EUROPEAN
LANGUAGE
GRID

EUROPEAN
LANGUAGE
EQUALITY

European Language Grid overview

Maria Giagkou
Institute for Language and Speech Processing, ATHENA Research and Innovation Centre

UniDive, WG3 meeting, September 8, 2023

EUROPEAN
LANGUAGE
GRID

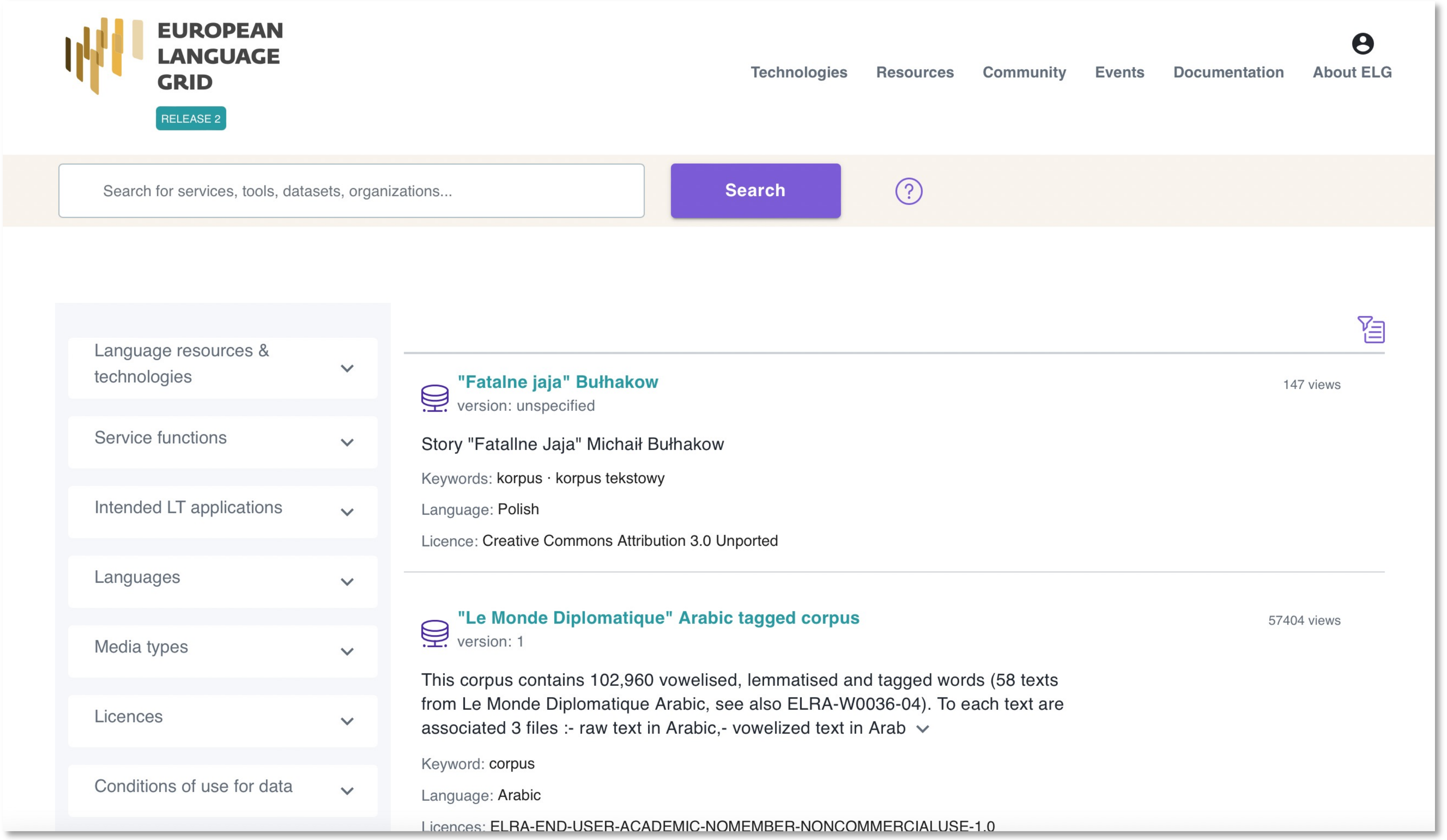
European Language Grid project

- Project duration: 2019-2022
- Roots: META-NET Strategic Research Agenda for Multilingual Europe (2020)
  - 24 official EU languages, more than 60 official languages at the national and regional levels
  - Fragmented LT landscape: thousands for SMEs directly or indirectly active in the LT sector, hundreds of research and academic institutions
  - A "European Service Platform for Language Technologies" recommended
- Main aim: tackle the fragmentation of the European LT landscape
- Landmark outcome: the **European Language Grid platform**

EUROPEAN
LANGUAGE
GRID

European Language Grid platform

- Accessible at https://live.european-language-grid.eu/
- A platform for commercial and non-commercial LTs
- In support of developers and integrators of LTs as well as users/consumers of LT
- Enabling the European LT community to **upload** services and data sets, to **deploy** them and to connect with, and **use** resources made available by others



"Fatalne jaja" Bułhakow
version: unspecified
147 views

Story "Fatallne Jaja" Michaił Bułhakow

Keywords: korpus · korpus tekstowy

Language: Polish

Licence: Creative Commons Attribution 3.0 Unported

"Le Monde Diplomatique" Arabic tagged corpus
version: 1
57404 views

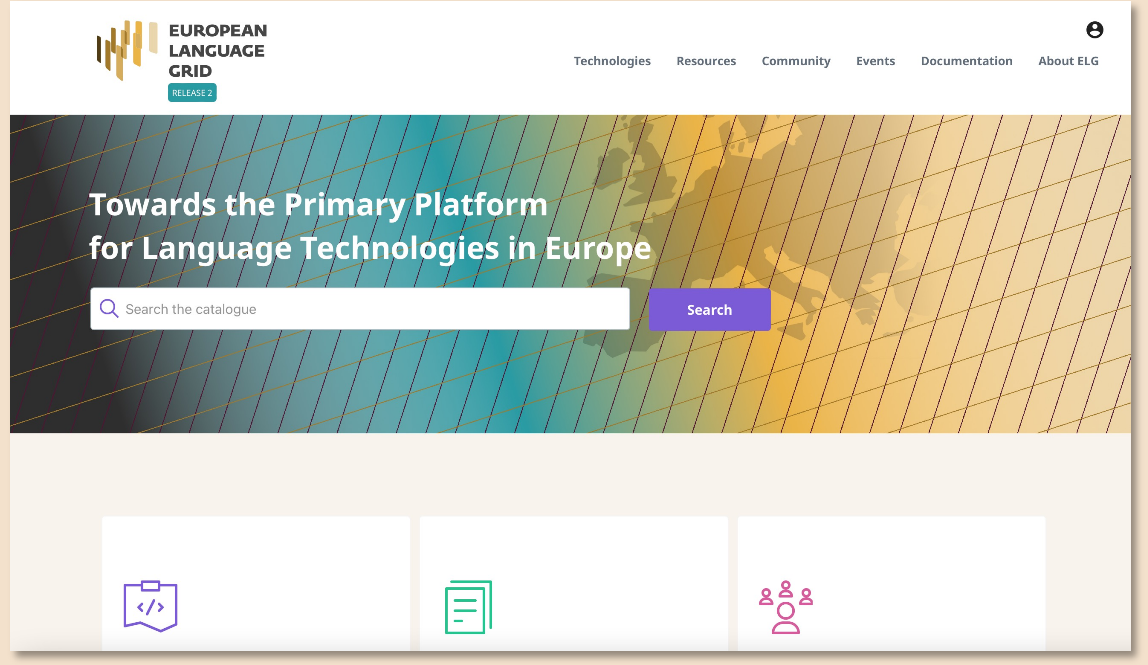
This corpus contains 102,960 vowelised, lemmatised and tagged words (58 texts from Le Monde Diplomatique Arabic, see also ELRA-W0036-04). To each text are associated 3 files :- raw text in Arabic,- vowelized text in Arab

Keyword: corpus

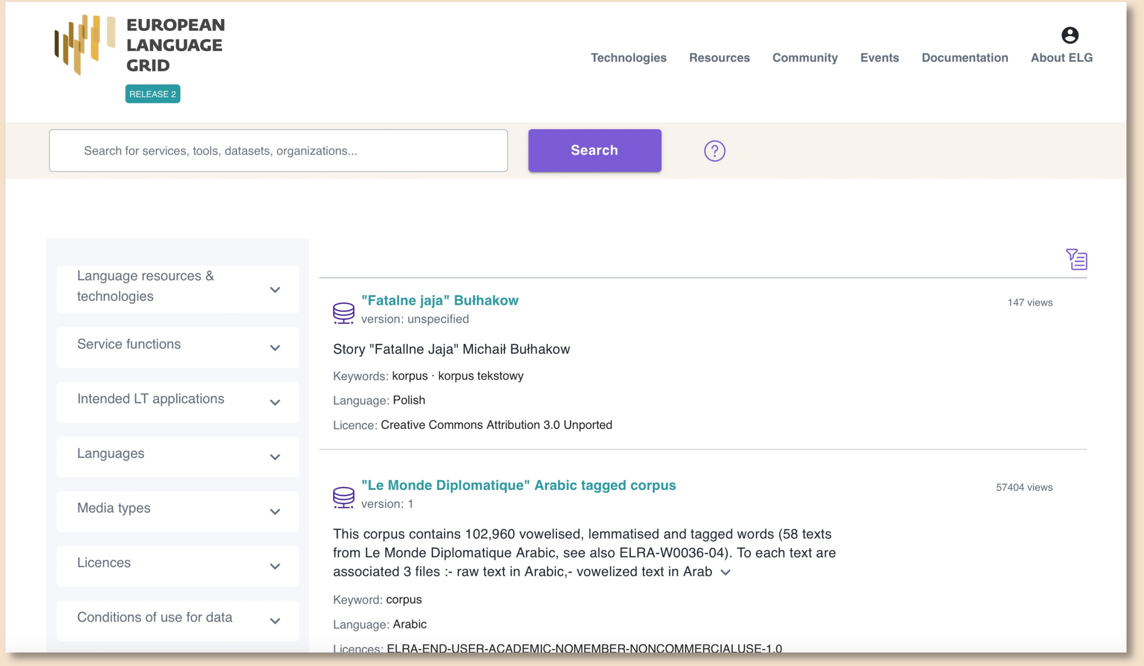
Language: Arabic

Licences: ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0

EUROPEAN
LANGUAGE
GRID



Towards the Primary Platform
for Language Technologies in Europe



"Fatalne jaja" Bułhakow
version: unspecified
147 views

Story "Fatallne Jaja" Michaił Bułhakow

Keywords: korpus · korpus tekstowy

Language: Polish

Licence: Creative Commons Attribution 3.0 Unported

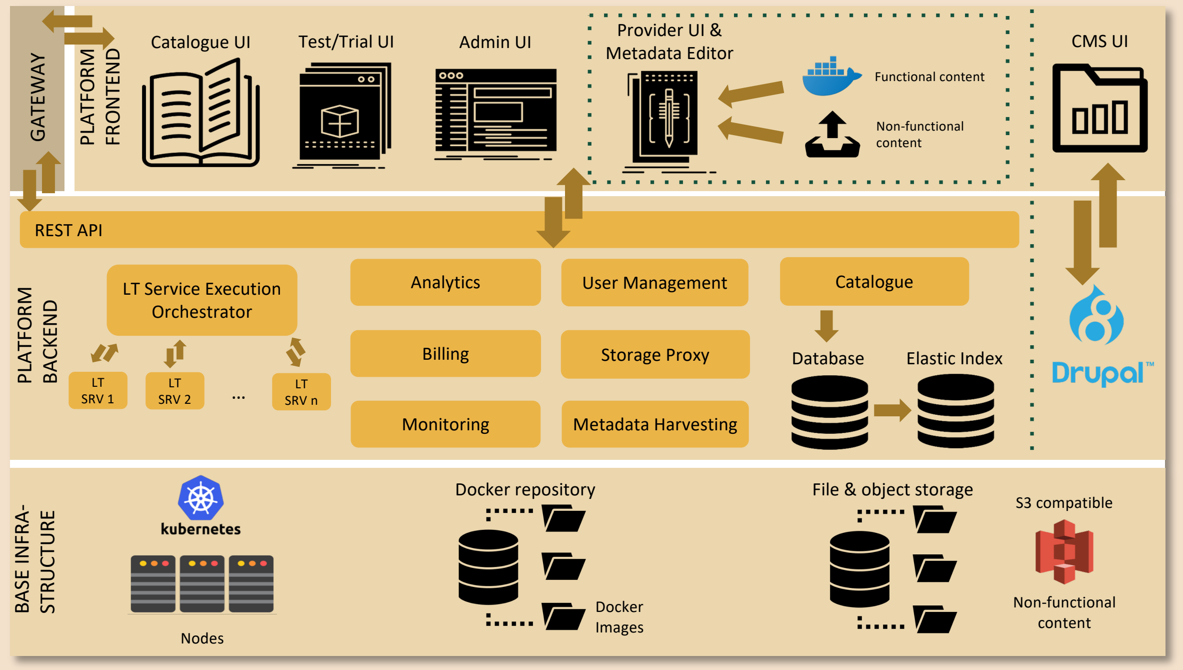
"Le Monde Diplomatique" Arabic tagged corpus
version: 1
57404 views

This corpus contains 102,960 vowelised, lemmatised and tagged words (58 texts from Le Monde Diplomatique Arabic, see also ELRA-W0036-04). To each text are associated 3 files :- raw text in Arabic,- vowelized text in Arab

