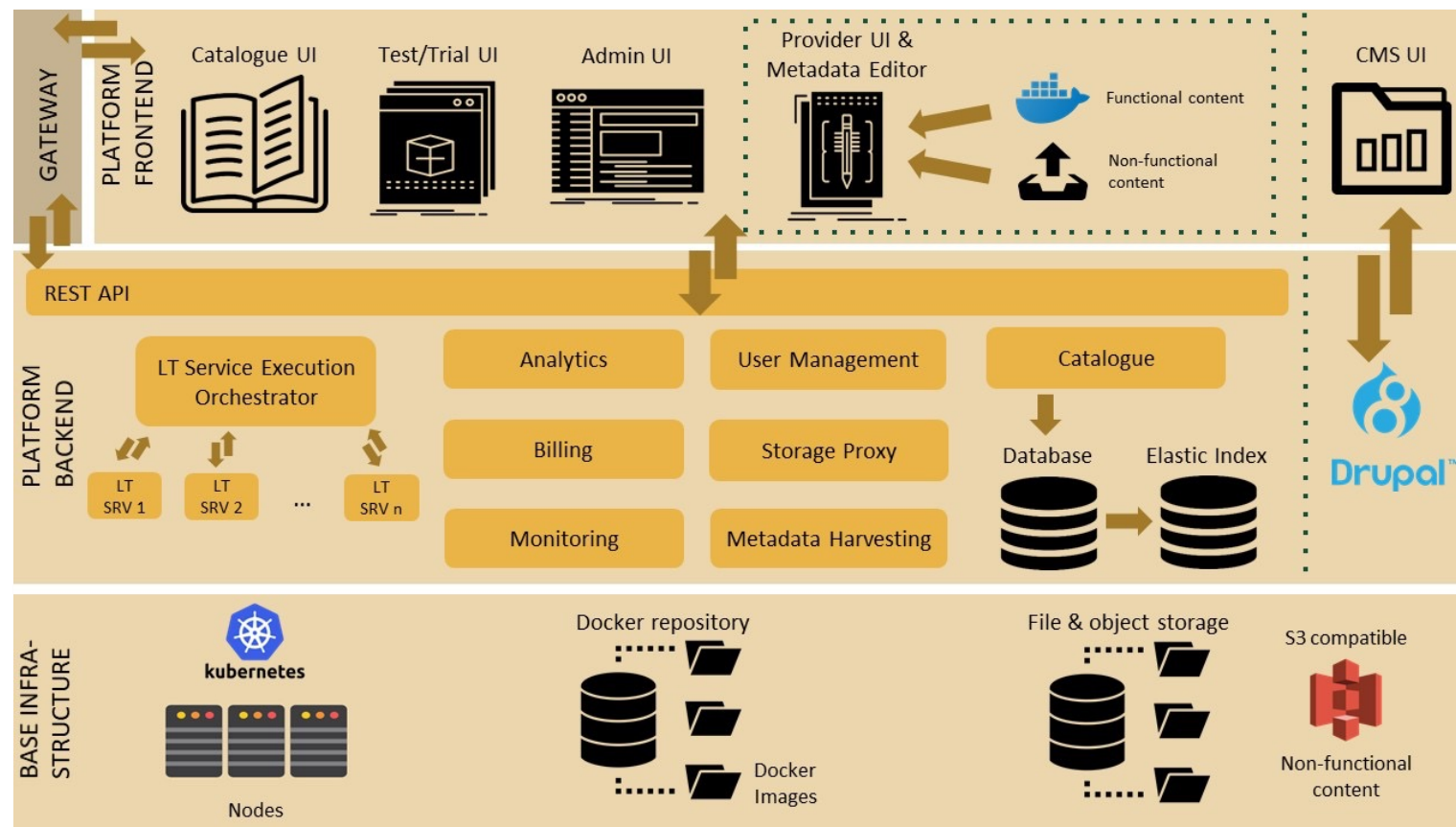


European Language Grid – META-FORUM 2022
Stelios Piperidis: European Language Grid – Platform

- ELG Platform Architecture
- ELG for LT consumers
- ELG for Providers of LT
- ELG in the wider LT and AI Ecosystem

ELG Platform Architecture

- Robust scalable, reliable, widely used technologies
- Docker containers for all LT services and ELG platform components
- Kubernetes for container orchestration
- Scalable with the growing demand and supply of resources
- Flexible, allowing for integration of new functionalities
- Foundations for interoperable data and services spaces



- search and browse the ELG Catalogue
 - for different types of language processing services and data
 - related projects and organisations in Europe
 - using simple and advanced free text search
 - using facets for resource type, language, service function, media type, conditions of use, hosted or not, license, related entities, source
- View detailed information on landing pages
- View statistics of number of views and downloads (for ELG hosted resources)
- Download data in accordance with licensing terms
- Export metadata records in multiple formats
- View which datasets they have downloaded, their versions, ...

