D8.6 SPECIFICATIONS FOR LIAISING EASIER WITH ELG

Revision: v.1.0

Work package	WP 8
Task	Task 8.6
Due date	31/12/2023
Submission date	21/12/2023
Deliverable lead	ATHENA
Version	1.0
Authors	Dimitris Galanis, Penny Labropoulou, Dimitris Gkoumas, Stelios Piperidis (ATHENA), Christian Tismer (NURO)
Reviewers	Uwe Reichel (AUD), Mike Ludemann (NURO)

Abstract	D8.6 reports on the specifications for integrating resources (datasets and tools/services) that have been developed in the framework of the EASIER project into the European Language Grid (ELG, https://live.european-language-grid.eu). Following an overview of the ELG catalogue, the deliverable focuses on the functionalities of ELG that cater for the specificities posed by sign language resources and technologies and a short account of the currently integrated EASIER resources into ELG.
Keywords	Sign language resources, European Language Grid, catalogue of resources, Language Resources, Language Technologies

Grant Agreement No.: 101016982
Call: H2020-ICT-2020-2
Topic: ICT-57-2020
Type of action: RIA

WWW.PROJECT-EASIER.EU

Document Revision History

Version	Date	Description of change	List of contributor(s)
V0.8	29/11/2023	Version with ILSP input	Dimitris Galanis, Penny Labropoulou, Dimitris Gkoumas, Stelios Piperidis (ATHENA)
V1.0	19/12/2023	Version with input from all	Dimitris Galanis, Penny Labropoulou, Dimitris Gkoumas, Stelios Piperidis (ATHENA), Christian Tismer (NURO)

DISCLAIMER

The information, documentation and figures available in this deliverable are written by the "Intelligent Automatic Sign Language Translation" (EASIER) project's consortium under EC grant agreement 101016982 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

COPYRIGHT NOTICE

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:	**R**	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to EASIER project and Commission Services	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

EXECUTIVE SUMMARY

Deliverable D8.6 reports on the current status of the integration of resources produced in the framework of the EASIER project into the European Language Grid (ELG, https://live.european-language-grid.eu/).

The ELG platform provides access to commercial and non-commercial Language Technology (LT) resources (i.e., datasets, lexica, models, etc. as well as LT tools/services) for all European languages. The deposition of resources developed within the EASIER project in the ELG platform will increase their visibility and uptake, while the integration of services following the ELG technical specifications will make them readily available for use through user-friendly interfaces and customized APIs that can enable their integration in end-user applications.

The deliverable gives all information required for integrating Language Resources and Technologies (LRTs) into the ELG platform, with a focus on the features that cater for sign language data resources and technologies and includes a summary on the status of EASIER resources already added in the ELG platform as well as future plans.

Funded by the Horizon 2020
Framework Programme of the European Union

Table of Contents

Funded by the Horizon 2020 Framework Programme of the European Union

LIST OF FIGURES

Funded by the Horizon 2020
Framework Programme of the European Union

ABBREVIATIONS

API	Application Programming Interface
DOI	Digital Object Identifier
ELG	European Language Grid
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
IE	Information Extraction
JSON	JavaScript Object Notation
LR	Language Resource
LRTs	Language Resources and Technologies
LT	Language Technology
OCR	Optical Character Recognition
UI	User Interface
URL	Uniform Resource Locator

1 INTRODUCTION

This document reports on

a. the specifications set for adding resources developed in the framework of the EASIER project into the European Language Grid (ELG) infrastructure.

b. the particular requirements set by sign language resources on the ELG platform and the way these are accommodated, and

c. the current status of the integration of the project resources into the ELG platform.

The ELG platform provides access to commercial and non-commercial Language Technology (LT) resources (i.e., datasets, lexica, models, etc. as well as LT tools/services) for all European languages. The deposition of resources developed within the EASIER project in the ELG platform will increase their visibility and uptake, while the integration of services following the ELG technical specifications will make them readily available for use through user-friendly interfaces and customized APIs that can enable their integration in end-user applications.

Section 2 gives an overview of the ELG platform and metadata model as a background for the activities required for this task. Section 3 focuses on the ELG functionalities that cater for specificities posed by sign language resources and technologies. Section 4 reports on the project resources integrated in ELG and Section 5 concludes with a summary of the current situation and future plans.

Funded by the Horizon 2020 Framework Programme of the European Union

2 Background: overview of the ELG infrastructure

2.1 Introduction

The European Language Grid (ELG) project started in January 2019 and was completed in June 2022. The platform developed in the context of the project [[1]] aims to become the primary hub for Language Technology (LT) in Europe. It is based on a scalable cloud infrastructure and provides access to commercial and non-commercial LT resources (i.e., datasets, lexica, models, etc. as well as LT tools/services) for all European languages. Many of these services are deployed within ELG and can be tested/used via the platform. The ELG functionalities allow the European LT community to upload their resources into the platform in an easy and efficient way and link them to other resources. In addition, the ELG platform offers information for and about the LT domain and activities, such as information on projects that have, for instance, funded LT products, and organizations (companies, universities, research institutes, etc.) involved in LT.

The ELG catalogue[2] is the central point that provides access to all resources and related entities and integrates all the functionalities that are offered to users. The current release of the ELG catalogue includes components for

- uploading, documenting (with the appropriate metadata), storing, and managing, resources;
- browsing and searching the catalogue, as well as viewing detailed pages of resources (with their metadata);
- executing LT services;
- downloading resources stored in ELG in accordance with their licensing terms;
- user management that supports all user roles defined in ELG (e.g., provider, consumer, administrator);
- etc.

2.2 Architecture overview

An overview of the ELG platform architecture is shown in Figure 1. The platform consists of three main layers: the **base infrastructure**, the **platform backend** and the **platform frontend** (user interface).

[1][] Piperidis, S., Labropoulou, P., Galanis, D., Deligiannis, M., Rehm, G. (2023). The European Language Grid Platform: Basic Concepts. In: Rehm, G. (eds) European Language Grid. Cognitive Technologies. Springer, Cham. https://doi.org/10.1007/978-3-031-17258-8_2

