EUROPEAN
LANGUAGE
GRID

D3.3
Platform GUI (interim release)

Authors:	Andis Lagzdinš, Uldis Silinš, Ēriks Ajausks (TILDE), Penny Labropoulou, Dimitris Gkoumas (ILSP)
Dissemination Level:	Public
Date:	31-03-2021

About this document

Project	ELG – European Language Grid
Grant agreement no.	825627 – Horizon 2020, ICT 2018-2020 – Innovation Action
Coordinator	Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2019, 42 months (GA amendment version: AMD-825627-7)
Deliverable number	D3.3
Deliverable title	Platform GUI (interim release)
Type	Other
Number of pages	28
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 31-03-2021 – Actual: 31-03-2021
WP number and title	WP3: Grid Platform – Interactive Interface and Information System
Task number and title	Task 3.2: Design, development and deployment of ELG front-end Task 3.3: Set up and maintenance of CMS for the ELG
Authors	Andis Lagzdinš, Uldis Silinš, Ēriks Ajausks (TILDE), Penny Labropoulou, Dimitris Gkoumas (ILSP)
Reviewers	Ian Roberts (USFD), Ulrich Germann (UEDIN)
Consortium	Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Germany
	Institute for Language and Speech Processing (ILSP), Greece
	University of Sheffield (USFD), United Kingdom
	Charles University (CUNI), Czech Republic
	Evaluations and Language Resources Distribution Agency (ELDA), France
	Tilde SIA (TILDE), Latvia
	Hensoldt Analytics (HENS), Austria
	Expert System Iberia SL (EXPSYS), Spain
	University of Edinburgh (UEDIN), United Kingdom
EC project officers	Philippe Gelin
For copies of reports and other ELG-related information, please contact:	DFKI GmbH European Language Grid (ELG) Alt-Moabit 91c D-10559 Berlin Germany Dr. Georg Rehm, DFKI GmbH georg.rehm@dfki.de Phone: +49 (0)30 23895-1833 Fax: +49 (0)30 23895-1810 http://european-language-grid.eu © 2021 ELG Consortium

Table of Contents

List of Figures 4
List of Tables 4
List of Abbreviations 5
Abstract 6
1 Introduction 6
2 Central Portal 7
2.1 Functionality and Web Interface 7
3 Content Management System 11
3.1 Added Functionality and Web Interfaces 11
4 Catalogue UI 11
4.1 Technical Description 11
4.2 Functionality and Web Interfaces 13
4.2.1 Catalogue Browse and Search Page 14
4.2.2 Landing Page for Items 15
4.2.3 Dashboard 16
4.2.4 Creating Items with the Interactive Editor 17
4.2.5 Managing items 22
5 Authentication Solution 24
5.1 Technical Description 24
5.2 Keycloak GUI customization 24
6 UX implementation 28
7 Conclusion 28

List of Figures

Figure 1: Language technologies silo page 7
Figure 2: Language and resources data 8
Figure 3: Recently updated Language and resources data 8
Figure 4: Community silo page 9
Figure 5: Latest organisations added on the community silo page 9
Figure 6: User profile page 10
Figure 7: Invitation to become a provider 10
Figure 8: Landing page with the map of Europe 11
Figure 9: Provider’s request form in CMS admin area 11
Figure 10: Data exchange with and without code splitting 12
Figure 11: Performance metrics for R1 13
Figure 12: Performance metrics for R2 13
Figure 13: List view of catalogue contents 14
Figure 14: Open facet for language 15
Figure 15: Display of publication status 15
Figure 16: My grid 16
Figure 17: Editor form for corpora 17
Figure 18: Selection of item type 18
Figure 19: Search of existing records 18
Figure 20: View of a multilingual element with help tips and examples 19
Figure 21: Selecting the language for a multilingual element 19
Figure 22: Hidden elements 20
Figure 23: Opening a hidden element in order to add metadata 20
Figure 24: Lookup field 20
Figure 25: Use of accordion for long elements 21
Figure 26: Error message 21
Figure 27: My items 22
Figure 28: Actions on my items 23
Figure 29: Actions on multiple items 23
Figure 30: Actions area on the landing page of an item 24
Figure 31: Login form 25
Figure 32: New user registration form 25
Figure 33: Password recovery form 26
Figure 34: Styled form for password change 26
Figure 35: User account edit form 27
Figure 36: User submitted attributes in Keycloak Management Console 27
Figure 37: Provider's journey 28

List of Tables

Table 1: Actions allowed on metadata record by publication status 22

List of Abbreviations

API	Application Programming Interface
CMS	Content Management System
CSS	Cascading Style Sheets
ELG	European Language Grid
GPL	GNU General Public License
GUI	Graphical User Interface
HTML	Hypertext Markup Language
JWT	JSON Web Token
LT	Language Technologies
MVP	Minimum Viable Product
POS	Part of Speech
REST	Representational State Transfer
SEO	Search Engine Optimization
SME	Small and Medium Size Enterprises
SPA	Single Page Application
SSR	Server-Side Rendering
UI	User Interface
URL	Uniform Resource Locator
UX	User Experience
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines
WYSIWYG	What You See Is What You Get

Abstract

This document gives an overview of the ELG platform graphical user interface. In particular, it summarises the work done for the interim release. Since the submission of D3.2, we have created a provider's journey, performed visual improvements for the central portal and created silo pages for the sections on technologies, resources and community in the catalogue. We have improved the procedure for a potential provider to become a user. Work has been carried out to customise themes for Keycloak. Furthermore, we have added new functionalities in the catalogue. The four major functionalities that have been added comprise: the Search and browse page, the single landing page for each catalogue item, the dashboard and the item management and editing capabilities. This document summarises all listed features.

1 Introduction

The most recent GUI interface is deployed in Release 2 of the platform at https://live.european-language-grid.eu. This document focuses on the changes made since the first release and thus supplements rather than replaces Deliverable D3.2.

Recall from D3.2 that the Platform UI and the back-end that drives it comprises the following components:

- The **Central Portal** is the network of web pages that provide user role-specific **journeys** through the ELG's offerings. User roles include, among others, *Consumers* who use the ELG to find information or use services, and *Providers*, who supply resources or services. A journey is a use-case specific network of paths that describes the user experience as they navigate through the Platform in order to achieve their objectives. Journey design is a core component of UX design. In particular, we added a Provider role and Journey to facilitate contributions to the ELG.
- The **Content Management System** drives the dynamic generation of web pages that users of the portal interact with.
- The **Catalogue UI** allows users to interact with the ELG Catalogue.
- The **Authentication Mechanism** authenticates users and grants or denies access rights to the platform depending on the user's role within the platform.

Sections 2–5 describe changes to these components since the last release. Section 6 describes the newly added provider journey. Section 7 wraps up this deliverable.

The frontend architecture has not changed since the initial release. For an overview of the frontend, see the architecture description in Deliverable D3.2.

2 Central Portal

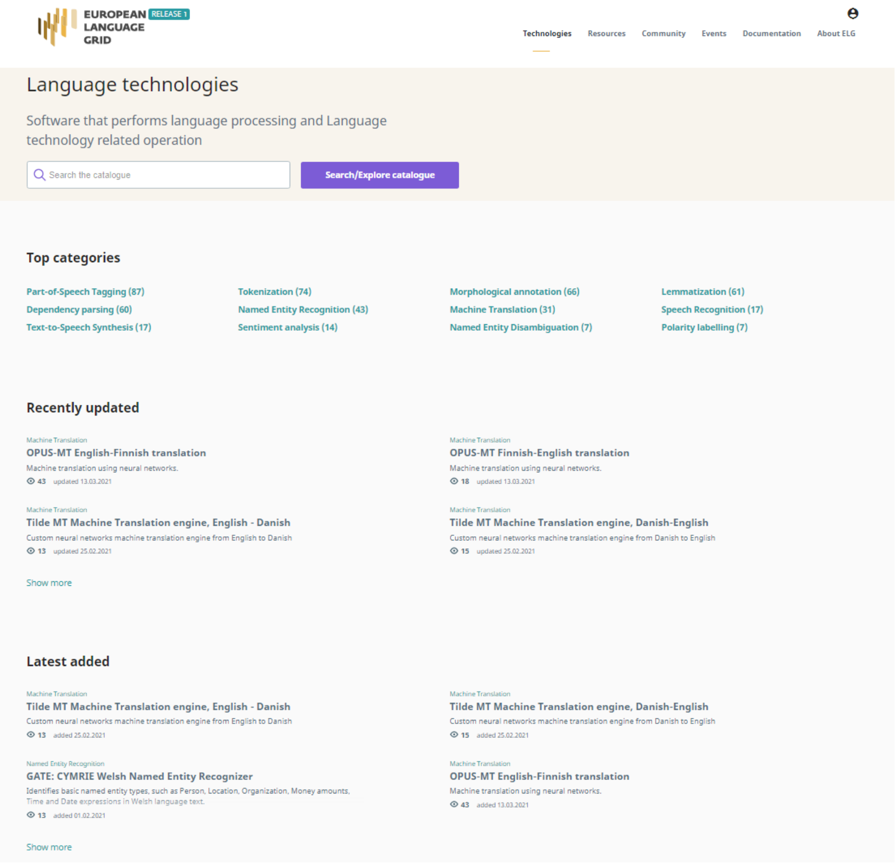
2.1 Functionality and Web Interface

New content

In order to facilitate the users' journey through the platform, we added silo pages to the ELG front-end. Silo pages are a convenient way of separating content into categories. They give the user an overview and good starting point to explore. The following silo pages were added:

- Technologies

The Technology page shows top technology categories used in the catalogue. Each category is clickable and leads the user the catalogue search page with the relevant facet[1] pre-selected. Besides the top categories, the user can see recently updated and added services.



Language technologies

Software that performs language processing and Language technology related operation

Top categories

Part-of-Speech Tagging (87)
Dependency parsing (60)
Text-to-Speech Synthesis (17)
Tokenization (74)
Named Entity Recognition (43)
Sentiment analysis (14)
Morphological annotation (66)
Machine Translation (31)
Named Entity Disambiguation (7)
Lemmatization (61)
Speech Recognition (17)
Polarity labelling (7)

Recently updated

Machine Translation
OPUS-MT English-Finnish translation
Machine translation using neural networks.
43 updated 13.03.2021

Machine Translation
OPUS-MT Finnish-English translation
Machine translation using neural networks.
18 updated 13.03.2021

Machine Translation
Tilde MT Machine Translation engine, English - Danish
Custom neural networks machine translation engine from English to Danish
13 updated 25.02.2021

Machine Translation
Tilde MT Machine Translation engine, Danish-English
Custom neural networks machine translation engine from Danish to English
15 updated 25.02.2021

Show more

Latest added

Machine Translation
Tilde MT Machine Translation engine, English - Danish
Custom neural networks machine translation engine from English to Danish
13 added 25.02.2021

Machine Translation
Tilde MT Machine Translation engine, Danish-English
Custom neural networks machine translation engine from Danish to English
15 added 25.02.2021

Named Entity Recognition
GATE: CYMRIE Welsh Named Entity Recognizer
Identifies basic named entity types, such as Person, Location, Organization, Money amounts, Time and Date expressions in Welsh language text.
13 added 01.02.2021

Machine Translation
OPUS-MT English-Finnish translation
Machine translation using neural networks.
43 added 13.03.2021

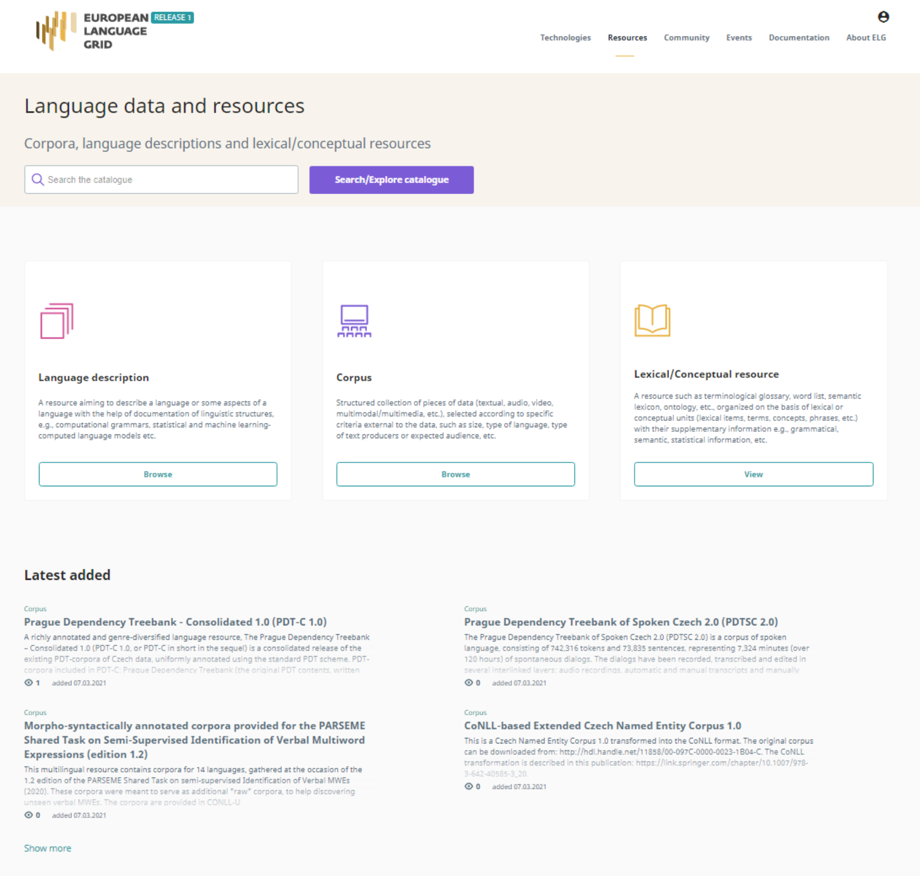
Show more

Figure 1: Language technologies silo page

- Resources

The Language Data and Resources silo page shows three main resource categories that can be used as a shortcut to the catalogue search results.