ELG Data Consumers can ...

The screenshot displays the ELG Catalogue interface. At the top, the 'EUROPEAN LANGUAGE GRID' logo is visible. The main search results for 'Italian newspaper' show 51 results, with the first two being 'CHANGE-IT dataset' and 'MLCC Multilingual and Parallel Corpora'. The 'CHANGE-IT dataset' is highlighted, showing its version (1.0.0) and a description: 'The CHANGE-IT dataset contains approximately 152,000 article-headline pairs, collected from two Italian newspapers situated at opposite ends of the political spectrum, namely la Repubblica (left) and Il Giornale (right)'. Below this, there are tabs for 'Overview' and 'Download'. The 'Distribution' section is expanded, showing the dataset is downloadable, has a text feature, a size of 43,328 text, and a data format of text/plain. The license is Creative Commons Attribution Non Commercial Share Alike 4.0 International, with links to the license pages. At the bottom, there is a 'Cite resource' section with the citation: 'De Mattei, Lorenzo; Gatt, Albert; Dell'Orletta, Felice; Nissim, Malvina; Cafagna, Michele (2021). CHANGE-IT dataset. Version 1.0.0. [Dataset (Text corpus)]. De Mattei, Lorenzo; Gatt, Albert; Dell'Orletta, Felice; Nissim, Malvina; Cafagna, Michele https://doi.org/10.57771/j8ah-m778'. There is also a 'Cite all versions' section with a similar citation. The interface includes a 'Go to catalogue' button and a 'Share' section with social media icons.

Consumers of ELG Services can ...

The screenshot displays the ELG Consumer's Grid interface. At the top, there's a search bar and navigation links like 'My grid', 'Test Consumer', 'Catalogue', 'Documentation & Media', and 'About'. Below the search bar, a sidebar on the left allows filtering by 'Language resources & technologies' and 'Service functions'. The main area shows '78 search results' for 'Named Entity Recognition'. Two services are highlighted: 'BERTNER English' (version 1.0.0) and 'Cogito Discover Named Entity Recognizer' (version 14.3.0). The 'Cogito' service page is expanded, showing an overview, download/run options, and code samples. The code samples include a cURL command, a Python snippet for using the ELG SDK, and a command-line interface for running the service.