[2] https://live.european-language-grid.eu/

Funded by the Horizon 2020 Framework Programme of the European Union

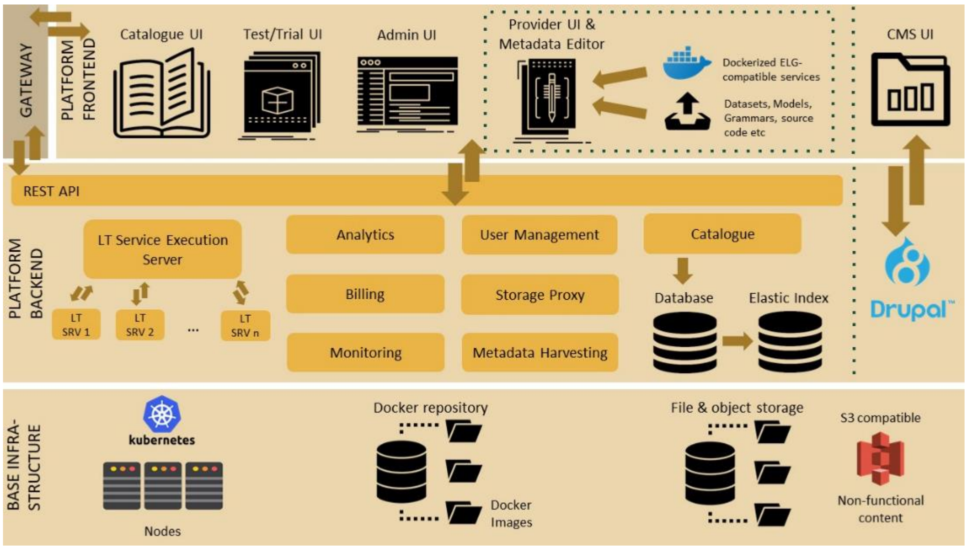


Figure 1: ELG architecture

The base infrastructure is the layer on which all ELG components/applications are deployed as Docker containers[3]; it is based on a Kubernetes[4] cluster. Containers are instantiations of images and can be thought of as lightweight virtual machines. An image of an application contains the application itself and all required dependencies that are required to run it; e.g., the operating system, frameworks, settings, configuration files and libraries, etc. Kubernetes is an open-source container orchestration framework originally developed by Google which has become the de facto standard for managing containerized applications.

The platform backend layer consists of all the components that enable the operation of the ELG platform, i.e., the core components (such as the catalogue database and index), the component for executing LT services (LT Service Execution Server) and the platform support and management components; e.g., the user management module.

The platform frontend layer consists of the platform user interfaces including the catalogue User Interface (UI, i.e., the [5]browse/search page with all the metadata records, the view pages of the metadata records), the ELG interactive metadata editor for registering/documenting resources, the user dashboard, etc. The catalogue UI also embeds the trial UIs used to test LT services. It consumes REST services exposed by the ELG platform backend e.g., catalogue application, LT Service execution server etc. The front-end layer also includes the static pages maintained in the Content Management System, which aim to provide information on the project and the LT domain.

[3] https://www.docker.com/

[4] https://kubernetes.io/

[5][] Labropoulou, P, Gkirtzou, K, Gavriilidou, M., et al. (2020). Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3428–3437, Marseille, France. European Language Resources Association.

Funded by the Horizon 2020 Framework Programme of the European Union

2.3 The ELG Metadata Model

The **ELG metadata model** (or ELG-SHARE[6]) is used for the description of all entities of interest to the ELG target users [2]). It constitutes the backbone of the ELG catalogue, which brings together language processing services and tools, LRs (data sets of different types and media, models, lexica, terminologies, etc.) as well as actors and activities related to LT (Figure 2).

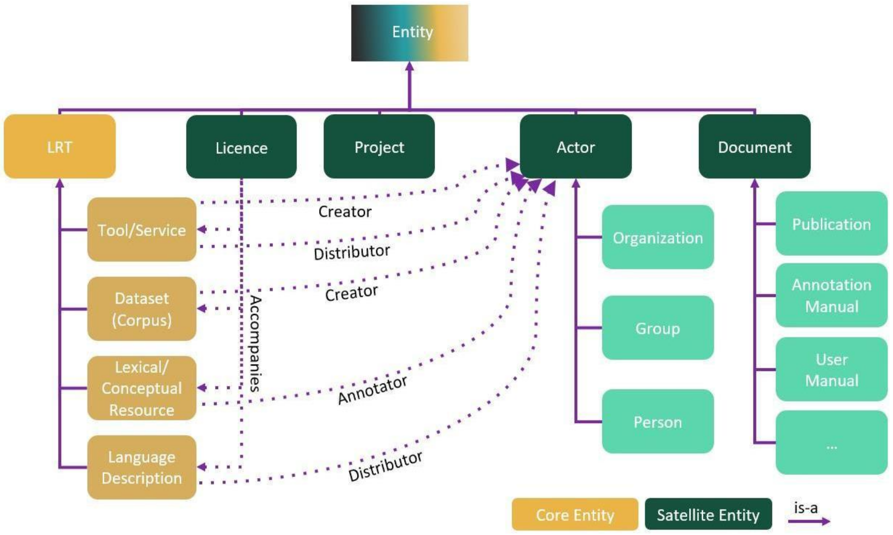


Figure 2: Overview of the ELG-SHARE entities

The model caters for the description of **ELG core entities**, i.e.,

- *LT tools/services*, covering all software that performs language processing and/or any LT-related operation (e.g., basic processing tools, applications, web services etc. that perform annotation, machine translation systems, speech recognizers, etc.).

- *Corpora* (data sets), defined for our purposes as structured collections of pieces of data (textual, audio, video, multimodal/multimedia, etc.), typically of considerable size and selected according to criteria external to the data (e.g., size, type of language, type of producer or expected audience, etc.) to represent as comprehensively as possible the object of study.

- *Lexical/conceptual resources*, i.e., resources (such as terminological glossaries, word lists, semantic lexica, ontologies, gazetteers, 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.).

- *Language descriptions*, i.e., resources aiming to describe a language or some aspect(s) of a language via a systematic documentation of linguistic structures (e.g., computational grammars, statistical and machine learning-computed language models).

The ELG model also provides for **satellite entities** involved in the production and usage of LTs/LRs and, in general, LT activities, i.e., actors (*organizations*, *groups*, *persons*), *documents* (e.g., user manuals, publications, etc.), *projects* and *licences/terms of use*.

[6] https://european-language-grid.readthedocs.io/en/stable/all/A2_Metadata/Metadata.html

Funded by the Horizon 2020 Framework Programme of the European Union

The model includes a large number of metadata elements grouped along three key concepts: *resource type*, *media type* and *distribution*. The *resource type* element distinguishes LRTs in the four classes presented above. *Media type* refers to the form/physical medium of a data resource (or of its parts, in the case of multimodal resources), i.e., text, audio, image, video and numerical text (used for biometrical, geospatial and other numerical data). Finally, *distribution*, following the DCAT vocabulary[7], refers to the physical form of the resource that can be distributed and deployed by consumers; for instance, software resources may be distributed as web services, executable files or source code files, while data resources as PDF, CSV or plain text files or through a user interface. Administrative and descriptive metadata are mostly common to all LRTs, while technical metadata differ across resource and media types as well as distributions. Figure 3 provides an excerpt of the metadata model with the elements required for tools/services.

