, according to the ELRC regional distribution. This is only relevant to ELRC operations and not to ELRC-external projects.





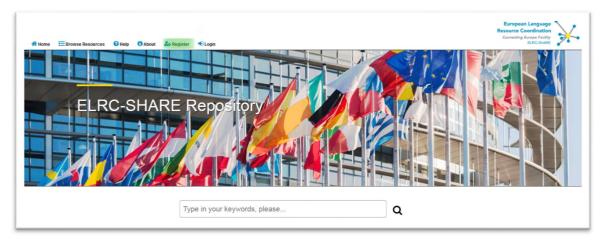


Important note:

If you represent a CEF-funded project, please 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.

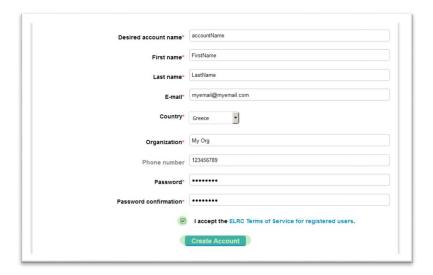
ELRC will by no means edit or otherwise manipulate an external project's datasets and/or metadata hosted in ELRC-SHARE.

To join the ELRC-SHARE users group start by registering: To register, go to the ELRC-SHARE repository (www.elrc-share.eu) and click on **Register**.



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





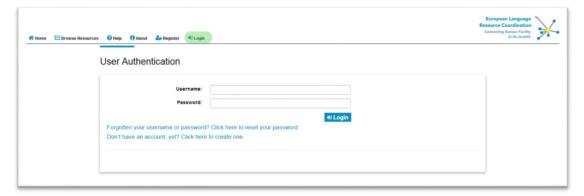
Click on the **ELRC Terms of Service for registered users** link to read the terms of use. If you accept the terms of use, check the relevant checkbox.

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 that is indicated in the email message.

Once registered, contact the ELRC-SHARE repository administrators at elrc-share@ilsp.gr, stating their registered username, affiliation and role in the context of ELRC or of a CEF-funded project. Once granting of Editor permissions is approved by the ELRC consortium, the user is notified by email and he/she can log in to the repository as Editor.

To log in to your account, visit https://elrc-share.eu and click on **Login** at the top menu. Fill in your username and password, as set during registration





2 Basic concepts and background

2.1 Types of resources on ELRC-SHARE

The ELRC-SHARE repository hosts:

- (1) Language data, of the following types:
 - a. multilingual/monolingual corpora,
 - b. Lexical/conceptual resources,
 - c. language descriptions (aka models, e.g. translation and language models)
- (2) **Tools/services**, i.e. **natural language processing tools** (e.g. lemmatizers, part-of-speech taggers, dependency parsers etc.)

2.2 Publication status of resources

The **publication status of a resource** determines the visibility of its description (metadata) and the availability of the actual dataset for download. The publication status can be:

- Internal: It is the initial status of a metadata record upon upload of a dataset. A record in internal status is usually minimally described, thus its metadata should be enriched and, subsequently, ingested by a repository Editor. Internal resources are the least visible and accessible, i.e. only the Editors to whom they are assigned for metadata completion can access and edit them through the documentation environment and they are never displayed on the public inventory of resources.
- Ingested: It is the intermediate publication status of resources, indicating that the
 metadata have been completed by the Editor, although some information may still be
 left incomplete or undecided. Ingested resources are accessible not only by their
 owners, but also by users registered to ELRC-SHARE as European Commission
 members.
- **Published**: It is the status of resources which have been fully described with metadata and validated. Published records are available through the public inventory for search/browse and download (if under a permissive licence) by the general public (i.e. even by unregistered users).

2.3 The ELRC-SHARE metadata schema

The ELRC-SHARE metadata schema is an application profile of the META-SHARE schema, appropriately modified for the requirements of ELRC, e.g. pruned resource types (e.g. audio), adapted legal component (e.g. PSI compliance element, licence values added), additional elements (e.g. Appropriateness for DSI).

The ELRC-SHARE schema includes the following mandatory metadata categories:

- Administrative information
 - o *Identification* (e.g. resource name, description)
 - Distribution (e.g. licensing, access form of the resource)
 - Contact Person
 - o Metadata (e.g. creator and creation date of the metadata record)
- Technical information for all resource types



- o Linguality (i.e. number of languages included in the resource)
- o Languages
- Sizes
- o Text Formats (e.g. plain text, PDF, XML, TMX etc.)
- Technical information for specific resource types
 - o for lexical/conceptual resources
 - Lexical conceptual resource type (e.g. terminological glossary, bilingual dictionary etc.)
 - o for language descriptions
 - Language description type (e.g. grammar).
 - for tools
 - *Tool description type* (e.g. lemmatiser)
 - Additional technical information, where available

In addition, the following metadata categories are optional:

- Administrative information
 - Resource Documentation
 - Resource Creation
 - Funding project
 - o *Relations* (e.g. between an original dataset and its updated version, or its derivate/processing output)
 - Appropriateness for DSI
- Technical information
 - Character Encodings
 - Annotations (only for corpora)
 - Domains
 - Text Classifications
 - Creation
 - o Encoding Information for Language Description (e.g. linguistic level of encoding)
 - Encoding Information for Lexical Conceptual Resource (e.g. types of encoding for its contents, such as lemmas, grammatical information, translation equivalents etc.)