Search results for Named Entity Recognition:

- BERTNER English** (version: 1.0.0)
The Named Entity Recognise entities: (1) based o has been implemented i
Keywords: Named Entity Re
Language: English
Licence: GNU Affero Gener
- BERTNER German** (version: 1.0.0)
The Named Entity Recognise entities: (1) based o has been implemented i

Cogito Discover Named Entity Recognizer (version: 14.3.0)
ELG compatible service

Code samples:

```
cURL
curl -X POST -H 'Content-Type: application/json' -d '{"text": "The G-20 Summit hosted by Italy kicked off Saturday in Rome, where leaders from the world's major economies discussed issues of mutual concern, including pandemic recovery and climate change. The red carpet was rolled out at "La Nuvola," Rome's convention center, as Italian Prime Minister Mario Draghi welcomed U.S. President Joe Biden and other leaders amid strict COVID-19 protocols. On day one, G-20 leaders voiced their support for a global corporate minimum tax deal agreed to by finance ministers from 136 countries earlier this month after four years of negotiations led by the Organization for Economic Cooperation and Development."}' http://localhost:8080/api/v1/ner

Python
from elg import Service
service = Service.from_id(474)
result = service(request_input="input string, ELG request object, or path to a file", request_type="text")

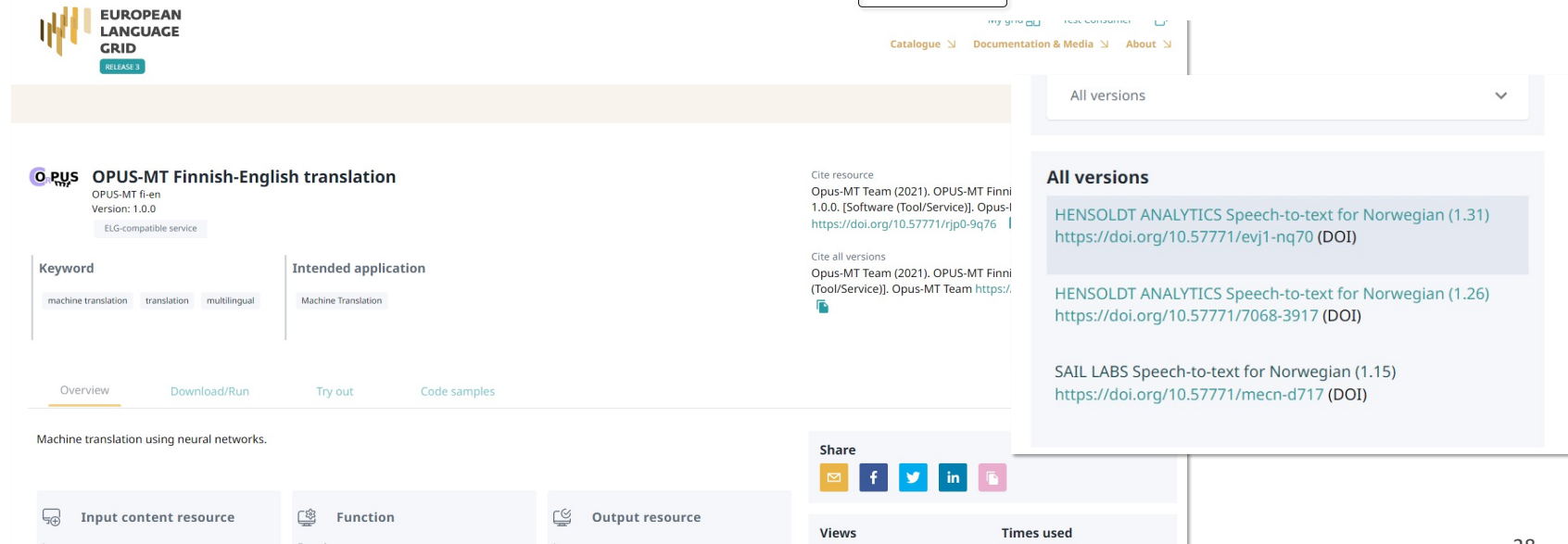
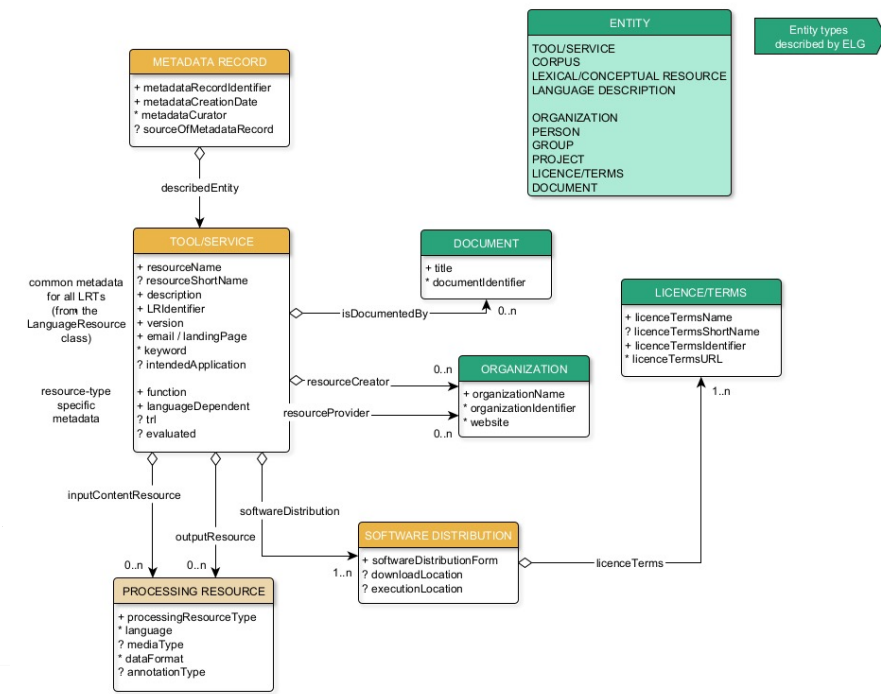
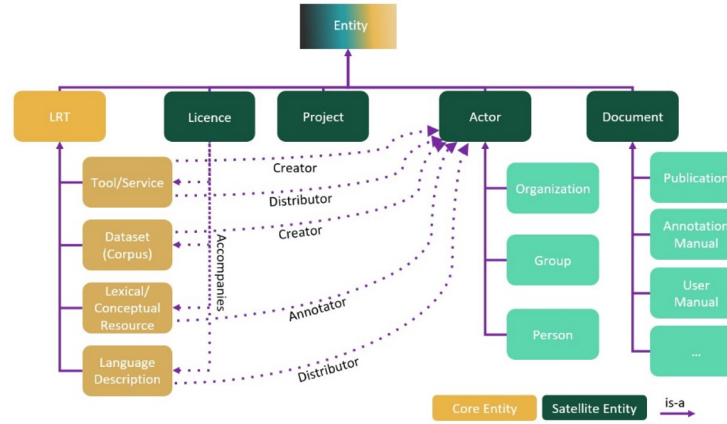
The ELG Python SDK can be installed using pip:
pip install elg

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

- Try out and test language processing services
 - registration/authentication is required
- Call a service from the command line directly (via its common REST API) and integrate it in their own workflows
- Current APIs support
 - information extraction & annotation for text & speech
 - text-to-text generation (machine translation, summarization)
 - classification for text and speech
 - speech recognition
 - speech synthesis
 - image recognition
- Python-based API for accessing the ELG catalogue, searching and directly fetching datasets to feed them into their own pipelines, running ELG services
- Get an overview of their daily usage, quotas, downloads in the **ELG Consumer's Grid**

LT Providers can ...

- register and get authenticated as providers, ensuring **trustability**
- **document their assets** by providing formal descriptions compliant with a dedicated metadata schema
 - describe & provide access (remote or integrated in ELG **LRTs**)
 - describe **organisations & projects**
 - **linked with each other**
 - Get a **persistent identifier (DOI)**
 - Provide **new versions of a dataset or tool/service**



LT Providers can ...

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

EUROPEAN LANGUAGE GRID
RELEASE 3

My grid My items Feedback

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.

