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

European Language Resource Coordination
Connecting Europe Facility

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.: V2.1

Date: 2017-11-20

Contents

1 Introduction

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

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

(a) resulting from a process of transformation from resources uploaded by contributors,
(b) resulting from a process of transformation from crawled data,
(c) input into the repository by editors entirely from scratch.

All metadata records, regardless of their source, must comply with the ELRC-SHARE metadata schema, in order to be valid and become publicly available through the inventory.

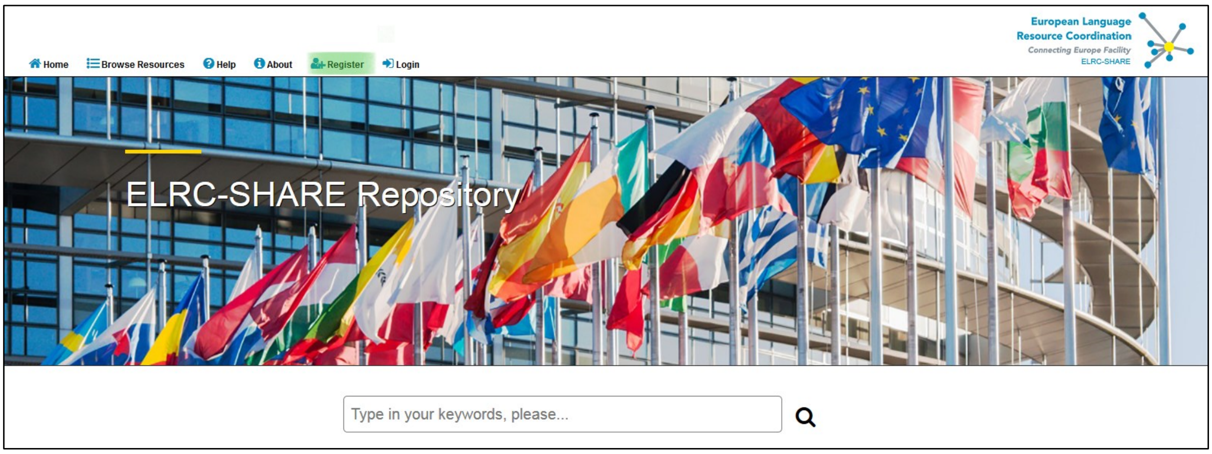
The process of editing the metadata record is the same for all cases. However, the information that appears on the initial metadata record differs depending on the source.

2 Becoming an editor

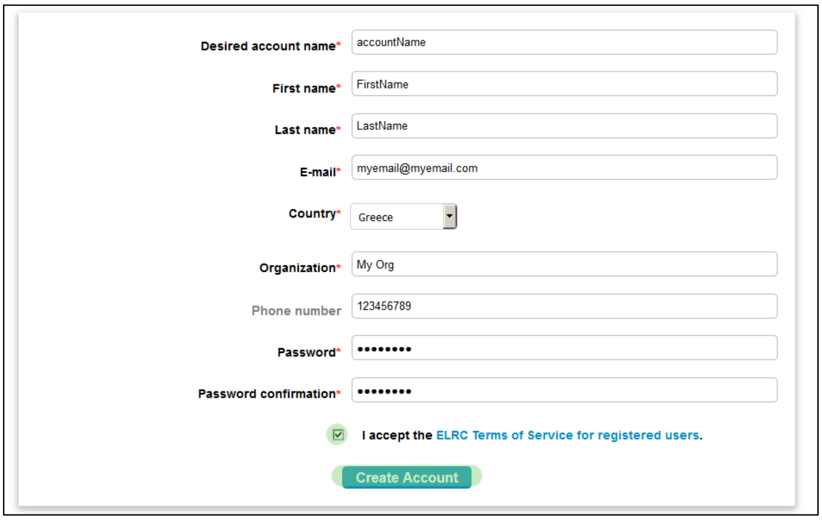
Register → Activate account → Request editor status → Login

2.1 Registration procedure

To register, go to the ELRC-SHARE repository (www.elrc-share.eu) and click on **Register**.



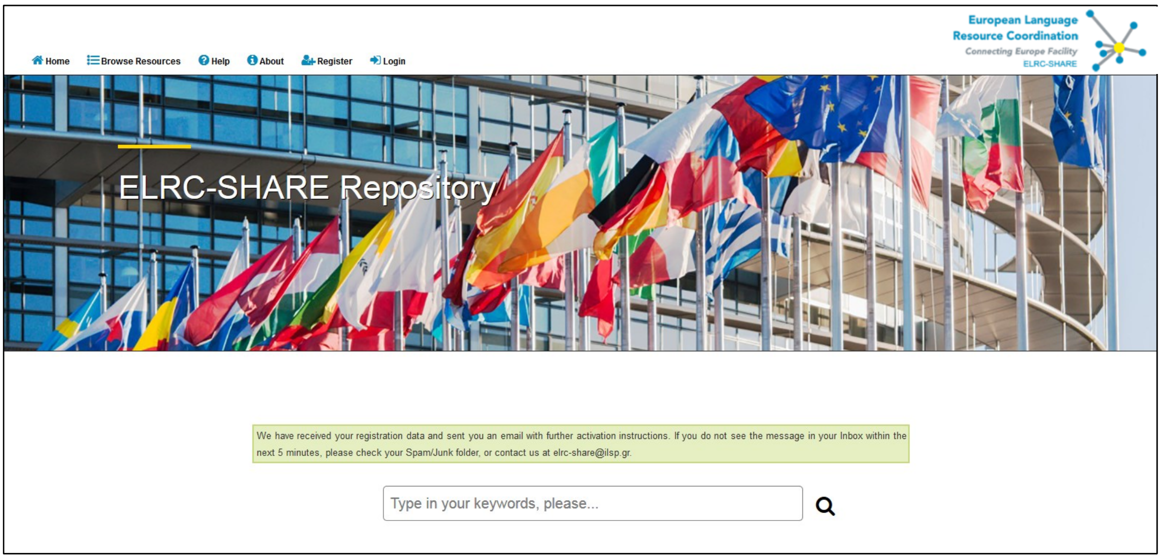
On the registration page, fill in all the required information.


Desired account name* accountName
First name* FirstName
Last name* LastName
E-mail* myemail@myemail.com
Country* Greece
Organization* My Org
Phone number 123456789
Password* ••••••••
Password confirmation* ••••••••
I accept the ELRC Terms of Service for registered users.
Create Account

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.

2.2 Obtaining editor status

Only ELRC consortium members and National Anchor Points (NAPs) can request for editor status. ELRC representatives who are responsible for data collection in a country or, in exceptional cases, the NAP who represents it are responsible for managing and editing metadata records of resources originating from this country.

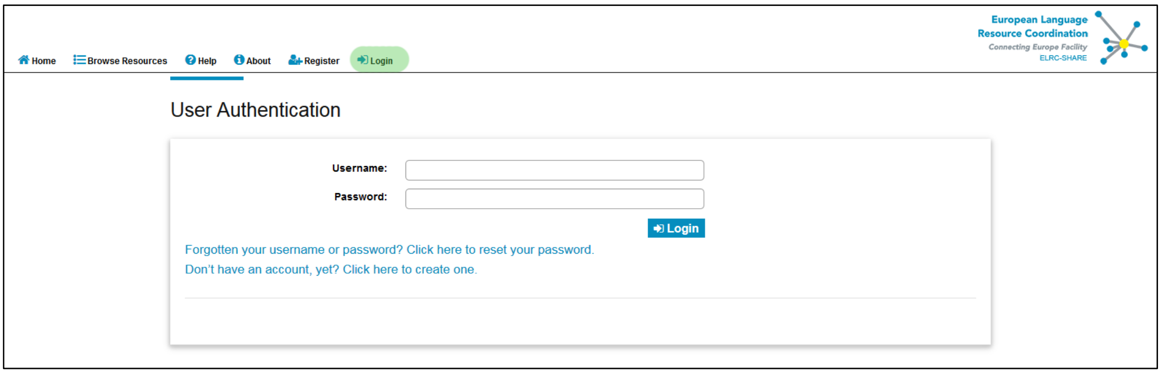
Users who wish to obtain editor status should, **once registered**, contact the ELRC-SHARE repository administrators at elrc-share@ilsp.gr, stating their affiliation and role in the context of ELRC. Once this has been approved by the ELRC consortium, they will be notified by email.

Note the use of "editor status" herein actually describes three sub-groups of users with editor rights in ELRC-SHARE: Global editors (ELRC consortium members), 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 LRs	All LRs	Own LRs	All LRs	Own LRs	All LRs	Own LRs	All LRs	Own LRs	All LRs
Global Editors	■		■		■				■	
Legal Reviewers	■	■	■	■	■	■			■	■
NAPs	■		■						■	