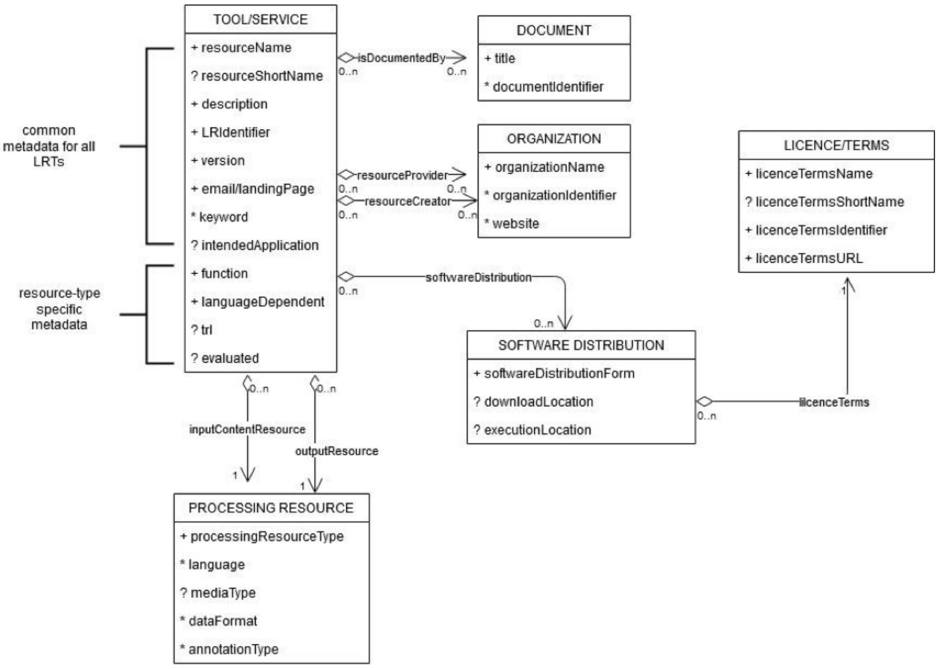


FIGURE 3: EXCERPT OF THE ELG METADATA MODEL (FOCUSING UPON TOOLS/SERVICES)

The abundance of information in the model makes the task of creating metadata records quite tedious. To ensure flexibility and uptake, we distinguish *mandatory*, *recommended* and *optional* metadata elements. Minimal metadata records with only the mandatory and strictly recommended elements are, thus, possible. The criteria used in ELG to determine the optionality status of elements include: required for discovery, especially features considered of high interest to consumers; considered indispensable for accessing the resources and, in the case of ELG-compatible services, ensuring proper deployment through the platform; supporting usage of resources; deemed valuable for experiments and projects and essential for achieving interoperability with existing metadata models used in the wider LT and neighbouring communities.

The minimal version of the ELG model includes the following metadata categories of information:

[7] https://www.w3.org/TR/vocab-dcat-2/

Funded by the Horizon 2020 Framework Programme of the European Union

- for all types of resources: resource name; identifier; a short description of the contents; version identifier; a contact point (email or landing page); keywords; licence, access location and distribution form for each distributable form of the resource;

- in addition, specifically for tools/services: service/application type (i.e., function/task they perform); specifications for the input resource that a tool can process with regard to language(s), media type and formats; where applicable, information on the output resource, again for language(s), media type and format(s), as well as annotation/extraction types (e.g., lemmas, named entities, sentiment tags, etc.); in addition, for ELG-compatible services, docker image location and execution endpoint;

- in addition, specifically for all data LRs (i.e., corpora, lexical/conceptual resources, models, etc.): language of contents; size and data format per distribution; subtype (e.g., raw/annotated corpus, or ontology, lexicon, term list, etc. for lexical/conceptual resources); whether they contain personal or sensitive data and, if yes, whether they are anonymized; for models, the function/task they perform;

- for organisations and projects: only names are mandatory.

2.4 LT Services

2.4.1 LT Service integration

One of the main goals of ELG is to provide a wide range of LT services ready-to-be deployed and integrated in end user applications through the platform. However, LT developers use a variety of frameworks, programming languages and exchange formats. This makes the deployment and integration of LT services (to the ELG platform) a tedious task. To solve these issues ELG was based on (a) **containerisation** for packaging each LT service in a bundle that can be easily deployed and managed, and (b) **Generic LT processing HTTP APIs** and exchange formats.

2.4.1.1 Containerisation

In order to be integrated to the ELG platform, an LT service must be containerised and the respective image must be uploaded into a Docker Registry, such as the GitLab[8], DockerHub[9] or Azure container registry[10] and made accessible to anyone. Docker images that are private (i.e., access is limited only to its owners) can also be deployed to ELG, by configuring the Kubernetes cluster with the respective access tokens. As depicted in Figure 4, three different options for the provision of LT tools at ELG's Kubernetes are supported:

- **LT tools packaged in one image:** One Docker image is created that contains the application that exposes the ELG-compatible endpoint.

- **LT tools running remotely outside the ELG infrastructure**: For these tools a proxy image is created that exposes one (or more) ELG-compatible endpoints; the container

[8] https://about.gitlab.com/

[9] https://hub.docker.com/

[10] https://azure.microsoft.com/en-us/products/container-registry

communicates with the actual LT service that runs externally, i.e., outside the ELG infrastructure.

- **LT tools requiring an adapter**: For tools that already offer an application that exposes a non ELG-compatible endpoint (HTTP-based or other), a second adapter image should be created that exposes an ELG compatible endpoint and that acts as proxy to the container that hosts the actual LT tool.

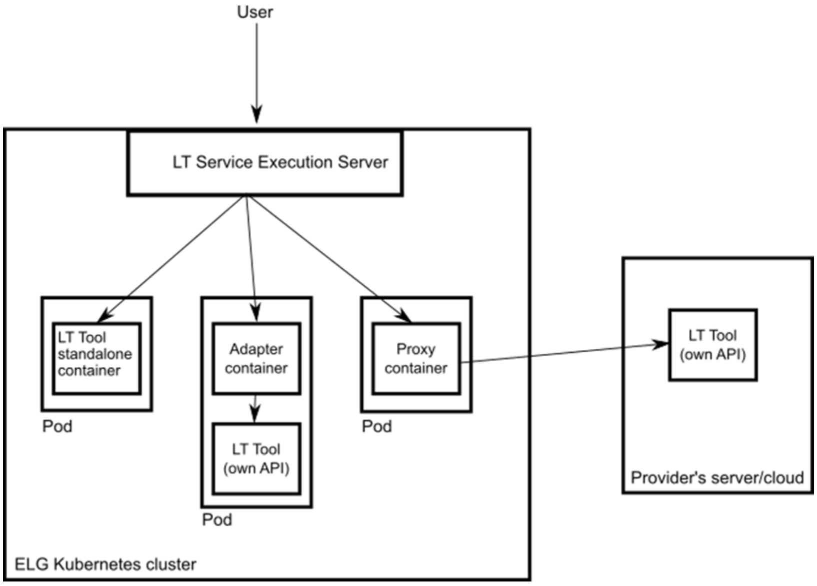


FIGURE 4: INTEGRATION OF LT SERVICES IN ELG

2.4.1.2 Generic LT processing HTTP APIs

The application/service that is packaged (in a container) should expose an ELG-compatible HTTP endpoint that makes its integration with the LT Service Execution Server easy. The server acts as a gateway for accessing/using all the services. Seven broad classes of HTTP APIs were created, each class supports a separate set of LT services. The classes are the following:

- **Information Extraction (IE) & text analysis** services that take text input and produce standoff annotations over that text.

- **Text-to-Text** services, most notably Machine Translation, but also summarisation, anonymisation, etc. They take text input and return new text that is derived from the input.

- **Text classification** services that take text input and classify it somehow (e.g., language identification, “fake news” detection, etc.)

- **Speech recognition services** which accept audio and return a text transcription.