Grammar
Computational grammars, rule-based descriptions

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

EUROPEAN LANGUAGE GRID
RELEASE 3

My grid My items Feedback

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

Daily usage

calls remaining: 1%

bytes (for processing text) remaining: 1%

bytes (for processing binary data) remaining: 1%

MY ITEMS MY USAGE MY DOWNLOADS

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

551 search results

Items	Count
+ LicenceTerms	(394)
+ Tool/Service	(74)
+ Person	(59)
+ Project	(11)
+ Organization	(8)
+ Corpus	(5)

Status	Count
+ published	(481)
+ syntactically valid	(64)
+ draft	(4)
+ submitted	(2)

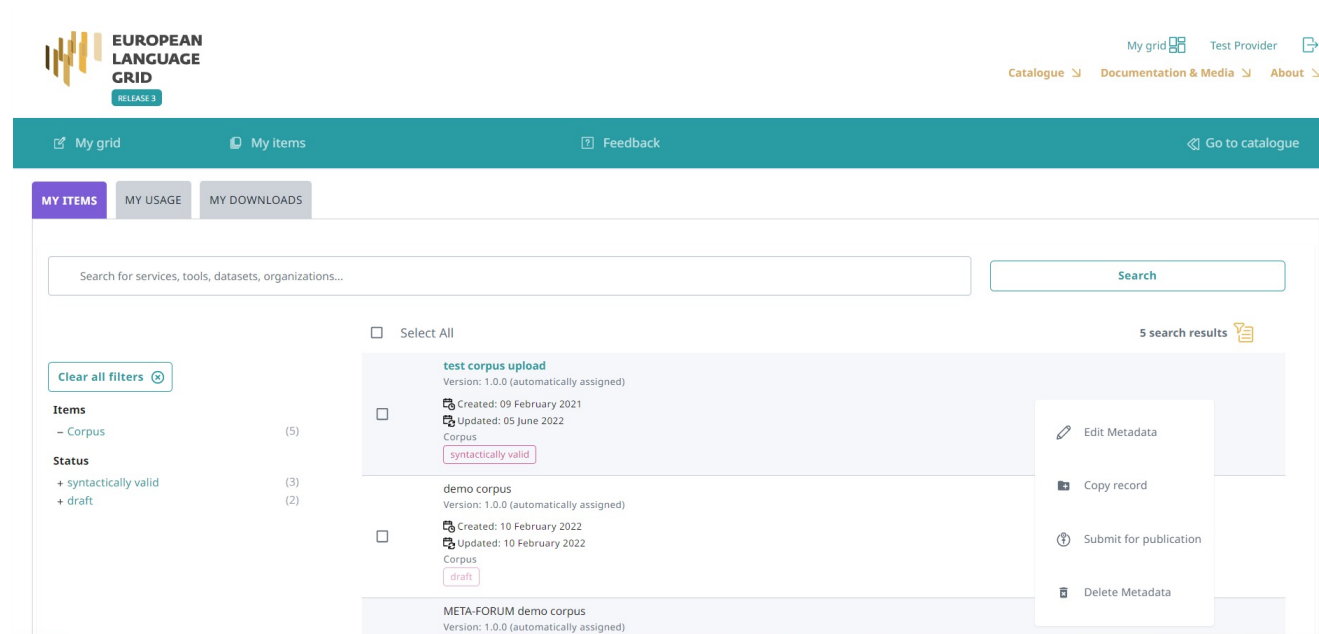
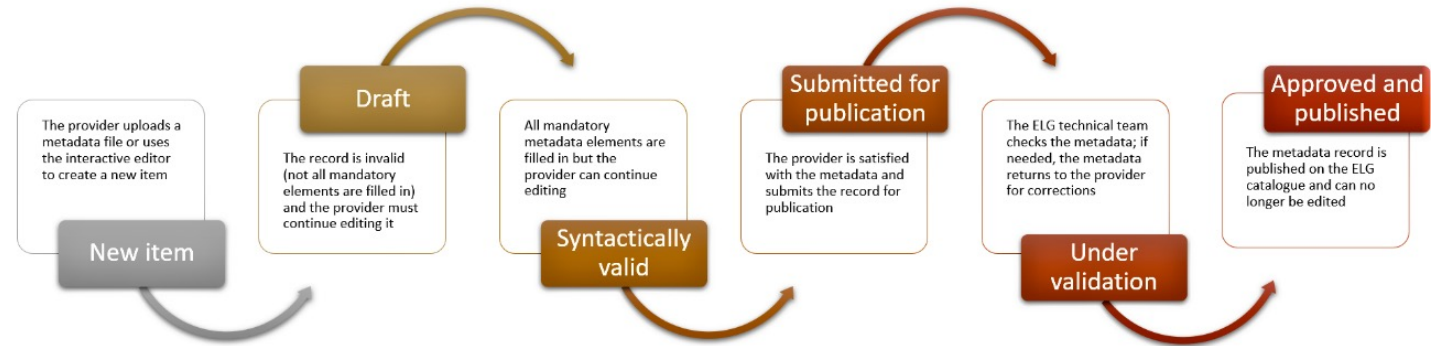
test corpus upload
Version: 1.0.0 (automatically assigned)
Created: 09 February 2021
Updated: 05 June 2022
Corpus
syntactically valid

Evaluation and Language Resources Distribution Agency
Created: 12 March 2020
Updated: 01 June 2022
Organization
published

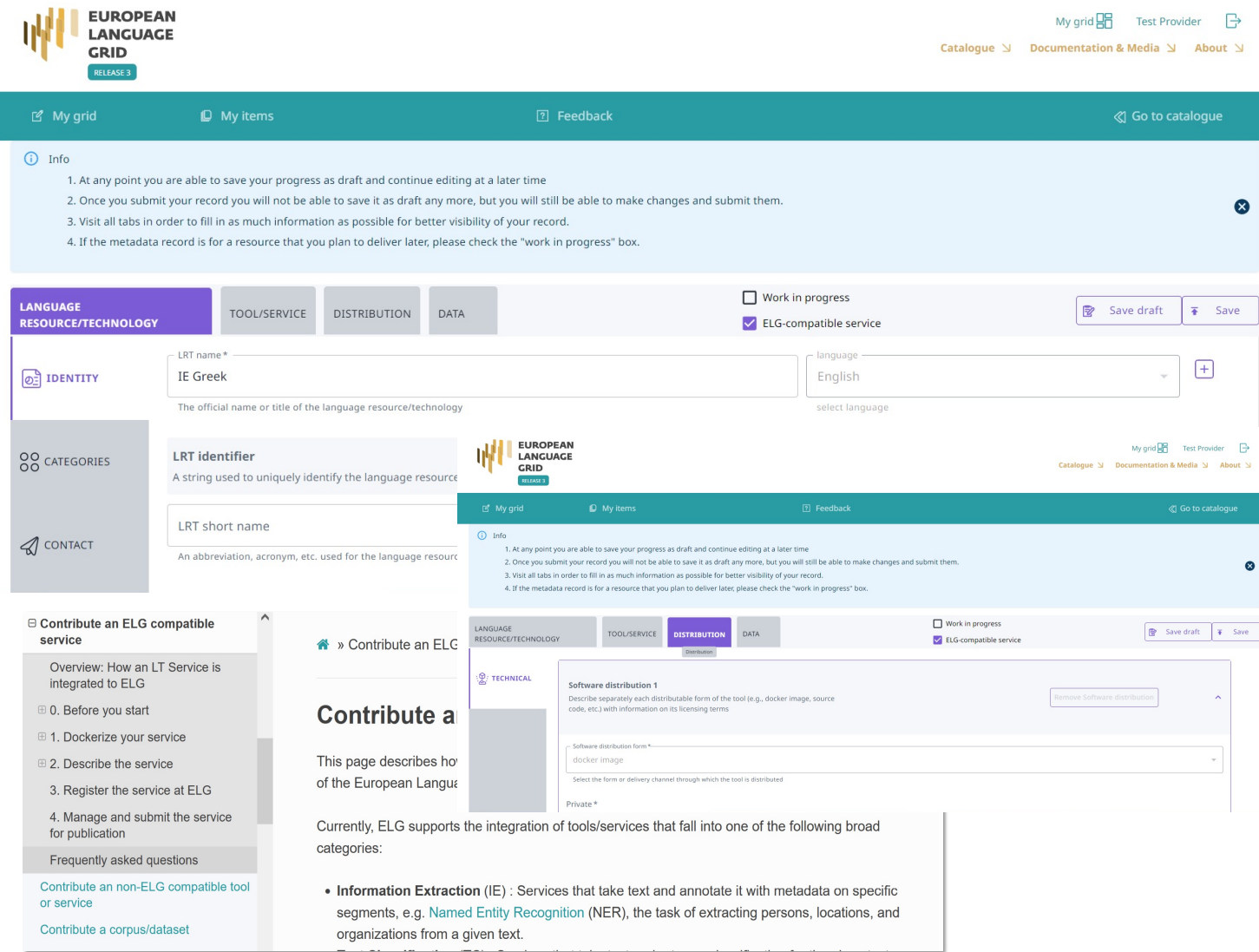
European Language Grid
Created: 28 February 2020

Provider's Grid

- LT Providers assisted by a rigorous publication process (**ELG publication lifecycle**) for
- **managing** the descriptions of their assets/items (data, resources, services, etc.)
- uploading, managing files