Keyword: corpus

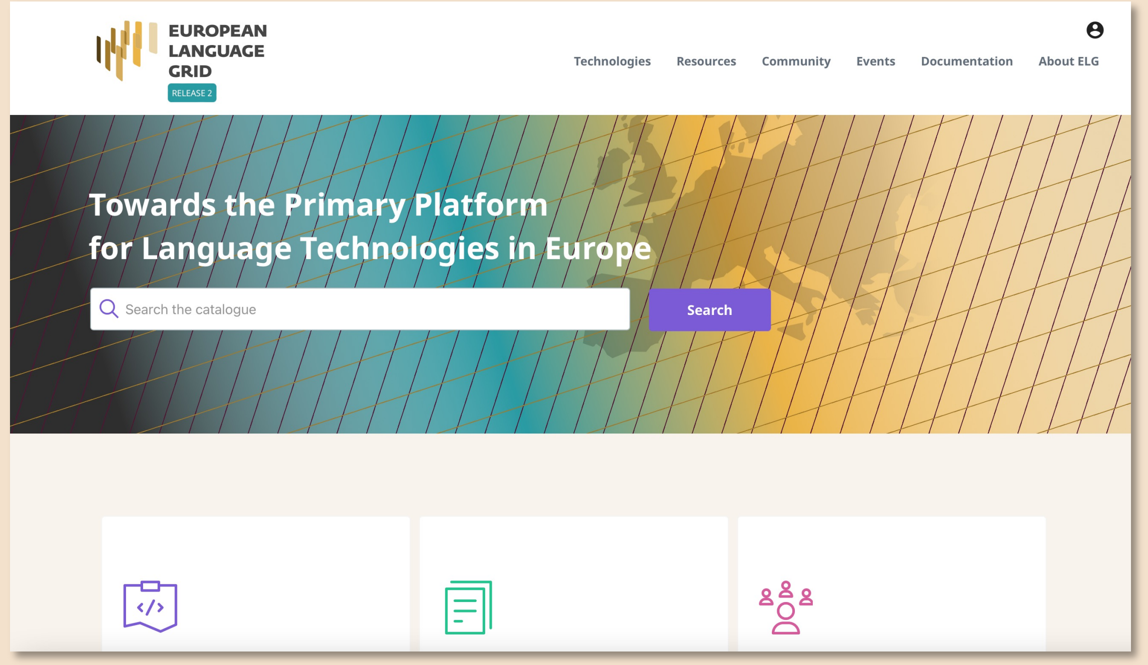
Language: Arabic

Licences: ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0

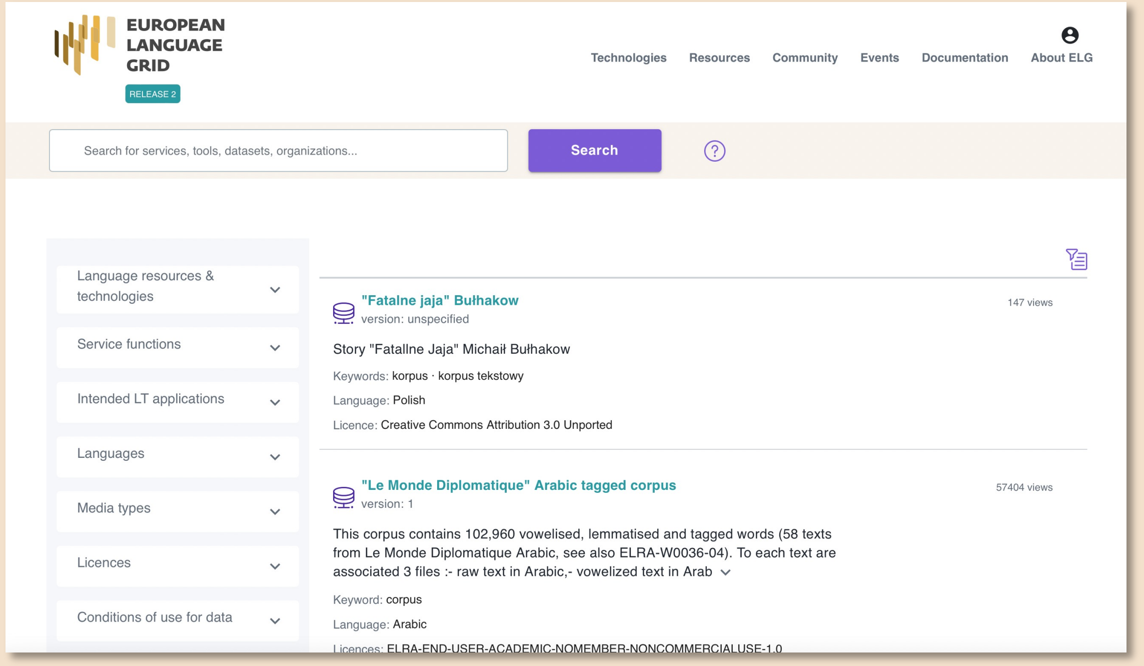

GATEWAY
PLATFORM FRONTEND
Catalogue UI
Test/Trial UI
Admin UI
Provider UI & Metadata Editor
Functional content
Non-functional content
CMS UI
REST API
PLATFORM BACKEND
LT Service Execution Orchestrator
LT SRV 1
LT SRV 2
...
LT SRV n
Analytics
Billing
Monitoring
User Management
Storage Proxy
Metadata Harvesting
Catalogue
Database
Elastic Index
Drupal™
BASE INFRA-STRUCTURE
kubernetes
Nodes
Docker repository
Docker Images
File & object storage
S3 compatible
Non-functional content

European Language Grid is a joint data, tools and services sharing platform

EUROPEAN
LANGUAGE
GRID



Towards the Primary Platform
for Language Technologies in Europe



"Fatalne jaja" Bułhakow
version: unspecified

147 views

Story "Fatallne Jaja" Michaił Bułhakow

Keywords: korpus · korpus tekstowy

Language: Polish

Licence: Creative Commons Attribution 3.0 Unported

"Le Monde Diplomatique" Arabic tagged corpus
version: 1

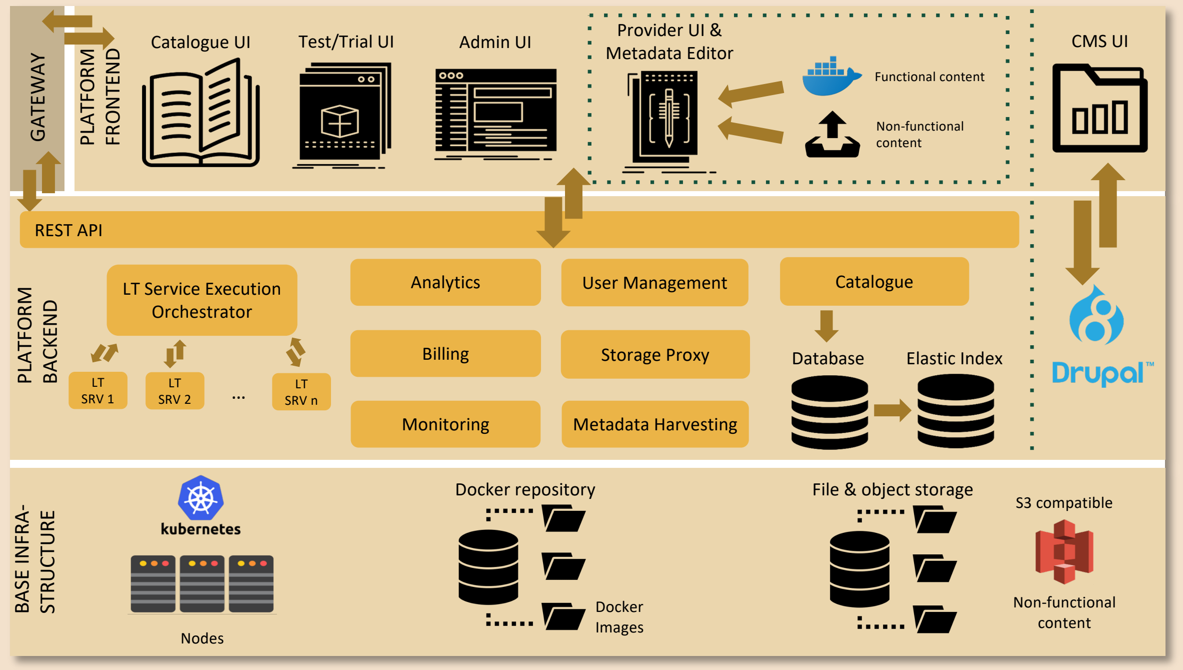
57404 views

This corpus contains 102,960 vowelised, lemmatised and tagged words (58 texts from Le Monde Diplomatique Arabic, see also ELRA-W0036-04). To each text are associated 3 files :- raw text in Arabic,- vowelized text in Arab

Keyword: corpus

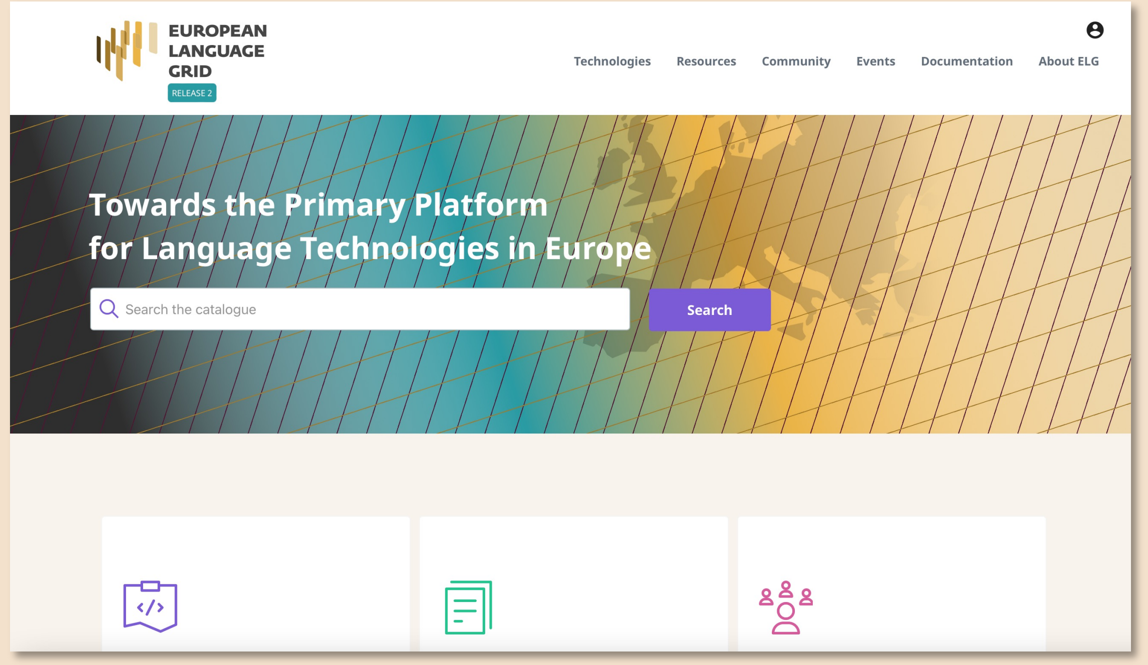
Language: Arabic

Licences: ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0

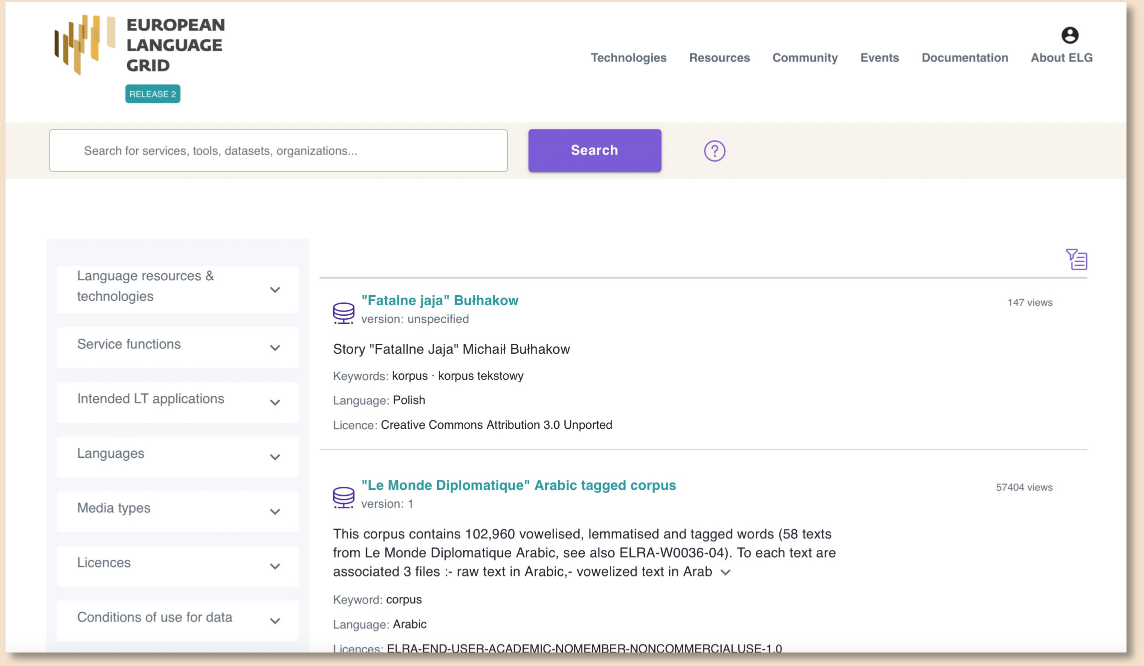


European Language Grid is a marketplace and **one-stop-shop** for the whole European Language Technology community

EUROPEAN
LANGUAGE
GRID



Towards the Primary Platform
for Language Technologies in Europe



"Fatalne jaja" Bułhakow
version: unspecified
147 views

Story "Fatallne Jaja" Michaił Bułhakow

Keywords: korpus · korpus tekstowy

Language: Polish

Licence: Creative Commons Attribution 3.0 Unported

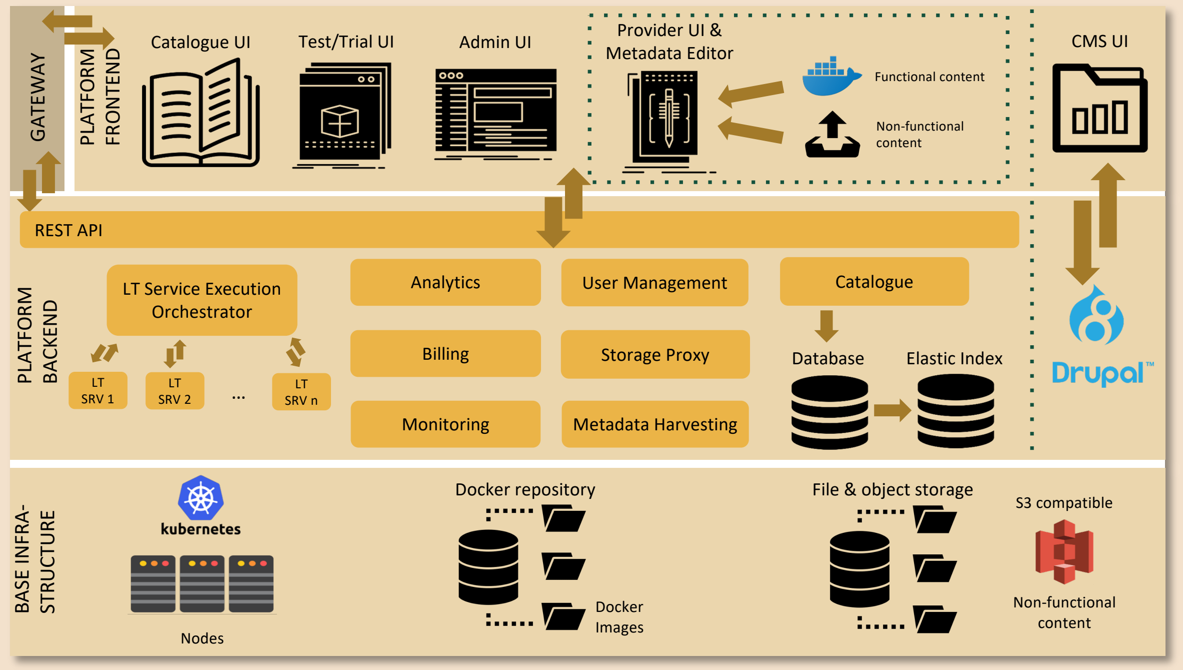
"Le Monde Diplomatique" Arabic tagged corpus
version: 1
57404 views

This corpus contains 102,960 vowelised, lemmatised and tagged words (58 texts from Le Monde Diplomatique Arabic, see also ELRA-W0036-04). To each text are associated 3 files :- raw text in Arabic,- vowelized text in Arab