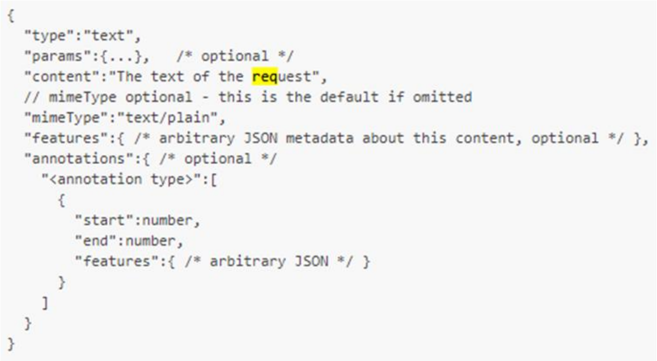
- **Audio annotation services** that take audio and return standoff annotations over certain time segments of the audio stream.

Funded by the Horizon 2020 Framework Programme of the European Union

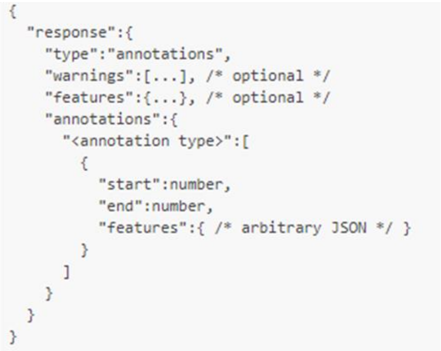
- **Text-to-speech services** take text and return audio.

- **Image optical character recognition (OCR)** services take image data and return text extracted from the image.

For instance, an IE tool/service consumes/produces JSON messages in the following format:



```
{
  "type":"text",
  "params":{...},   /* optional */
  "content":"The text of the request",
  // mimeType optional - this is the default if omitted
  "mimeType":"text/plain",
  "features":{ /* arbitrary JSON metadata about this content, optional */ },
  "annotations":{ /* optional */
    "<annotation type>":[
      {
        "start":number,
        "end":number,
        "features":{ /* arbitrary JSON */ }
      }
    ]
  }
}
```



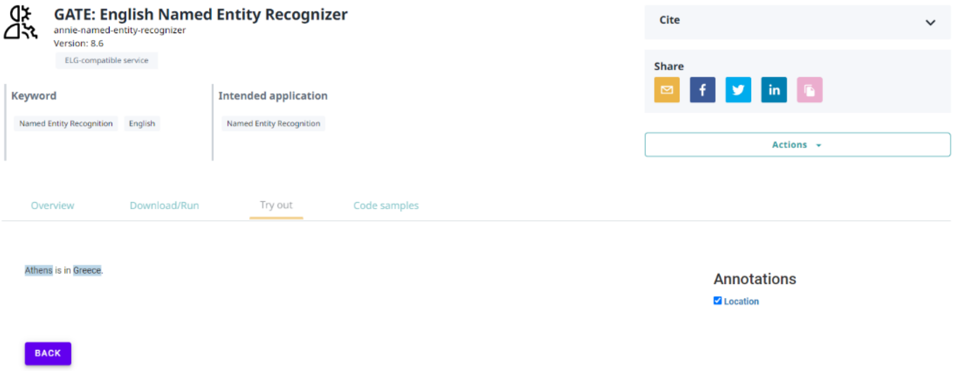
```
{
  "response":{
    "type":"annotations",
    "warnings":[...], /* optional */
    "features":{...}, /* optional */
    "annotations":{
      "<annotation type>":[
        {
          "start":number,
          "end":number,
          "features":{ /* arbitrary JSON */ }
        }
      ]
    }
  }
}
```

Similar JSON-based messages have been defined for the other classes too.

2.4.2 Using services

After being integrated an LT service can be used by any ELG registered user using:

- The respective “Try out” tab, a specially designed GUI where users can provide a sample input and see the results output by the service. Depending on the service class, they can type in or paste some text, upload or record audio, and get the results rendered in a task-specific viewer. For instance, Figure 5 below shows the results received when a user provided a sample text and called an IE tool, a Named Entity Recognizer.



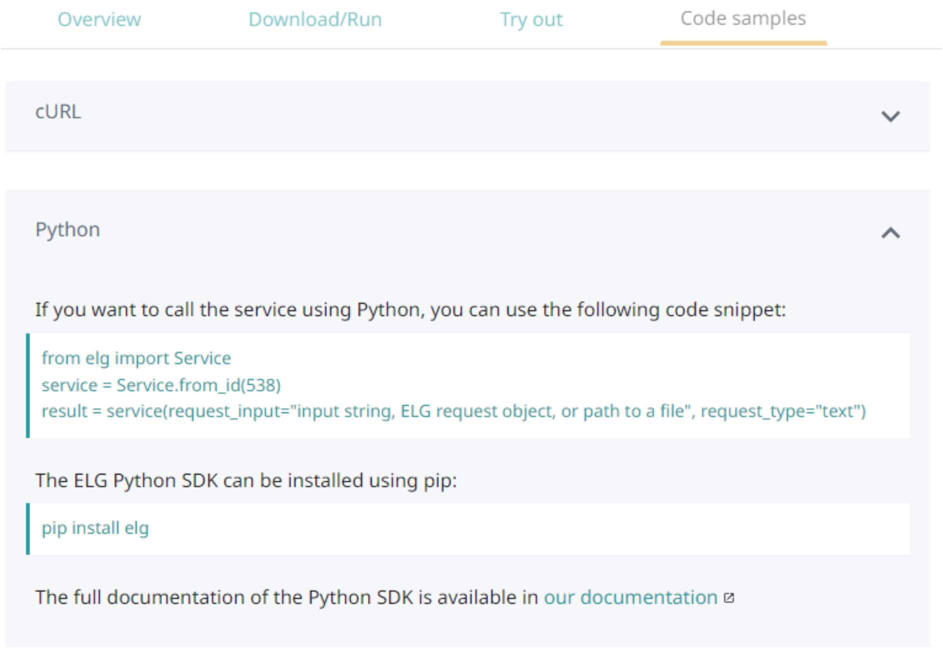
GATE: English Named Entity Recognizer
annie-named-entity-recognizer
Version: 8.6

Athens is in Greece.

FIGURE 5: "TRY OUT" TAB FOR ELG-COMPATIBLE SERVICE

- The programmatic interfaces that are available: The “Code samples” tab (Figure 6) provides for each service examples on how to call a service via Python or curl.

Funded by the Horizon 2020 Framework Programme of the European Union



If you want to call the service using Python, you can use the following code snippet:

```
from elg import Service
service = Service.from_id(538)
result = service(request_input="input string, ELG request object, or path to a file", request_type="text")
```

The ELG Python SDK can be installed using pip:

```
pip install elg
```

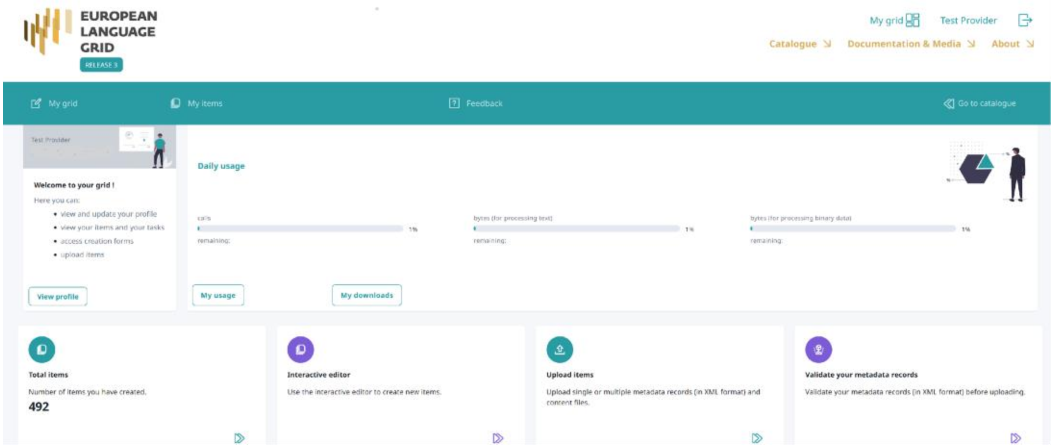
The full documentation of the Python SDK is available in our documentation