2.3 Logging in

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



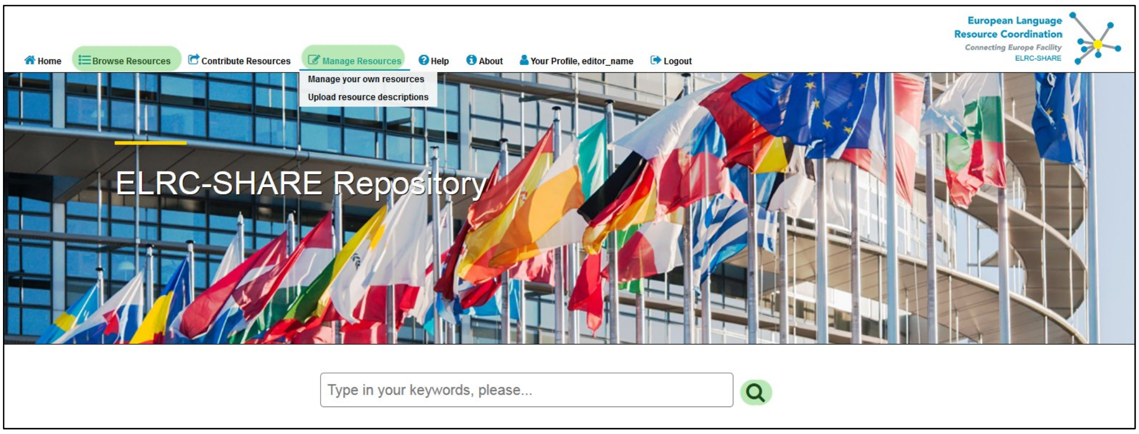
3 Editor - the basics

3.1 Editor menu

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

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

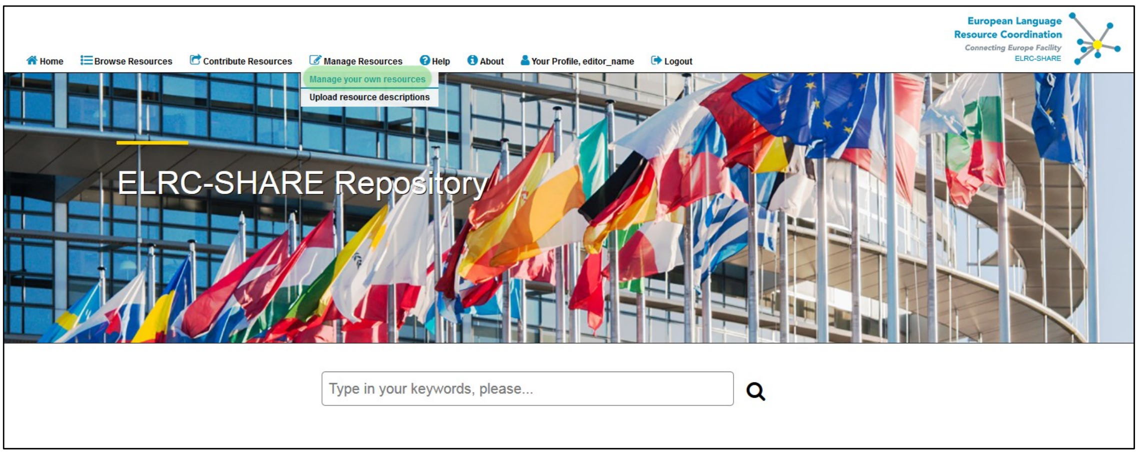
Editors can also use the **search** functionality to look for specific published metadata records.



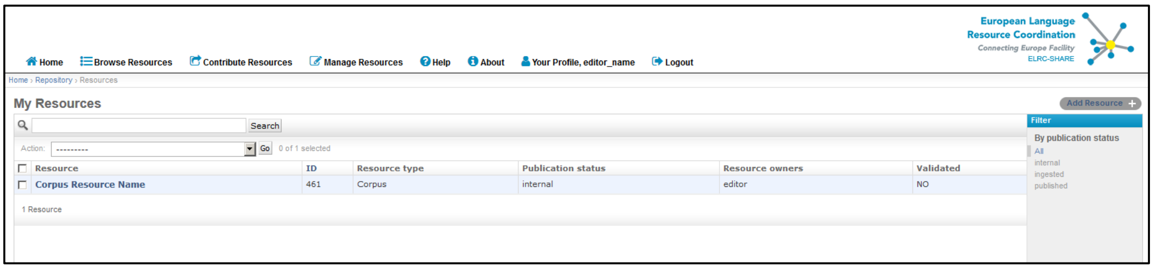
3.2 Managing form

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).

To access the managing form, click **Manage your own resources** from the drop-down menu item **Manage resources**.



In the list of available resources, editors can see the **name**, **ID, resource type**, **publication status**, the **owners** and the current validation status of each resource, namely the users that granted editing rigths.



The resource type can be one of the following types:

- **corpus**: for sets of documents in one or more languages, including, for instance,
  - sets of public administration official documents (e.g. ministerial decisions, legal acts, Board decisions etc.),
  - sets of journal, newspaper, bulletin, blog articles, etc.
  - parallel corpora, i.e. sets of original document with its translations (without any processing)
  - translation memories, i.e. translated documents aligned with their originals.
- **language description**: for resources such as:
  - grammars, i.e. sets of rules that formalize a language
  - language and translation models, which contain statistical information that assigns a probability to a piece of unseen text (based on some training data).
- **lexical/conceptual resource**: for resources such as:
  - terminological lexica, glossaries etc., i.e. including lists of terms, with or without any other information (e.g. definition, examples, translation equivalents, linguistic information etc.)
  - lists of words, such as person names, names of places, names of products etc.
  - lexica and dictionaries containing words with linguistic information (e.g. part of speech, inflectional information, etc.).
- **tool/Service**
  - language processing tools, web services, applications such as linguistic annotators (e.g. Part-of-speech taggers, parsers, etc.), semantic annotators (e.g. Named Entity Recognisers), aligners, etc.

The publication status can be:

- **internal**: used as the initial status for all metadata records created from scratch by editors or uploaded from resource descriptions of contributors; in the latter case, the records contain basic information and they must be enriched by editors with information about the resource obtained, for instance, during interviews with the contributors and/or accessing and checking the resource itself;
- **ingested**: after enriching a metadata record, editors can ingest the metadata record to indicate that the first validation step has been completed; some information may still be

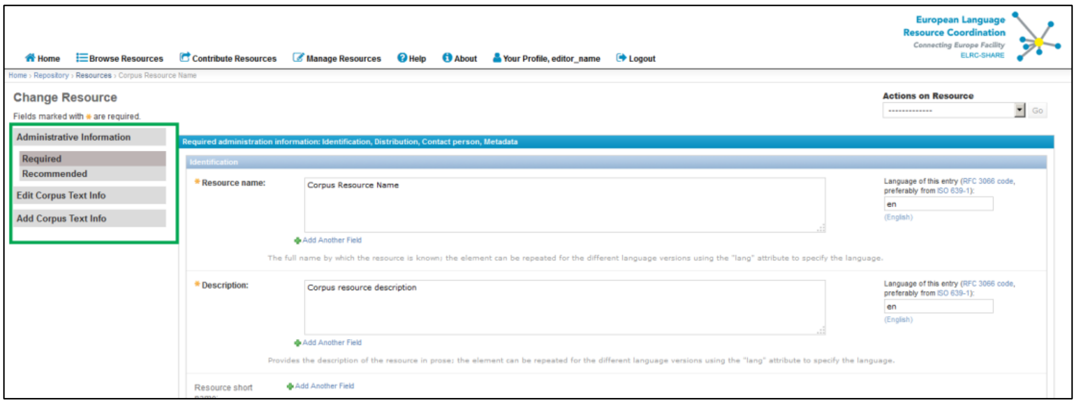
left incomplete or undecided (e.g. the value of the *licence* element may be set to *under review* for resources with unclear legal status, in which case editors should notify the legal helpdesk for further instructions); it is also the initial status for metadata records transformed from the metadata of crawled resources;

- **published**: as soon as the description of the resource has been finalized, editors can publish the metadata record. Published records are displayed on the public inventory of resources; for the time being, however, downloading is not allowed even for resources with permissive licenses.

3.3 Editing form & Schema basics

The **editing form** is where you can describe and upload a resource; you can access it through the managing form, either by clicking the button **Add Resource** or by selecting an existing resource description - see Section 4.

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**.



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

- Administrative information
  - *Identification* (e.g. resource name, description)
  - *Distribution* (e.g. licensing, access form of the resource)
  - *Contact Person*
  - *Metadata* (e.g. creator and creation date of the metadata record)
- Technical information for all resource types
  - *Linguality* (i.e. number of languages included in the resource)
  - *Languages*
  - *Sizes*
  - *Text Formats* (e.g. plain text, PDF, XML, TMX etc.)
- Technical information for specific resource types
  - for lexical/conceptual resources
    - *Lexical conceptual resource type* (e.g. terminological glossary, bilingual dictionary etc.)

- for language descriptions
    - *Language description type* (e.g. grammar).

In addition, the following metadata categories are **optional**, however **recommended**:

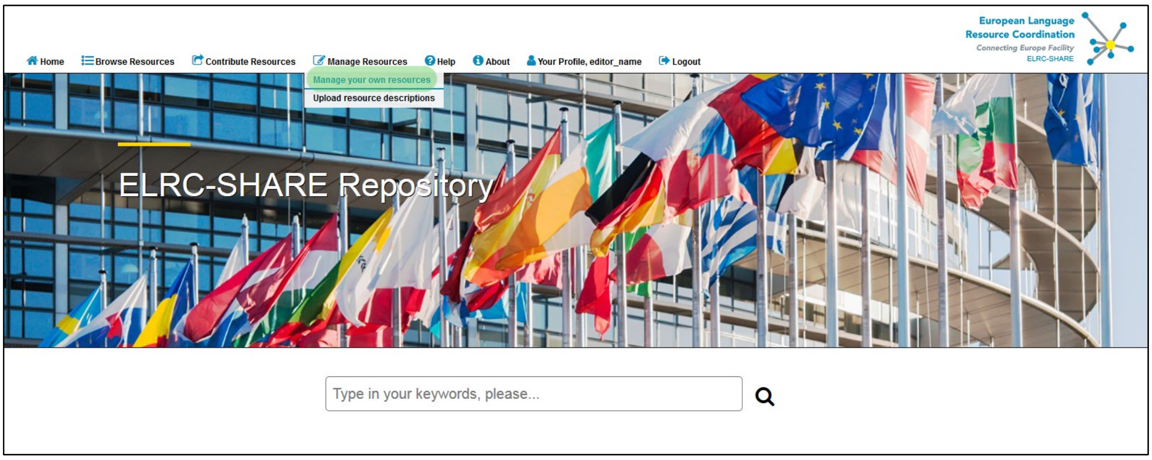
- Administrative information
    - *Version*
    - *Resource Documentation*
    - *Resource Creation*
    - *Validations*
    - *Relations*
- Technical information
    - *Character Encodings*
    - *Annotations* (only for corpora)
    - *Domains*
    - *Text Classifications*
    - *Creation*
    - *Encoding 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.*)

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.