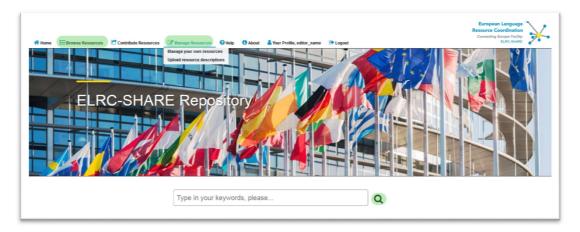
3 The Editor user interface

3.1 Editor menu

Users that are granted editor rights have access to the following menu items (along with the default items for simple registered users):

- Manage Resources
 - Manage your own resources
 - Upload resource descriptions

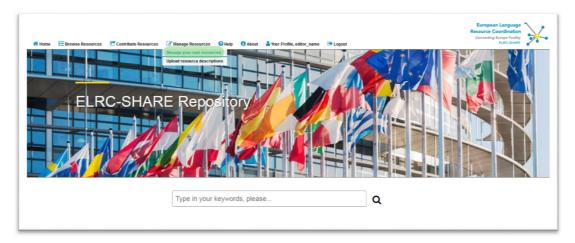




3.2 Managing form

The managing form is the central point from which you can proceed to adding new resource descriptions, editing existing resource descriptions and changing their publication status (e.g. making them publicly available or deleting them).

Access the managing form from the top menu **Manage resources** > **Manage your own resources**.



The managing form lists all the resources owned by the Editor (**resource name**, **ID**, **resource type**, **publication status**, **owners and validation status**). Click on the "ID" row header to sort the resources in ascending or descending ID number.

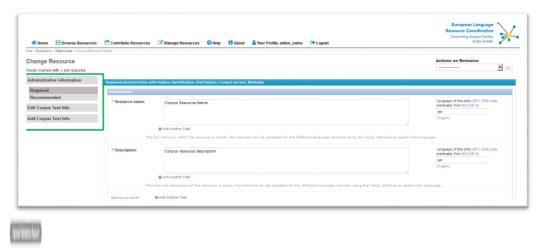




3.3 Editing form

The **editing form** is the outlet of the ELRC-SHARE metadata schema. It is the environment where you can describe a resource and upload the attached dataset.

The editing form displays all the metadata elements that can be filled in for each resource organized in pages: administrative information for all resources and, depending on the resource type, additional technical information. The pages are accessed via the **left side menu**.



Inside each category there are mandatory and optional metadata elements. Mandatory ones are displayed in bold typeface and are marked with a yellow asterisk (*). You can see an explanation for metadata elements when you mouse over each element.



4 Adding new resources (metadata records)

Besides using the modes of contribution available for simple registered users (through the online Contribution Form) or through the ELRC-SHARE-client API (see <u>Walkthrough for contributors</u>). Editors can additionally create new resources directly through the documentation environment.

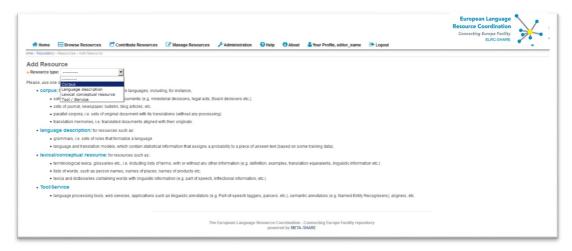
4.1 Create a new resource from scratch

From the top menu select **Manage resources > Manage your own resources,** to go to the Managing form.

On the **Managing form** click on the **Add Resource** button.

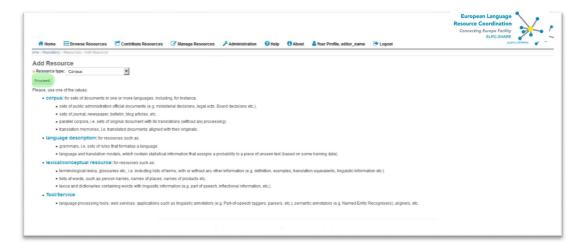


Select the appropriate resource type.



Click on Proceed.

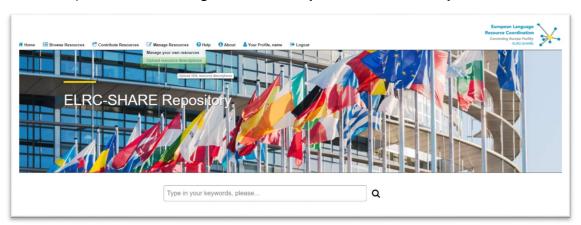




You will be directed to the metadata **editing form** of the specific resource you just created.