[1] See Section 6.2.1 for details on the new faceted search interface



Language data and resources

Corpora, language descriptions and lexical/conceptual resources

Language description

A resource aiming to describe a language or some aspects of a language with the help of documentation of linguistic structures, e.g., computational grammars, statistical and machine learning-computed language models etc.

Corpus

Structured collection of pieces of data (textual, audio, video, multimodal/multimedia, etc.), selected according to specific criteria external to the data, such as size, type of language, type of text producers or expected audience, etc.

Lexical/Conceptual resource

A resource such as terminological glossary, word list, semantic lexicon, ontology, etc., organized on the basis of lexical or conceptual units (lexical items, terms, concepts, phrases, etc.) with their supplementary information e.g., grammatical, semantic, statistical information, etc.

Latest added

Corpus
Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)
A richly annotated and genre-diversified language resource, The Prague Dependency Treebank – Consolidated 1.0 (PDT-C 1.0, or PDT-C in short in the sequel) is a consolidated release of the existing PDT-corpora of Czech data, uniformly annotated using the standard PDT scheme. PDT-corpora included in PDT-C: Prague Dependency Treebank (the original PDT contents, written
1 added 07.03.2021

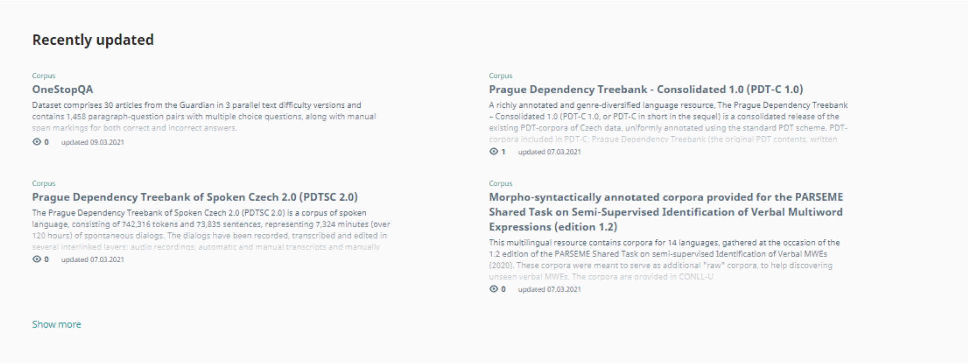
Corpus
Prague Dependency Treebank of Spoken Czech 2.0 (PDTSC 2.0)
The Prague Dependency Treebank of Spoken Czech 2.0 (PDTSC 2.0) is a corpus of spoken language, consisting of 742,316 tokens and 73,835 sentences, representing 7,324 minutes (over 120 hours) of spontaneous dialogs. The dialogs have been recorded, transcribed and edited in several interlinked layers: audio recordings, automatic and manual transcripts and manually
0 added 07.03.2021

Corpus
Morpho-syntactically annotated corpora provided for the PARSEME Shared Task on Semi-Supervised Identification of Verbal Multiword Expressions (edition 1.2)
This multilingual resource contains corpora for 14 languages, gathered at the occasion of the 1.2 edition of the PARSEME Shared Task on semi-supervised Identification of Verbal MWEs (2020). These corpora were meant to serve as additional "raw" corpora, to help discovering unseen verbal MWEs. The corpora are provided in CONLL-U
0 added 07.03.2021

Corpus
CoNLL-based Extended Czech Named Entity Corpus 1.0
This is a Czech Named Entity Corpus 1.0 transformed into the CoNLL format. The original corpus can be downloaded from: http://hdl.handle.net/11858/00-097C-0000-0023-1B04-C. The CoNLL transformation is described in this publication: https://link.springer.com/chapter/10.1007/978-3-642-40585-3_20.
0 added 07.03.2021

Figure 2: Language and resources data

Afterwards, the user can see recently added and updated resources, ensuring quick access to the latest changes since their last visit.



Recently updated

Corpus
OneStopQA
Dataset comprises 30 articles from the Guardian in 3 parallel text difficulty versions and contains 1,458 paragraph-question pairs with multiple choice questions, along with manual span markings for both correct and incorrect answers.
0 updated 09.03.2021

Corpus
Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)
A richly annotated and genre-diversified language resource, The Prague Dependency Treebank – Consolidated 1.0 (PDT-C 1.0, or PDT-C in short in the sequel) is a consolidated release of the existing PDT-corpora of Czech data, uniformly annotated using the standard PDT scheme. PDT-corpora included in PDT-C: Prague Dependency Treebank (the original PDT contents, written
1 updated 07.03.2021

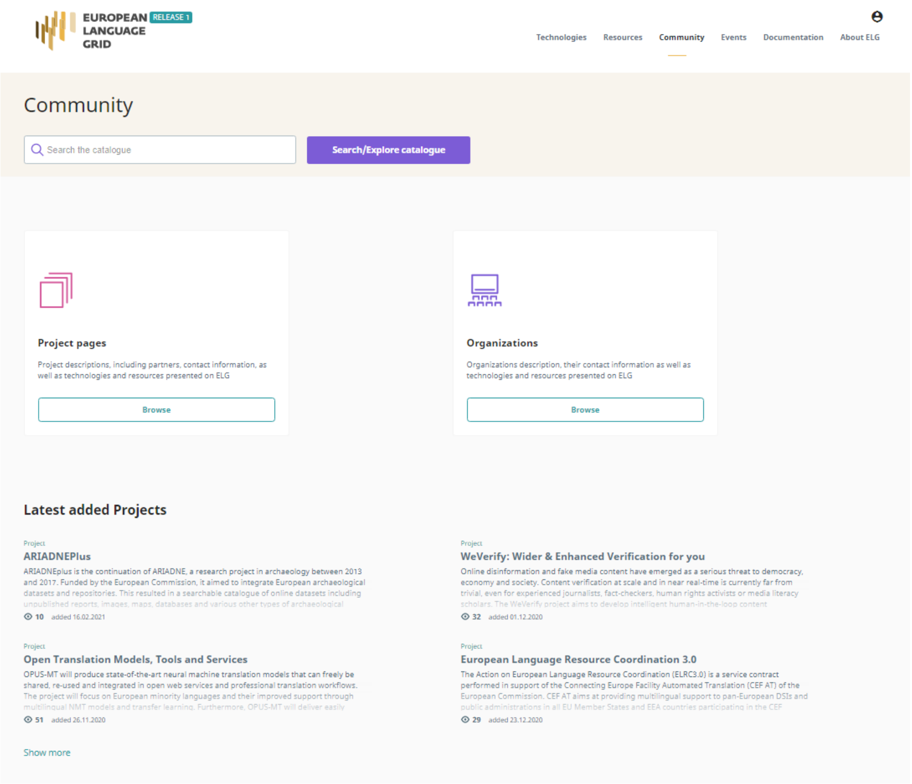
Corpus
Prague Dependency Treebank of Spoken Czech 2.0 (PDTSC 2.0)
The Prague Dependency Treebank of Spoken Czech 2.0 (PDTSC 2.0) is a corpus of spoken language, consisting of 742,316 tokens and 73,835 sentences, representing 7,324 minutes (over 120 hours) of spontaneous dialogs. The dialogs have been recorded, transcribed and edited in several interlinked layers: audio recordings, automatic and manual transcripts and manually
0 updated 07.03.2021

Corpus
Morpho-syntactically annotated corpora provided for the PARSEME Shared Task on Semi-Supervised Identification of Verbal Multiword Expressions (edition 1.2)
This multilingual resource contains corpora for 14 languages, gathered at the occasion of the 1.2 edition of the PARSEME Shared Task on semi-supervised Identification of Verbal MWEs (2020). These corpora were meant to serve as additional "raw" corpora, to help discovering unseen verbal MWEs. The corpora are provided in CONLL-U
0 updated 07.03.2021

Figure 3: Recently updated Language and resources data

- Community

The Community silo page (Figure 4) has the same structure and content as the previous two, in this way providing the user with a consistent user interface that helps find the right item.



Community

Project pages

Project descriptions, including partners, contact information, as well as technologies and resources presented on ELG

Organizations

Organizations description, their contact information as well as technologies and resources presented on ELG

Latest added Projects

Project
ARIADNEPlus
ARIADNEplus is the continuation of ARIADNE, a research project in archaeology between 2013 and 2017. Funded by the European Commission, it aimed to integrate European archaeological datasets and repositories. This resulted in a searchable catalogue of online datasets including unpublished reports, images, maps, databases and various other types of archaeological
10 added 16.02.2021

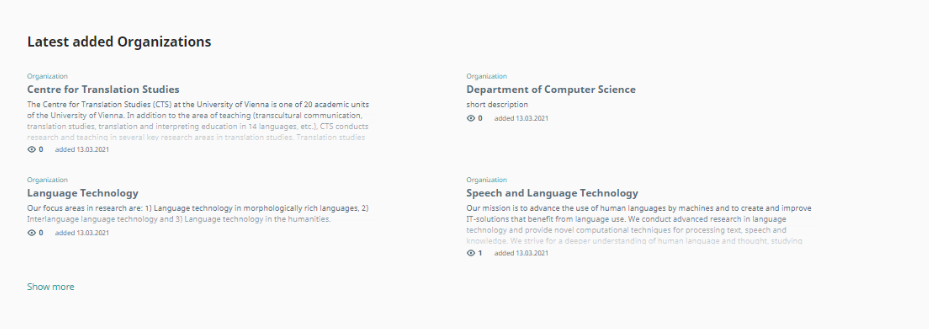
Project
WeVerify: Wider & Enhanced Verification for you
Online disinformation and fake media content have emerged as a serious threat to democracy, economy and society. Content verification at scale and in near real-time is currently far from trivial, even for experienced journalists, fact-checkers, human rights activists or media literacy scholars. The WeVerify project aims to develop intelligent human-in-the-loop content
32 added 01.12.2020

Project
Open Translation Models, Tools and Services
OPUS-MT will produce state-of-the-art neural machine translation models that can freely be shared, re-used and integrated in open web services and professional translation workflows. The project will focus on European minority languages and their improved support through multilingual NMT models and transfer learning. Furthermore, OPUS-MT will deliver easily
51 added 26.11.2020

Project
European Language Resource Coordination 3.0
The Action on European Language Resource Coordination (ELRC3.0) is a service contract performed in support of the Connecting Europe Facility Automated Translation (CEF AT) of the European Commission. CEF AT aims at providing multilingual support to pan-European DSIs and public administrations in all EU Member States and EEA countries participating in the CEF
29 added 23.12.2020

Figure 4: Community silo page

At the bottom of the page, users can find the latest added organisations.



Latest added Organizations

Organization
Centre for Translation Studies
The Centre for Translation Studies (CTS) at the University of Vienna is one of 20 academic units of the University of Vienna. In addition to the area of teaching (transcultural communication, translation studies, translation and interpreting education in 14 languages, etc.), CTS conducts research and teaching in several key research areas in translation studies. Translation studies
0 added 13.03.2021

Organization
Department of Computer Science
short description
0 added 13.03.2021

Organization
Language Technology
Our focus areas in research are: 1) Language technology in morphologically rich languages, 2) Interlanguage language technology and 3) Language technology in the humanities.
0 added 13.03.2021

Organization
Speech and Language Technology
Our mission is to advance the use of human languages by machines and to create and improve IT-solutions that benefit from language use. We conduct advanced research in language technology and provide novel computational techniques for processing text, speech and knowledge. We strive for a deeper understanding of human language and thought, studying
1 added 13.03.2021

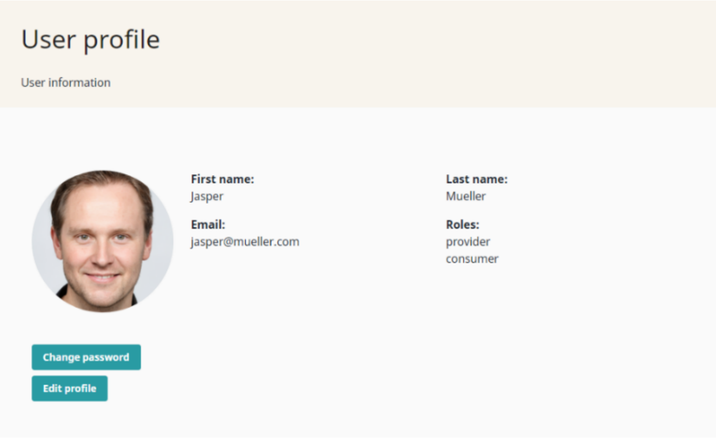
Figure 5: Latest organisations added on the community silo page

Integration to external pages was added as well:

- Events are now on the managed projects webpage
- Documentation is managed in the separated documentation solution

User account overview page

The user can now add a profile picture and see their roles Figure 6. This is crucial as we have added provider functionality and it is necessary to see whether this role has been granted. The profile page is also one of the entry points for requesting the provider's role.