Keyword: corpus

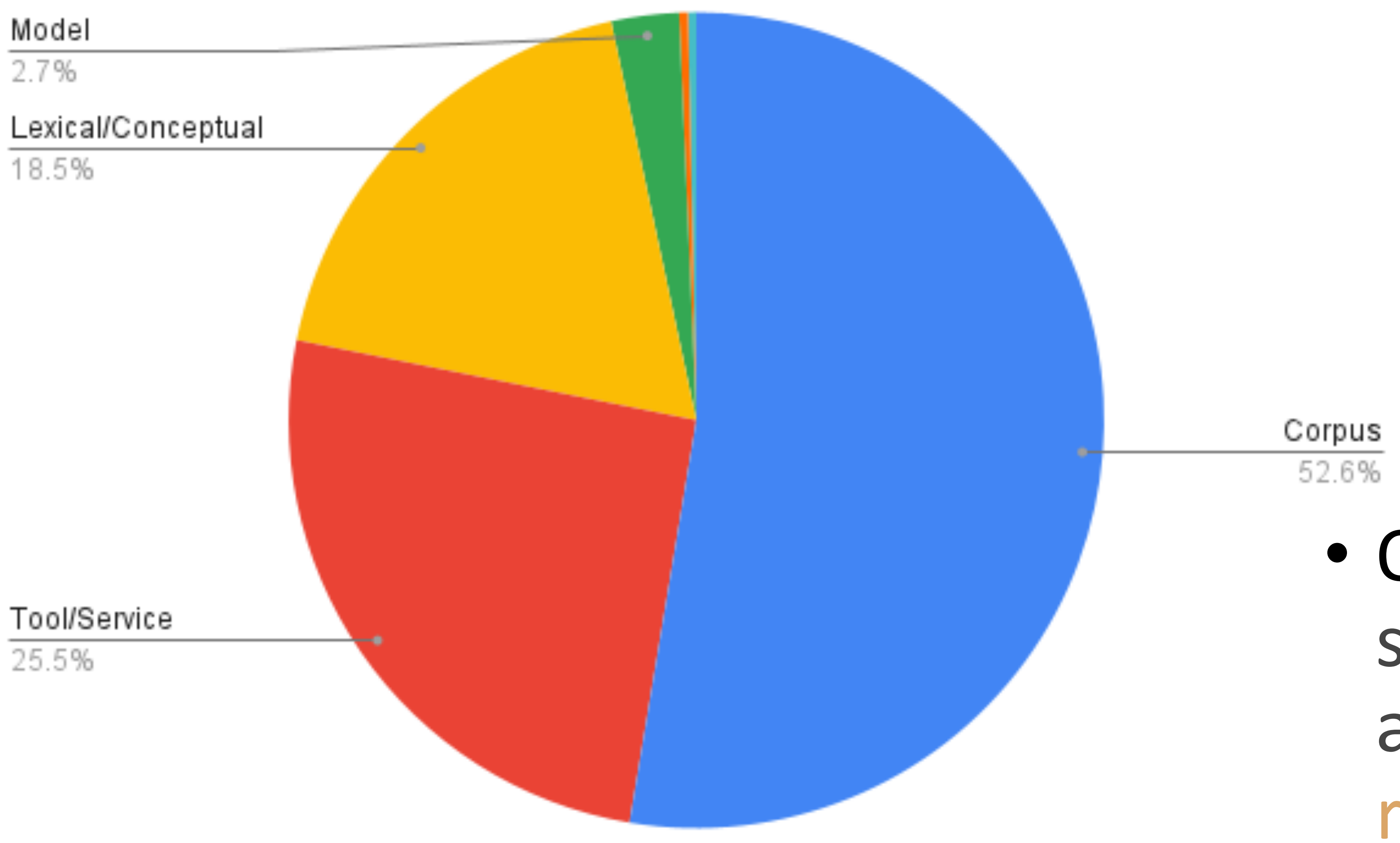
Language: Arabic

Licences: ELRA-END-USER-ACADEMIC-NOMEMBER-NONCOMMERCIALUSE-1.0



European Language Grid is the yellow pages of the European Language Technology community

ELG catalogue in numbers (2023/09/05)



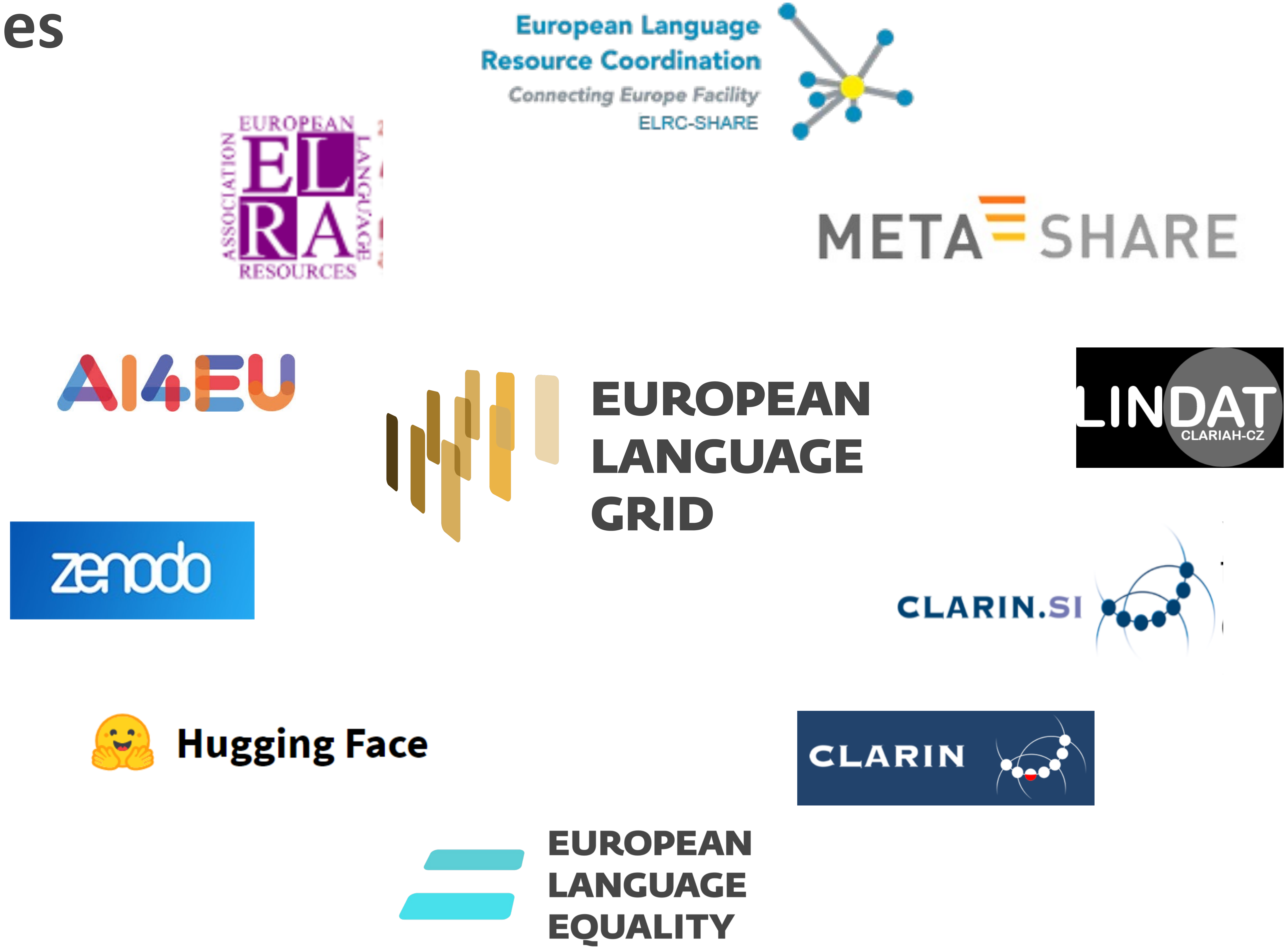
15.199 Language Resources & Technologies (LRTs), including

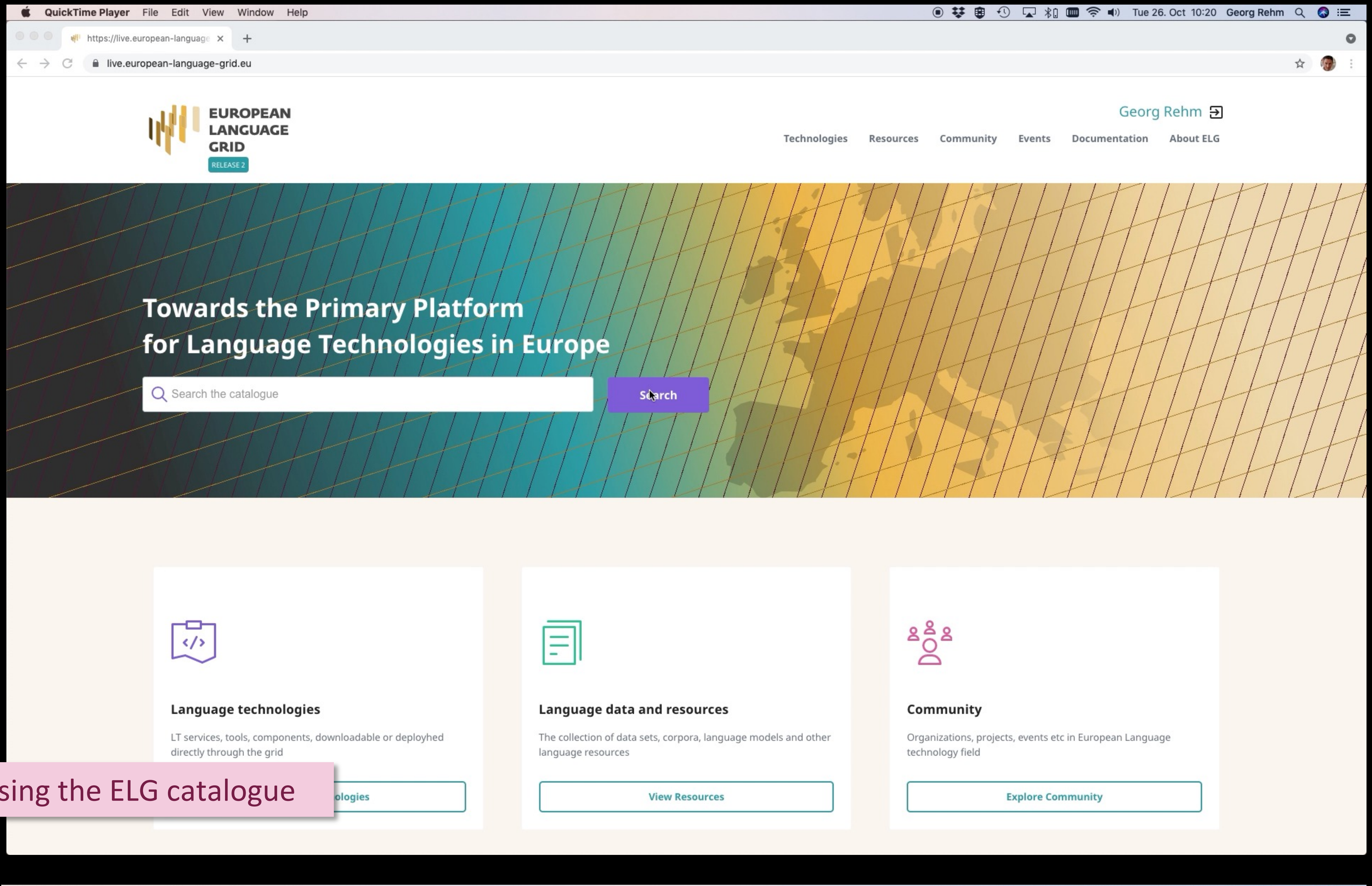
- **As per type:**
  - 11.319 data resources, e.g. corpora (collections of data), models, lexical/conceptual resources (e.g., lexica, ontologies, terminologies, etc.), computational grammars
  - 3.880 language processing tools and services (e.g., MT services, speech synthesis/analysis tools, IE services, etc.),
- Of which, 1307 LRTs integrated in the ELG infrastructure: 1187 services running in the ELG cloud and 120 datasets uploaded and hosted in ELG; remaining records are "metadata-only records" providing information on the LRTs and re-directing consumers to the location where they can access them

- Related entities: 1775 Organizations involved in LT & 513 Projects

ELG catalogue contents: Sources

- **Bridges** with existing platforms and infrastructures
  - Mainly in terms of metadata-based descriptions
  - Based on **open protocols** (OAI-PMH), or **APIs** offered by the platform or infrastructure providers
  - Respecting their own policies
- ELG also as infrastructural arm of ELE
  - using a mixture of automatic and collaborative population of the ELG catalogue


European Language Resource Coordination
Connecting Europe Facility
ELRC-SHARE
EUROPEAN LANGUAGE RESOURCES ASSOCIATION
ELRA
META SHARE
AI4EU
EUROPEAN LANGUAGE GRID
LINDAT
CLARIAH-CZ
zenodo
CLARIN.SI
Hugging Face
CLARIN
EUROPEAN LANGUAGE EQUALITY


QuickTime Player File Edit View Window Help
Tue 26. Oct 10:20 Georg Rehm
live.european-language-grid.eu
EUROPEAN LANGUAGE GRID
RELEASE 2
Georg Rehm
Technologies Resources Community Events Documentation About ELG
Search

Towards the Primary Platform
for Language Technologies in Europe

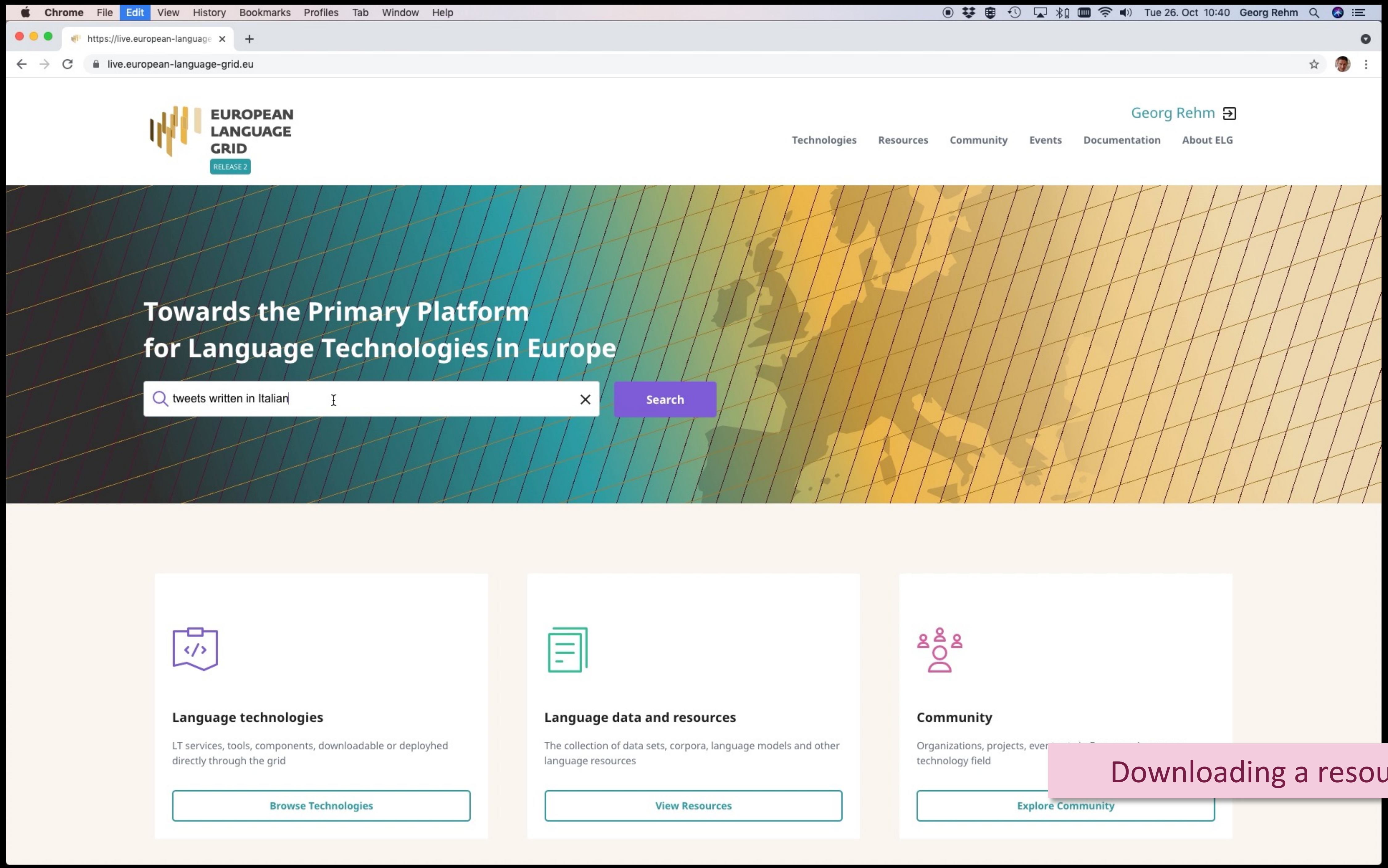
Search the catalogue

Language technologies
LT services, tools, components, downloadable or deployed directly through the grid