4.2 Upload a resource description (in .xml)

From the top menu select Manage resources > Upload resource descriptions.

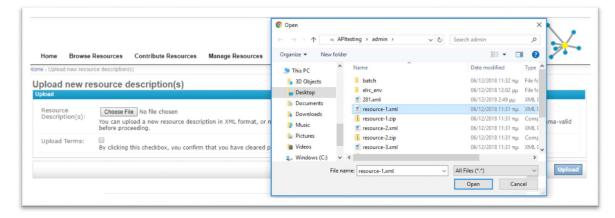


You will be redirected to the **Upload new resource description(s)** form.

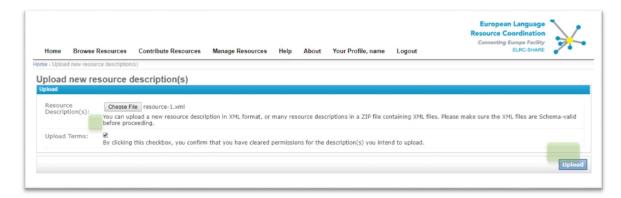
To upload a resource description

a. Click Browse / Choose file (naming depends on your browser) In the appearing window, browse your local folders for the intended .xml file containing the description of the resource or for zipped file (.zip) containing multiple .xml files. In the latter case you will create as many new resources as the number of .xml files contained in your .zip archive. Select the appropriate file and click Open.





- b. Check the **Upload Terms** checkbox,
- c. Click Upload



Three ELRC-SHARE schema-valid .xml templates are available for download in the <u>ELRC-SHARE Help page</u>. You can use the one(s) appropriate depending on the resource type, to edit them and use them for uploading your resource(s) description(s). Note that the .xml templates provided are minimal, i.e. they include the mandatory metadata only.

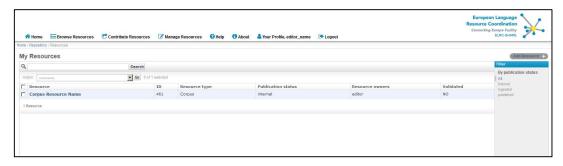
5 Describing resources / editing metadata records

5.1 Selecting a resource for editing

From the top menu select **Manage resources > Manage your own resources,** to go to the Managing form.

On the **Managing form**, from the list of available resources, click on the name of the resource you wish to edit.





You will be directed to the **editing form** where you can complete the resource metadata.

5.2 Editing of metadata records from contributors

Editors are required to edit the metadata records by **filling in at least the mandatory fields** (e.g. licence, text format, size etc.) and, if possible, by providing recommended information that may facilitate further processing of the resource. For example, though optional, it is highly recommended to fill in the fields on **Domain** classification and **DSI relevance**, as these descriptors are of high importance to the EC.

Additionally, for reasons of monitoring and reporting to the EC, all Editors should fill in the **Funding project** element (see section 5.2.1.6).

Note that when editing metadata records for resources that have been previously created (for instance through the online Contribution form) specific fields are pre-populated with the information provided by the resource creator².

⁻

 $^{^2}$ This note is relevant only to ELRC consortium members. It is not relevant to users of ELRC-SHARE that represent external projects.

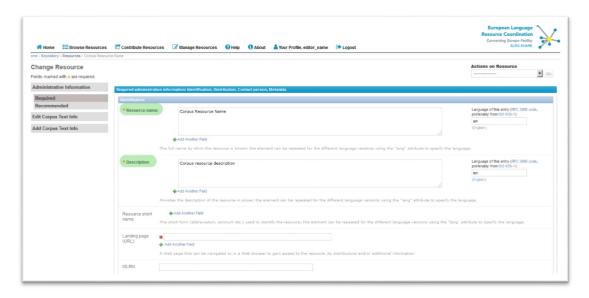
For resources that are contributed to ELRC through the online contribution form the following process is triggered:

[•] the resource description is validated by the ELRC-SHARE administrator, an initial metadata record is created and the information entered by the contributor is fed into the appropriate metadata fields;

[•] a notification email is sent to the ELRC editor that has been designated as representative of the contributor's country of origin (as declared on his/her registration form);

[•] the editor can log into the repository with his/her credentials in order to view, edit and publish the metadata record.





5.2.1 Administrative information

The Administrative Information page contains the basic identifying descriptors of a resource. Among them, **Resource name**, **Description**, **Availability**, **Licence** and **Contact person** are mandatory fields.

5.2.1.1 Resource name

The full name by which the resource is known or you would like it to be known. It is recommended that the resource name is indicated in English, however you can also provide it in another language by clicking on **Add another field**. Make sure you appropriately indicate the input language of each of the Resource name fields in the **Language of this entry** field.