User profile

User information

First name:
Jasper

Last name:
Mueller

Email:
jasper@mueller.com

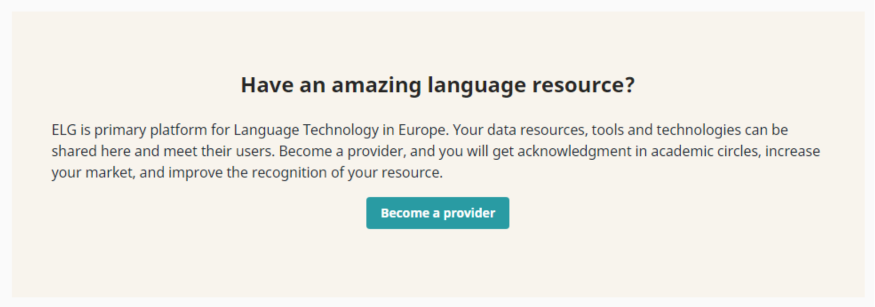
Roles:
provider
consumer

Figure 6: User profile page

Provider’s role requests

Added functionality to request the provider’s role after signing up.

To be able to add new services, a user must either claim an existing resource from the catalogue or obtain provider rights by signing up for it. When the user visits the portal, there is a banner on the landing page that encourages them to sign in and become a provider (Figure 7). The same banner is located on the user’s profile page. When a non-provider signs up for the role of provider, they are informed that we will get in touch within a couple of days. This is a standard signing up process that can be found in other portals, and does not require extra effort from the user.



Have an amazing language resource?

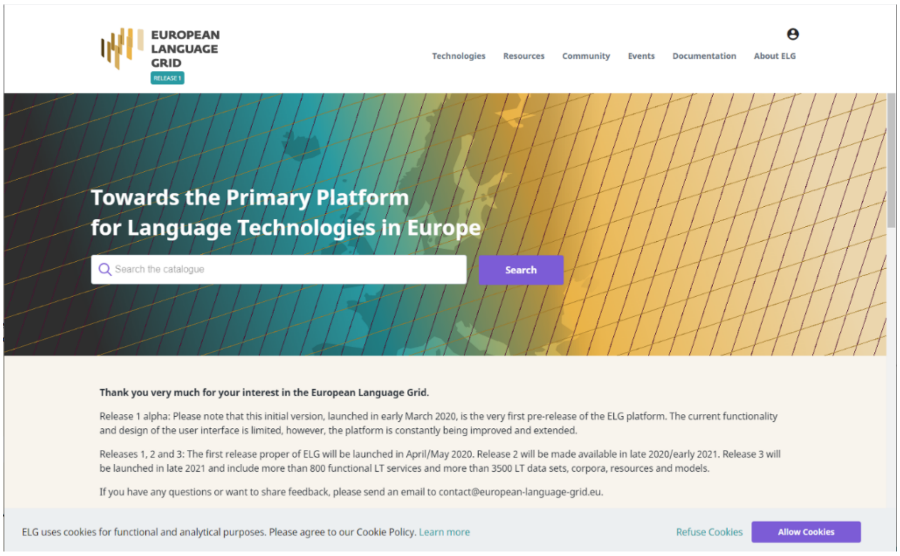
ELG is primary platform for Language Technology in Europe. Your data resources, tools and technologies can be shared here and meet their users. Become a provider, and you will get acknowledgment in academic circles, increase your market, and improve the recognition of your resource.

Figure 7: Invitation to become a provider

Visual improvements

- Sliding landing page

The landing page leaves an important first impression for the new visitor, and this impression is decisive with regard to whether the user wants to further interact with the page. To better represent the common interface, according to the graphical designer’s recommendations, we have added a map of Europe that slides across the header, mimicking the rotation of the Earth as indicated in Figure 8.



Towards the Primary Platform
for Language Technologies in Europe

Thank you very much for your interest in the European Language Grid.

Release 1 alpha: Please note that this initial version, launched in early March 2020, is the very first pre-release of the ELG platform. The current functionality and design of the user interface is limited, however, the platform is constantly being improved and extended.

Releases 1, 2 and 3: The first release proper of ELG will be launched in April/May 2020. Release 2 will be made available in late 2020/early 2021. Release 3 will be launched in late 2021 and include more than 800 functional LT services and more than 3500 LT data sets, corpora, resources and models.

If you have any questions or want to share feedback, please send an email to contact@european-language-grid.eu.

ELG uses cookies for functional and analytical purposes. Please agree to our Cookie Policy. Learn more

Figure 8: Landing page with the map of Europe

Mobile view adjustments

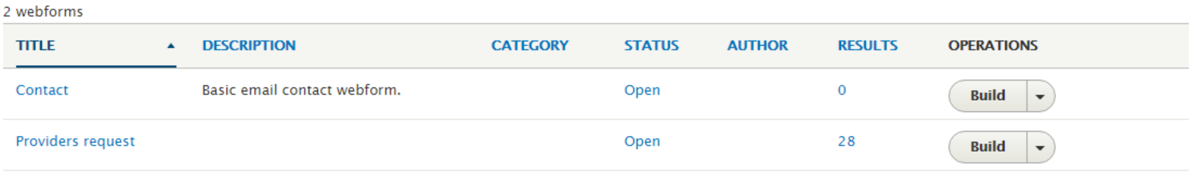
Every new element that is added to the existing page, as well as any new page, is tested and adjusted to different screen sizes starting from mobile devices to large widescreen monitors. No new features are added to the mobile view, but it is double-checked that the user can use the portal from a phone as easily as from a desktop.

3 Content Management System

3.1 Added Functionality and Web Interfaces

In addition to existing functionality, webform submission was enabled with automatic email notifications.

This was part of the request provider's role scenario. To receive and store user requests, we decided to use the portal's CMS as it can send out user notifications as well. When the user presses the "request provider's role" button, there is no need to fill out any other data – the system gathers the logged-in user's data and posts it to CMS for storage purposes. At the same time, our user managers receive an email notification and can act upon the posted data.



2 webforms

TITLE ▲	DESCRIPTION	CATEGORY	STATUS	AUTHOR	RESULTS	OPERATIONS
Contact	Basic email contact webform.		Open		0	Build ▾
Providers request			Open		28	Build ▾

Figure 9: Provider's request form in CMS admin area

4 Catalogue UI

4.1 Technical Description

For the implementation of the catalogue UI of Release 2 (R2), we continue to follow the architectural and technical directives used for R1 described in D3.2.

The most notable architectural change that we have introduced in this release is the notion of code splitting (https://reactjs.org/docs/code-splitting.html). Until now, the ELG catalogue pages were provided as one large bundle file. All the code was delivered in one single file to the client.

As the ELG catalogue pages started to grow, the size of the bundle file also grew, making this approach inefficient. When users access the ELG catalogue pages, they have to wait until this file is fully downloaded, and only then can they interact with the page.

One more drawback to this approach is that because all of the code is merged into one single file, users have to download portions of code that they will never make use of. For instance, unregistered users that do not have access to dashboard pages still have to wait and download the related code, even though they will never use it. This creates a bad user experience, especially for users that are on slow or congested mobile networks.

By using the code splitting technique, we mitigate all of these issues. The idea behind this is that the code is split into smaller autonomous files and the users are only provided with the files that are required to load the pages that they are accessing. Thus, users that navigate to the catalogue's main page will only have to download the files that are needed to interact with this page. Moreover, most browsers would cache these files so any subsequent visit to these pages would make use of the cached version of the files.

Figure 10 depicts the total amount of data in MB that a mobile user will have to use to load the ELG catalogue's main page when code splitting is used and when it is not. The chart depicts the transferred resources when they are compressed and when they are not. As we can see, the use of code splitting has a huge impact and dramatically improves the initial load size.

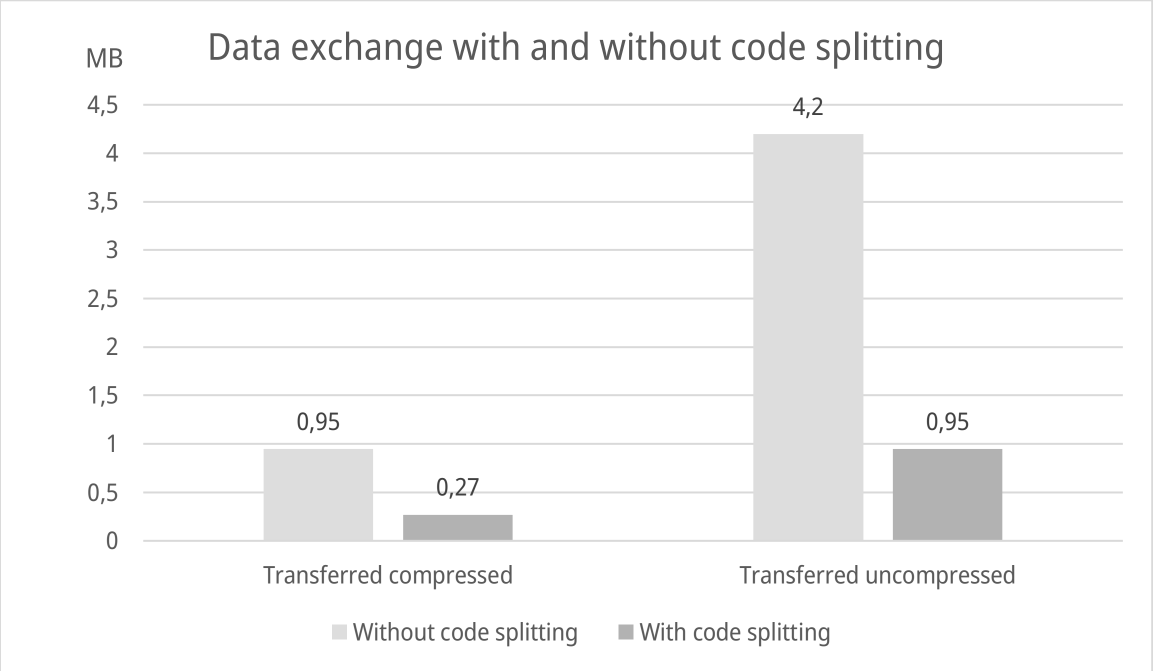
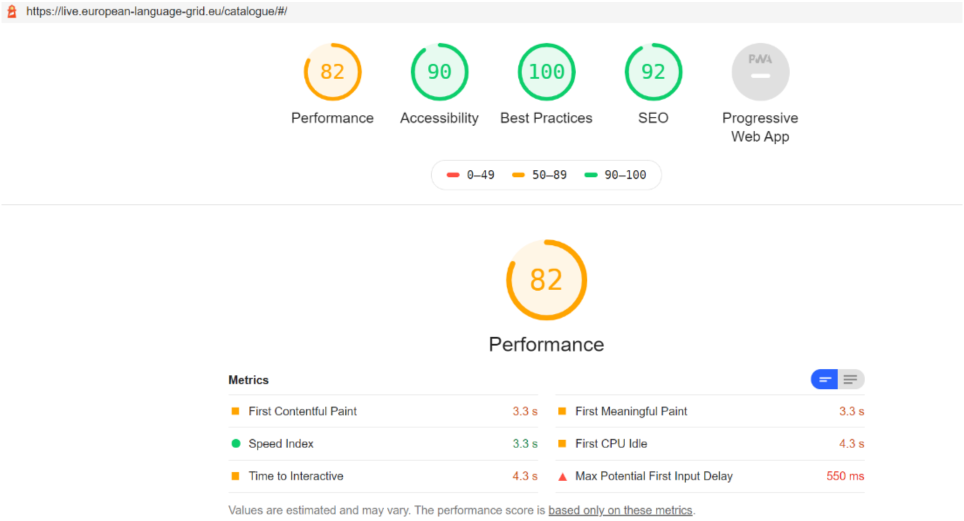


Figure 10: Data exchange with and without code splitting

In order to have some baselines of quality measures for our application we have used Google Lighthouse[2] , an Open Source auditing tool for performance and accessibility. Figure 11 and Figure 12 show the performance

[2] https://developers.google.com/web/tools/lighthouse/

metrics for R1 and R2. For upcoming releases, we aim to improve these metrics and focus more on features such as SEO and Server-Side Rendering.



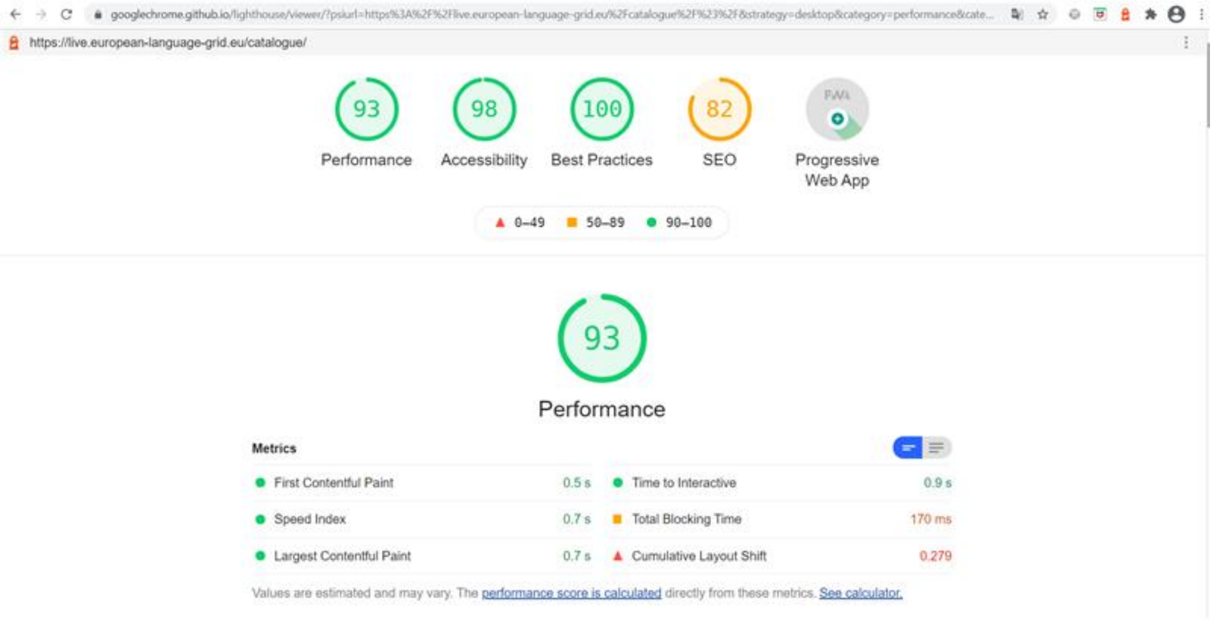
https://live.european-language-grid.eu/catalogue/#/

Metrics

■	First Contentful Paint	3.3 s	■	First Meaningful Paint	3.3 s
●	Speed Index	3.3 s	■	First CPU Idle	4.3 s
■	Time to Interactive	4.3 s	▲	Max Potential First Input Delay	550 ms

Values are estimated and may vary. The performance score is based only on these metrics.