Language data and resources
The collection of data sets, corpora, language models and other language resources
View Resources

Community
Organizations, projects, events etc in European Language technology field
Explore Community

Browsing the ELG catalogue


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 26. Oct 10:40 Georg Rehm
live.european-language-grid.eu
EUROPEAN LANGUAGE GRID
RELEASE 2
Georg Rehm
Technologies Resources Community Events Documentation About ELG
Search

Towards the Primary Platform
for Language Technologies in Europe

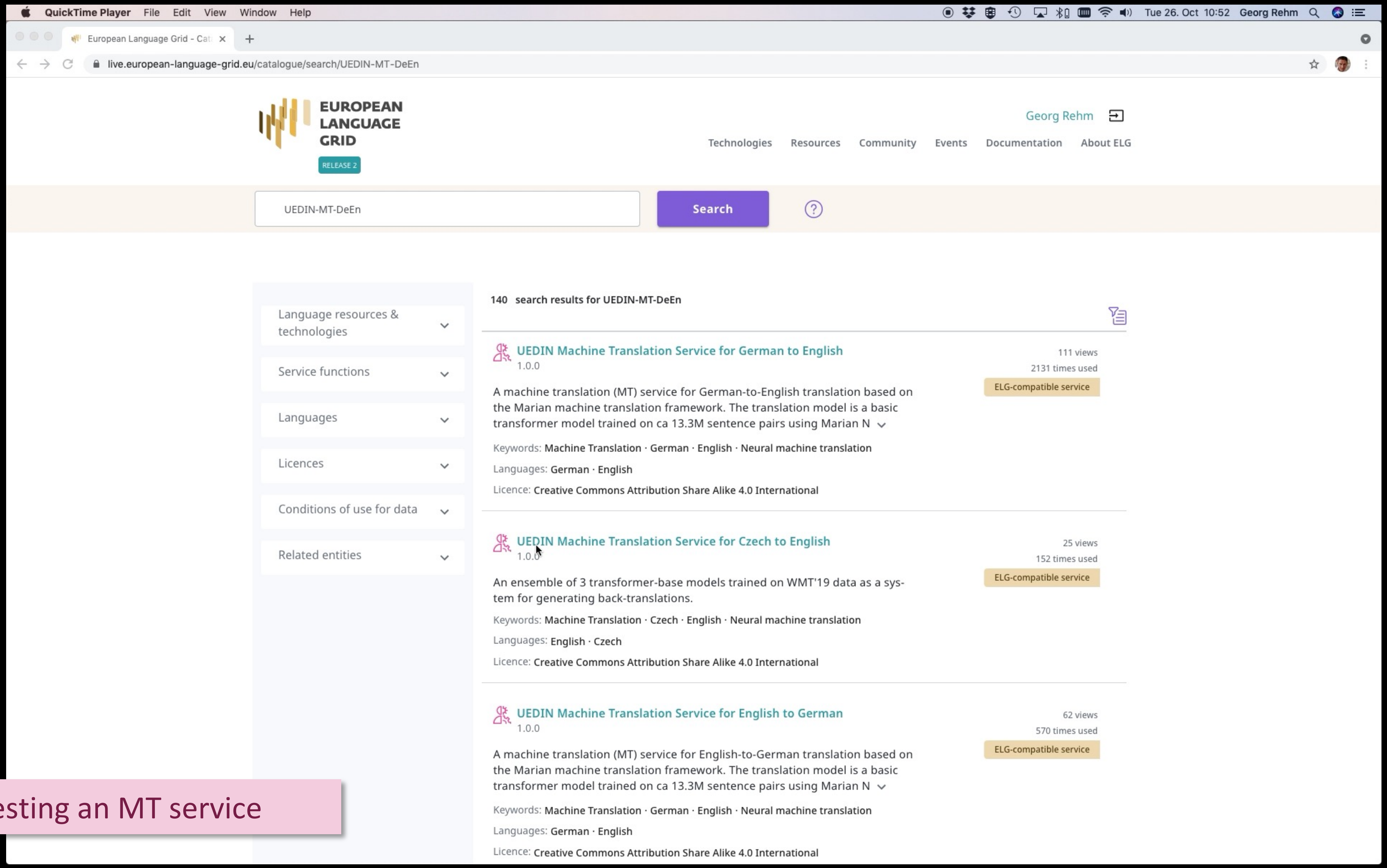
tweets written in Italian

Language technologies
LT services, tools, components, downloadable or deployed directly through the grid
Browse Technologies

Language data and resources
The collection of data sets, corpora, language models and other language resources
View Resources

Community
Organizations, projects, technology field
Explore Community

Downloading a resource


QuickTime Player File Edit View Window Help
Tue 26. Oct 10:52 Georg Rehm
European Language Grid - Cat
live.european-language-grid.eu/catalogue/search/UEDIN-MT-DeEn
EUROPEAN LANGUAGE GRID
RELEASE 2
Georg Rehm
Technologies Resources Community Events Documentation About ELG
Search

UEDIN-MT-DeEn

Language resources & technologies
Service functions
Languages
Licences
Conditions of use for data
Related entities

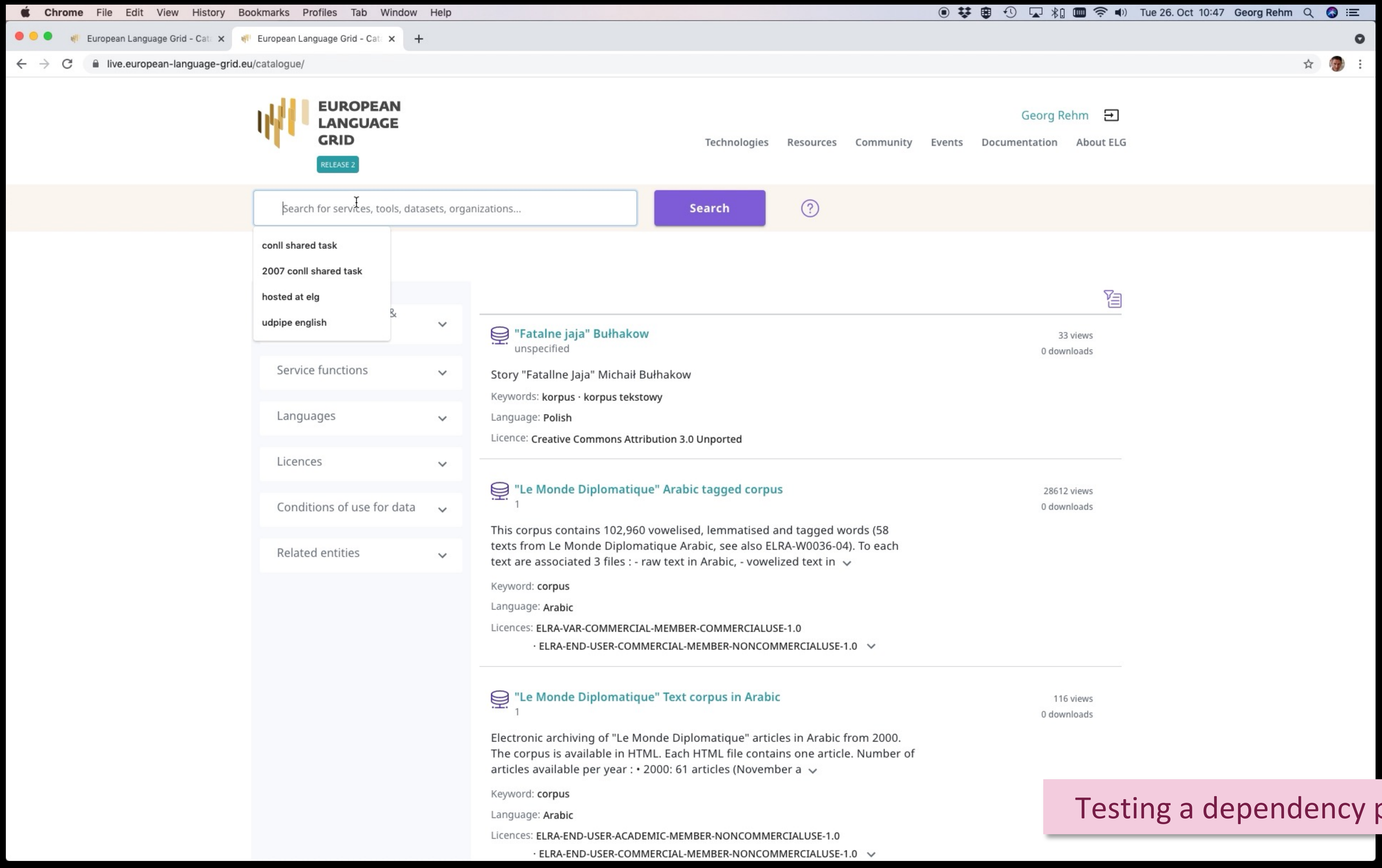
140 search results for UEDIN-MT-DeEn

UEDIN Machine Translation Service for German to English
1.0.0
111 views
2131 times used
ELG-compatible service
A machine translation (MT) service for German-to-English translation based on the Marian machine translation framework. The translation model is a basic transformer model trained on ca 13.3M sentence pairs using Marian N
Keywords: Machine Translation · German · English · Neural machine translation
Languages: German · English
Licence: Creative Commons Attribution Share Alike 4.0 International

UEDIN Machine Translation Service for Czech to English
1.0.0
25 views
152 times used
ELG-compatible service
An ensemble of 3 transformer-base models trained on WMT'19 data as a system for generating back-translations.
Keywords: Machine Translation · Czech · English · Neural machine translation
Languages: English · Czech
Licence: Creative Commons Attribution Share Alike 4.0 International

UEDIN Machine Translation Service for English to German
1.0.0
62 views
570 times used
ELG-compatible service
A machine translation (MT) service for English-to-German translation based on the Marian machine translation framework. The translation model is a basic transformer model trained on ca 13.3M sentence pairs using Marian N
Keywords: Machine Translation · German · English · Neural machine translation
Languages: German · English
Licence: Creative Commons Attribution Share Alike 4.0 International

Testing an MT service


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Tue 26. Oct 10:47 Georg Rehm
European Language Grid - Cat
European Language Grid - Cat
live.european-language-grid.eu/catalogue/
EUROPEAN LANGUAGE GRID
RELEASE 2
Georg Rehm
Technologies Resources Community Events Documentation About ELG
Search

Search for services, tools, datasets, organizations...

conll shared task
2007 conll shared task
hosted at elg
udpipe english

Service functions
Languages
Licences
Conditions of use for data
Related entities

"Fatalne jaja" Bułhakow
unspecified
33 views
0 downloads
Story "Fatallne Jaja" Michaił Bułhakow
Keywords: korpus · korpus tekstowy
Language: Polish
Licence: Creative Commons Attribution 3.0 Unported

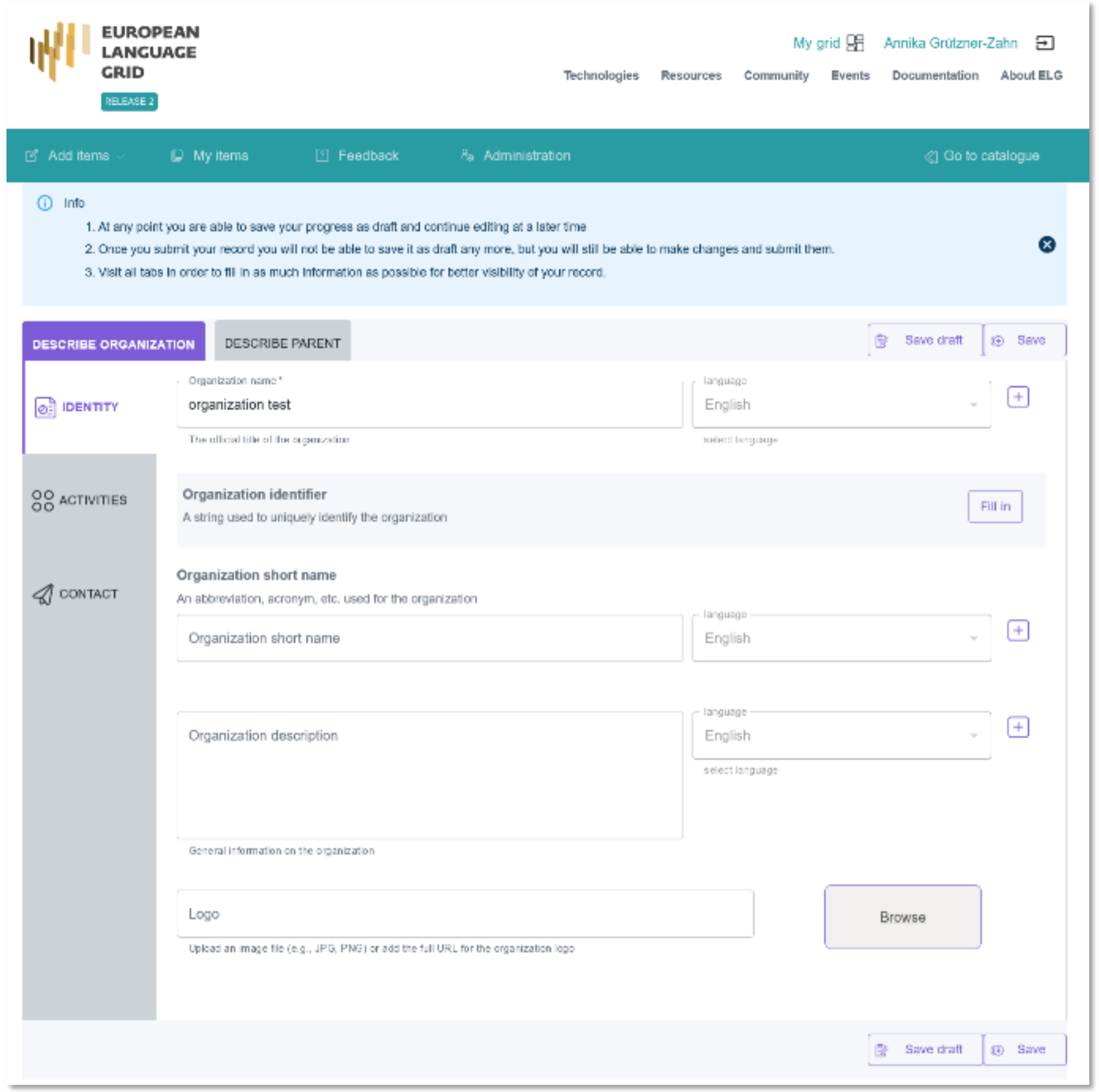
"Le Monde Diplomatique" Arabic tagged corpus
1
28612 views
0 downloads
This corpus contains 102,960 vowelised, lemmatised and tagged words (58 texts from Le Monde Diplomatique Arabic, see also ELRA-W0036-04). To each text are associated 3 files : - raw text in Arabic, - vowelized text in
Keyword: corpus
Language: Arabic
Licences: ELRA-VAR-COMMERCIAL-MEMBER-COMMERCIALUSE-1.0
· ELRA-END-USER-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0

"Le Monde Diplomatique" Text corpus in Arabic
1
116 views
0 downloads
Electronic archiving of "Le Monde Diplomatique" articles in Arabic from 2000. The corpus is available in HTML. Each HTML file contains one article. Number of articles available per year : • 2000: 61 articles (November a
Keyword: corpus
Language: Arabic
Licences: ELRA-END-USER-ACADEMIC-MEMBER-NONCOMMERCIALUSE-1.0
· ELRA-END-USER-COMMERCIAL-MEMBER-NONCOMMERCIALUSE-1.0

Testing a dependency parser

Provider's grid

- Providers can create new items
  - by validating and uploading a schema compliant metadata record (single or batch)
  - using an **interactive metadata editor**



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.