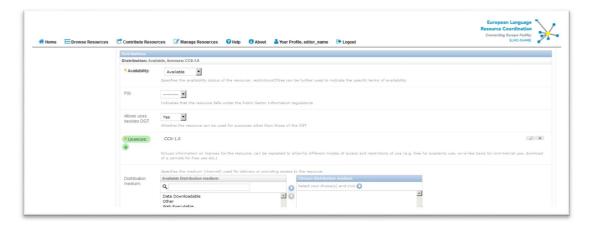
5.2.1.2 Description

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.). It is recommended that the resource description is provided in English, however you can also provide it in another language by clicking on **Add another field**. Make sure you appropriately indicate the input language of each of the Description fields in the **Language of this entry** field.

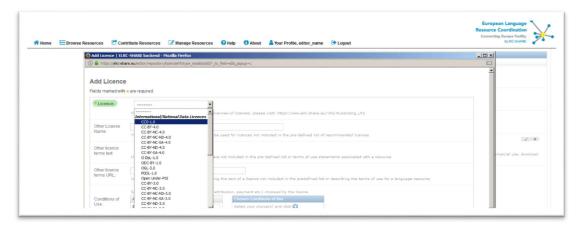
5.2.1.3 Distribution (Licence)

Click on the ricon of the *Licences* field inside the *Distributions* section field.

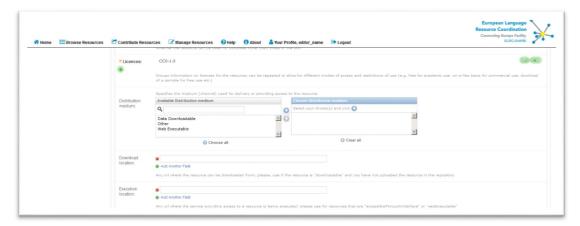




In the "Add Licence" window that appears, select one of the available licenses from the drop down list and click **Save**.



To add a new licence value, click on the picon of the *Licences* field inside the *Distributions* section. To edit or delete an existing licence, click on the or the icon respectively.





A number of additional elements in the Distribution section are optional. Among them the following are **highly recommended and should be checked/filled in by the editor**:

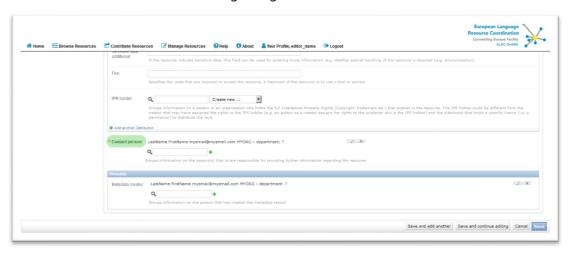
- **PSI**: Indicate whether the resource falls in the regime of the Public Sector Information Directive
- **Allows uses besides DGT**: This field is by default set to "No". Always check the value of the field in accordance with the specified licence.
- Attribution text: If the licence requires attribution, we recommend that you use the following attribution text template, where appropriate:
 <Resource Name> was created for the <Project name> by <Person X, Institute of X> with primary data copyrighted by <Person/Body Y> and is licensed under licence
- IPR holder

value, licence link>

As soon as you have filled in the metadata elements, click **Save** at the bottom of the page.

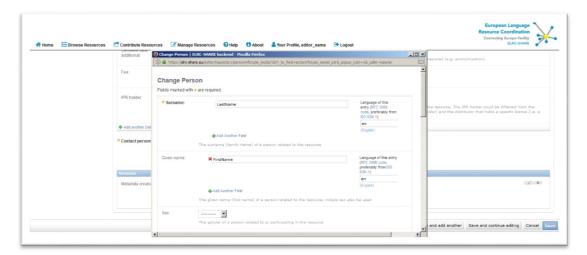
5.2.1.4 Contact person

The *Contact person* contains information about the contributor, which the editors can use to contact them for more information regarding the resource.

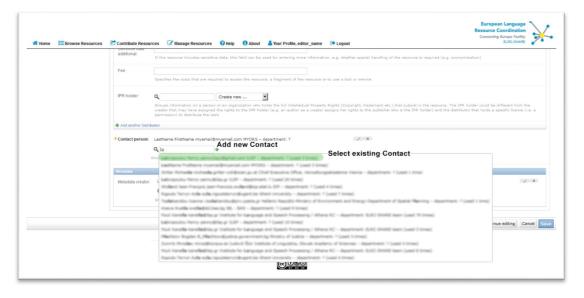


To view/edit details of a contact person, click on the icon.





You can also add another contact person either by selecting an existing one or creating a new one by clicking the \P icon.