Figure 11: Performance metrics for R1



googlechrome.github.io/lighthouse/viewer/?psiurl=https%3A%2F%2Flive.european-language-grid.eu%2Fcatalogue%2F%23%2F&strategy=desktop&category=performance&cate...

https://live.european-language-grid.eu/catalogue/

Metrics

●	First Contentful Paint	0.5 s	●	Time to Interactive	0.9 s
●	Speed Index	0.7 s	■	Total Blocking Time	170 ms
●	Largest Contentful Paint	0.7 s	▲	Cumulative Layout Shift	0.279

Values are estimated and may vary. The performance score is calculated directly from these metrics. See calculator.

Figure 12: Performance metrics for R2

4.2 Functionality and Web Interfaces

The catalogue UI is the main point through which users interact with resources hosted or described in the ELG. These include:

- Functional content, i.e., LT services and applications that are fully integrated into the platform and that can be used through trial UIs supplied by ELG and executed following ELG specifications.
- Non-functional content, i.e., metadata records for
  - data resources (e.g., corpora, datasets, lexica, terminologies, Machine Learning models, computational grammar, etc.) that are available through ELG and/or other catalogues, and

  - tools/services that are only available in a downloadable form from the ELG platform or that are available through other catalogues or websites.
- Metadata records for companies, research organisations, projects, etc., that are active in the LT domain and can promote themselves through ELG, while providing an overview of the LT landscape.

ELG targets a wide range of users with different requirements and expectations, broadly divided into the following groups (cf. Deliverables D2.1, D2.3 and D7.1 for more information):

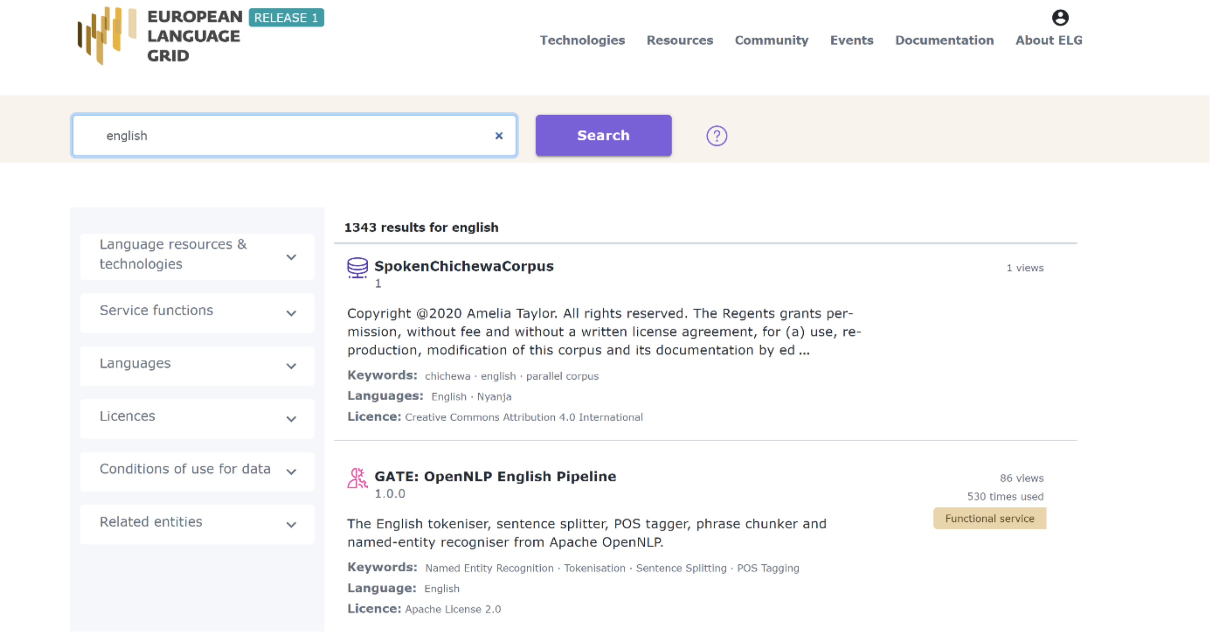
- **content providers** and **developers** and **integrators** are providers and consumers of LT resources, as described above,
- **information providers** and **information seekers** are providers and consumers of LT-related (meta)-information, as described above,
- **citizens**, i.e., individuals who want to inform themselves about LT and who understand the scope of ELG,
- **ELG platform and content administrators**, i.e., the ELG technical team that administers and monitors the day-to-day operation and performance of the platform.

The needs and requirements of each user group have influenced the platform architecture (D2.2), the metadata schema used to describe the entities of the catalogue (D2.3) and, of course, its design.

R2 of the Catalogue UI builds upon, enhances and extends the functionalities and features of the previous release. In R1 the focus was on the consumer's side, while R2 incorporates functionalities for the provider and, to a lesser degree, the ELG administrators. In the following sections, we present the main added features, while an overview of the previous features can be found in D3.2.

4.2.1 Catalogue Browse and Search Page

The look and feel of the catalogue's main page (Figure 13), which is used for browsing and searching the catalogue items, remains largely the same as for R1, with a few stylistic improvements.



1343 results for english

SpokenChichewaCorpus
1
1 views

Copyright @2020 Amelia Taylor. All rights reserved. The Regents grants permission, without fee and without a written license agreement, for (a) use, reproduction, modification of this corpus and its documentation by ed ...

Keywords: chichewa · english · parallel corpus
Languages: English · Nyanja
Licence: Creative Commons Attribution 4.0 International

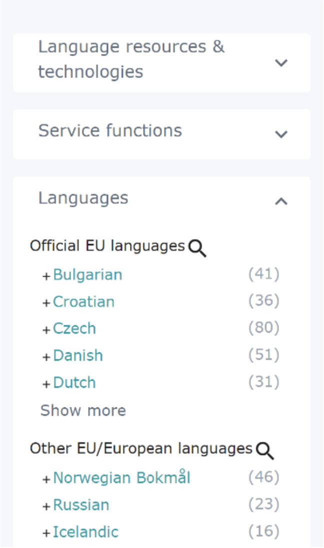
GATE: OpenNLP English Pipeline
1.0.0
86 views
530 times used
Functional service

The English tokeniser, sentence splitter, POS tagger, phrase chunker and named-entity recogniser from Apache OpenNLP.

Keywords: Named Entity Recognition · Tokenisation · Sentence Splitting · POS Tagging
Language: English
Licence: Apache License 2.0

Figure 13: List view of catalogue contents

The most important changes implemented in R2 for this page target the display of facets on the left side of the page. First, we have introduced accordions to collapse the facets; this allows us to handle the additional facets that were requested by users, while at the same time making all facets visible at once. A second level ("subfacets") of display is used for the "language" facet (Figure 14), which includes a long list of values; their grouping helps users to have a better view of the values and find what they are looking for more easily.



Language resources & technologies

Service functions

Languages

Official EU languages

+Bulgarian	(41)
+Croatian	(36)
+Czech	(80)
+Danish	(51)
+Dutch	(31)

Show more

Other EU/European languages

+Norwegian Bokmål	(46)
+Russian	(23)
+Icelandic	(16)

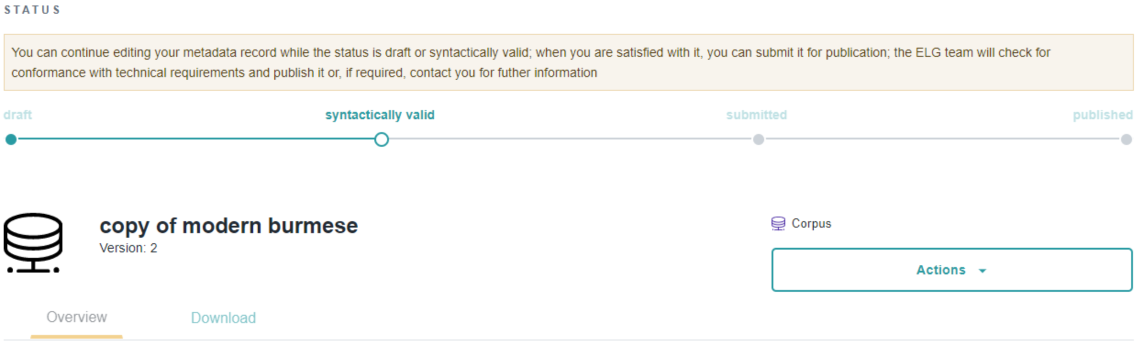
Figure 14: Open facet for language

4.2.2 Landing Page for Items

The layout of the landing page (Figure 15) for items remains the same as described in D3.2, with a few stylistic changes.

In addition, the landing page of items for logged in providers and administrators (Figure 14) is enriched with:

- a dedicated area for "Actions", which allows them to manage the metadata records, as described in detail in Section 6.2.4;
- the publication status of the item.



STATUS

You can continue editing your metadata record while the status is draft or syntactically valid; when you are satisfied with it, you can submit it for publication; the ELG team will check for conformance with technical requirements and publish it or, if required, contact you for futher information

copy of modern burmese

Version: 2

Figure 15: Display of publication status

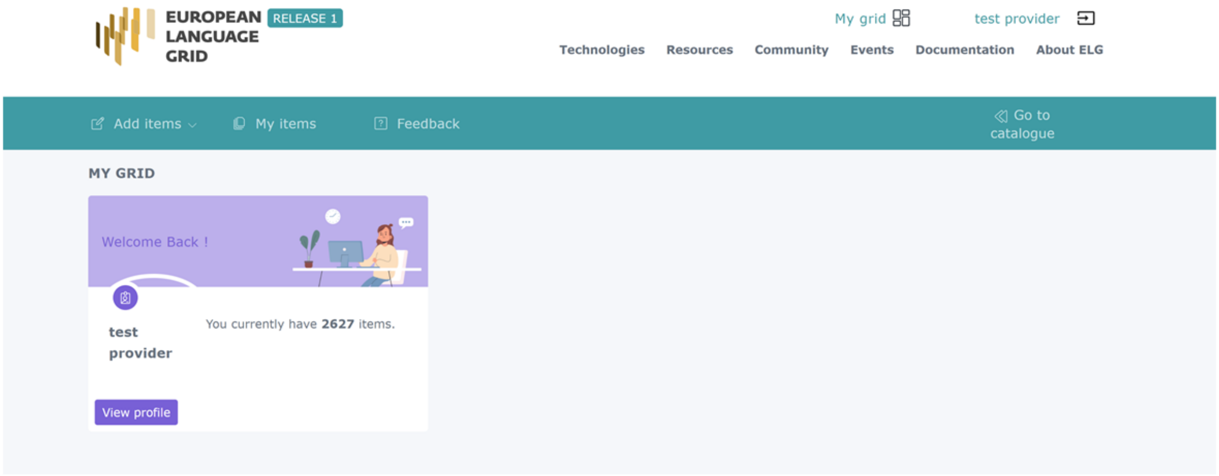
An item (i.e., a metadata record and, optionally, content files uploaded with it), contributed in ELG, goes through a set of statuses before its publication in the catalogue ("ELG publication lifecycle").

The statuses are:

- **new item**: A provider creates an item, by creating a metadata record through the interactive editor or by uploading a metadata file and, optionally, content files;
- **draft**: When using the interactive editor, the provider can save the metadata record, even without filling in all mandatory elements; only compliance as to the data type of the elements is checked (e.g. elements that take a URL must be filled in with the accepted pattern);
- **syntactically valid**: The metadata record complies with the ELG metadata schema and all mandatory elements are filled in. The provider can still continue to edit it until satisfied with the description, and can then submit it for publication;
- **submitted for publication**: The record is no longer editable. Depending on the item type and the source (see Table 1 below), the item is validated at the metadata, technical and legal level. The validation aims to check the consistency of the description and, where required, the technical compliance of the item to the ELG specifications; it doesn’t include any qualitative evaluation. The validation is currently performed by the ELG consortium members. When validators identify a problem, they contact the provider for further information and may ask the provider to edit the metadata; in such cases, the status of the item is changed to “syntactically valid” and the provider is notified to make the appropriate amendments.

4.2.3 Dashboard

To support users in accessing and managing the catalogue items and performing the actions that are allowed depending on their user role, we make use of a dashboard, which in ELG we call a **grid** (Figure 16).



You currently have **2627** items.

