

Resource Uploader Manual

Author: Neeme Kahusk
Version: 1.0
Date: 2018-09-21
number of words: 722

Introduction	1
Users	1
Formats of the Resources	2
Minimal Metadata	2
Types of Resources	3
Corpus	3
Language Description	3
Lexical Conceptual Resource	3
Tool Service	4
How to Find a Suitable Licence?	4
Additional Information	4

Introduction

The Registry of Language Resources of the Center of Estonian Language Resources ([CELR](#)) is created to find and disseminate various language resources. They may be created some specific tool in mind, but they can be used for reference, like dictionaries browsable in the World Wide Web.

Every registered user has the opportunity to upload resource descriptions. One should be a registered user (see [Users](#)), and have knowledge about main characteristics of the resource (see [Minimal Metadata](#)).

Users

The easiest and most preferable way to log in to Meta-Share portal is to use credentials provider by his/her organization. In this case use [SSO login](#) button (see Fig. [SSO login](#)). If you can not find the organization of your affiliation, you can create user by pressing button [Register](#).

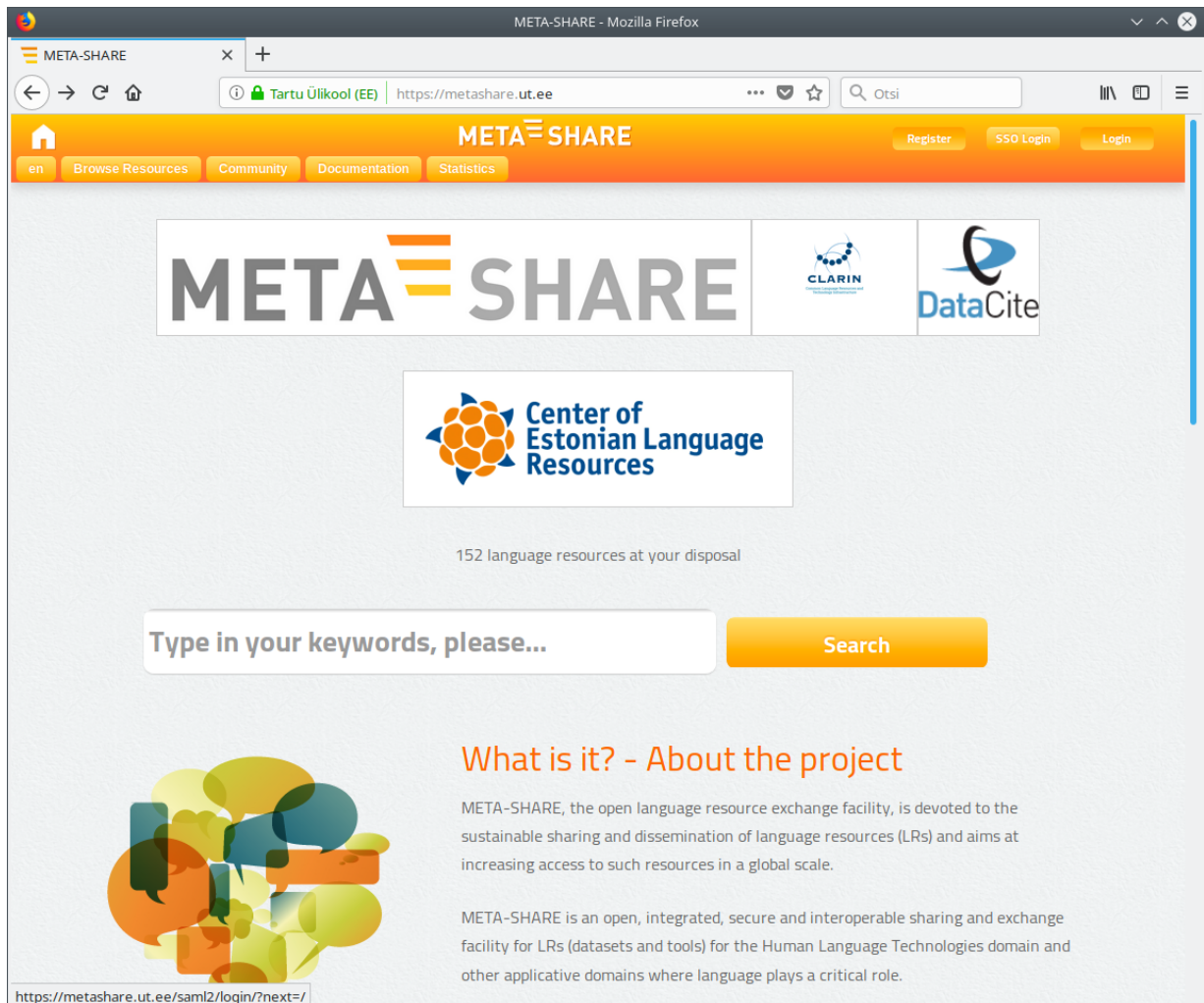


Fig. SSO login: Meta-Share portal home. The cursor points to the SSO login button.

Formats of the Resources

Even when technically it is possible to load up a resource in whatever format you like, we strongly suggest to prefer formats, that correspond to CLARIN's basic principles of resource formats (see [Standards and Formats at CLARIN home page](#)). According to them, open standards are preferred over proprietary standards, well-documented over less documented, verifiable and proven over not verifiable and not used in practice. If there is a choice between text only and binary format, the first should be preferred. In the case of digitisation of an analogue signal, using lossless compression is recommended.

Minimal Metadata

In spite of the type of the resource, these metadata that belong to the **Administrative Information** should be described:

Resource name

Resource name should be long enough to distinguish it from others, and short enough not to be a description. There can be one name for one language, we encourage to use at least a name in English.

Description

Provides the description of the resource in prose. There can be descriptions in different languages.

Distribution