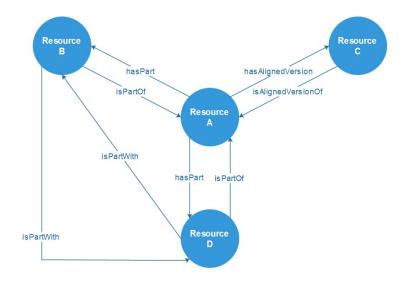
5.2.1.5 Relations between resources

Different resources in the ELRC-SHARE repository can be interrelated. Such relations apply mainly when a resource **B** is a derivative of a resource **A** after processing (e.g. cleaning, format conversion, alignment etc.). The relations available in the repository are classified into two sets 'Has' and 'Is':

HAS	IS			
hasPart	isPartOf			
Haspart	isPartWith			
hasVersion	isVersionOf			
hasAnnotatedVersion	isAnnotatedVersionOf			
hasAlignedVersion	isAlignedVersionOf			
	isRelatedTo			
hasConvertedVersion	isConvertedVersionOf			

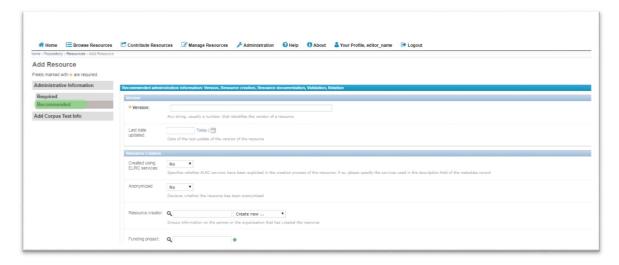
For further details, see the **ELRC-SHARE** metadata schema documentation.

Relations between resources **should be defined bidirectionally**, as shown in the following figure.

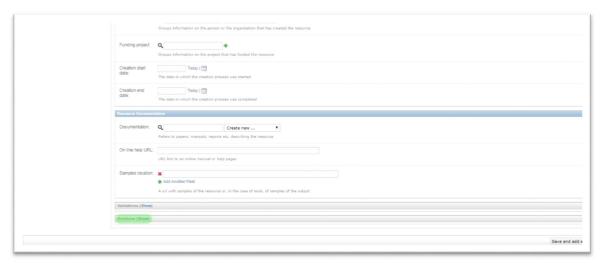


To add a relation for a resource, from the left-side menu select **Administrative Information** > **Recommended**.

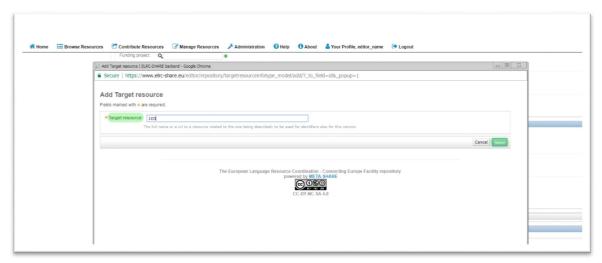




Scroll to the bottom of the page and click on the "Show" button of the "Relations".



Select the relation type from the list of available relation types and set the id of the related resource. In order to create a new related resource click on the button and input the resource id in the field "Target Resource". Click "Save".





When the popup window closes, click "Save" on the main page and repeat the process for the related resource where you should define the reverse relation.

5.2.1.6 *Indicating the Funding project*

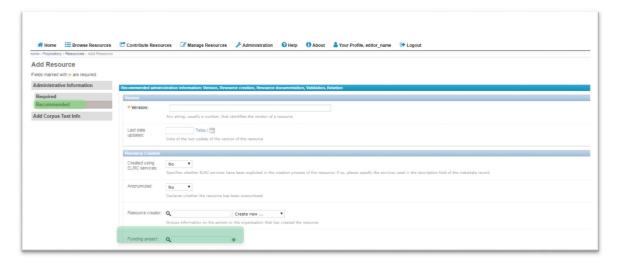


Important note:

If you represent a CEF-funded project, please note: Since different projects host their datasets on ELRC-SHARE, it is critical that you appropriately fill in the Funding project.

ELRC will by no means edit or otherwise manipulate an external project's datasets and/or metadata hosted in ELRC-SHARE.

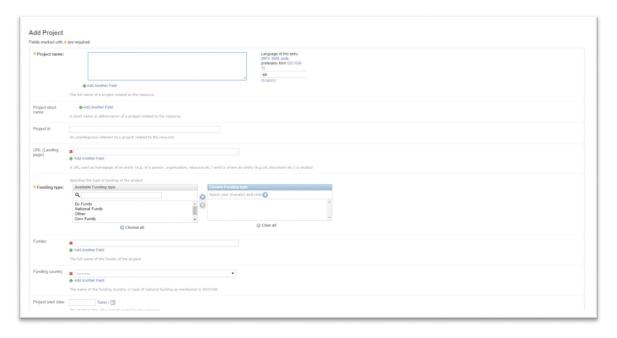
From the left side menu select **Administrative information > Recommended**.



Start typing in the **Funding project** field, and select from the list of available projects. If the project is not already available in the ELRC-SHARE database, click on the * button, to create a new project.

In the appearing **Add project** window, fill in at least the mandatory fields (**Project na**me and **Funding type**) and click on **Save**.