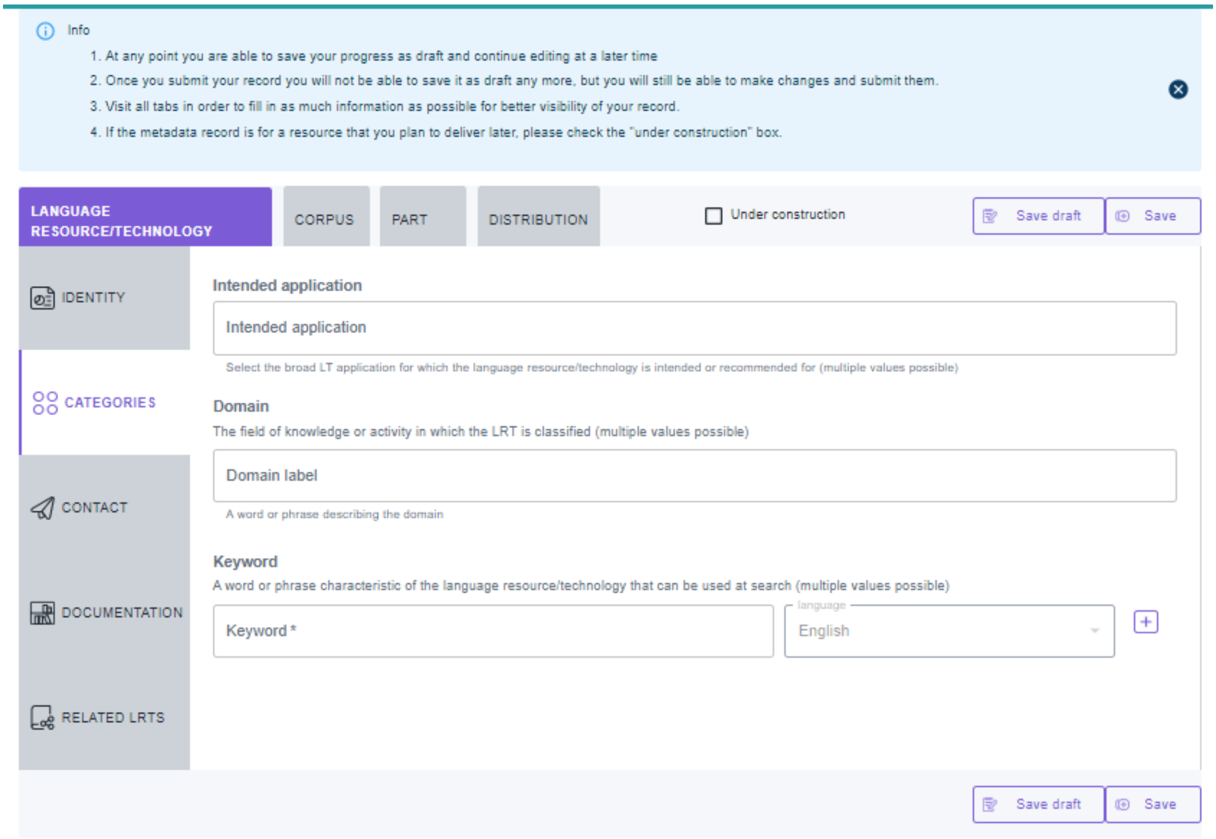
Figure 16: My grid

Although the look and feel is the same across user roles, the actions and menu items differ depending on the user role. In this release, we have focused on the provider's grid.

The grid is the central place through which providers have access to all the items they have created, as well as to the functionalities that will allow them to create new items, i.e., the upload of metadata records that we described in Deliverable D3.2, and the interactive editor that we describe in Section 6.2.4. This is possible through the menu items which are at the top of the page.

4.2.4 Creating Items with the Interactive Editor

For the ELG metadata editor, we have designed and implemented forms that enable users to formally describe their resources in compliance with the ELG metadata schema. Since the schema is rather complex, we attempt to meet the needs of our users by hiding this complexity, and present them with a full-fledged GUI. As a result, we refrain from using long forms that could overwhelm users; instead, we have adopted a more interactive design based on the organisation of the elements along a natural workflow split across multiple steps. The steps are not linear, meaning that users can proceed in any step, revisit, or revise previous steps at any time. For example, Figure 17 shows the form for creating the metadata record for a corpus.



Info

1. At any point you are able to save your progress as draft and continue editing at a later time
2. Once you submit your record you will not be able to save it as draft any more, but you will still be able to make changes and submit them.
3. Visit all tabs in order to fill in as much information as possible for better visibility of your record.
4. If the metadata record is for a resource that you plan to deliver later, please check the "under construction" box.

□ Under construction

Intended application

Intended application

Select the broad LT application for which the language resource/technology is intended or recommended for (multiple values possible)

Domain

The field of knowledge or activity in which the LRT is classified (multiple values possible)

Domain label

A word or phrase describing the domain

Keyword

A word or phrase characteristic of the language resource/technology that can be used at search (multiple values possible)

Keyword *

language: English

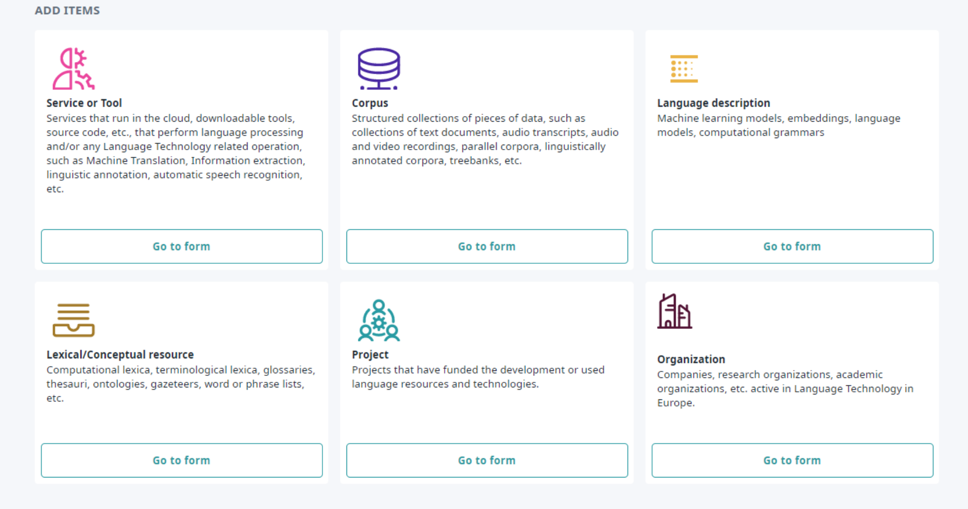
Figure 17: Editor form for corpora

By grouping and organising the elements, we achieve the following goals:

- **Focus**. By breaking up the workflow into multiple steps, users can focus on one topic at a time. They are thus potentially less distracted and overwhelmed by other screen elements competing for their attention. We obviously modularise the application in ways that make sense to users and minimise the need to refer to information contained in other areas.
- **In-context help**. As users have fewer items on the screen, we can clearly explain each one. For this, we supplement buttons and menus with text and examples that illustrate their proper use (utilising placeholders in form elements, tooltips, help examples). Having this information on the main screen is preferable because users dislike navigating to separate help screens.
- **Avoiding crowded screen elements**. A second advantage of having fewer items on each screen is that our layout is less burdened by space constraints. For example, we avoid menu drop downs and long

scrolling lists and instead, we present lists of selectable autocomplete items (where only related items are visible simultaneously). This makes selection faster and less error-prone.

- **Flexible workflow**. In our application, we let users work on different parts of the metadata schema in a sequence that makes sense to them. We also let them refer to supporting data sources in the process. We achieve this by letting users temporarily save their progress (as a "Draft") so they can come back at any time to continue editing the workflow.



ADD ITEMS

Service or Tool
Services that run in the cloud, downloadable tools, source code, etc., that perform language processing and/or any Language Technology related operation, such as Machine Translation, Information extraction, linguistic annotation, automatic speech recognition, etc.

Go to form

Corpus
Structured collections of pieces of data, such as collections of text documents, audio transcripts, audio and video recordings, parallel corpora, linguistically annotated corpora, treebanks, etc.

Go to form

Language description
Machine learning models, embeddings, language models, computational grammars

Go to form

Lexical/Conceptual resource
Computational lexica, terminological lexica, glossaries, thesauri, ontologies, gazeteers, word or phrase lists, etc.

Go to form

Project
Projects that have funded the development or used language resources and technologies.

Go to form

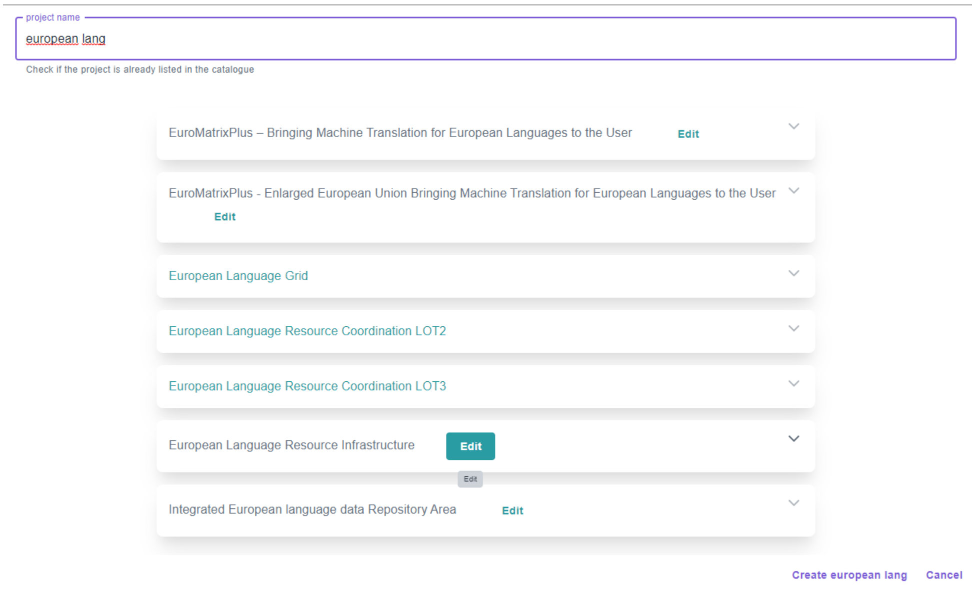
Organization
Companies, research organizations, academic organizations, etc. active in Language Technology in Europe.

Go to form

Figure 18: Selection of item type

Users can create a metadata record by accessing the main editor page and selecting the type of record they are interested in (Figure 18). According to the users' selections they will get forwarded to the corresponding editor page where they can fill the metadata information for the resource they want to register.

Each time a user creates a new record as a first step, they are required to fill in its title. To save users the time and effort required to find out if the record is already created by another user, the front-end application performs a lookup (using the appropriate backend endpoints) and returns matching records (Figure 19).



project name
european lang
Check if the project is already listed in the catalogue

EuroMatrixPlus – Bringing Machine Translation for European Languages to the User Edit

EuroMatrixPlus - Enlarged European Union Bringing Machine Translation for European Languages to the User Edit

European Language Grid

European Language Resource Coordination LOT2

European Language Resource Coordination LOT3

European Language Resource Infrastructure Edit

Edit

Integrated European language data Repository Area Edit

Create european lang Cancel

Figure 19: Search of existing records

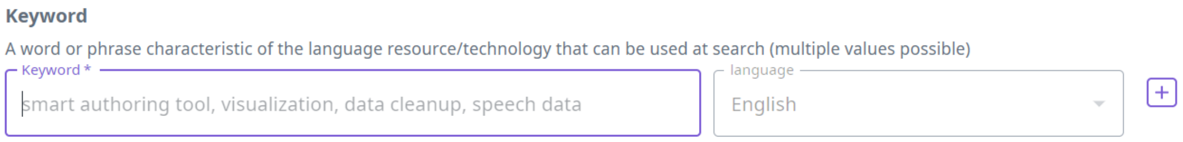
This initial step serves the following purposes:

- Users are informed of existing matching records and their status (i.e., whether it is published or in the process of being edited)
- For records that are published or in the process of being edited by other users, the record is locked (non-editable)
- If the matching record is one that the user has initiated at a previous visit, the “Edit” button is visible and the user can proceed to editing it.

After the initial lookup step, the user will be presented with the form for the respective item type. The editor interface is not meant to only be used by experts in the field of language technology, but our goal is to keep it simple, with straightforward steps, so that we can guide each user in filling out the input elements.

In order to accomplish the aforementioned goals, we make use of several backend endpoints.

More precisely, all the necessary information such as required fields, help texts and labels are sent to the frontend from the catalogue backend as JSON responses. All the required fields are displayed with a star near their label. The placeholders are displayed when the users focus on the elements and the help texts are always visible at the bottom or at the top of each field (Figure 20).



Keyword

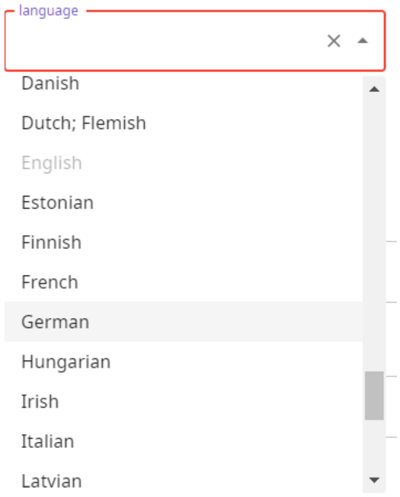
A word or phrase characteristic of the language resource/technology that can be used at search (multiple values possible)

Keyword *: smart authoring tool, visualization, data cleanup, speech data

language: English

Figure 20: View of a multilingual element with help tips and examples

For multilingual elements, the users are required to fill in the value for English, and then they can add the same element for another language by pressing the ⊞ button next to the field and selecting the language from the dropdown list (Figure 21).