- through a dedicated page (“my items”)
- creating new versions, copying records, exporting records in multiple formats



A few Details on Service Provision



The screenshot shows the European Language Grid (ELG) service provision interface. At the top, there's a header with the ELG logo and navigation links: 'My grid', 'Test Provider', 'Catalogue', 'Documentation & Media', and 'About'. Below the header is a teal navigation bar with 'My grid', 'My items', 'Feedback', and 'Go to catalogue'. A light blue info box contains instructions: 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 "work in progress" box.

The main form is divided into tabs: 'LANGUAGE RESOURCE/TECHNOLOGY', 'TOOL/SERVICE', 'DISTRIBUTION', and 'DATA'. The 'LANGUAGE RESOURCE/TECHNOLOGY' tab is active, showing fields for 'LRT name*' (IE Greek), 'language' (English), 'LRT identifier', and 'LRT short name'. There are checkboxes for 'Work in progress' and 'ELG-compatible service'. Buttons for 'Save draft' and 'Save' are present. A sidebar on the left contains 'CATEGORIES' and 'CONTACT'. A bottom sidebar shows a list of steps: 'Contribute an ELG compatible service', 'Overview: How an LT Service is integrated to ELG', '0. Before you start', '1. Dockerize your service', '2. Describe the service', '3. Register the service at ELG', '4. Manage and submit the service for publication', and 'Frequently asked questions'. The 'DISTRIBUTION' tab is also visible, showing a 'Software distribution 1' section with a 'Software distribution form*' dropdown set to 'docker image' and a 'Private*' checkbox.

- LT service providers need to provide a **Docker image** with their LT tool or service
- Docker images have to be uploaded to a **Docker registry** (e.g., GitLab, DockerHub)
- Different options for providing LT services
 - Hosted in ELG – LT tool packaged in one Docker image exposing an ELG-compliant endpoint, running in the ELG cloud
 - LT tool running outside the ELG infrastructure – proxy image with one or more ELG-compliant endpoints

ELG Support Services

The screenshot displays the European Language Grid (ELG) documentation website. The main header includes the ELG logo and the text 'Welcome to European Language Grid Documentation'. A sidebar on the left contains a search bar and a list of navigation links. The main content area is titled 'Contribute an ELG compatible service' and includes a 'Note' section, a list of steps for contributing a service, and a section for 'Examples' with links to 'Information Extraction service' and 'Machine Translation service'. The sidebar also lists various annexes and projects.