5.2.2 Resource type-specific information

Resource-type pages can be accessed by clicking on the menu items that appear on the left side menu, labelled according to the resource type:

- Edit Corpus Text Info
- Edit Language Description Text Info
- Edit Lexical Conceptual Resource Text Info
- Edit Language Description General Info
- Edit Lexical Conceptual Resource General Info
- Edit Tool / Service Info

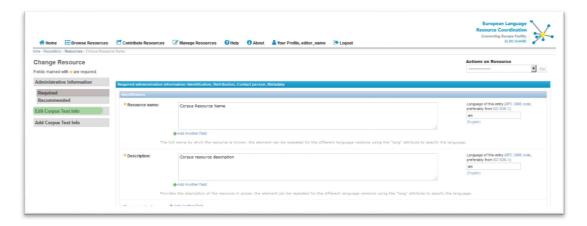
Mandatory information is included in all of the above pages, so you **must review all of them** in order to ensure that the metadata record is valid. The following paragraphs describe the process to be followed for editing **xTextInfo** as an example.

If the contributor has provided information about the language(s) of the resource (optional in the contribution form), a **CorpusTextInfo** / **LanguageDescriptionTextInfo** / **LexicalConceptualResourceTextInfo** record is created, where this information is initially stored. To edit and fill in missing information click **Edit xText Info** at the left side menu. Mandatory elements are again marked with *.

If the contributor has not provided any language information, you must create a new **xText Info** by clicking the relevant button, i.e. add **xTextInfo**.

Assuming that there is language information available for a resource of type **corpus**: Click on **Edit Corpus Text** Info.





The new window displays all the mandatory and optional fields. The language information is already filled in and the editor should, at the very least, provide the rest of the required fields, namely *Linguality type*, *Size* and *Size unit*, *Mime Type*.

Although optional, it is **highly recommended** to fill in information on the **Domain** of the resource and the **resource size**, since these are of particular interest to the EC.

Important note: Following the EC specifications, indicate the resource size consistently with the size units per resource type as follows:

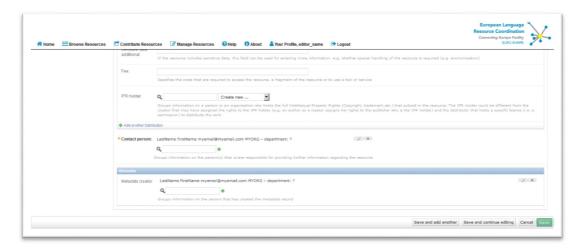
- For parallel aligned corpora: in Translation Units
- For monolingual corpora: in number of Sentences
- For other Lexical/conceptual resources: in number of Entries

If necessary, you can indicate more than one sizes and size units.

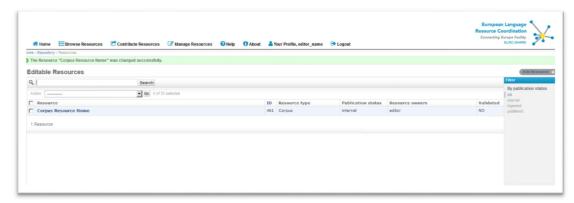
As soon as you have filled in all the required information, click **Save** at the bottom of the page. Do the same for all resource type-specific editing pages.

When you have finished editing a resource description, you should finally save the record by clicking **Save** at the bottom of the main (**Administrative Information**) page.





You are redirected to the managing form where a notification about successful modification of the metadata record is displayed.



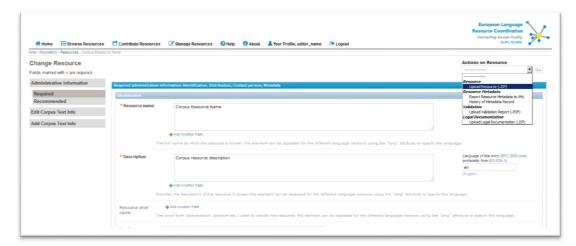
Always click on **Save** at the bottom of the main (**Administrative Information**) page, when a metadata record is finalised.

5.3 Uploading a dataset

For resources previously created, e.g. through the online Contribution form or through the ELRC-SHARE-client API, the dataset in .zip format is already attached to the metadata record. For resources that are created from scratch by the Editor through the documentation environment, follow the steps below to upload the dataset attached to a metadata record.



On the **Administrative information page**, select the **Upload Resource (.ZIP)** action from the **Actions on Resource** dropdown list and click **Go**³. In case data for this resource already exists, the action is labelled **Replace Resource (.ZIP)**.



You will be redirected to the upload resource form.