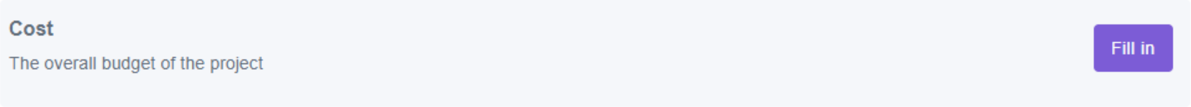


language

Danish
Dutch; Flemish
English
Estonian
Finnish
French
German
Hungarian
Irish
Italian
Latvian

Figure 21: Selecting the language for a multilingual element

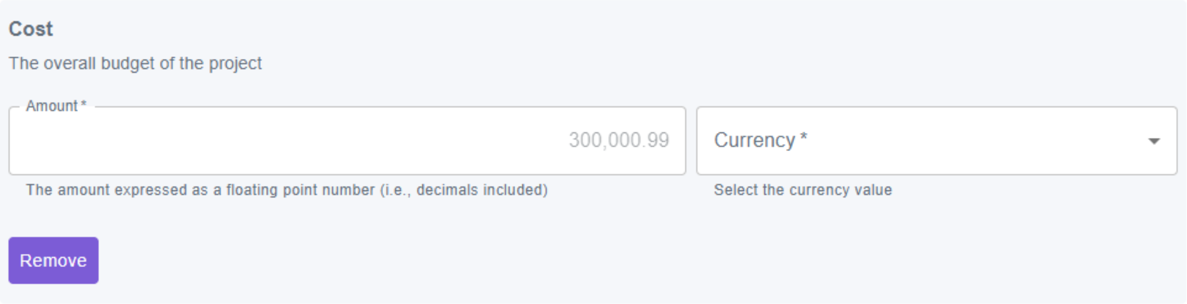
Other complex elements remain hidden (Figure 22) until the user decides to fill them in.


Cost

The overall budget of the project

Figure 22: Hidden elements

After we press the “Fill in” button, we can add the appropriate values (Figure 23). By clicking “remove”, the elements of this group will be cleared.


Cost

The overall budget of the project

Amount * 300,000.99

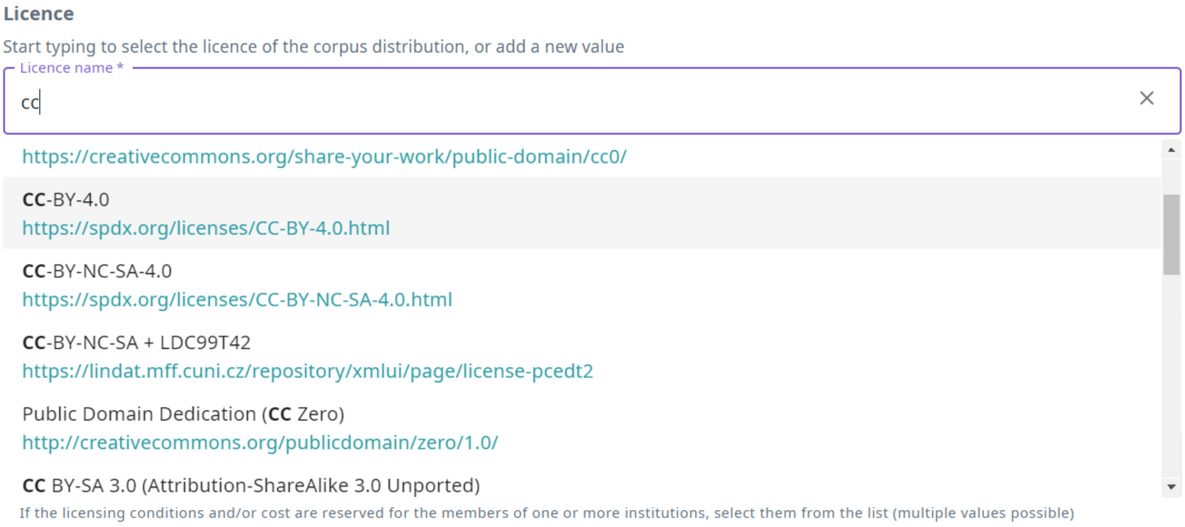
The amount expressed as a floating point number (i.e., decimals included)

Currency *

Select the currency value

Figure 23: Opening a hidden element in order to add metadata

Moreover, complex editor elements make use of backend lookup endpoints. These endpoints respond back with a list of possible suggestions that the user can choose from (Figure 24). Following this approach, users have the option of selecting an existing value instead of filling in all the required fields from scratch, making the whole process easier and faster.


Licence

Start typing to select the licence of the corpus distribution, or add a new value

Licence name *

cc

https://creativecommons.org/share-your-work/public-domain/cc0/

CC-BY-4.0
https://spdx.org/licenses/CC-BY-4.0.html

CC-BY-NC-SA-4.0
https://spdx.org/licenses/CC-BY-NC-SA-4.0.html

CC-BY-NC-SA + LDC99T42
https://lindat.mff.cuni.cz/repository/xmlui/page/license-pcedt2

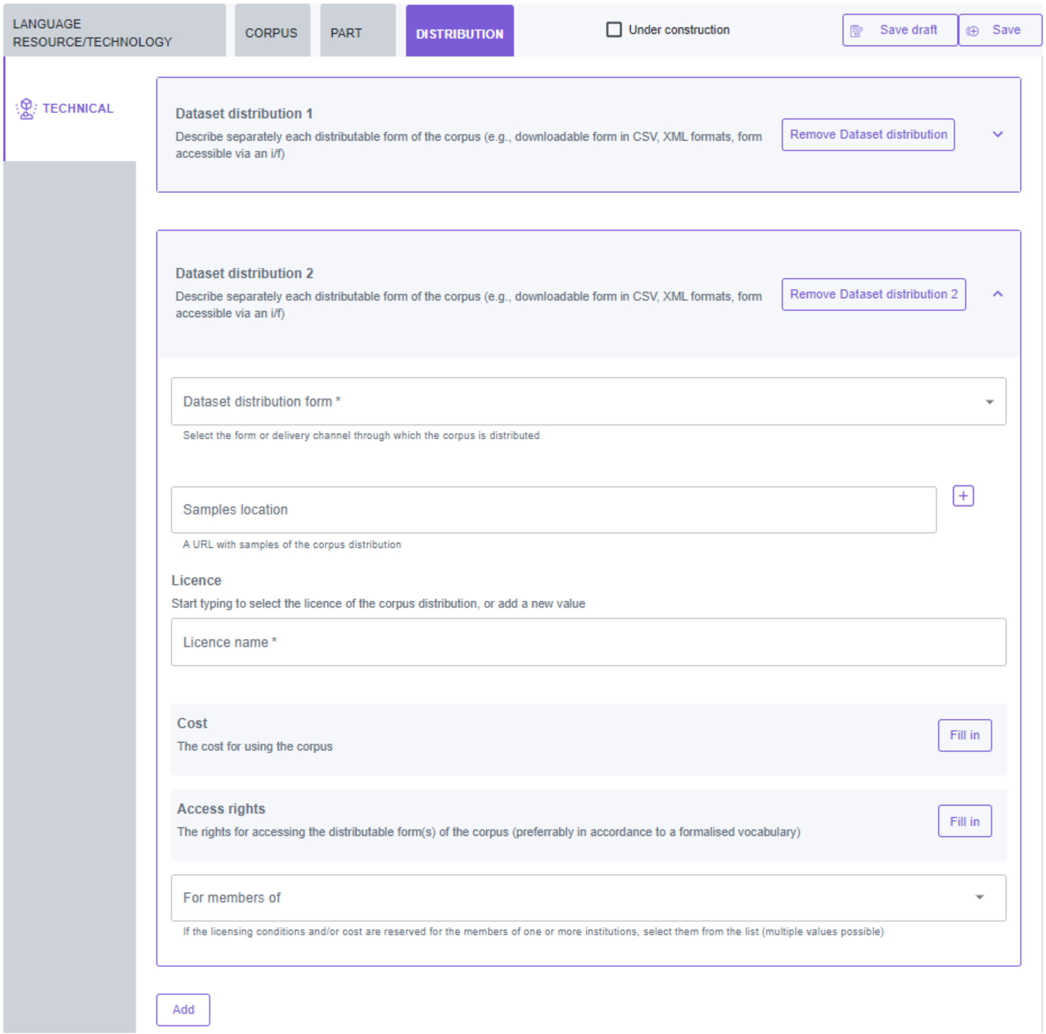
Public Domain Dedication (**CC** Zero)
http://creativecommons.org/publicdomain/zero/1.0/

CC BY-SA 3.0 (Attribution-ShareAlike 3.0 Unported)

If the licensing conditions and/or cost are reserved for the members of one or more institutions, select them from the list (multiple values possible)

Figure 24: Lookup field

Finally, when a group of elements gets rather long, we deploy Expansion panels (or Accordions) (Figure 25), not only to avoid the need for scrolling to the extent possible, but to support creation flow and allow the lightweight editing of an element. For example, for each corpus, users can describe multiple distributions. At any time, users can click the “Add” button and a new distribution group will be created in a new area. Of course, the individual groups can also be removed.



LANGUAGE RESOURCE/TECHNOLOGY
CORPUS
PART
DISTRIBUTION
□ Under construction
Save draft
Save
TECHNICAL

Dataset distribution 1

Describe separately each distributable form of the corpus (e.g., downloadable form in CSV, XML formats, form accessible via an i/f)

Remove Dataset distribution

Dataset distribution 2

Describe separately each distributable form of the corpus (e.g., downloadable form in CSV, XML formats, form accessible via an i/f)

Remove Dataset distribution 2

Dataset distribution form *

Select the form or delivery channel through which the corpus is distributed

Samples location

A URL with samples of the corpus distribution

Licence

Start typing to select the licence of the corpus distribution, or add a new value

Licence name *

Cost

The cost for using the corpus

Access rights

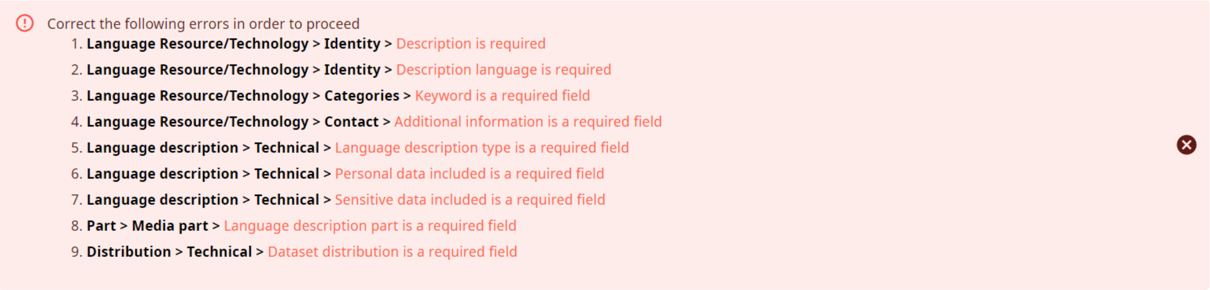
The rights for accessing the distributable form(s) of the corpus (preferrably in accordance to a formalised vocabulary)

For members of

If the licensing conditions and/or cost are reserved for the members of one or more institutions, select them from the list (multiple values possible)

Figure 25: Use of accordion for long elements

Before the form is submitted to the backend, the client performs via validation using the yup validation library (https://github.com/jquense/yup). By doing this, we make sure that all the required fields are filled in and that the user's input meets the constraints imposed by the schema. In the case of errors, appropriate messages are displayed describing the error and the location where this error occurred (Figure 26).



Correct the following errors in order to proceed

1. **Language Resource/Technology** > **Identity** > Description is required
2. **Language Resource/Technology** > **Identity** > Description language is required
3. **Language Resource/Technology** > **Categories** > Keyword is a required field
4. **Language Resource/Technology** > **Contact** > Additional information is a required field
5. **Language description** > **Technical** > Language description type is a required field
6. **Language description** > **Technical** > Personal data included is a required field
7. **Language description** > **Technical** > Sensitive data included is a required field
8. **Part** > **Media part** > Language description part is a required field
9. **Distribution** > **Technical** > Dataset distribution is a required field

Figure 26: Error message

4.2.5 Managing items

Users can manage the metadata records they have created through a dedicated interface, and in accordance with their user rights, perform the following actions:

- edit or update a metadata record
- submit it for publication
- delete a metadata record
- upload content files to a metadata record, and
- replace the content files that have been uploaded.

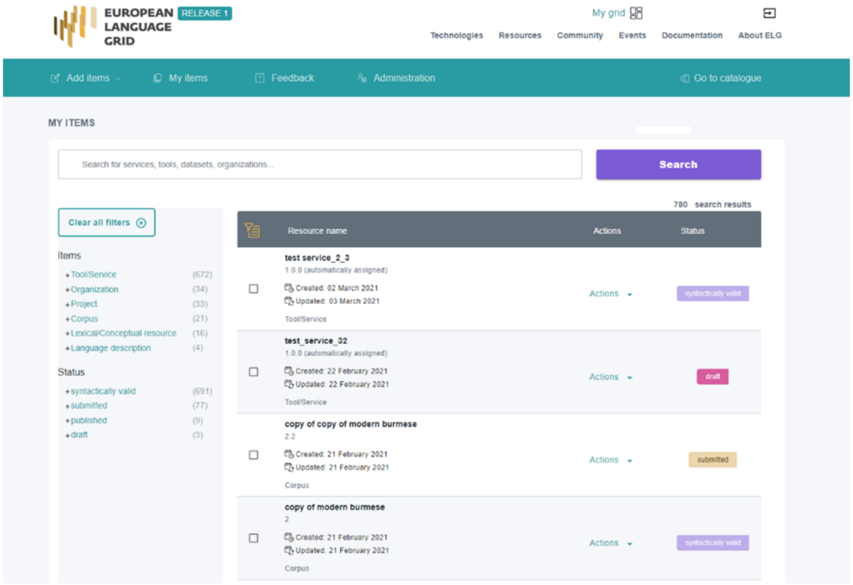
The allowable actions depend on which status of the ELG publication lifecycle the metadata record is in. Table 1 shows the actions that can be performed on an item in the various phases of the publication lifecycle.