Organization name *: organization test — language: English

The official title of the organization — select language

Organization identifier
A string used to uniquely identify the organization

Organization short name
An abbreviation, acronym, etc. used for the organization

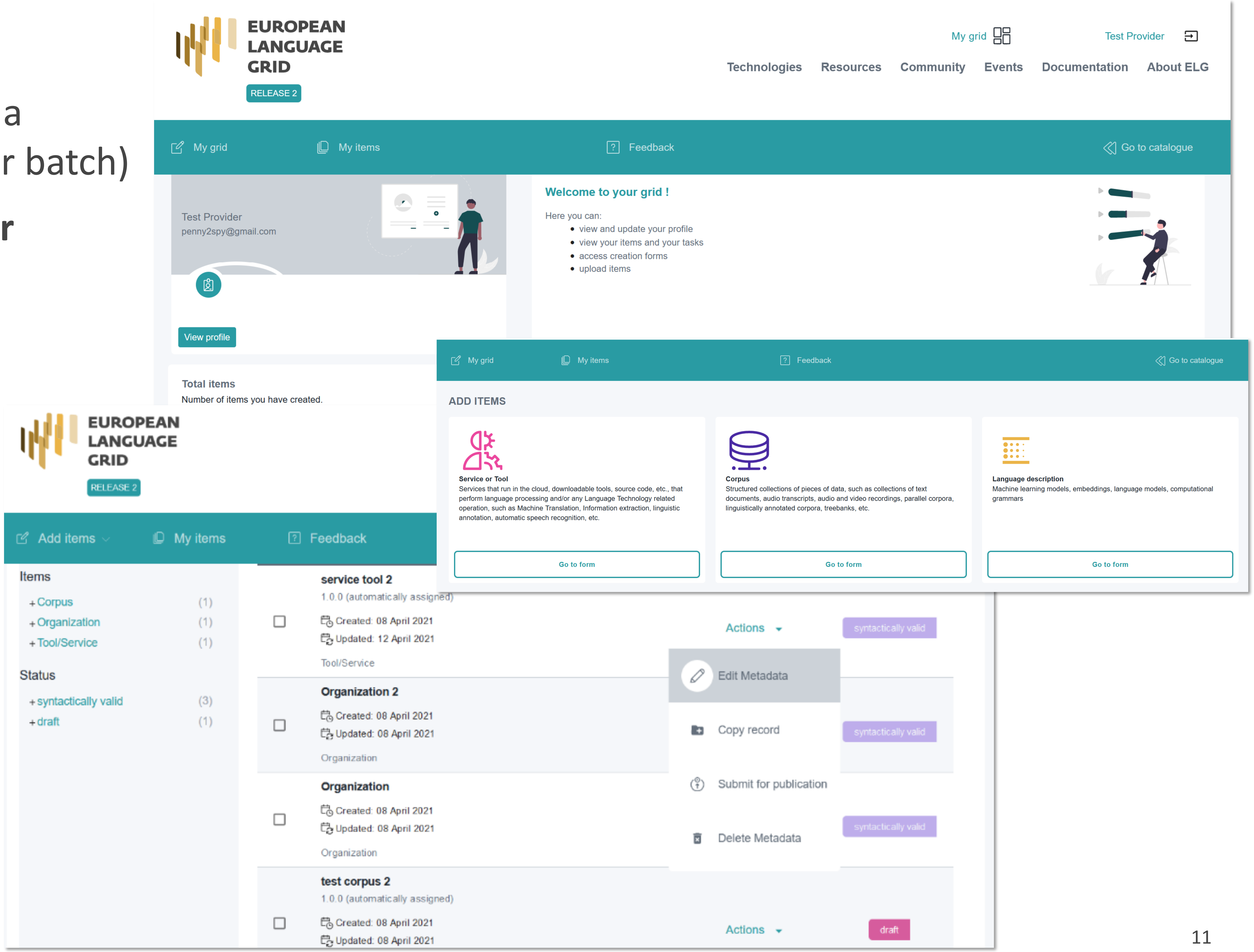
Organization short name — language: English

Organization description — language: English — select language

General information on the organization

Logo

Upload an image file (e.g., JPG, PNG) or add the full URL for the organization logo



Test Provider
penny2spy@gmail.com

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.

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.

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.

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

Items

+Corpus	(1)
+Organization	(1)
+Tool/Service	(1)

Status

+syntactically valid	(3)
+draft	(1)

service tool 2
1.0.0 (automatically assigned)
Created: 08 April 2021
Updated: 12 April 2021
Tool/Service

Organization 2
Created: 08 April 2021
Updated: 08 April 2021
Organization

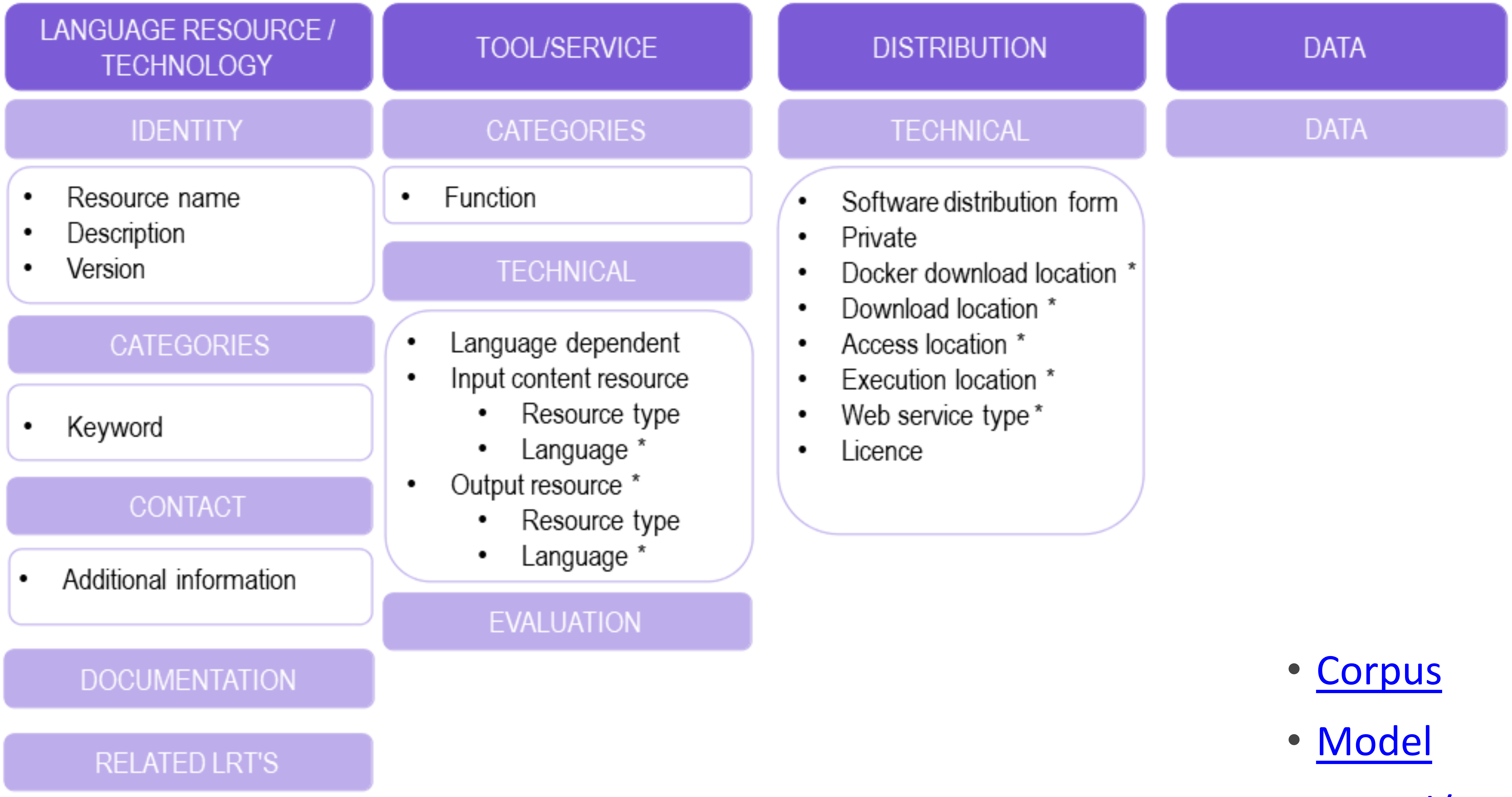
Organization
Created: 08 April 2021
Updated: 08 April 2021
Organization

test corpus 2
1.0.0 (automatically assigned)
Created: 08 April 2021
Updated: 08 April 2021

ELG metadata schema: main features

- **rich structured model** aiming to describe LRTs properties throughout their lifecycle, from design to (re-)use
- three versions:
  - maximal (full description) with required, recommended and optional properties
  - minimal (intended to support discovery) with required properties
  - relaxed (intended only for harvesting from more generic sources) with strictly required properties
- implementation: XSD re-using elements and values from OWL ontologies and controlled vocabularies in SKOS
  - MS-OWL (derived from the META-SHARE model)
  - OMTD-SHARE (catering for annotation/extraction types, data formats, methods & "LT taxonomy" used for service functions, intended application and LT areas for organisations and projects)
- Mappings with DataCite, DCAT, DC, (partially) schema.org

ELG minimal metadata schema



LANGUAGE RESOURCE / TECHNOLOGY

IDENTITY

- Resource name
- Description
- Version

CATEGORIES

- Keyword

CONTACT

- Additional information

DOCUMENTATION

RELATED LRT'S

TOOL/SERVICE

CATEGORIES

- Function

TECHNICAL

- Language dependent
- Input content resource
  - Resource type
  - Language *
- Output resource *
  - Resource type
  - Language *

EVALUATION

DISTRIBUTION

TECHNICAL

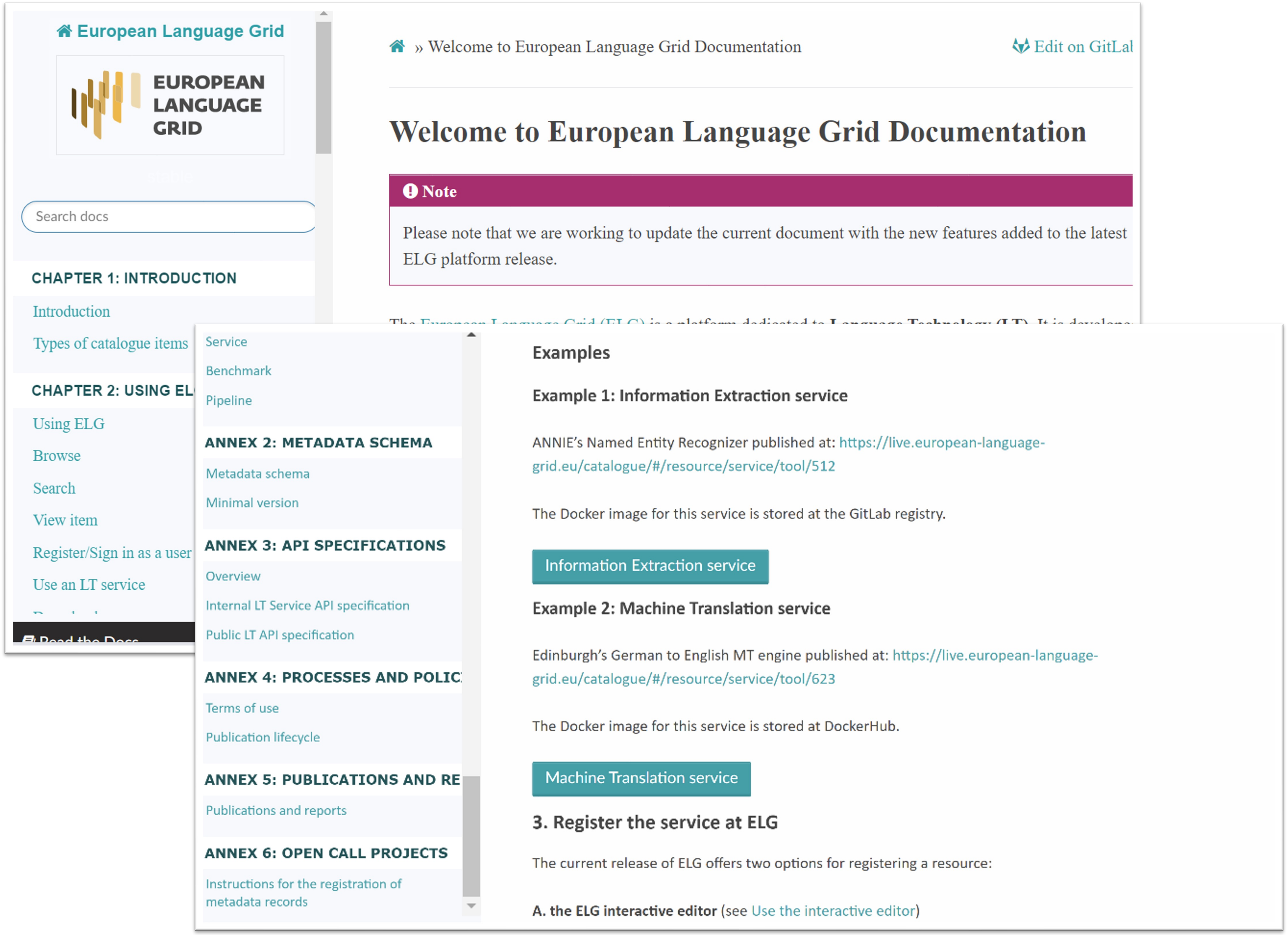
- Software distribution form
- Private
- Docker download location *
- Download location *
- Access location *
- Execution location *
- Web service type *
- Licence

DATA

DATA

- Corpus
- Model
- Lexical/Conceptual resource
- Tool/Service (not ELG-compatible)

ELG Support



European Language Grid

EUROPEAN LANGUAGE GRID

Search docs

CHAPTER 1: INTRODUCTION

Introduction

Types of catalogue items

CHAPTER 2: USING EL

Using ELG

Browse

Search

View item

Register/Sign in as a user

Use an LT service

» Welcome to European Language Grid Documentation

Edit on GitLal

Welcome to European Language Grid Documentation

Note

Please note that we are working to update the current document with the new features added to the latest ELG platform release.

Service

Benchmark

Pipeline

ANNEX 2: METADATA SCHEMA

Metadata schema

Minimal version

ANNEX 3: API SPECIFICATIONS

Overview

Internal LT Service API specification

Public LT API specification

ANNEX 4: PROCESSES AND POLIC

Terms of use

Publication lifecycle

ANNEX 5: PUBLICATIONS AND RE

Publications and reports

ANNEX 6: OPEN CALL PROJECTS

Instructions for the registration of metadata records

Examples

Example 1: Information Extraction service

ANNIE's Named Entity Recognizer published at: https://live.european-language-grid.eu/catalogue/#/resource/service/tool/512

The Docker image for this service is stored at the GitLab registry.