FIGURE 6: "CODE SAMPLES" TAB FOR AN ELG-COMPATIBLE SERVICE

2.5 REGISTERING ENTITIES INTO ELG

2.5.1 How to register entities in ELG

To contribute any type of entity in ELG, individuals[11] must register and obtain the "provider" role through the "sign in/sign up" icon at the top right side of the website (https://live.european-language-grid.eu/). Once logged in, they have access to their personal "grid" (Figure 7), which is the point through which they can perform and manage all their interactions with the ELG platform, in both their roles as consumers and providers.



Welcome to your grid !

Here you can:

- view and update your profile
- view your items and your tasks
- access creation forms
- upload items

Total items

Number of items you have created.

492

Interactive editor

Use the interactive editor to create new items.

Upload items

Upload single or multiple metadata records (in XML format) and content files.

Validate your metadata records

Validate your metadata records (in XML format) before uploading.

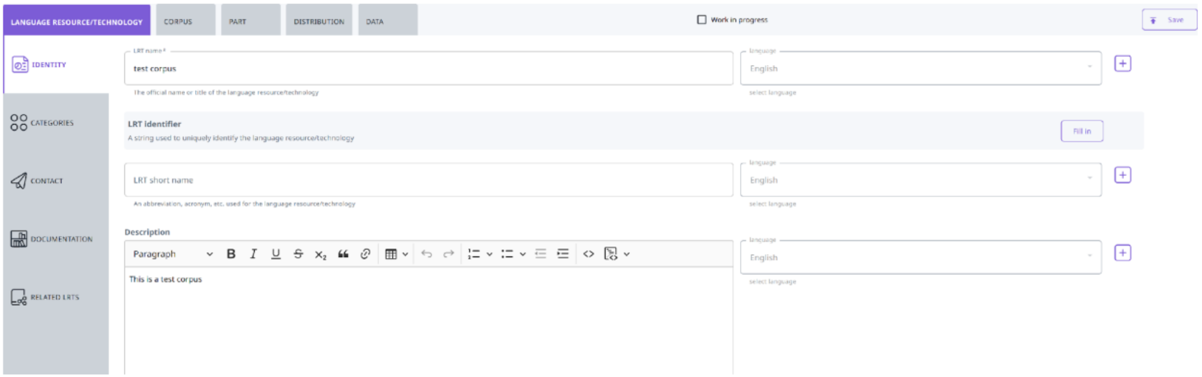
FIGURE 7: PROVIDER'S GRID

[11] ELG employs also mechanisms for importing metadata records from other repositories using standard harvesting protocols (e.g., OAI-PMH) and customized APIs; however, this functionality cannot be used in the case of the EASIER project and, thus, not described here.

Funded by the Horizon 2020
Framework Programme of the European Union

The interactive editor includes all mandatory and recommended elements (see Section 2.3) of the ELG metadata model organized in horizontal and vertical tabs, as depicted in Figure 8. Users can fill in the elements in any order they wish, navigating through the tabs, save the metadata record and continue editing it at later sessions. Upon saving of the record, syntactic validation is performed and the user is notified of any errors (e.g., absence of values for mandatory elements).

In addition to the metadata description, users can also upload files with the actual contents of their resources (e.g., the set of data files that constitute the corpus they are describing); in such case, the resource is stored in the ELG cloud infrastructure and consumers will be able to download the resource directly from it. Otherwise, providers must add in the appropriate metadata element a URL where the resource can be accessed from; in this case, ELG consumers will be re-directed to this remote location.



□ Work in progress

LRT name*: test corpus
The official name or title of the language resource/technology

language: English
select language

LRT identifier
A string used to uniquely identify the language resource/technology

LRT short name
An abbreviation, acronym, etc. used for the language resource/technology

language: English
select language

Description
This is a test corpus

language: English
select language

FIGURE 8: ELG INTERACTIVE EDITOR

When providers have filled in all the information they wish, they can “submit for publication” the metadata record. In compliance with the publication lifecycle for ELG records, the ELG technical team will be notified and proceed to the validation of the record, by checking that the information is as it should. For instance, ensuring that a proper licence has been assigned to the resource, or that the description adequately describes the resource, etc. When confirmed, the metadata record appears in the ELG public catalogue and the providers are notified by e-mail about the publication. In addition, resources with data stored in ELG and ELG-compatible resources are assigned a Digital Object Identifier (DOI) making the reference to and discovery of the resource easier.

Detailed information on the registration process with examples and instructions for each type of entity is provided at the ELG technical documentation in the chapter for providers: https://european-language-grid.readthedocs.io/en/stable/all/3_Contributing/Contributing.html.

2.5.2 Additional requirements for registering an ELG-compatible service

A similar process as above is followed when an ELG-compatible service is registered to the platform. However, in the case of services, one or more Docker images should also be provided. An ELG technical validator must deploy these images, check whether the HTTP endpoints/messages meet the ELG specifications and define which “Try out” UIs should be used. The technical validator communicates with the provider, if needed, and the required adjustments are made to the given images and packaged LT applications. Again, as above,

Funded by the Horizon 2020
Framework Programme of the European Union

when metadata and the Docker images are given the green light by the respective validators, the service is published in the central catalogue.

Funded by the Horizon 2020
Framework Programme of the European Union

3 ACCOMMODATING EASIER SPECIFICITIES IN ELG

With regards to the description of resources that contain or support the processing of sign language data, the ELG metadata schema includes a set of attributes that can accommodate their specificities.

The typical sign language resources can be described as multilingual multimodal resources, given that they usually contain data both in sign and spoken languages, in the form of videos (for the sign language contents) and texts (for the spoken language counterpart)[12]. Therefore, it is recommended that they are described with the two media parts (video and text) separately, allowing for a better representation of each part. For each of these parts, the appropriate 'language' and 'data format' values must be assigned. The 'language' element accepts values following the BCP47 recommendation[13], which prescribes language tags consisting of the ISO 639 codes[14] for the "language" subtag, as well as "region", "script" and "variant" subtags. The ISO 639 standard already caters for sign languages. The separate description of media parts enables the selection of video formats for the video parts and text formats for the text parts. Additional optional metadata elements can be used to describe the creation process of the resource, the domain of the contents (e.g., tourism, health, etc.), links to tools/services used for their creation and processing, technical details on the video recording setting, etc.