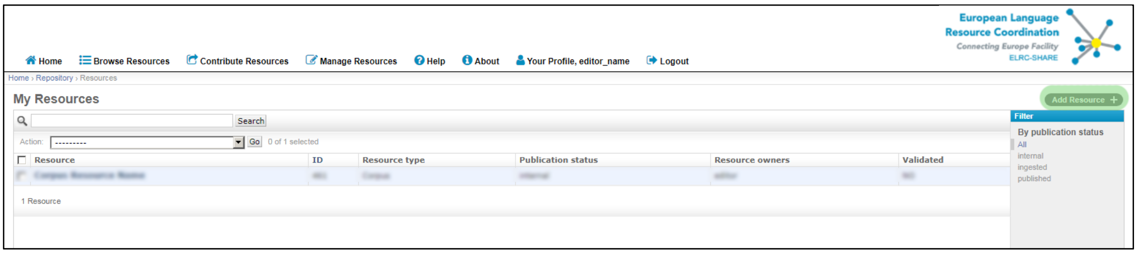
4 Adding and editing resources (metadata records)

4.1 Adding new resources

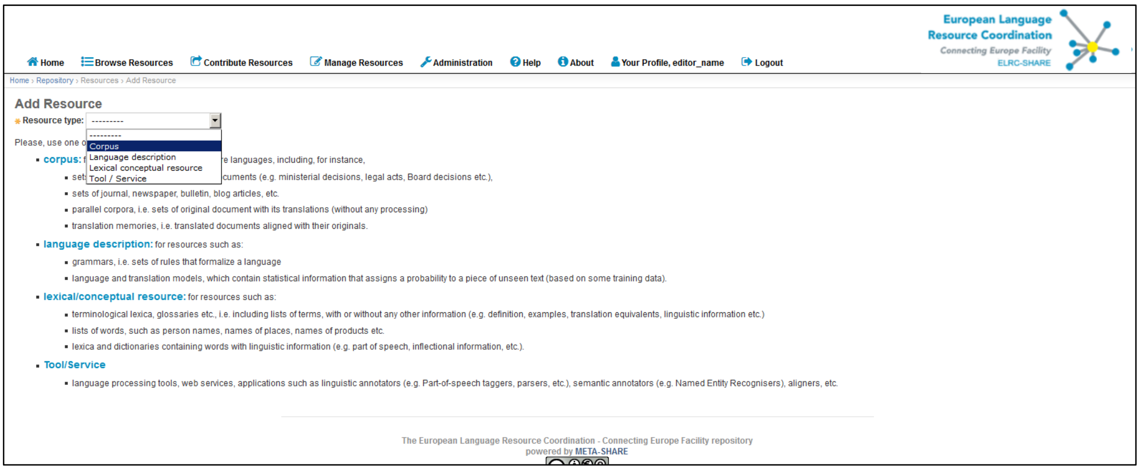
To add a new resource, click **Manage your own resources** from the drop-down menu item **Manage resources**.



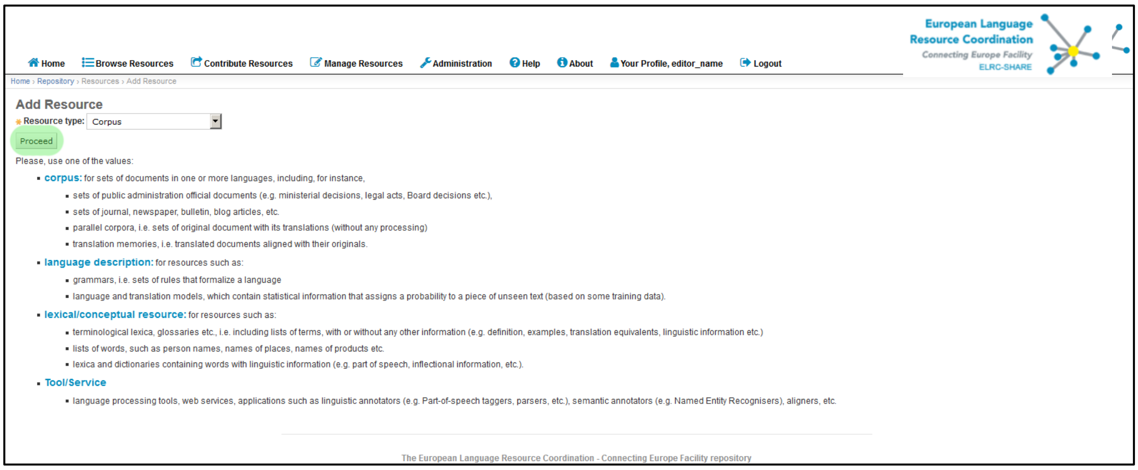
On the **Managing form**, click **Add Resource**.



Select the **resource type** that represents the resource at hand.



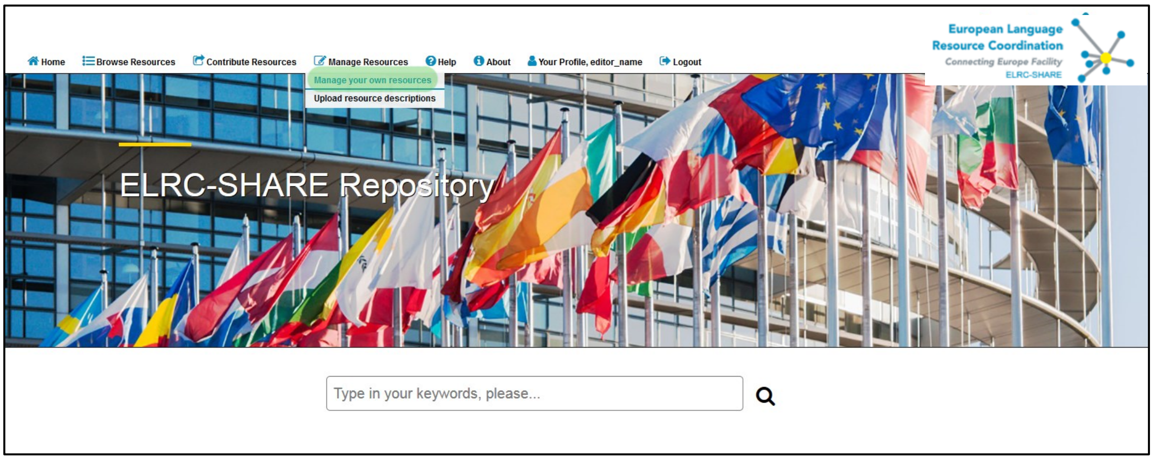
Click **Proceed**.



You will be directed to the **editing form**.

4.2 Selecting a resource for editing

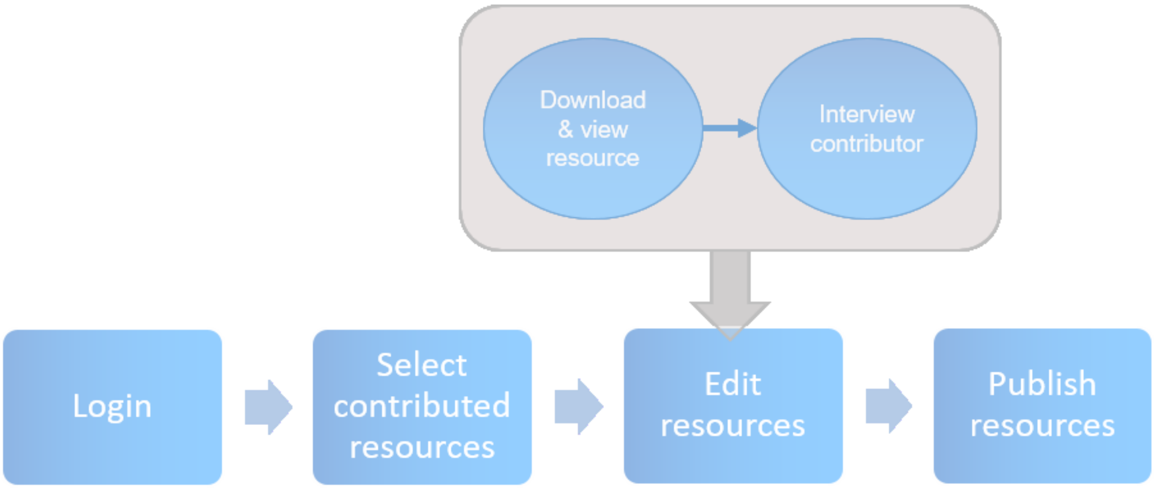
To edit an existing resource, click **Manage your own resources** from the drop-down menu item **Manage resources**.



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**.

5 Working on resources (metadata records) from contributors



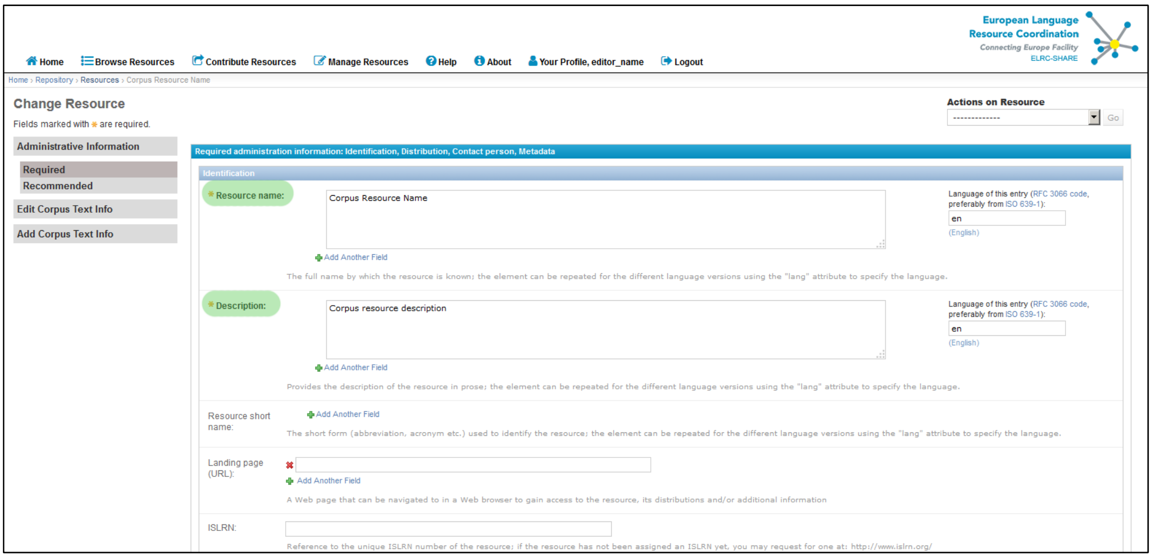
5.1 Pre-processing procedure

Contributors upload and describe resources with basic information through a simple web form. The following process is triggered afterwards:

- 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 Editing of metadata records from contributors

In order to edit a metadata record, please access the managing form as described in section 4 and select the appropriate record. The editing area is displayed, where specific fields are pre-populated with information obtained from the contribution form, namely *Resource name*, *Description* and *Contact person*.