Example 2: Machine Translation service

Edinburgh's German to English MT engine published at: https://live.european-language-grid.eu/catalogue/#/resource/service/tool/623

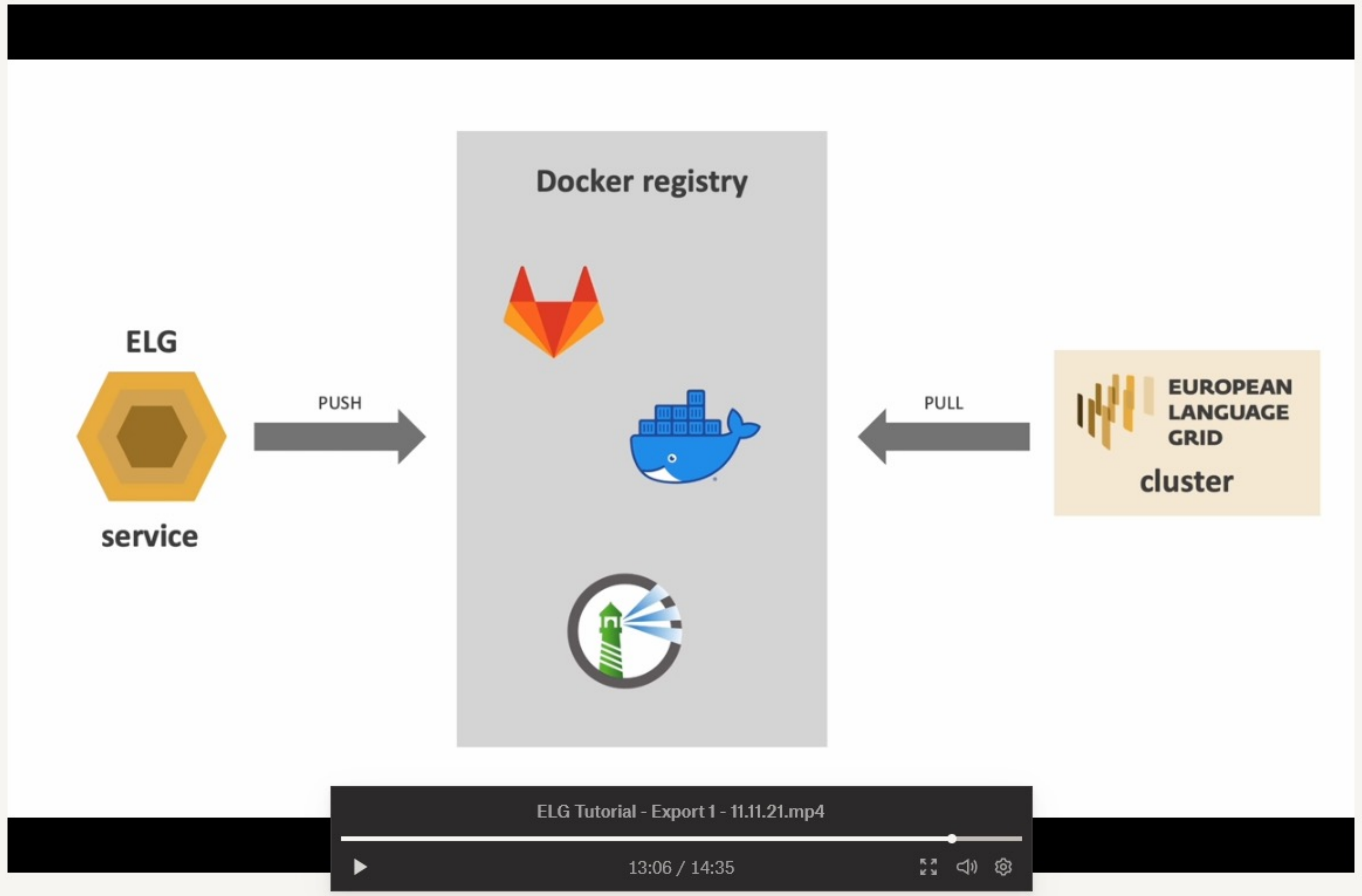
The Docker image for this service is stored at DockerHub.

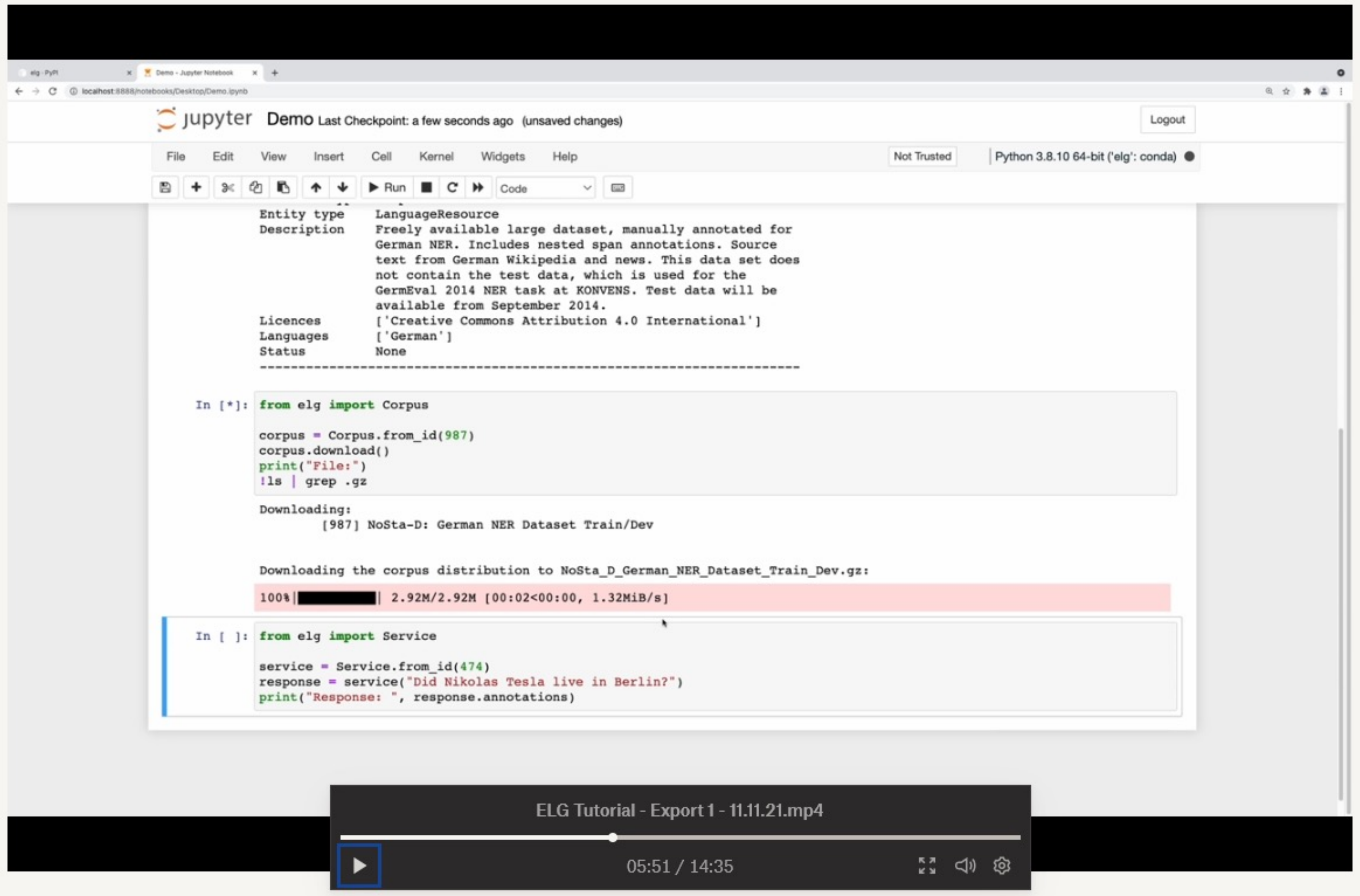
3. Register the service at ELG

The current release of ELG offers two options for registering a resource:

A. the ELG interactive editor (see Use the interactive editor)

- **Documentation**: https://european-language-grid.readthedocs.io
  - structured with the user in mind
    - Using, Contributing, Validating
  - examples and detailed technical guidance
  - recommendations on split of metadata records, data packaging and integration of services
  - continuously updated
- **Schema documentation** and ready-to-use templates and examples: https://gitlab.com/european-language-grid/platform/ELG-SHARE-schema



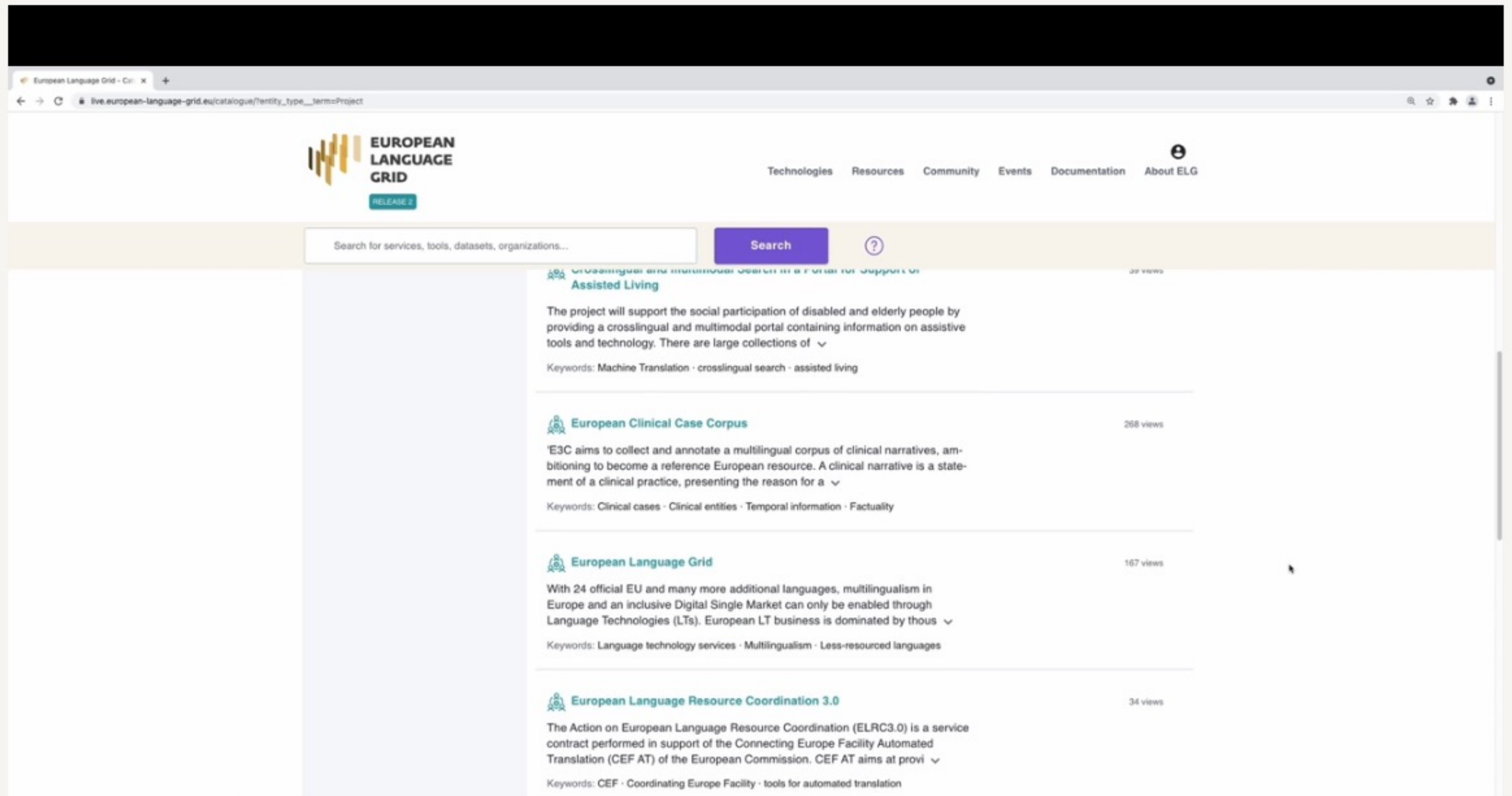

jupyter Demo Last Checkpoint: a few seconds ago (unsaved changes)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Not Trusted
Python 3.8.10 64-bit ('elg': conda)
Run
Code
ELG Tutorial - Export 1 - 11.11.21.mp4
05:51 / 14:35

```
Entity type    LanguageResource
Description    Freely available large dataset, manually annotated for
               German NER. Includes nested span annotations. Source
               text from German Wikipedia and news. This data set does
               not contain the test data, which is used for the
               GermEval 2014 NER task at KONVENS. Test data will be
               available from September 2014.
Licences       ['Creative Commons Attribution 4.0 International']
Languages      ['German']
Status         None
------------------------------------------------------------------
```

```
In [*]: from elg import Corpus

        corpus = Corpus.from_id(987)
        corpus.download()
        print("File:")
        !ls | grep .gz
```

```
Downloading:
        [987] NoSta-D: German NER Dataset Train/Dev


Downloading the corpus distribution to NoSta_D_German_NER_Dataset_Train_Dev.gz:
100%|██████████| 2.92M/2.92M [00:02<00:00, 1.32MiB/s]
```

```
In [ ]: from elg import Service

        service = Service.from_id(474)
        response = service("Did Nikolas Tesla live in Berlin?")
        print("Response: ", response.annotations)
```



Assisted Living

The project will support the social participation of disabled and elderly people by providing a crosslingual and multimodal portal containing information on assistive tools and technology. There are large collections of

Keywords: Machine Translation · crosslingual search · assisted living

European Clinical Case Corpus

268 views

'E3C aims to collect and annotate a multilingual corpus of clinical narratives, ambitioning to become a reference European resource. A clinical narrative is a statement of a clinical practice, presenting the reason for a

Keywords: Clinical cases · Clinical entities · Temporal information · Factuality

European Language Grid

167 views

With 24 official EU and many more additional languages, multilingualism in Europe and an inclusive Digital Single Market can only be enabled through Language Technologies (LTs). European LT business is dominated by thous

Keywords: Language technology services · Multilingualism · Less-resourced languages

European Language Resource Coordination 3.0

34 views

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 provi

Keywords: CEF · Coordinating Europe Facility · tools for automated translation

ELG tutorial video that explains how to make resources and tools available (https://youtu.be/29-V2EyMn4E).