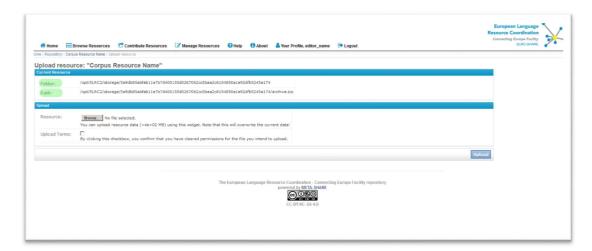
In this page the following information is displayed:

1. Current Resource Folder

The path of the current resource

2. Current Resource Path

The path of the current resource's data. If there is not any data already uploaded, this information will not be displayed.



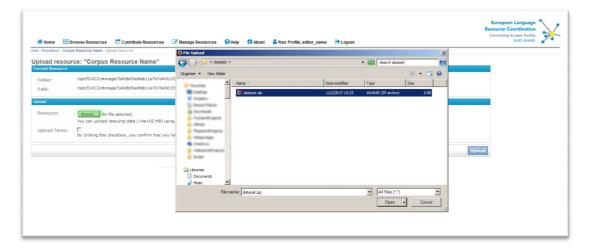
To upload a resource

d. Click Browse

In the appearing window, browse your local folders for the intended zipped file (.zip) containing the data you want to contribute, select it and click **Open**.

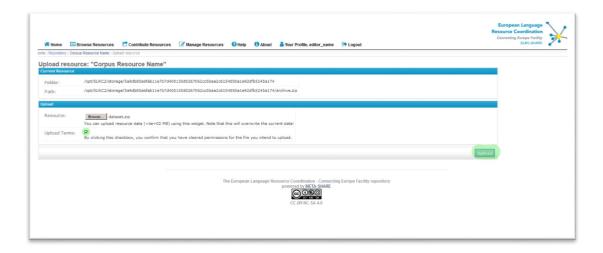
³ This button is disabled if no action is selected.





- e. Check the Upload Terms checkbox,
- f. Click **Upload**

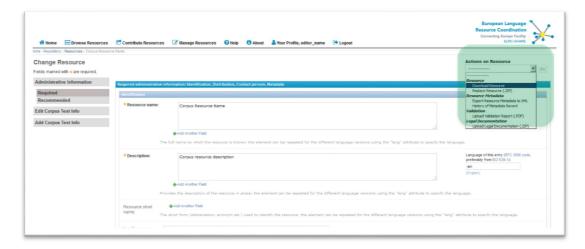
Any upload will replace any existing data that have been previously uploaded.



5.4 Downloading a resource

To directly download a resource, select the **Download Resource** action from the **Actions on Resource** drop-down menu and click **Go**.



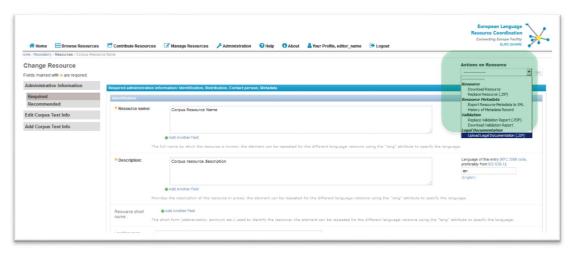




6 Adding legal documentation to language resources

Apart from the licensing specified in the distribution description, the repository provides the editor with the option to upload additional legal documentation.

To add legal documentation select the **Upload Legal Documentation (.ZIP)** action and click **Go**. In case legal documentation already exists, the action is labelled **Replace Legal Documentation (.ZIP)**.



You will be redirected to the "upload legal documentation" form.

In this page the following information are displayed:

1. Current Resource Folder

The path of the current resource

2. Current Legal Documentation Path

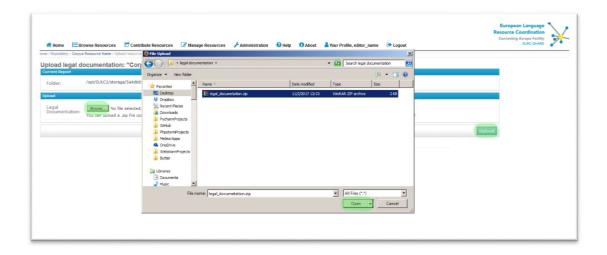
The path of the current resource's addition legal documentation. If there is not any legal documentation already uploaded, this information will not be displayed.



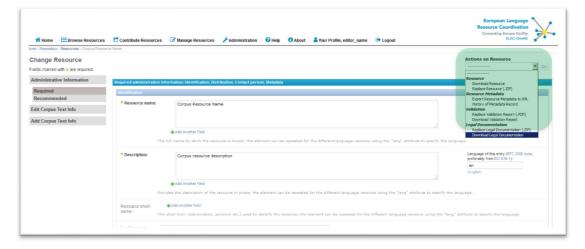
To upload legal documentation



- a. Click **Browse**
 - In the appearing window, browse your local folders for the intended compressed directory (.zip) containing the documents you want to upload, select it and click **Open**.
- b. Click Upload



After a successful upload, the legal documentation is available for download by selecting **Download Legal Documentation** action from the dropdown list and clicking **Go**.