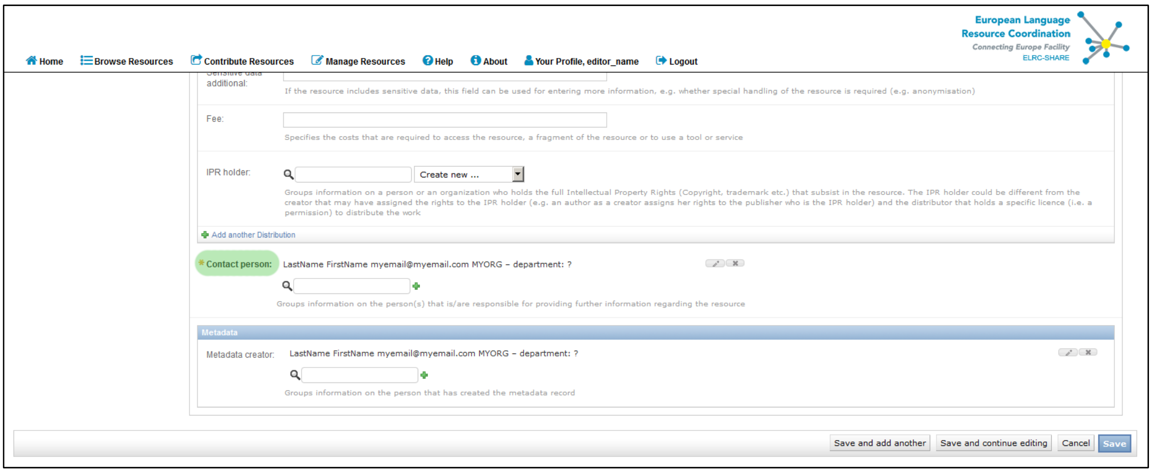
Editors are required to edit the metadata records by **filling in at least the mandatory fields** (e.g. licence, text format, size etc.) and, if possible, by providing recommended information that may facilitate further processing of the resource. For example, though optional, it is highly recommended to fill in the fields on **Domain** classification, **Public Sector Information** and **DSI relevance**, as these descriptors are of high importance to the EC. Further information can be obtained by interviewing the contributor (see Section 5.2.1 on how to find his/her contact details) and/or by downloading the resource and checking its contents (see Section 5.3).

Indicatively, a list of possible questions to the contributor during the interview could be:

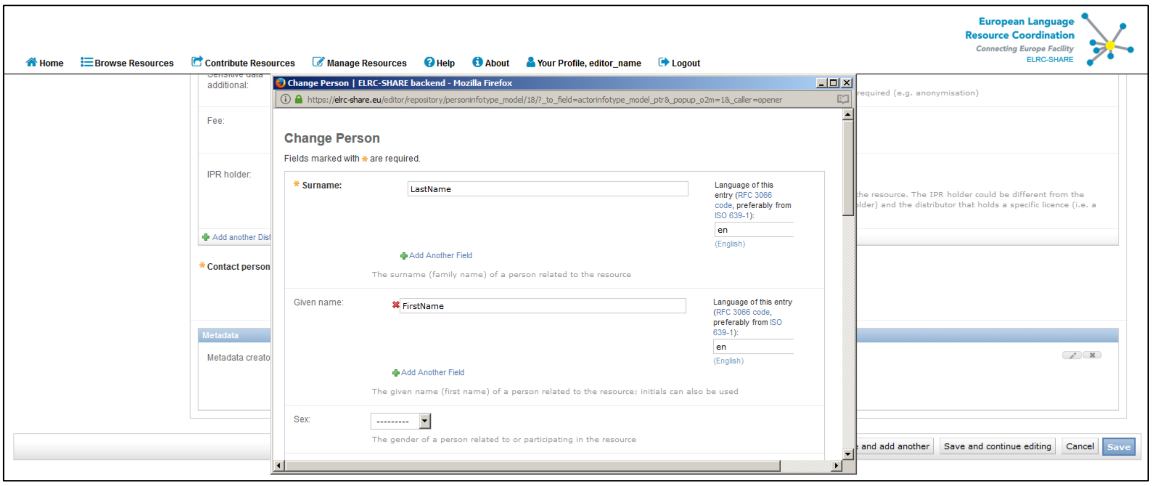
1. Was your donated LR translated by a professional translator / a language expert / a group of different translators? (if applicable, e.g. in the case of bilingual corpus or dictionary). If yes, was the translation done in-house (e.g. by the ministry) or out-sourced?
2. Do you have any concerns that the texts include personal information and need anonymization?
3. Do you have any concerns that the texts may include classified information?
4. Is the <organization of the contributor> the IPR holder of the resource (original and translation)?
5. If data fall under the PSI directive. If yes, what type of license is assigned to public data in your country?
6. If data in pdf/doc format: inform the contributor that we may convert them in tmx format and package it with his/her contribution in order to make it readily available for use by MT@EC.

5.2.1 Editing Contact person

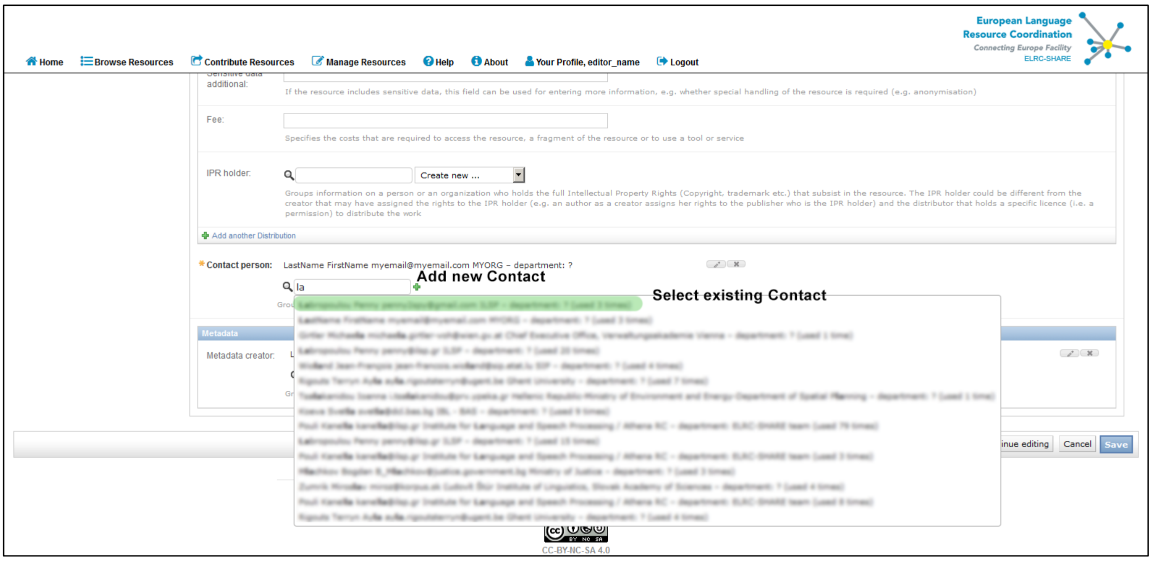
The *Contact person* contains information about the contributor, which the editors can use to contact them for more information regarding the resource.

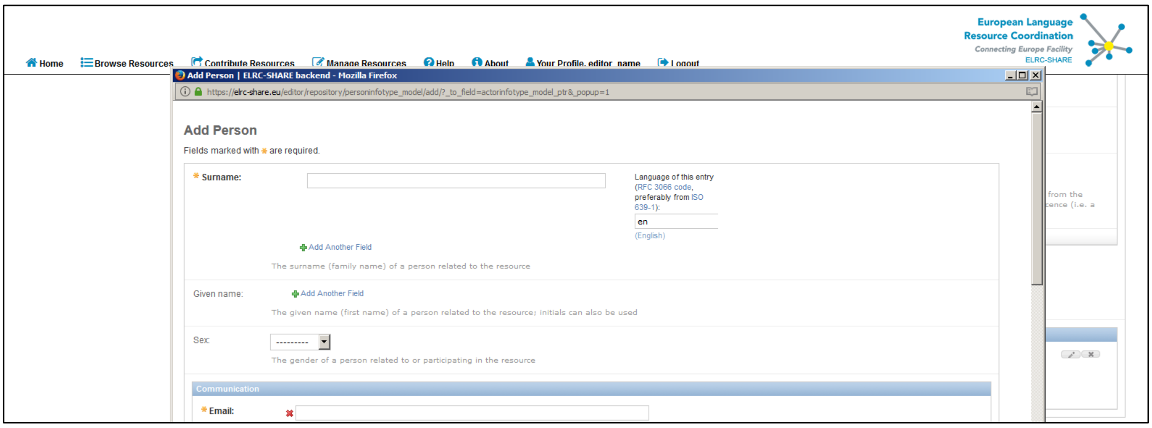


To view/edit details of a contact person, click on the icon.



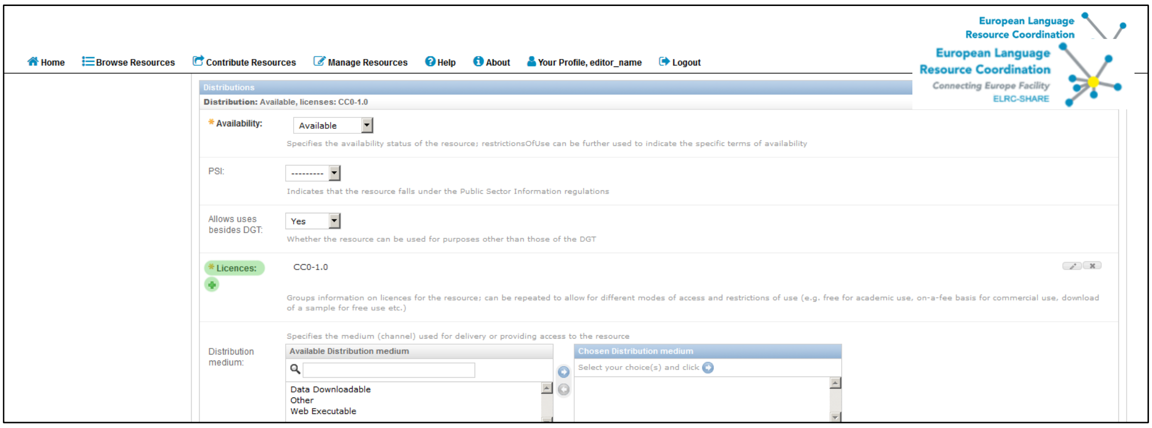
You can also add another contact person either by selecting an existing one or creating a new one by clicking the icon.