Groups information on the distribution of the resource. Also information about at least one licence is mandatory.

Contact person

At least one person who is responsible for providing additional information about the resource. Last name and e-mail address are mandatory.

Types of Resources

The resource can belong into one of these groups:

1. Corpus
2. Language Description
3. Lexical Conceptual Resource
4. Tool Service

Corpus

For corpora, the following information is needed (besides administrative info):

- **Linguality:** monolingual, bilingual or multilingual
- **identifier and name** for every **language** (the name of the language will be filled in automatically if proper identifier is filled in)
- **Size** and **size unit**. Size may be expressed for example as number of sentences, tokens, or megabytes.

If video or image component is included in corpus, at least **size** and **sizeunit** are required. If n-gram component is included, n-gram **base** and **order** is required.

Language Description

For Language Description, the following information is needed *per text part* (besides administrative info):

- **Language Description Type** grammar or else
- **Linguality:** monolingual, bilingual or multilingual
- **identifier and name** for every **language** (the name of the language will be filled in automatically if proper identifier is filled in)

If *video* or *image* component is included in Language Description, **reference to other medium** is required.

Lexical Conceptual Resource

For Lexical Conceptual Resource, the following information is needed (besides administrative info):

- **Lexical Conceptual Resource Type**
- **Linguality:** monolingual, bilingual or multilingual
- **identifier and name** for every **language** (the name of the language will be filled in automatically if proper identifier is filled in)
- **Size** and **size unit**. Size may be expressed for example as number of sentences, words, or megabytes.

If *video* component is included in Lexical Conceptual Resource **video type** is required.

Tool Service

For Tool Service, the following information is needed (besides administrative info):

- **Tool Service Type**
- Whether the Tool or Service is **Language dependent** or not.

How to Find a Suitable Licence?

Use [Licence calculator](#) or check out [FAQ](#).

Additional Information

Additional information can be found at:

- [FAQ](#)
- Mail to info@keeleressursid.ee