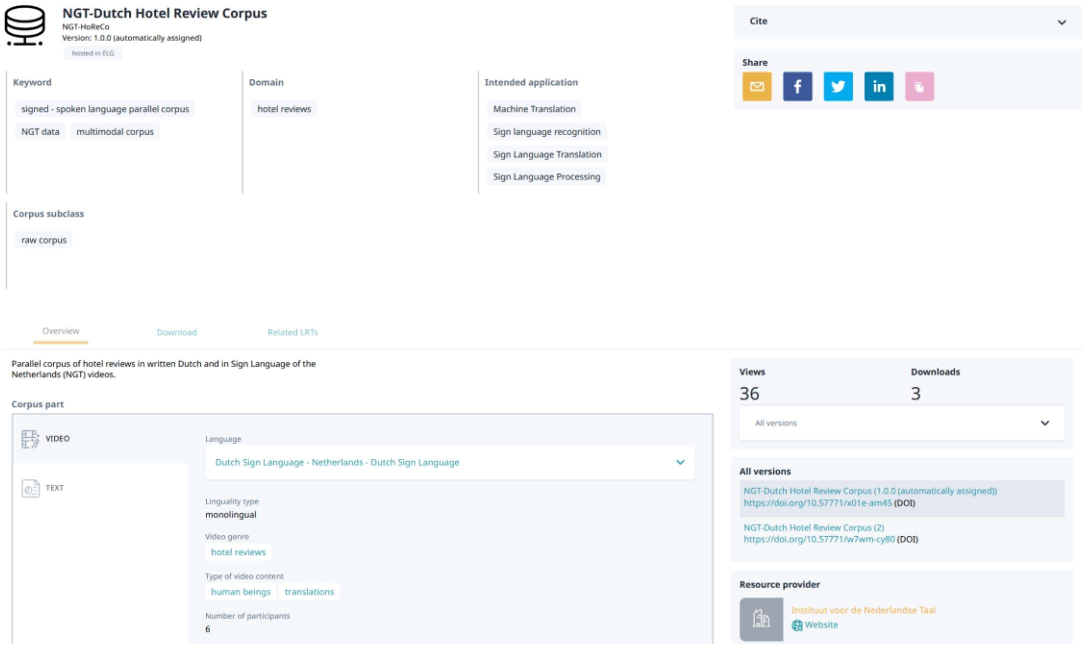
An important element that must be filled in for sign language data is the inclusion of personal and sensitive data, especially in the case of data that show human signers rather than avatars. In such cases, the consent of human signers for the publication of the data is of utmost importance.

[12] Of course, audio parts can also be contained in sign language resources and described on their own.

[13] https://www.rfc-editor.org/info/bcp47

[14] https://www.iso.org/iso-639-language-code

Funded by the Horizon 2020 Framework Programme of the European Union



NGT-Dutch Hotel Review Corpus
NGT-HoReCo
Version: 1.0.0 (automatically assigned)
hosted in ELG

Keyword: signed - spoken language parallel corpus; NGT data; multimodal corpus

Domain: hotel reviews

Intended application: Machine Translation; Sign language recognition; Sign Language Translation; Sign Language Processing

Corpus subclass: raw corpus

Parallel corpus of hotel reviews in written Dutch and in Sign Language of the Netherlands (NGT) videos.

Corpus part: VIDEO; TEXT

Language: Dutch Sign Language - Netherlands - Dutch Sign Language

Linguality type: monolingual

Video genre: hotel reviews

Type of video content: human beings; translations

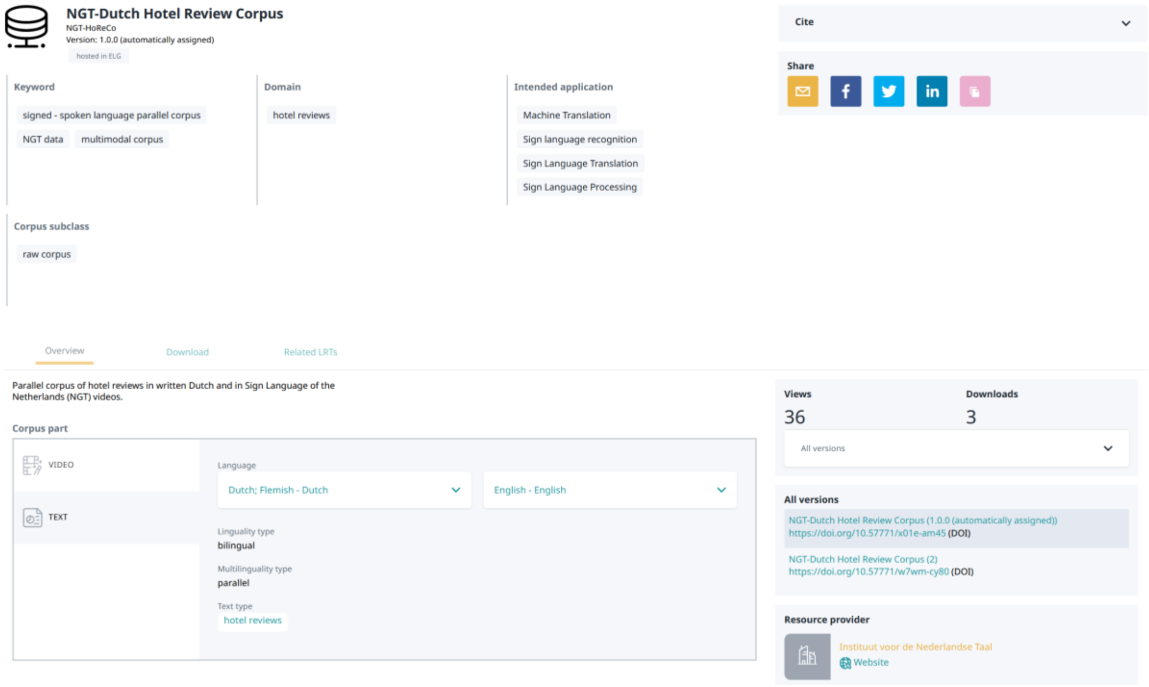
Number of participants: 6

Views: 36
Downloads: 3

All versions
NGT-Dutch Hotel Review Corpus (1.0.0 (automatically assigned)) https://doi.org/10.57771/x01e-am45 (DOI)
NGT-Dutch Hotel Review Corpus (2) https://doi.org/10.57771/w7wm-cy80 (DOI)

Resource provider
Instituut voor de Nederlandse Taal
Website

FIGURE 9: EXCERPT OF VIEW PAGE FOR A SIGN LANGUAGE CORPUS - VIDEO PART



NGT-Dutch Hotel Review Corpus
NGT-HoReCo
Version: 1.0.0 (automatically assigned)
hosted in ELG

Keyword: signed - spoken language parallel corpus; NGT data; multimodal corpus

Domain: hotel reviews

Intended application: Machine Translation; Sign language recognition; Sign Language Translation; Sign Language Processing

Corpus subclass: raw corpus

Parallel corpus of hotel reviews in written Dutch and in Sign Language of the Netherlands (NGT) videos.