EUROPEAN
LANGUAGE
EQUALITY

European Language Equality (ELE)

Objective: *development of a strategic research, innovation and deployment agenda to achieve digital language equality in Europe by 2030*

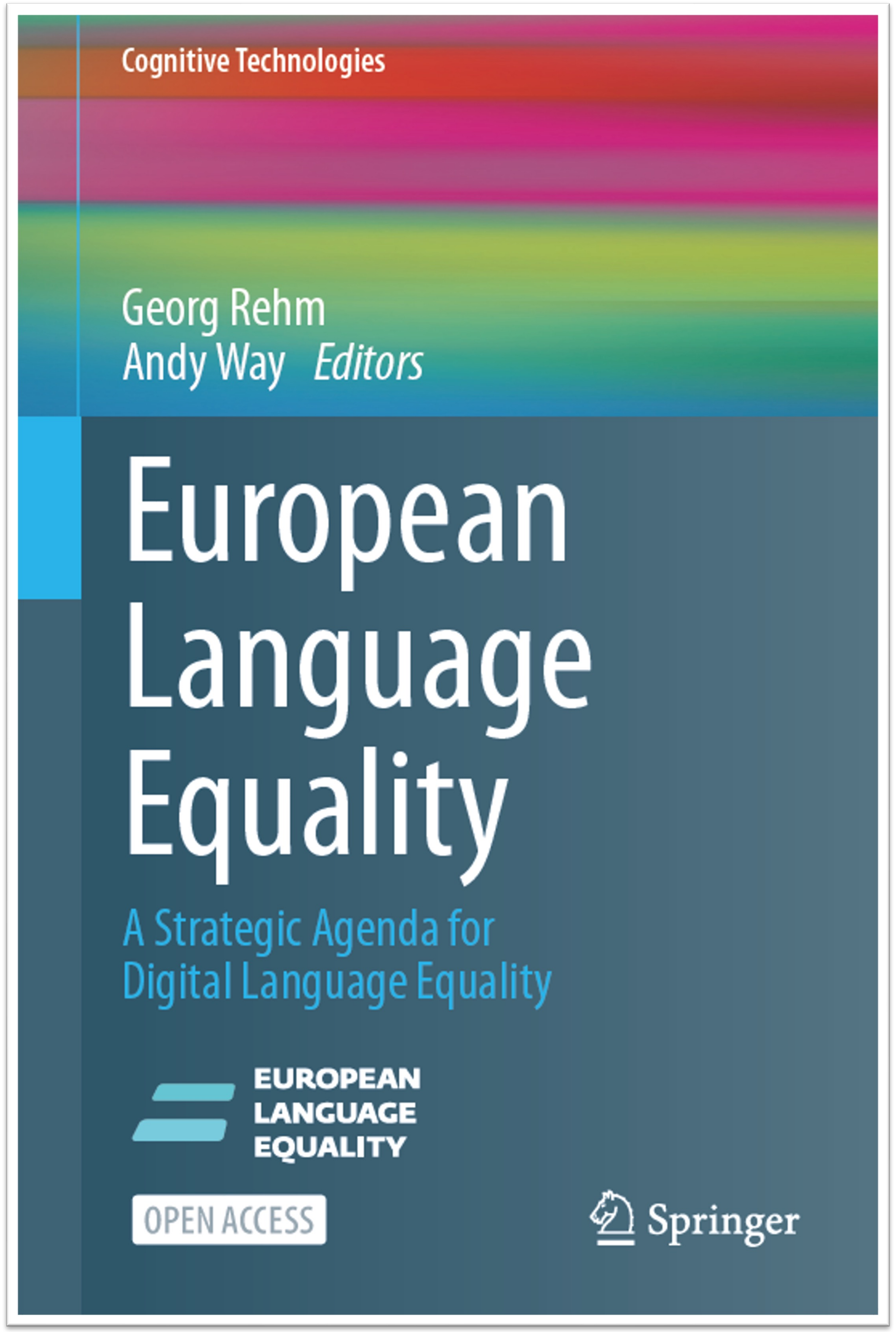
Runtime: 18 + 12 months (ELE1 and ELE2)

Ended 30 June 2023

http://www.european-language-equality.eu

ELE Some of the main outcomes and findings

- 35 language reports, >40 languages, ~100 authors (D1.4-D1.36, D1.38-D1.39) (https://european-language-equality.eu/deliverables/)
- Strategic Research, Innovation and Implementation Agenda and Roadmap (https://european-language-equality.eu/agenda/)
- ELE book (https://link.springer.com/content/pdf/10.1007/978-3-031-28819-7.pdf?pdf=button)



Cognitive Technologies

Georg Rehm
Andy Way *Editors*

European Language Equality

A Strategic Agenda for Digital Language Equality

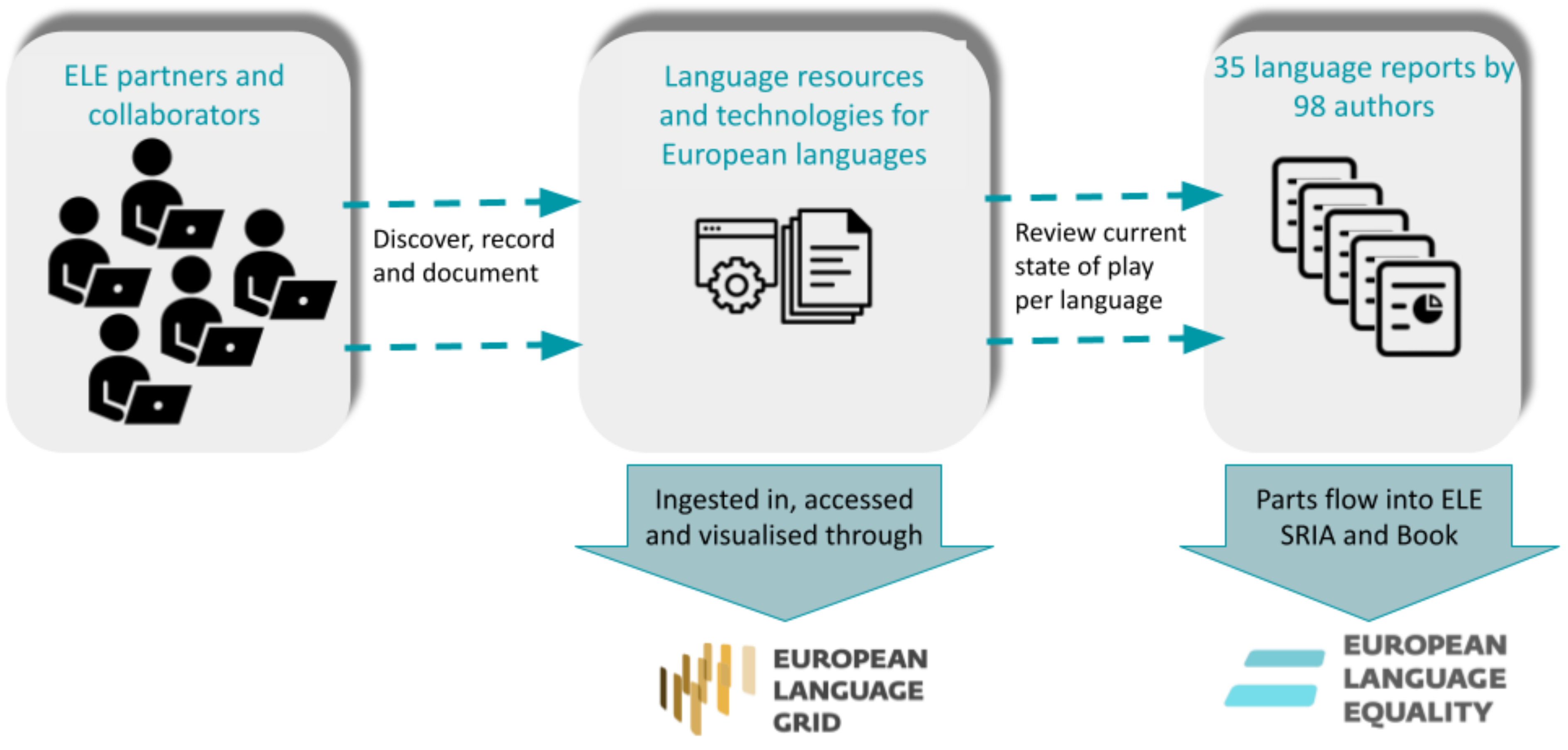
EUROPEAN LANGUAGE EQUALITY

OPEN ACCESS

Springer

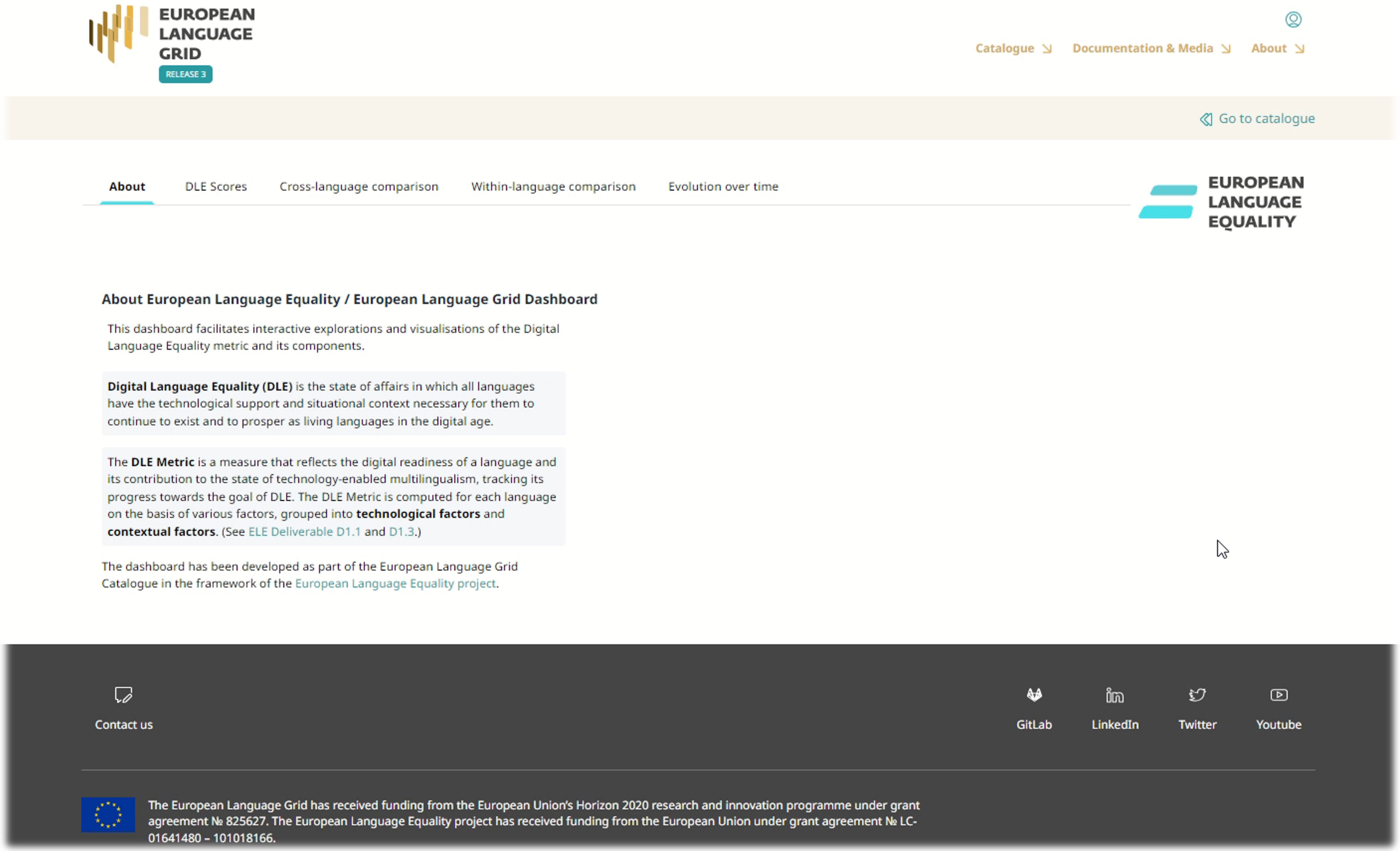
ELE Some of the main outcomes and findings

- Evidence-based investigation of the level of technology support per language
- Aggregate (meta)data through multiple strategies, including **manual population by ELE partners**
- Import these metadata to the ELG Catalogue and use the contents of the Catalogue to:
  - Compute the Digital Language Equality metric
  - DLE computations and visualisations through the DLE dashboard (https://live.european-language-grid.eu/catalogue/dashboard)


ELE partners and collaborators
Discover, record and document
Language resources and technologies for European languages
Review current state of play per language
35 language reports by 98 authors
Ingested in, accessed and visualised through
EUROPEAN LANGUAGE GRID
Parts flow into ELE SRIA and Book
EUROPEAN LANGUAGE EQUALITY

Dashboard walkaround

- DLE technological and contextual scores
- Cross-language comparisons
  - As per resource types and features
  - Through
    - Histograms
    - Heatmaps
    - Radial bars
- Within-language comparisons
- Evolution of resources over time



About European Language Equality / European Language Grid Dashboard

This dashboard facilitates interactive explorations and visualisations of the Digital Language Equality metric and its components.

Digital Language Equality (DLE) is the state of affairs in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age.

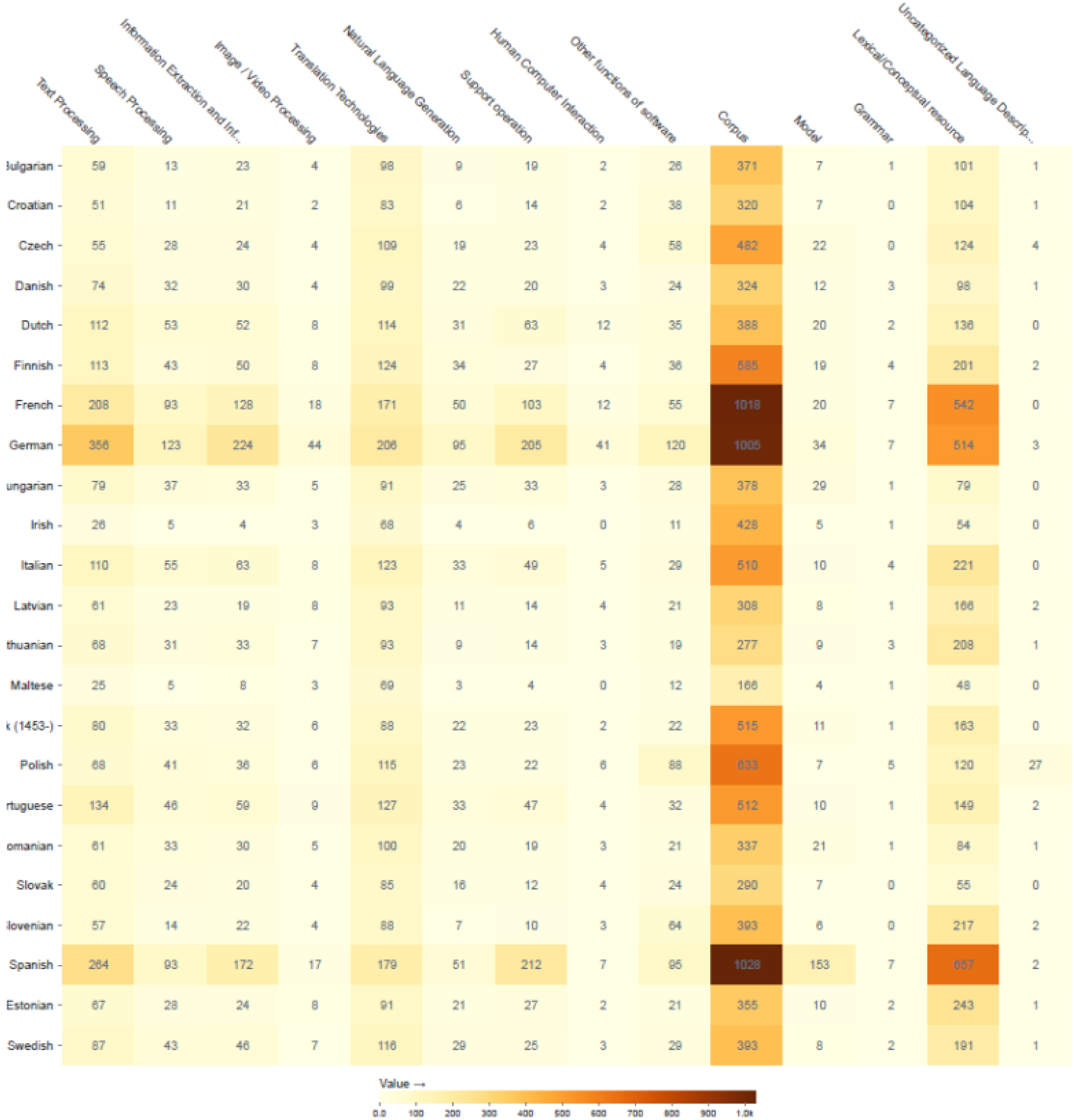
The **DLE Metric** is a measure that reflects the digital readiness of a language and its contribution to the state of technology-enabled multilingualism, tracking its progress towards the goal of DLE. The DLE Metric is computed for each language on the basis of various factors, grouped into **technological factors** and **contextual factors**. (See ELE Deliverable D1.1 and D1.3.)

The dashboard has been developed as part of the European Language Grid Catalogue in the framework of the European Language Equality project.

The European Language Grid has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 825627. The European Language Equality project has received funding from the European Union under grant agreement № LC-01641480 – 101018166.