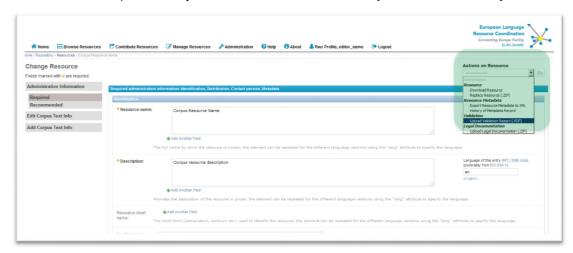
You can also repeat the above process if you want to replace existing legal documentation with a new one.

7 Adding validation report to language resources

The repository provides the editor with the option to upload a validation report **in pdf format** for a given resource.



To add a validation report select the **Upload Validation Report (.PDF)** action and click **Go**. In case a validation report already exists, the action is labelled **Replace Validation Report (.PDF)**.



You will be redirected to the "upload validation" form.

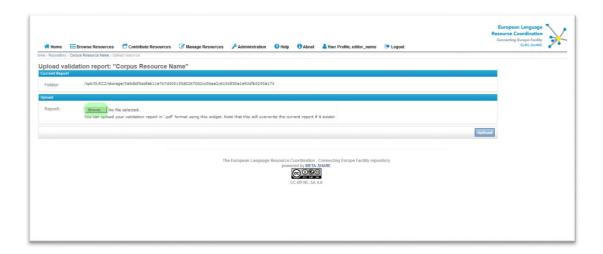
In this page the following information are displayed:

1. Current Resource Folder

The path of the current resource

2. Current Validation Path

The path of the current resource's validation report. If there is not any report already uploaded, this information will not be displayed.



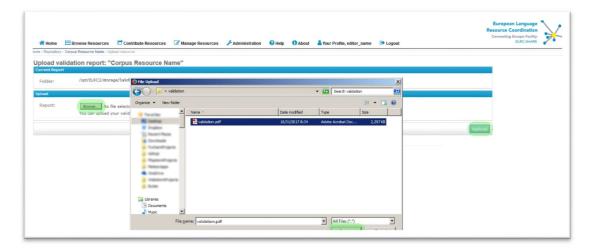
To upload a report

a. Click Browse

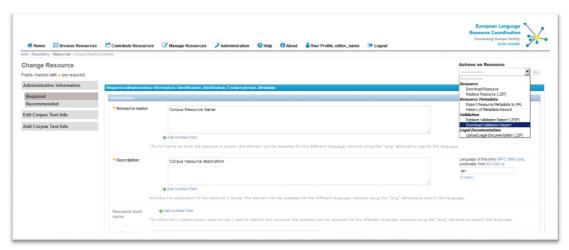
In the appearing window, browse your local folders for the intended pdf file (.pdf) containing the validation report you want to upload, select it and click **Open**.

b. Click **Upload**





After a successful upload, the report is available for download by selecting **Download Validation Report** action from the dropdown list and clicking **Go**.



You can also repeat the above process if you want to replace an existing report with a new one.



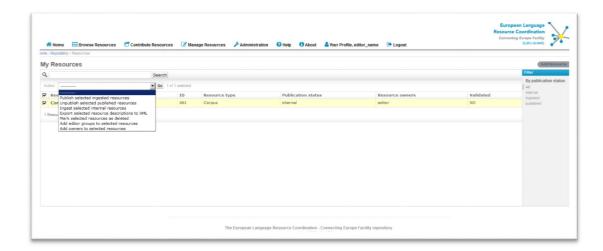
8 Publishing resource descriptions (metadata records)

When you have finished editing a metadata record and consider it ready for publication, go to the managing form and select it from the list; you can now **ingest** and, right afterwards, **publish** it (see section 2.2).

On the **Managing form**, select from the list the metadata record(s) that you want to publish, by clicking on the box left to the resource name; from the drop-down menu select the appropriate action:

- Ingest selected internal resources
- Publish selected ingested resources.

Tip: Keep metadata records that you have worked on as ingested while you seek further information, e.g. on licensing issues.





Important note:

If you represent a CEF-funded project, please note: As long as your resources are in internal publication status the EC members cannot view them. It is your responsibility to ingest and/or publish your metadata records, so that they are accessible by the EC.

ELRC will by no means edit or otherwise manipulate an external project's datasets and/or metadata hosted in ELRC-SHARE.

You can also use the same functionality for **exporting metadata records in XML format** or **marking metadata records as "unpublished"** in order to withdraw them from the public inventory.



9 Working on resources from crawling

Resources that have been produced by the ILSP-FC web crawler can be uploaded to the repository via the *ILSP-FC_Upload* tool. The tool converts the metadata file (.md.xml) that is automatically generated by the ILSP-FC web crawler into an ELRC-SHARE valid xml description and, at the same time, packages the actual data in a .zip file ready to be uploaded. If you wish to access the tool, please contact elrc-share@ilsp.gr.