Corpus part: VIDEO; TEXT

Language: Dutch; Flemish - Dutch; English - English

Linguality type: bilingual

Multilinguality type: parallel

Text type: hotel reviews

Views: 36
Downloads: 3

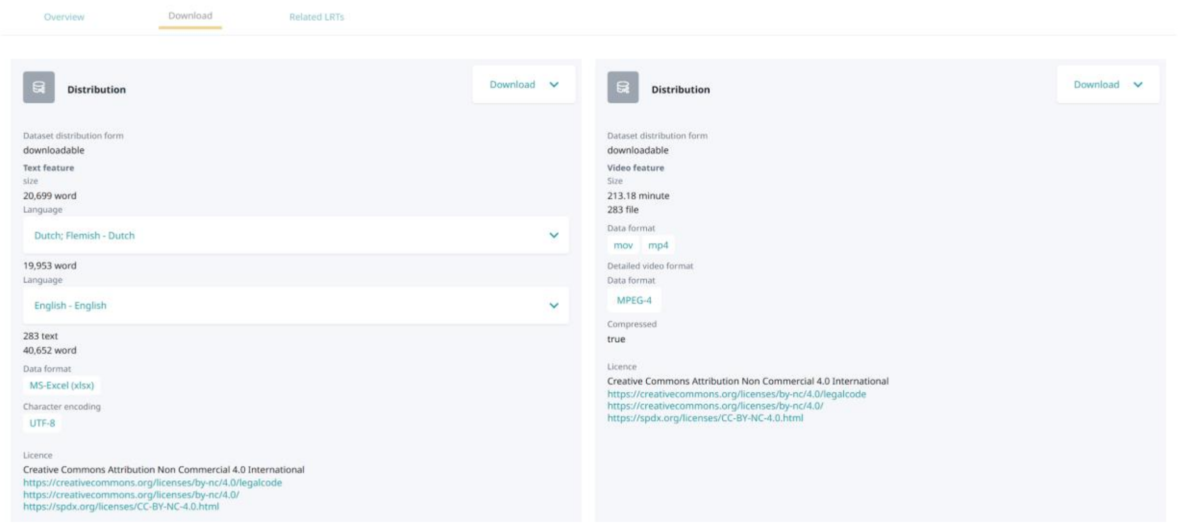
All versions
NGT-Dutch Hotel Review Corpus (1.0.0 (automatically assigned)) https://doi.org/10.57771/x01e-am45 (DOI)
NGT-Dutch Hotel Review Corpus (2) https://doi.org/10.57771/w7wm-cy80 (DOI)

Resource provider
Instituut voor de Nederlandse Taal
Website

FIGURE 10: EXCERPT OF VIEW PAGE FOR A SIGN LANGUAGE CORPUS - TEXT PART

Figure 9 and Figure 10 show the view page for a sign language corpus hosted in ELG (https://live.european-language-grid.eu/catalogue/corpus/21535), the first one showing the overview page including information on the video part and the second one for the text. Figure 11 shows the download page, with information on the licensing terms and offering access to the data files themselves.

Funded by the Horizon 2020 Framework Programme of the European Union



Distribution

Dataset distribution form
downloadable
Text feature
size
20,699 word
Language
Dutch; Flemish - Dutch
19,953 word
Language
English - English
283 text
40,652 word
Data format
MS-Excel (xlsx)
Character encoding
UTF-8

Licence
Creative Commons Attribution Non Commercial 4.0 International
https://creativecommons.org/licenses/by-nc/4.0/legalcode
https://creativecommons.org/licenses/by-nc/4.0/
https://spdx.org/licenses/CC-BY-NC-4.0.html

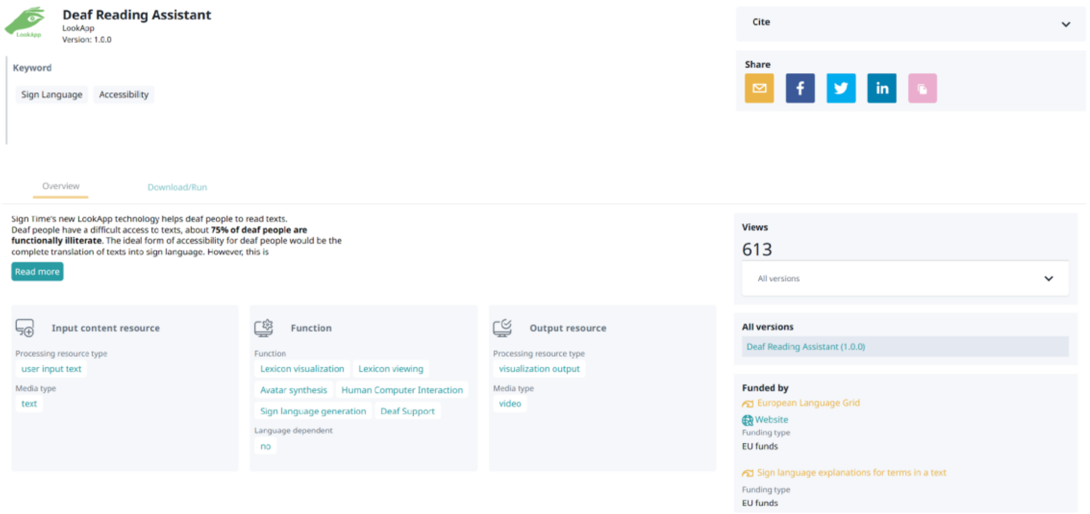
Distribution

Dataset distribution form
downloadable
Video feature
Size
213.18 minute
283 file
Data format
mov mp4
Detailed video format
Data format
MPEG-4
Compressed
true

Licence
Creative Commons Attribution Non Commercial 4.0 International
https://creativecommons.org/licenses/by-nc/4.0/legalcode
https://creativecommons.org/licenses/by-nc/4.0/
https://spdx.org/licenses/CC-BY-NC-4.0.html

FIGURE 11: EXCERPT OF VIEW PAGE FOR A SIGN LANGUAGE CORPUS (ACCESS TO DATA FILES)

Tools/services involving sign languages can also be adequately described with the ELG metadata. More specifically, depending on the processing type and the function they perform, the relevant information for input and output can be added. For example, translation services from a spoken language into a sign language can be described as taking as input "text" in the specific spoken language and exporting as output "video" in the sign language.



Deaf Reading Assistant
LookApp
Version: 1.0.0

Keyword
Sign Language Accessibility

Sign Time's new LookApp technology helps deaf people to read texts.
Deaf people have a difficult access to texts, about **75% of deaf people are functionally illiterate**. The ideal form of accessibility for deaf people would be the complete translation of texts into sign language. However, this is

Input content resource
Processing resource type
user input text
Media type
text

Function
Function
Lexicon visualization Lexicon viewing
Avatar synthesis Human Computer Interaction
Sign language generation Deaf Support
Language dependent
no