Language resources and technologies

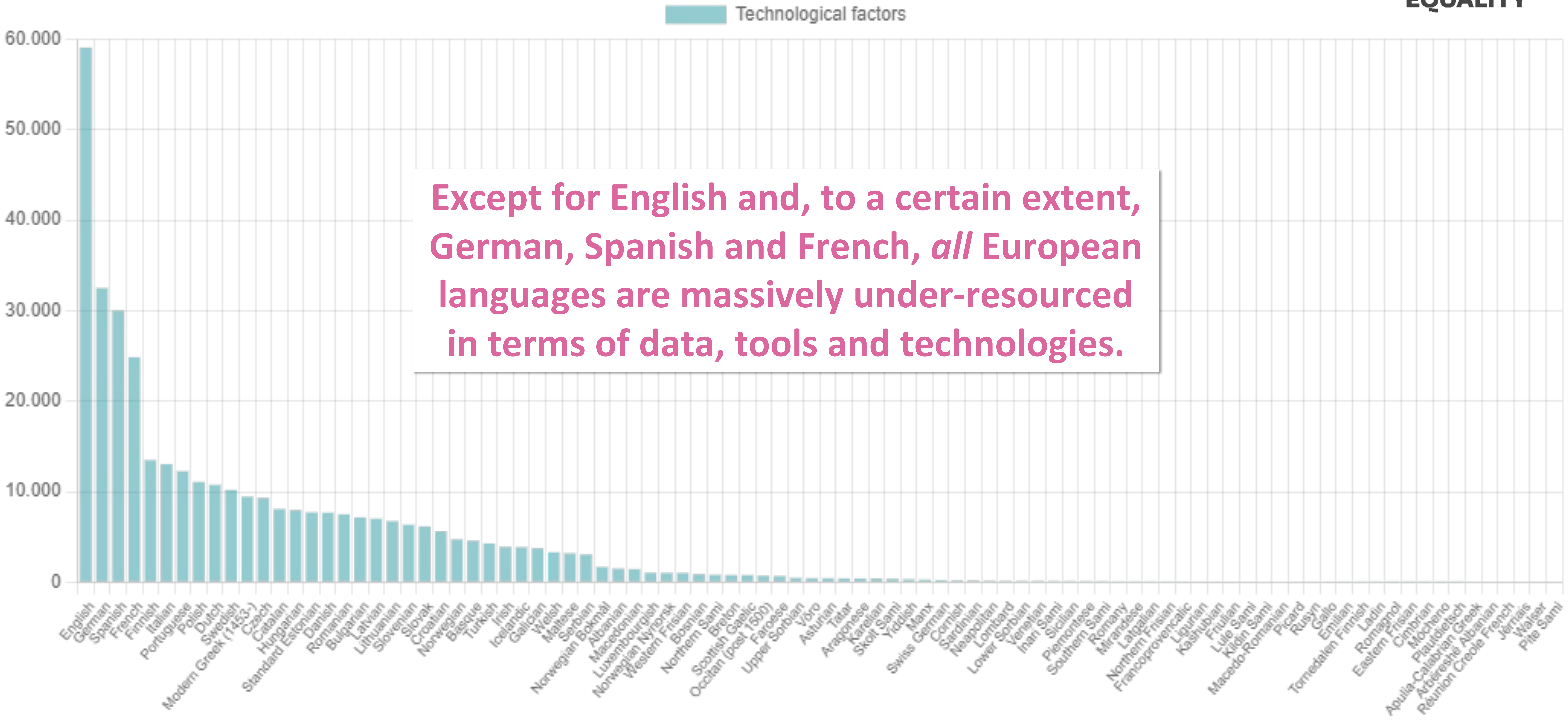
- Corpora and lexical resources are the most numerous resource types, across all languages
- All languages seem to be better supported with translation technologies and text processing tools among all tools and services functions
- The least populated technologies are HCI, NLG and image/video processing



	Text Processing	Speech Processing	Information Extraction and Inf...	Image / Video Processing	Translation Technologies	Natural Language Generation	Support operation	Human Computer Interaction	Other functions of software	Corpus	Model	Grammar	Lexical/Conceptual resource	Uncategorized Language Descrip...
ulgarian	59	13	23	4	98	9	19	2	26	371	7	1	101	1
Croatian	51	11	21	2	83	6	14	2	38	320	7	0	104	1
Czech	55	28	24	4	109	19	23	4	58	482	22	0	124	4
Danish	74	32	30	4	99	22	20	3	24	324	12	3	98	1
Dutch	112	53	52	8	114	31	63	12	35	388	20	2	136	0
Finnish	113	43	50	8	124	34	27	4	36	585	19	4	201	2
French	208	93	128	18	171	50	103	12	55	1018	20	7	542	0
German	356	123	224	44	206	95	205	41	120	1005	34	7	514	3
ungarian	79	37	33	5	91	25	33	3	28	378	29	1	79	0
Irish	26	5	4	3	68	4	6	0	11	428	5	1	54	0
Italian	110	55	63	8	123	33	49	5	29	510	10	4	221	0
Latvian	61	23	19	8	93	11	14	4	21	308	8	1	166	2
thuanian	68	31	33	7	93	9	14	3	19	277	9	3	208	1
Maltese	25	5	8	3	69	3	4	0	12	166	4	1	48	0
k (1453-)	80	33	32	6	88	22	23	2	22	515	11	1	163	0
Polish	68	41	36	6	115	23	22	6	88	633	7	5	120	27
rtuguese	134	46	59	9	127	33	47	4	32	512	10	1	149	2
omanian	61	33	30	5	100	20	19	3	21	337	21	1	84	1
Slovak	60	24	20	4	85	16	12	4	24	290	7	0	55	0
lovenian	57	14	22	4	88	7	10	3	64	393	6	0	217	2
Spanish	264	93	172	17	179	51	212	7	95	1028	153	7	657	2
Estonian	67	28	24	8	91	21	27	2	21	355	10	2	243	1
Swedish	87	43	46	7	116	29	25	3	29	393	8	2	191	1

DLE metric – technological scores

EUROPEAN LANGUAGE EQUALITY



Except for English and, to a certain extent, German, Spanish and French, *all* European languages are massively under-resourced in terms of data, tools and technologies.

ELG

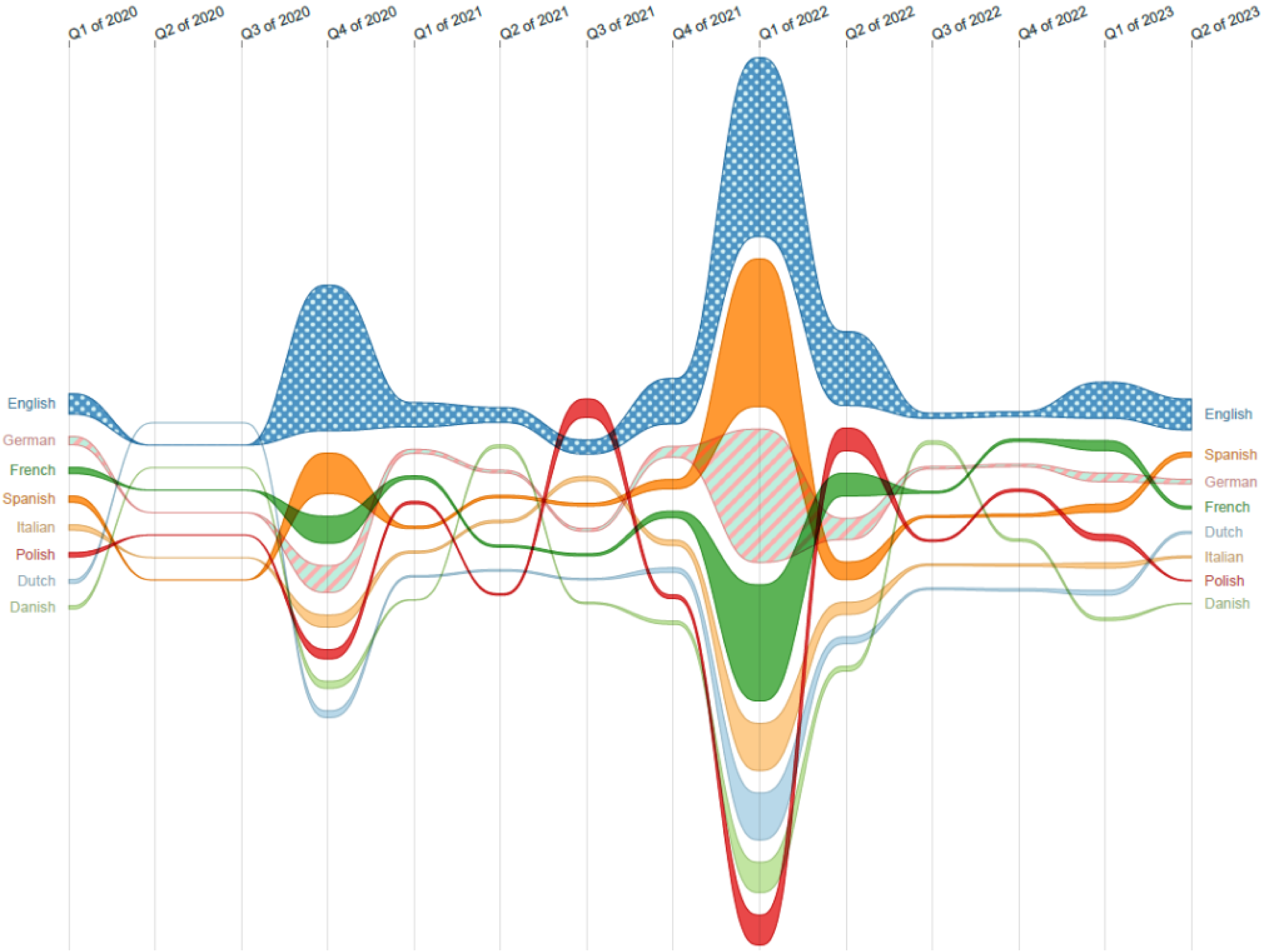
DLE metric – contextual scores

EUROPEAN
LANGUAGE
EQUALITY

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
🟦 = Official EU language
🟥 = Official national languages without being an official EU language
🟨 = Regional and minotrity languages
English
German
French
Italian
Spanish
Dutch
Swedish
Polish
Danish
Croatian
Hungarian
Greek
Czech
Finnish
Norwegian
Slovak
Portuguese
Romanian
Estonian
Icelandic
Slovene
Bulgarian
Latvian
Lithuanian
Serbian
Irish
Luxembourgish
Maltese
Albanian
Faroese
Turkish
Alsatian
Saami, South
Bosnian
Catalan
Frisian, Western
Romani
Saami, Lule
Welsh
Jerriais
Karelian
Saami, North
Saami, Inari
Saami, Skolt
Scottish Gaelic
Vöro
Yiddish
Basque
Galician
Romagnol
Rusyn
Sorbian, Upper
Walser
Frisian, Eastern
Frisian, Northern
Kashubian
Latgalian
Macedonian
Réunion Creole
Saami, Pite
Tornedalian Finnish
Aragonese
Asturian
Mirandese
Occitan
Southern Italian
Aromanian
Friulian
Ligurian
Lombard
Mocheno
Picard
Plautdietsch
Sardinian
Sicilian
Sorbian, Lower
Tatar
Venetian
Arberesh
Breton
Cimbrian
Cornish
Ladin
Emilian
Franco-Provencal
Griko
Manx
Piemontese
Gallo
Saami, Kildin

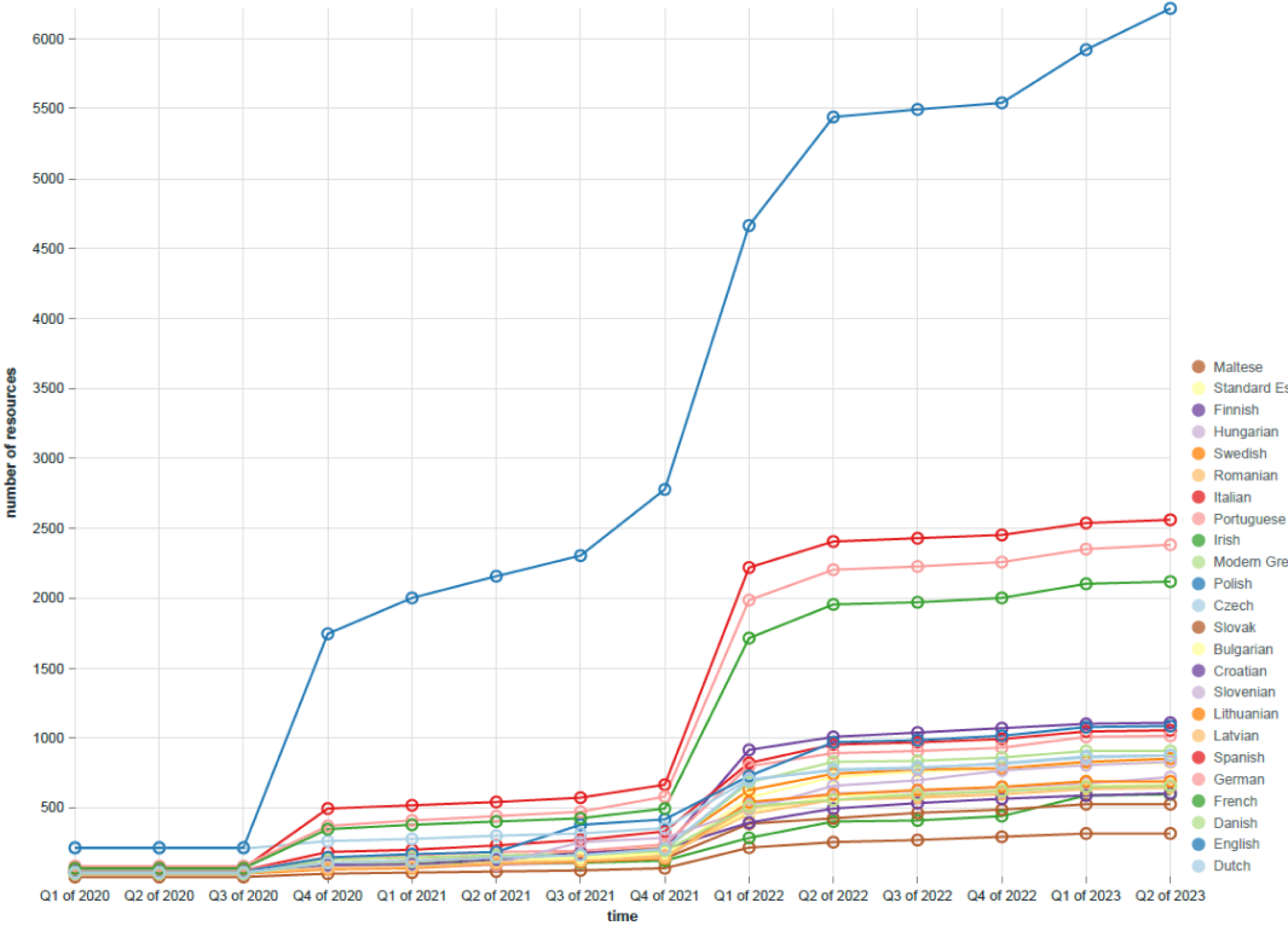
Nearly all European languages exist in a socio-political ecosystem that does *not* incentivise, encourage or foster the development of technologies for these languages.

Evolution over time (1/2)



Evolution over time (2/2)

- All languages are progressing
- English shows steeper progress in Q1-Q2 2023
- Distance has increased



EUROPEAN
LANGUAGE
GRID

EUROPEAN
LANGUAGE
EQUALITY

Thank you!

Maria Giagkou
Institute for Language and Speech Processing / “Athena” Research and Innovation Center
www.ilsp.gr / www.athenarc.gr
mgiagkou@athenarc.gr