Status	Edit	Submit for publication	Delete	Upload content	Replace content
Draft	Y	N	Y	Y	Y
Syntactically valid	Y	Y	Y	Y	Y
Submitted for publications	N	n.a.	N	N	N
Published	N	n.a.	N	N	N

Table 1: Actions allowed on metadata record by publication status

There are currently two access modes where users can manage items: a dedicated page that we call My items, from where they can select one or multiple items, and the view page of each metadata record.

Since the application will be used repeatedly, we provide users with an area for performing common actions, e.g., a user has created a draft metadata record, and would like to continue editing it. Or a user would like to submit the metadata for validation. Each user is provided with one-click access to his/her personal "My items" page (Figure 27), where they can perform single or bulk actions on the metadata created in the past.



780 search results

Items	
+Tool/Service	(672)
+Organization	(34)
+Project	(33)
+Corpus	(21)
+Lexical/Conceptual resource	(16)
+Language description	(4)

Status	
+syntactically valid	(691)
+submitted	(77)
+published	(9)
+draft	(3)

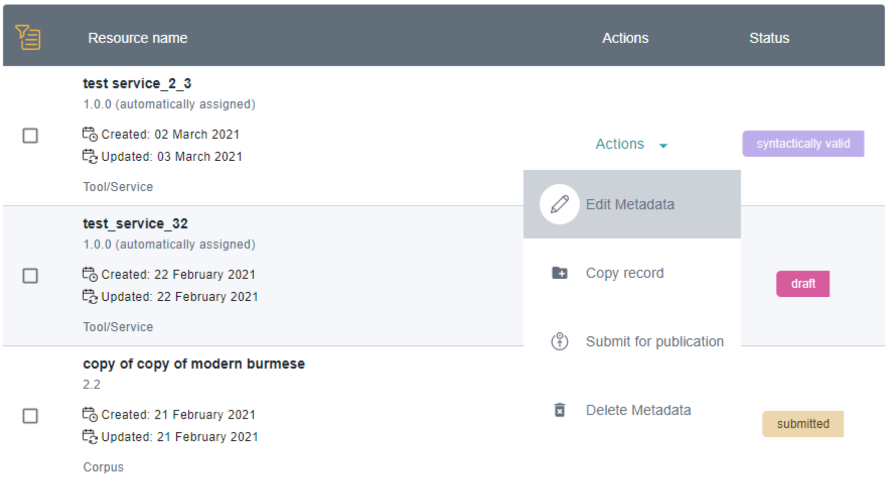
	Resource name	Actions	Status
□	test service_2_3 1.0.0 (automatically assigned) Created: 02 March 2021 Updated: 03 March 2021 Tool/Service	Actions ▾	syntactically valid
□	test_service_32 1.0.0 (automatically assigned) Created: 22 February 2021 Updated: 22 February 2021 Tool/Service	Actions ▾	draft
□	copy of copy of modern burmese 2.2 Created: 21 February 2021 Updated: 21 February 2021 Corpus	Actions ▾	submitted
□	copy of modern burmese 2 Created: 21 February 2021 Updated: 21 February 2021 Corpus	Actions ▾	syntactically valid

Figure 27: My items

The **My items** page (Figure 27) is a focused version of the catalogue filtering records according to the role of each user. This page also implements browse and search functionalities like the main catalogue page, i.e.:

- browse through the items and see all the entries
- sort items by name or date of update
- search for entries using free text search (e.g., search for an LT service, a corpus, a project or an organisation by its name; or search for all entries that mention “machine translation” or “English”),
- filter the catalogue or refine search results by status and type through the facets on the left side bar.

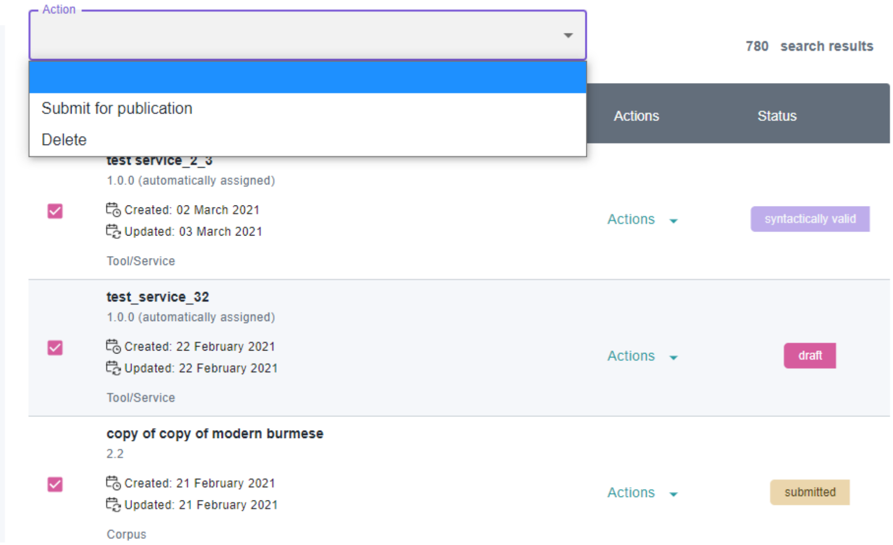
On the “My items” page, providers can perform actions on a selected item. To do this, they simply click on “**Actions**” of each record and they are presented with a choice of options (Figure 28).



	Resource name	Actions	Status
□	test service_2_3 1.0.0 (automatically assigned) Created: 02 March 2021 Updated: 03 March 2021 Tool/Service	Actions	syntactically valid
□	test_service_32 1.0.0 (automatically assigned) Created: 22 February 2021 Updated: 22 February 2021 Tool/Service		draft
□	copy of copy of modern burmese 2.2 Created: 21 February 2021 Updated: 21 February 2021 Corpus		submitted

Figure 28: Actions on my items

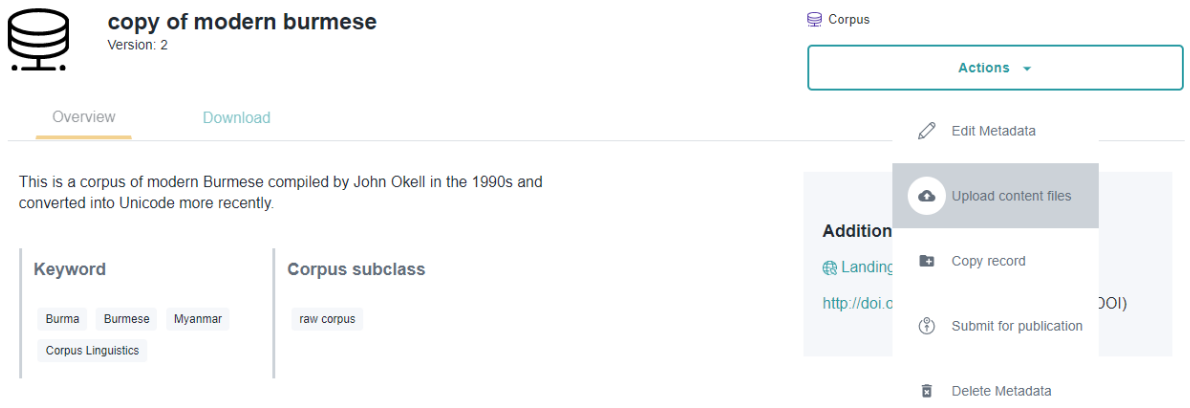
Single actions are straightforward from this interface, but one of the major benefits of this page is that a user can perform bulk actions on selected items. This feature saves a lot of time for providers who have multiple records. Bulk actions are performed by simply clicking the checkboxes on the left of each element, and then a dropdown menu appears at the top that lists all the allowed actions for the selected items (Figure 29).



		Actions	Status
☑	test service_2_3 1.0.0 (automatically assigned) Created: 02 March 2021 Updated: 03 March 2021 Tool/Service	Actions	syntactically valid
☑	test_service_32 1.0.0 (automatically assigned) Created: 22 February 2021 Updated: 22 February 2021 Tool/Service	Actions	draft
☑	copy of copy of modern burmese 2.2 Created: 21 February 2021 Updated: 21 February 2021 Corpus	Actions	submitted

Figure 29: Actions on multiple items

The same actions can also be performed on the landing page of an item (Figure 30).



This is a corpus of modern Burmese compiled by John Okell in the 1990s and converted into Unicode more recently.

Figure 30: Actions area on the landing page of an item

5 Authentication Solution

5.1 Technical Description

Keycloak technologies have not changed much since the initial release (see D3.2). Since Release 1 the following main extensions have been made:

1. Custom themes were added for login, account and emails;
2. We developed a custom module for synchronisation with the Catalogue backend. This is described in Deliverable D2.5;
3. Global roles were introduced, migrating from application-specific ones.

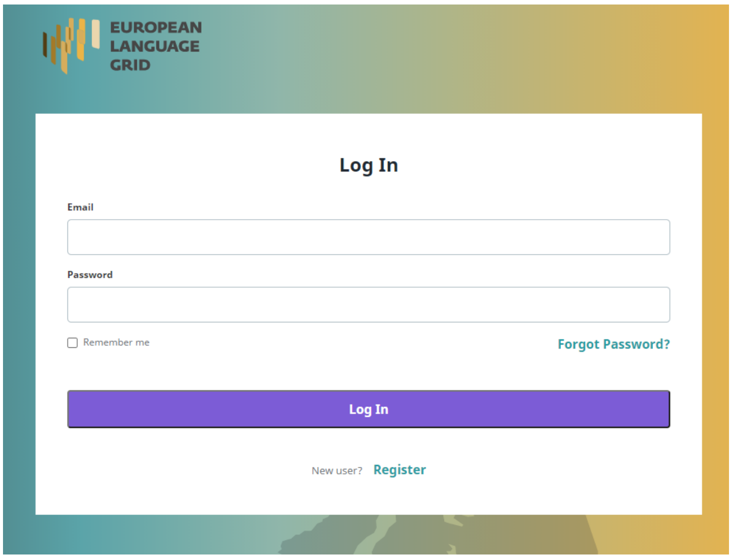
5.2 Keycloak GUI customization

In order to customise end user interface elements provided by Keycloak, it provides custom theme creation. As recommended in the Keycloak documentation, we extended the built-in theme, by adding a new theme and setting this built-in theme as a base. We added three new themes:

1. Login theme – theme for different login screens, password recovery, account linking, account registration, accepting terms, offline login and similar.
2. Account theme – theme for administering existing users' accounts, password change, session lists.
3. Email theme – for adjusting email texts and email designs sent by the Keycloak system.

Custom themes are built and packed into a Docker image and published. When Keycloak initialises, it downloads the image with custom themes and maps to the Keycloak filesystem. Custom themes are set in the top ELG realm level, so all design changes apply to all components of the system – central portal, catalogue, python clients, etc.

Customised login screen (Figure 31): the live environment still allows registration with locally administered accounts only, however, if needed, logins can also be performed with external identity providers, e.g., Google. In the case that the user is not yet registered, they can go to the registration page or reset the password if they have forgotten it.



Log In

Email

Password

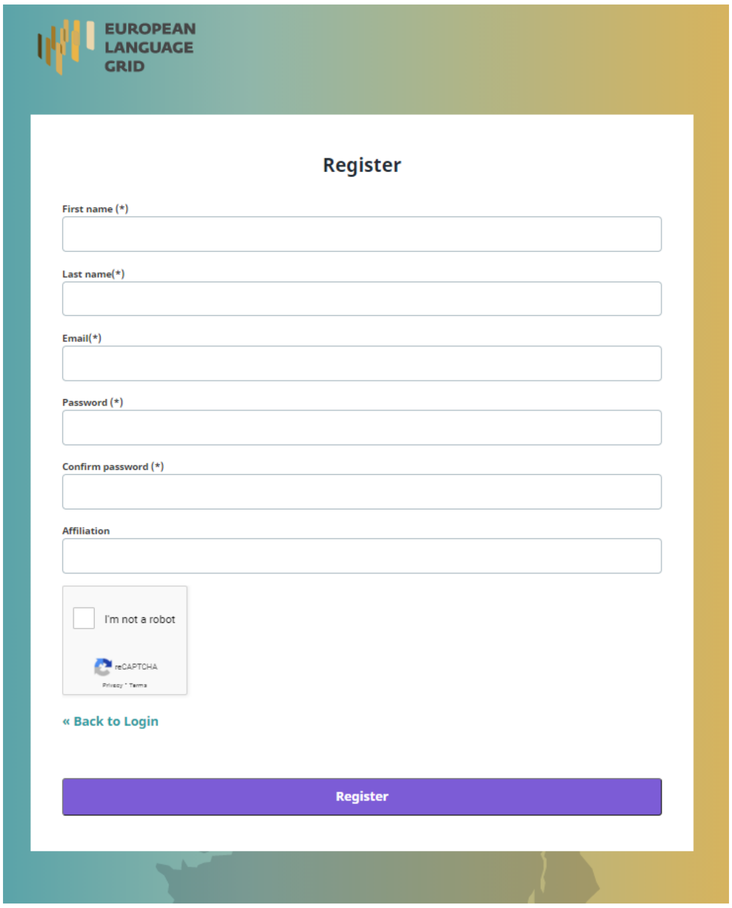
□ Remember me

Forgot Password?

New user? Register

Figure 31: Login form

Registration form (Figure 32): a self-registration form is now available in the live environment. The default user registration data model was extended and new fields added. During registration, not all fields are visible to the end-user; some fields will be visible later in the account editing form.