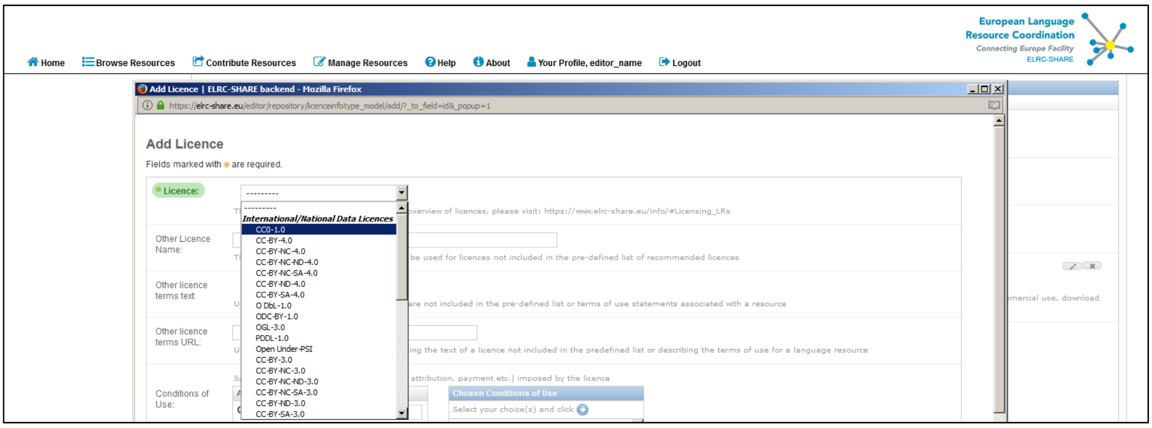


5.2.2 Adding Distribution

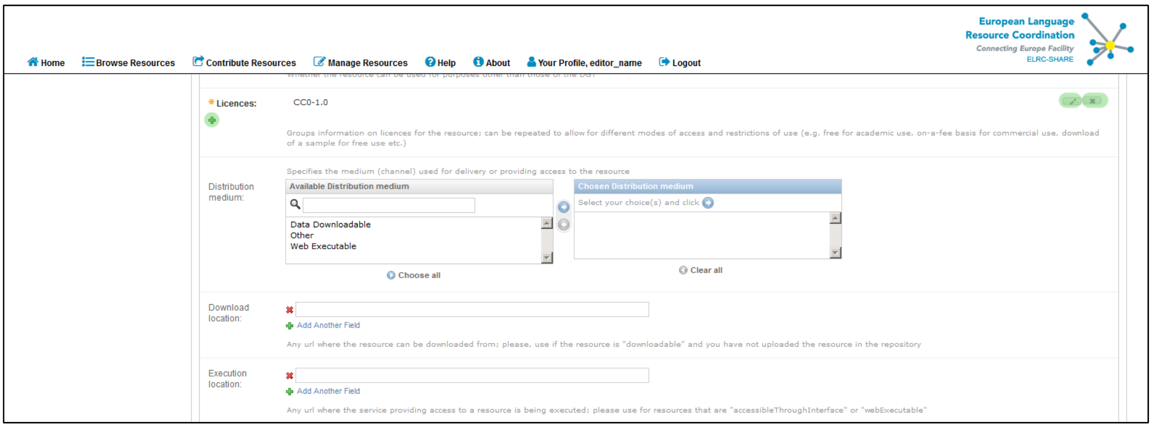
Click on the ✚ icon of the *Licences* field inside the *Distributions* section field.



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



If you want to add another licence, click on the icon of the *Licences* field inside the *Distributions* section field. 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 ELRC editor**:

- **PSI**: Indicate whether the resource falls in the regime of the Public Sector Information Directive, according to the respective guidelines provided by ELDA
- **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, use the following text template for the attribution text:
  <Resource Name> *was created for the European Language Resources Coordination Action (ELRC) by* ***<Person X, Institute of X>*** *with primary data copyrighted by* ***<Person/Body Y>*** *and is licensed under* ***<licence value, licence link>***
- **IPR holder**

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

5.2.3 Adding/Editing resource-type specific pages

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

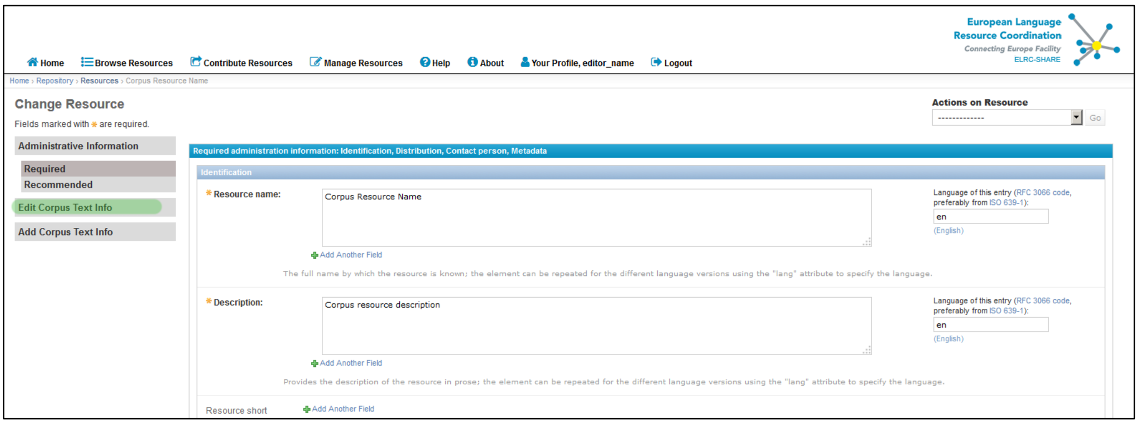
Mandatory information is included in all of the above pages, so you **must review all of them** in order to ensure that the metadata record is valid. The following paragraphs describe the process to be followed for editing **xTextInfo** as an illustrative case.

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 DG Translation specifications, indicate the resource size consistently with the size units per resource type as follows:

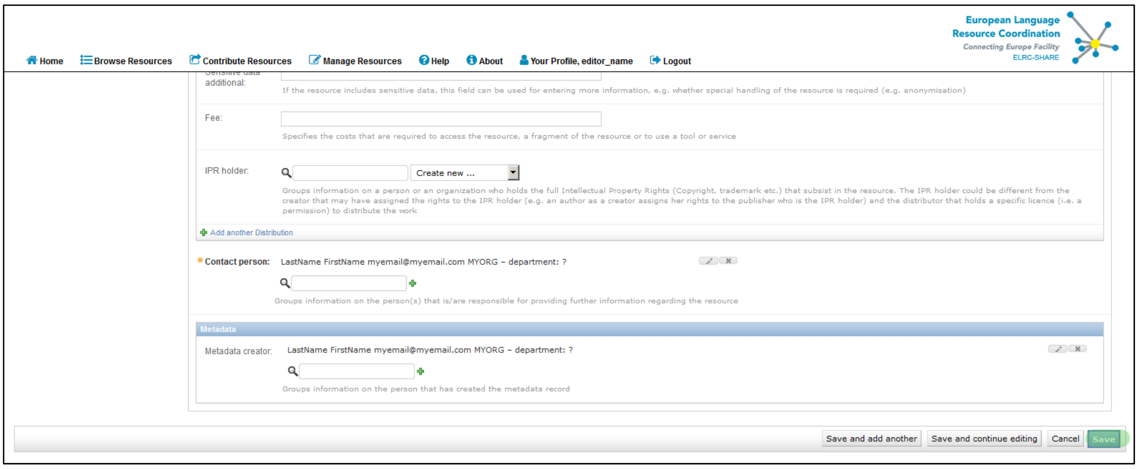
- For parallel aligned corpora: in Translation Units
- For monolingual corpora: in number of Tokens
- For Terminologies: in number of Terms
- For other Lexical/conceptual resources: in number of Entries

In all cases, avoid indicating size in Mb/Kb, number of files and number of sentences.

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 is displayed.

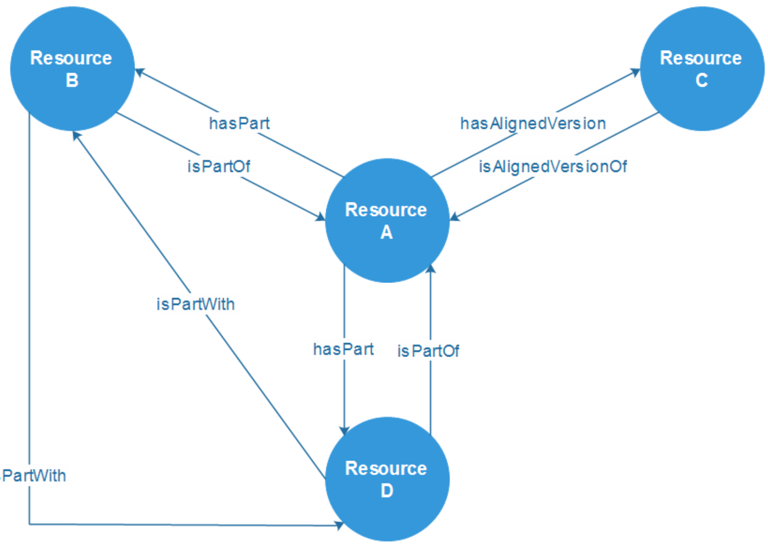
5.2.4 Indicating relations between resources

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*':

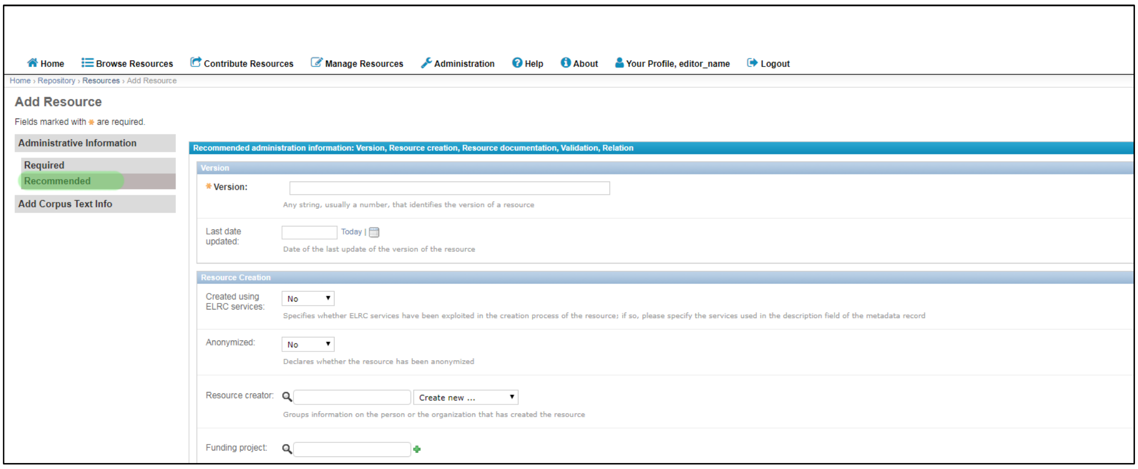
HAS	IS
hasPart	isPartOf
	isPartWith
hasVersion	isVersionOf
hasAnnotatedVersion	isAnnotatedVersionOf
hasAlignedVersion	isAlignedVersionOf
	isRelatedTo
hasConvertedVersion	isConvertedVersionOf