Output resource
Processing resource type
visualization output
Media type
video

Views
613
All versions

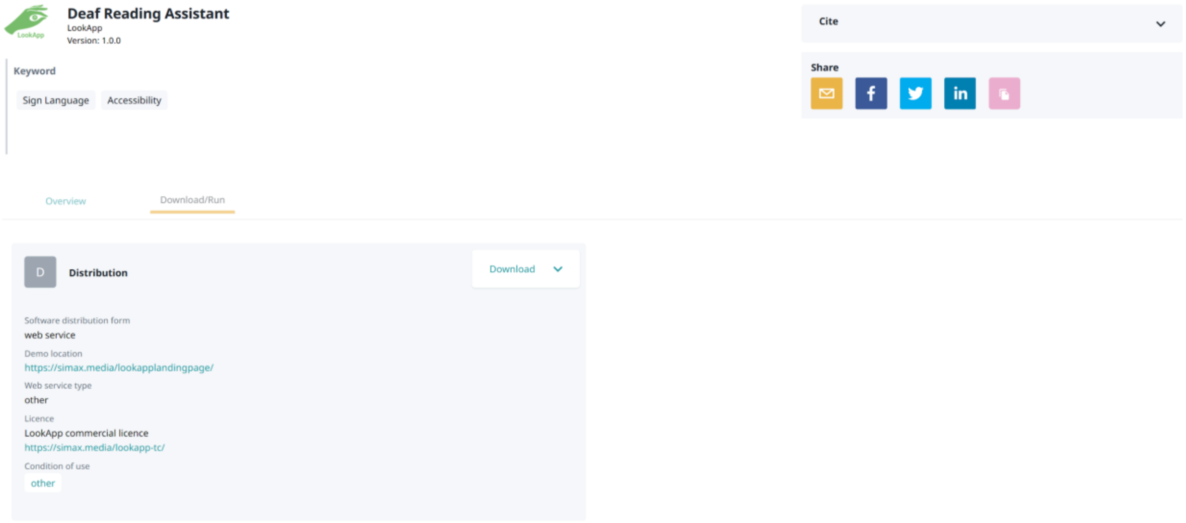
All versions
Deaf Reading Assistant (1.0.0)

Funded by
European Language Grid
Website
Funding type
EU funds

Sign language explanations for terms in a text
Funding type
EU funds

FIGURE 12: EXCERPT OF VIEW PAGE FOR A SIGN LANGUAGE APPLICATION (OVERVIEW TAB)

Figure 12 shows the landing page for an application (https://live.european-language-grid.eu/catalogue/tool-service/9387) that links text to a sign language dictionary, and shows for each word of the text selected by the user the relevant video entry from the dictionary. The application can be accessed at an external location, while a demo is also available (Figure 13).



Deaf Reading Assistant
LookApp
Version: 1.0.0

Keyword
Sign Language
Accessibility

Distribution

Software distribution form
web service

Demo location
https://simax.media/lookapplandingpage/

Web service type
other

Licence
LookApp commercial licence
https://simax.media/lookapp-tc/

Condition of use
other

Figure 13: Excerpt of view page for a sign language application (access tab)

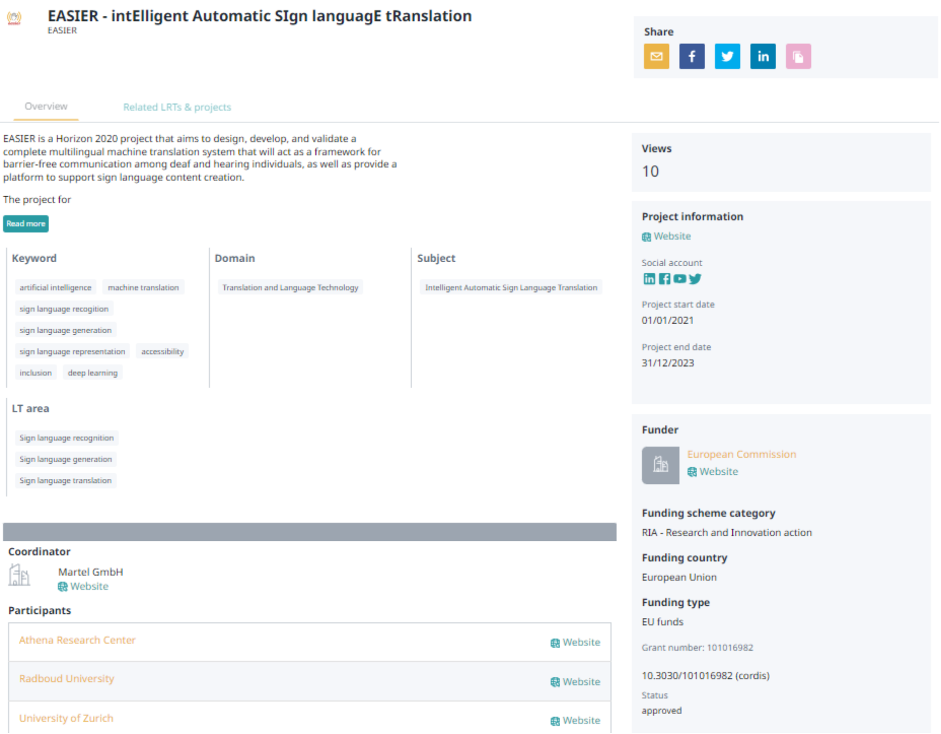
For integrating sign language tools/services as ELG-compatible services, providers must follow the procedure described in Section 2.5.2.

Funded by the Horizon 2020
Framework Programme of the European Union

4 INTEGRATION OF EASIER RESOURCES INTO ELG

To improve the access to all resources developed in the framework of the EASIER project, a page dedicated to the description of the project has been created: https://live.european-language-grid.eu/catalogue/project/23037 (Figure 14).

The page includes all information about the project and acts as the central hub with links to the respective ELG pages of the consortium partners as well as to the resources they have provided. Hyperlinking enables also the navigation from the organisation pages to the resources and to the project, as well as vice versa, so that users can locate the information they seek for through any of these pages. Furthermore, the resources are discoverable via the free text and faceting functionalities, as all other resources in ELG.



EASIER - intElligent Automatic SIgn languagE tRanslation
EASIER

EASIER is a Horizon 2020 project that aims to design, develop, and validate a complete multilingual machine translation system that will act as a framework for barrier-free communication among deaf and hearing individuals, as well as provide a platform to support sign language content creation.

The project for

Keyword: artificial intelligence, machine translation, sign language recogition, sign language generation, sign language representation, accessibility, inclusion, deep learning

Domain: Translation and Language Technology

Subject: Intelligent Automatic Sign Language Translation

LT area: Sign language recognition, Sign language generation, Sign language translation

Coordinator
Martel GmbH
Website

Participants

Athena Research Center	Website
Radboud University	Website
University of Zurich	Website

Views
10

Project information
Website

Social account

Project start date
01/01/2021

Project end date
31/12/2023

Funder
European Commission
Website

Funding scheme category
RIA - Research and Innovation action

Funding country
European Union

Funding type
EU funds

Grant number: 101016982

10.3030/101016982 (cordis)

Status
approved

FIGURE 14: LANDING PAGE OF EASIER IN ELG CATALOGUE

Currently EASIER resources are in the process of being integrated in the ELG catalogue. At the moment of writing, this includes a set of one corpus and 6 tools/services, while more resources will be added as they are finalized.

Funded by the Horizon 2020 Framework Programme of the European Union

5 CONCLUSIONS

Deliverable D8.6 reports on the work performed in the framework of Task 8.6 in order to integrate resources developed in the EASIER project through the ELG catalogue, both at the ELG side (ensuring that sign language specificities can be appropriately handled and displayed at the ELG catalogue) and the EASIER side (by including EASIER resources in the ELG catalogue). The work is still ongoing as more resources from the EASIER project are finalized and added to the catalogue.