Register

First name (*)

Last name(*)

Email(*)

Password (*)

Confirm password (*)

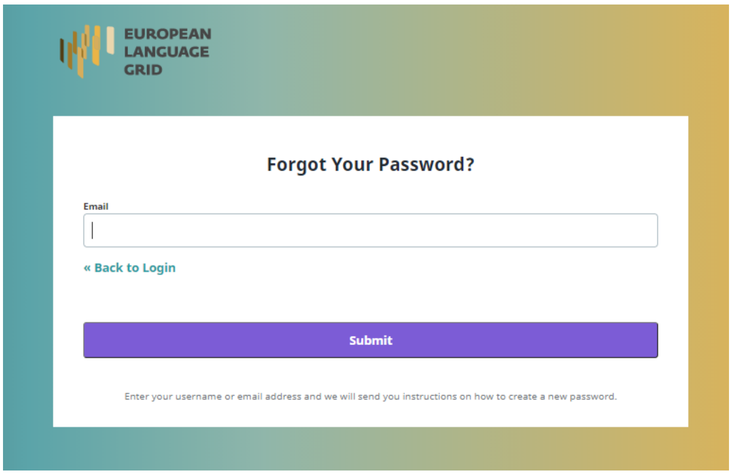
Affiliation

□ I'm not a robot

« Back to Login

Figure 32: New user registration form

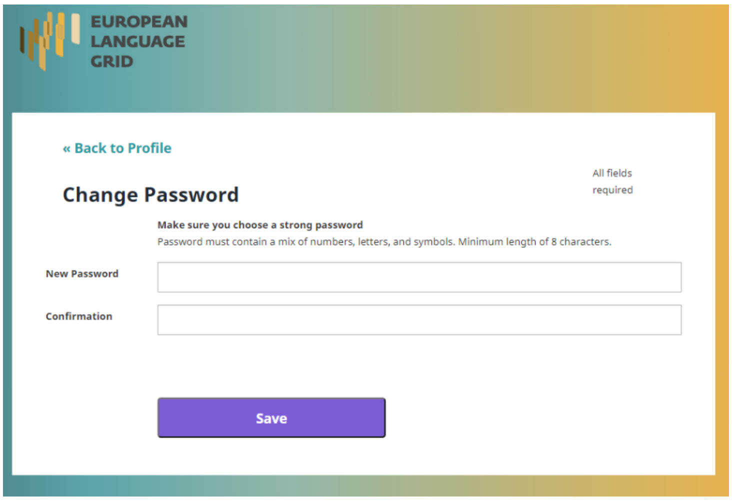
Password recovery form (Figure 33): In the case that users forget their credentials, they can restore their password by providing the e-mail address that was registered during sign-up. Then, the user will receive an e-mail from Keycloak to reset the password. This email message is also adjusted to fit the ELG design guidelines.



Enter your username or email address and we will send you instructions on how to create a new password.

Figure 33: Password recovery form

The Password change form was adjusted according to the design guidelines, more text messages were added.

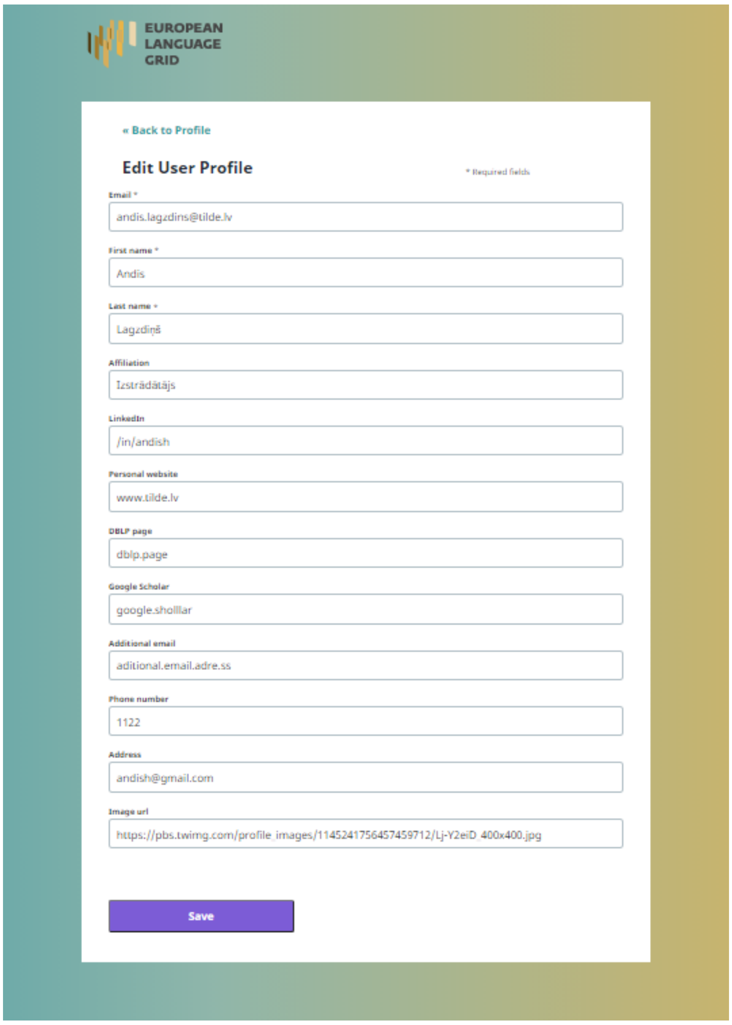


Make sure you choose a strong password

Password must contain a mix of numbers, letters, and symbols. Minimum length of 8 characters.

Figure 34: Styled form for password change

In the R2 version of the ELG platform, a user account editing form is enabled. Here, the user can edit all data categories defined by the extended user data model. The user can also set their profile image here.



« Back to Profile

Edit User Profile

* Required fields

Email *
andis.lagzdins@tilde.lv

First name *
Andis

Last name *
Lagzdiņš

Affiliation
Izstrādātājs

LinkedIn
/in/andish

Personal website
www.tilde.lv

DBLP page
dblp.page

Google Scholar
google.sholllar

Additional email
aditional.email.adre.ss

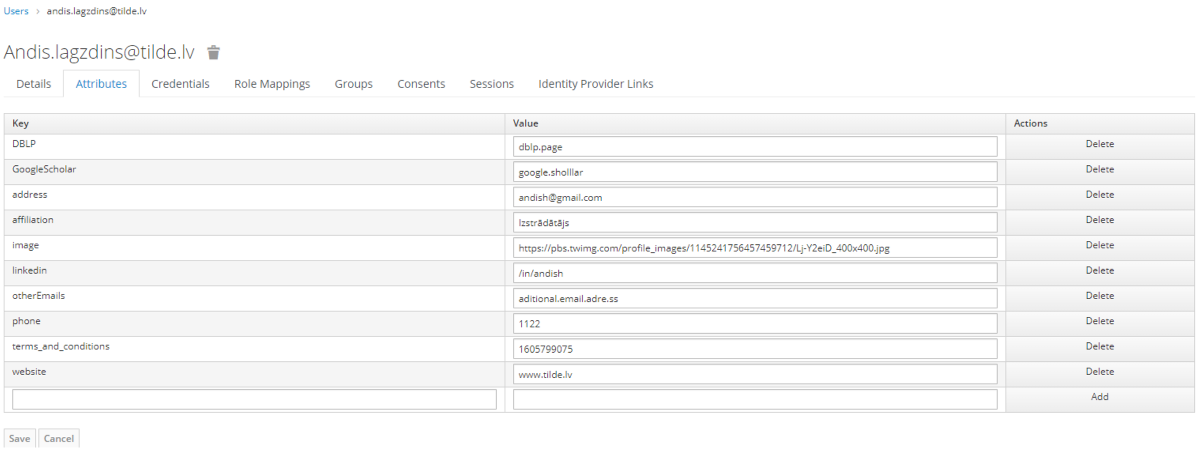
Phone number
1122

Address
andish@gmail.com

Image url
https://pbs.twimg.com/profile_images/1145241756457459712/Lj-Y2eiD_400x400.jpg

Figure 35: User account edit form

Once a user has submitted their data, it is saved in the Keycloak database as attributes. Later, these attributes are mapped to data categories in the authorisation token, and thus data included there are to be used in every platform's component in a decentralised manner.



Andis.lagzdins@tilde.lv

Key	Value	Actions
DBLP	dblp.page	Delete
GoogleScholar	google.sholllar	Delete
address	andish@gmail.com	Delete
affiliation	Izstrādātājs	Delete
image	https://pbs.twimg.com/profile_images/1145241756457459712/Lj-Y2eiD_400x400.jpg	Delete
linkedin	/in/andish	Delete
otherEmails	aditional.email.adre.ss	Delete
phone	1122	Delete
terms_and_conditions	1605799075	Delete
website	www.tilde.lv	Delete
		Add

Figure 36: User submitted attributes in Keycloak Management Console

6 UX implementation

We have carried out various workshops and stakeholders' meetings in order to brainstorm and create a provider's journey. Figure 37 which captures an extract of the provider's journey outlines the first steps and decisions the user takes before becoming a provider, as well as during the resource submission phase. As illustrated in Figure 37, each stage is divided into various categories – ELG components relating to the stage, processes that impact the stage, and questions raised by users while at a certain stage. For instance, in the awareness phase, the potential user hears about the ELG during various dissemination events of the project (an ELG conference, a newsletter, social media channels or other ELG events). All the events are carried out by ELG. Furthermore, during the awareness phase there are many non-ELG related channels, such as industry media, professional media, google search and others that might not be directly related to ELG direct awareness activities. For the purpose of visibility, each category is marked in a different color.

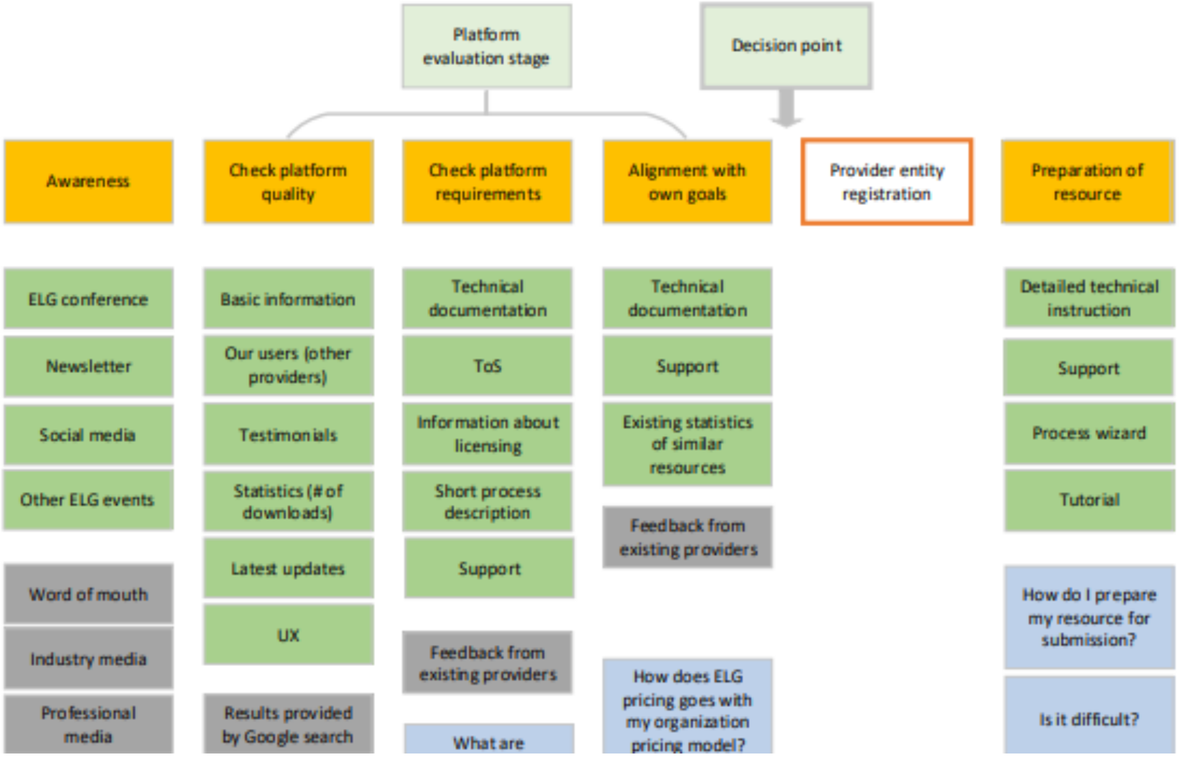

Platform evaluation stage
Decision point
Awareness
Check platform quality
Check platform requirements
Alignment with own goals
Provider entity registration
Preparation of resource
ELG conference
Newsletter
Social media
Other ELG events
Word of mouth
Industry media
Professional media
Basic information
Our users (other providers)
Testimonials
Statistics (# of downloads)
Latest updates
UX
Results provided by Google search
Technical documentation
ToS
Information about licensing
Short process description
Support
Feedback from existing providers
What are
Technical documentation
Support
Existing statistics of similar resources
Feedback from existing providers
How does ELG pricing goes with my organization pricing model?
Detailed technical instruction
Support
Process wizard
Tutorial
How do I prepare my resource for submission?
Is it difficult?

Figure 37: Provider's journey

7 Conclusion

In the deliverable D3.3 we report the progress of development of the ELG platform GUI. We present the overview of the overall front-end architecture, and provide the technical description and implementation details for the solutions designed for the central portal, CMS, the catalogue the UI, the implementation of UX and the authentication solution. In February 2022, the deliverable D3.4, "Platform GUI (final release)" will provide a final version of the current work.