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

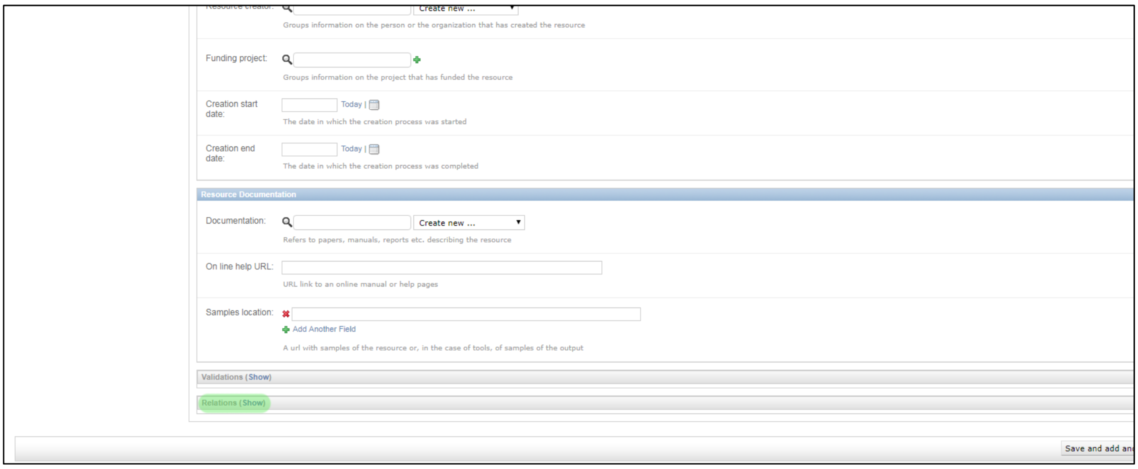
Relations between resources **should be defined bidirectionally**, as shown in the following figure.



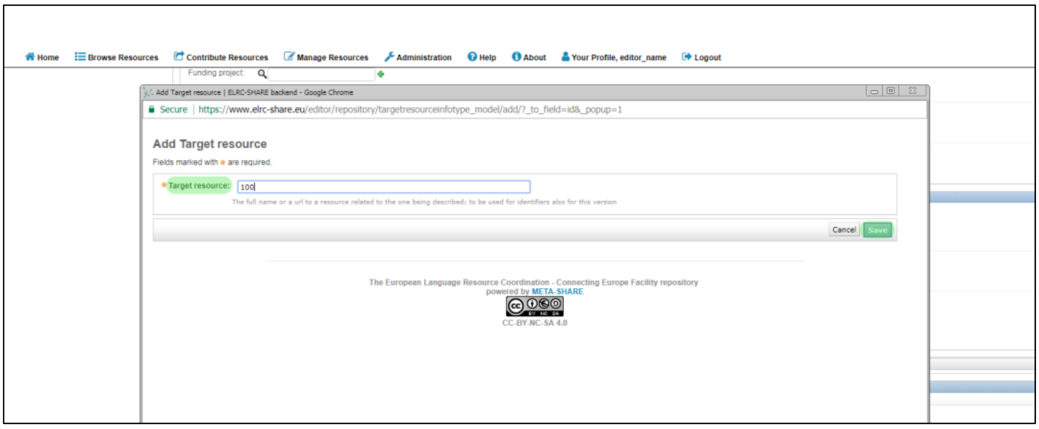
To add a relation for a resource, click on the "Recommended" item under "Administrative Information"



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



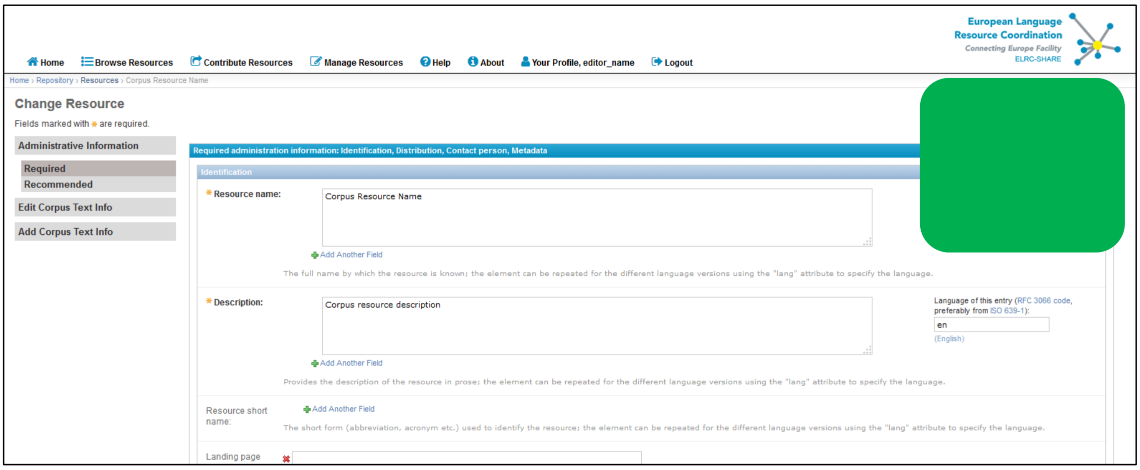
Select the relation type from the list of available relation types and set the id of the related resource. In order to create a new related resource click on the + button and input the resource id in the field “Target Resource”. Click “Save”.



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.

5.3 Downloading a resource

To directly download a resource, select the **Download Resource** action from the Actions on Resource drop-down menu and click **Go**[1].



5.4 Publishing metadata records

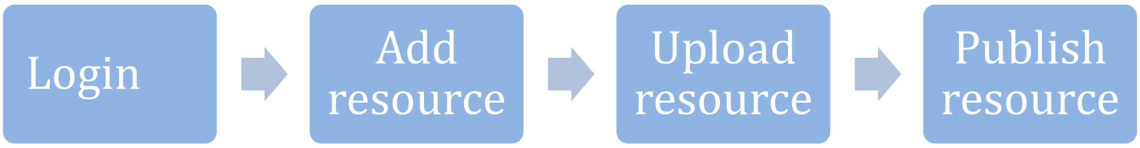
When you have finished editing a metadata record, you can proceed to its publication through the managing form - see section 8.

[1] This button is disabled if no action is selected.

6 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.

7 Working on resources from scratch

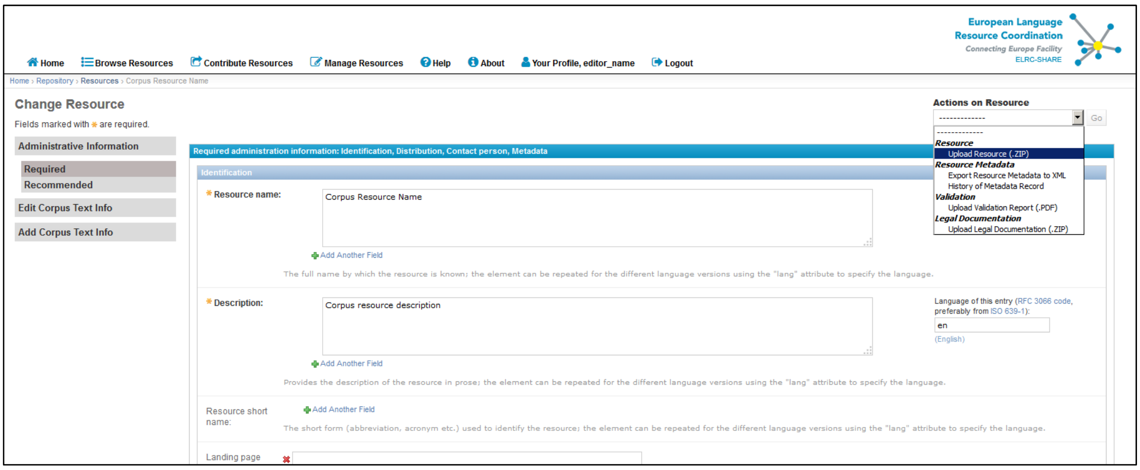


You can access the **editing form** for new resources as described in section 4.

On the Administrative information page, fill in **at least all the required information**.

From the left menu, click on **Add x Text Info** and **Add x General Info** in order to go to the editing page(s) for the resource type technical information; fill in at least the mandatory information on all these pages. Don't forget to click **Save** at the bottom of each page.

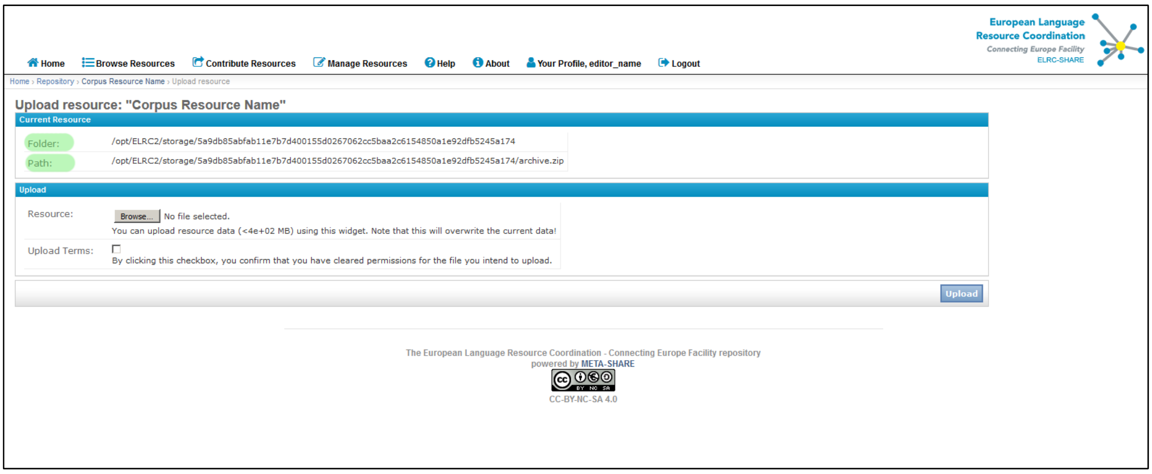
On the **Administrative information page**, select the **Upload Resource (.ZIP)** action from dropdown list and click **Go**. In case data for this resource already exists, the action is labelled **Replace Resource (.ZIP)**.



You will be redirected to the upload resource form.

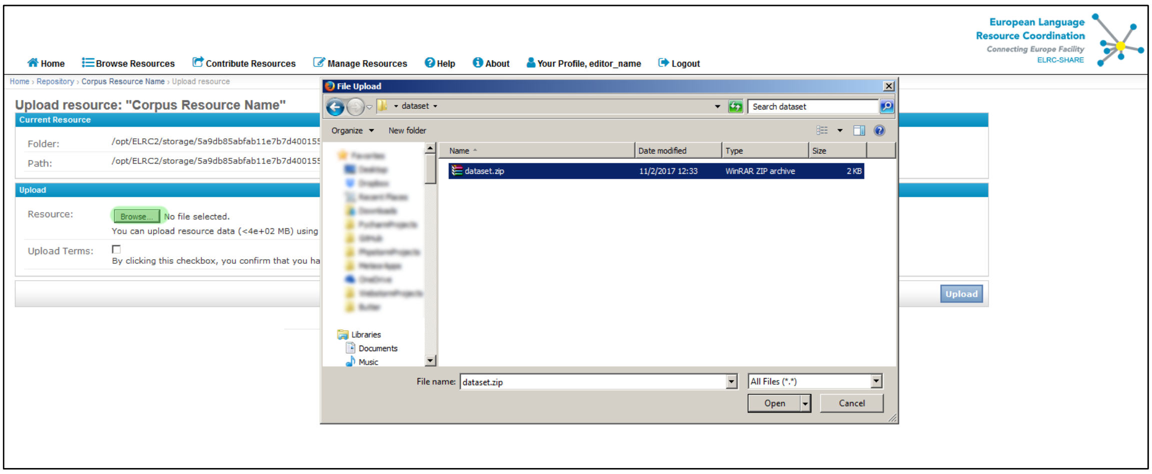
In this page the following information are displayed:

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.



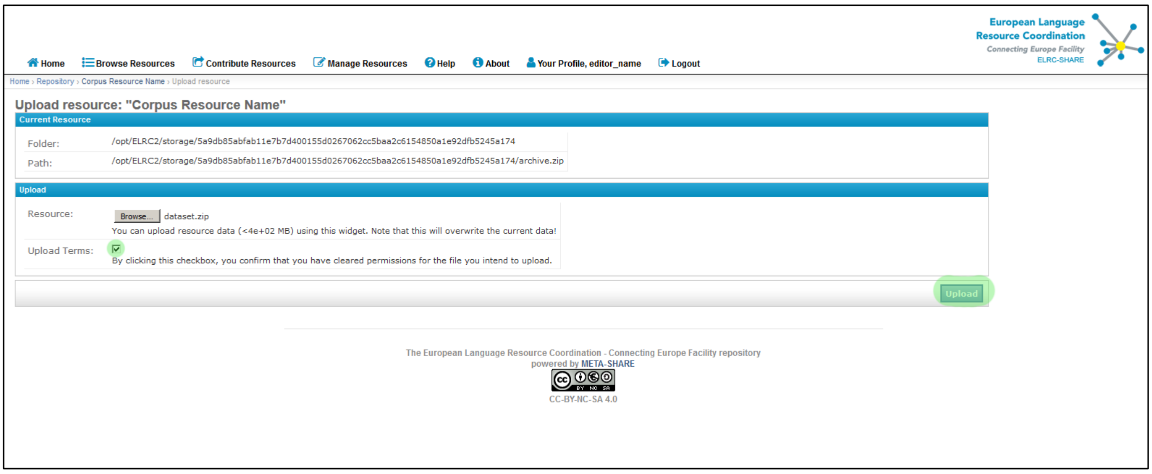
To upload a resource

a. 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**.



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

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



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.

Proceed to the **Managing form** in order to **ingest** and **publish** the record**.**

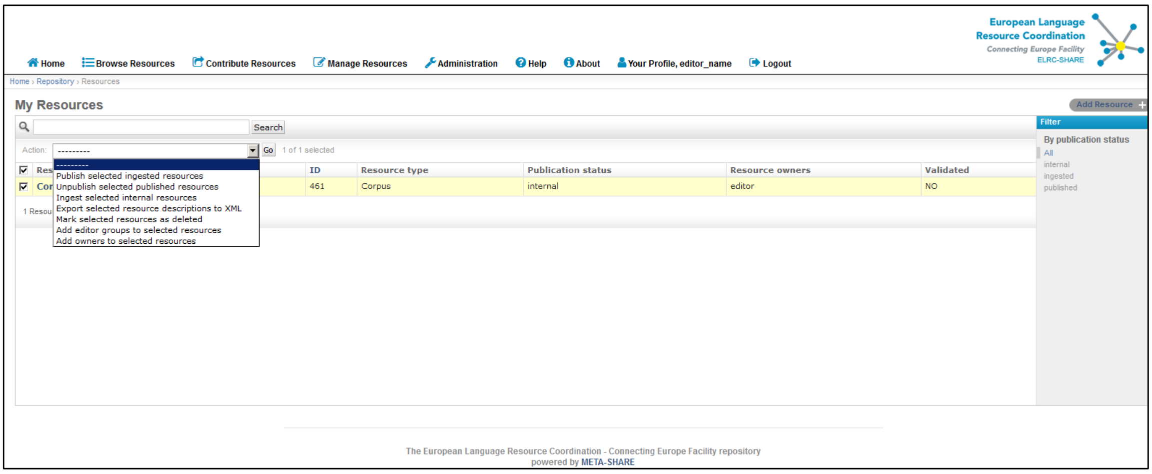
8 Publishing resource descriptions (metadata records)

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 3.2 for publication status).

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.

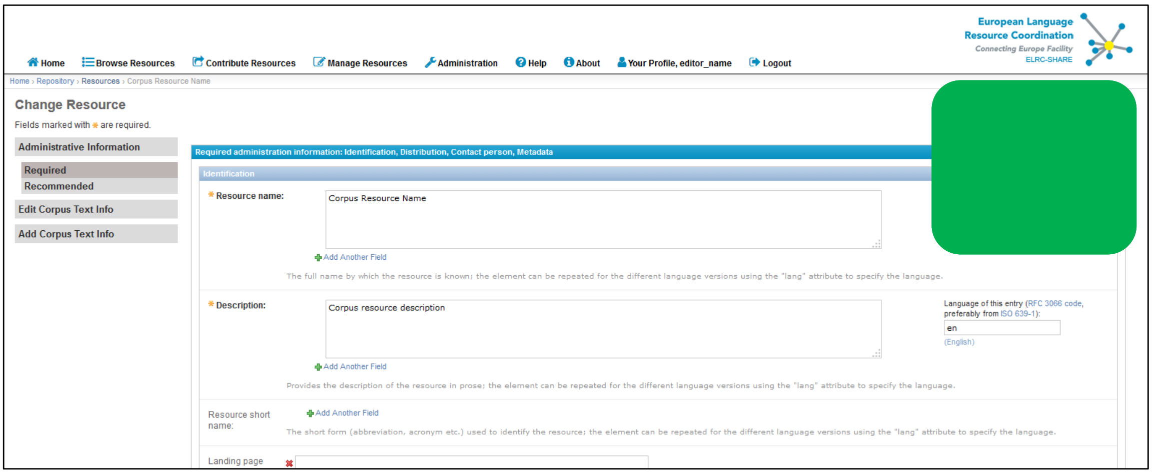


You can also use the same functionality for **exporting metadata records in XML format** or **marking metadata records as "unpublished"** in order to withdraw them from the public inventory.

9 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.

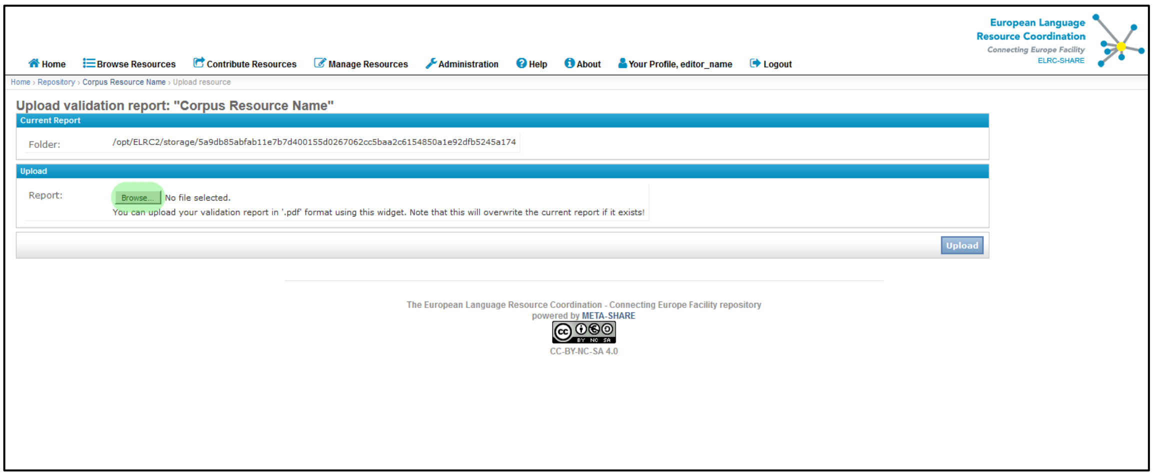
To add a validation report select the **Upload Validation Report (.PDF)** action and click **Go**. In case a validation report already exists, the action is labelled **Replace Validation Report (.PDF)**.



You will be redirected to the “upload validation” form.

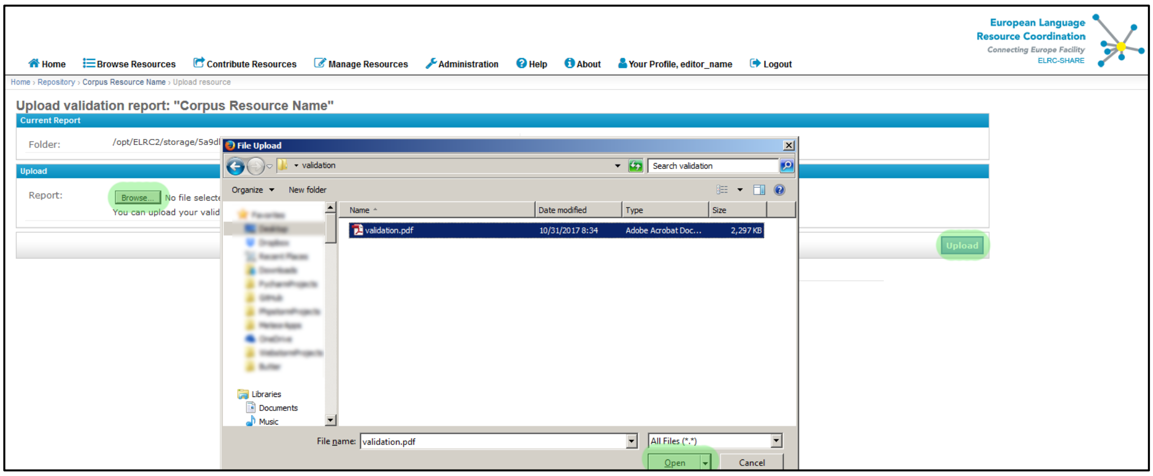
In this page the following information are displayed:

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.

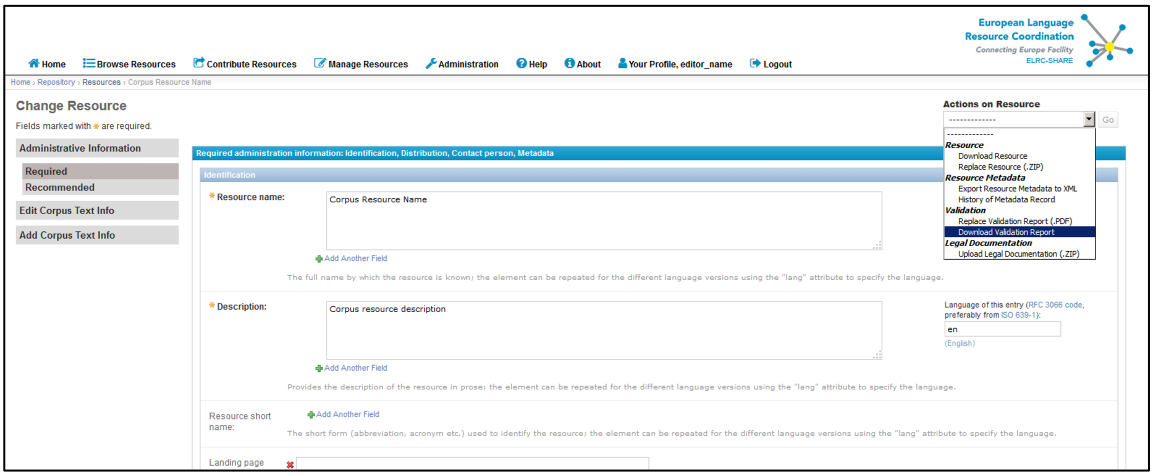


To upload a report

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**.

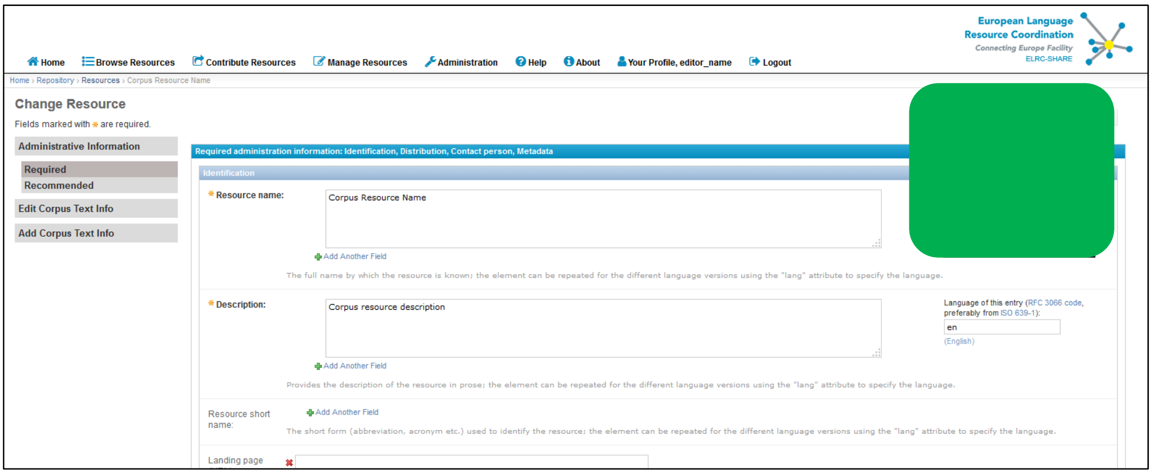


You can also repeat the above process if you want to replace an existing report with a new one.

10 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.

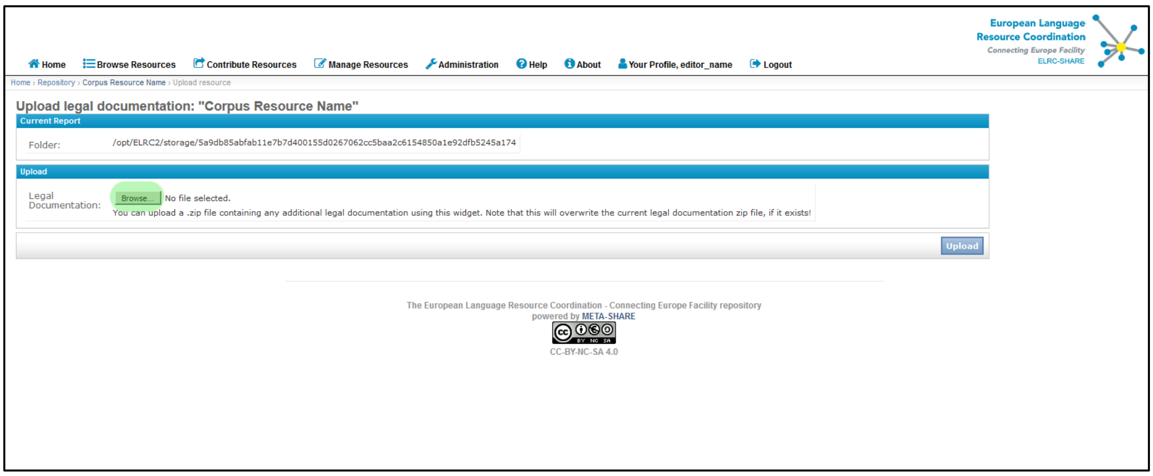
To add legal documentation select the **Upload Legal Documentation (.PDF)** action and click **Go**. In case legal documentation already exists, the action is labelled **Replace Legal Documentation (.ZIP)**.



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

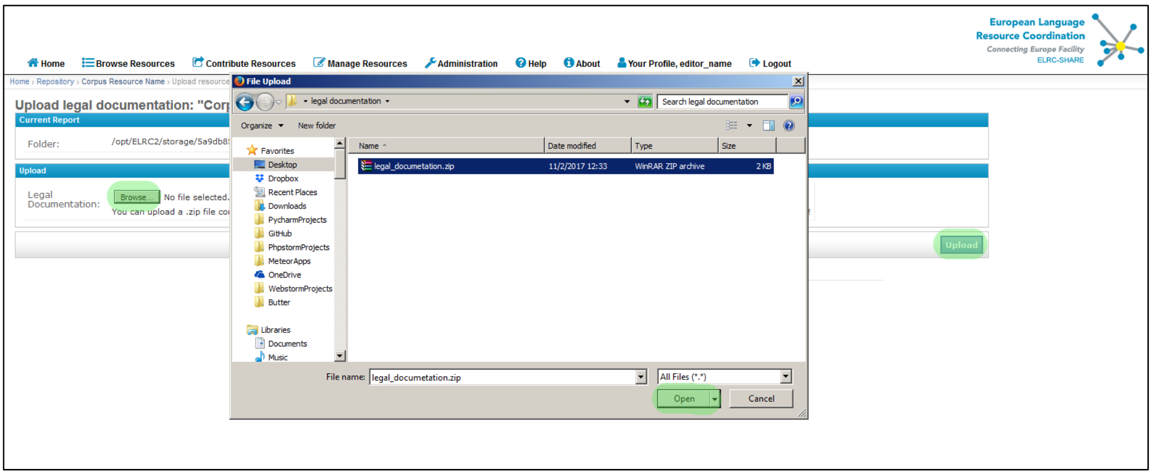
In this page the following information are displayed:

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.

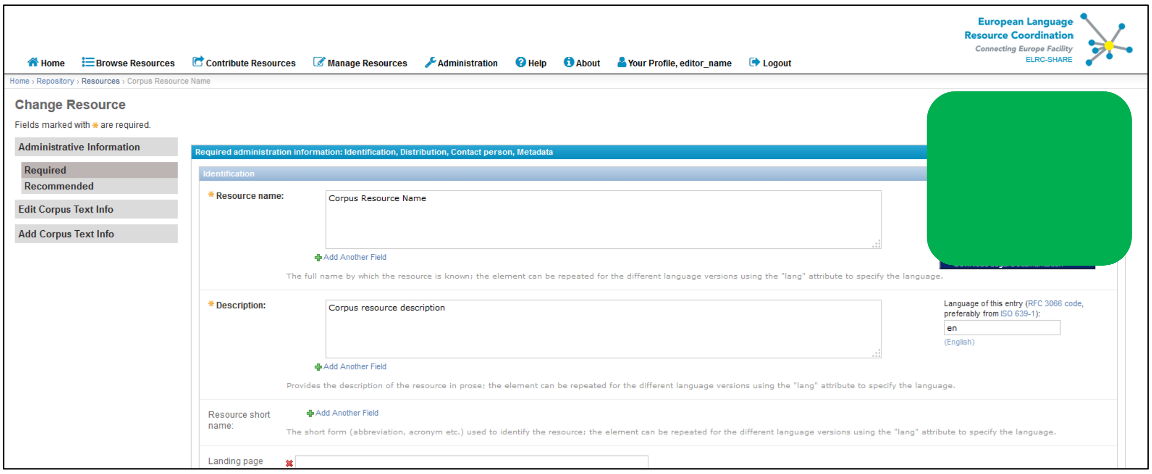


To upload legal documentation

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.