European Language Grid

» Welcome to European Language Grid Documentation

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 release.

» Contribute an ELG compatible service

Contribute an ELG compatible service

This page describes how to contribute a language technology service to run on the cloud platform of the European Language Grid.

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.

Information Extraction service

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.

Machine Translation service

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](#))

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 POLICIES

Terms of use

Publication lifecycle

ANNEX 5: PUBLICATIONS AND REPORTS

Publications and reports

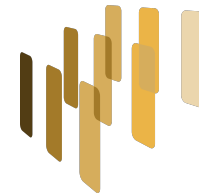
ANNEX 6: OPEN CALL PROJECTS

Instructions for the registration of metadata records

- **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>
- **Provide feedback** and get help:
<https://gitlab.com/european-language-grid/platform/elg-platform>
- **Helpdesk:** <https://www.european-language-grid.eu/contact/>

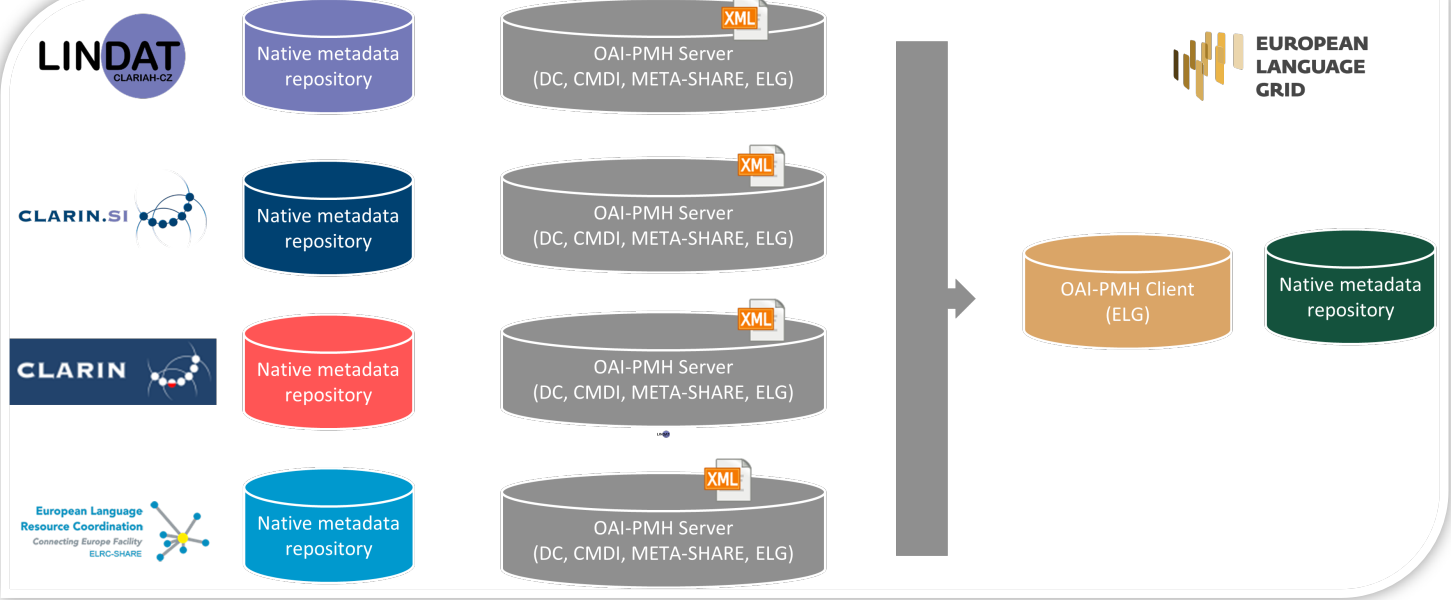
ELG in the wider LT and AI Ecosystem

- ELG has built **bridges** to 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

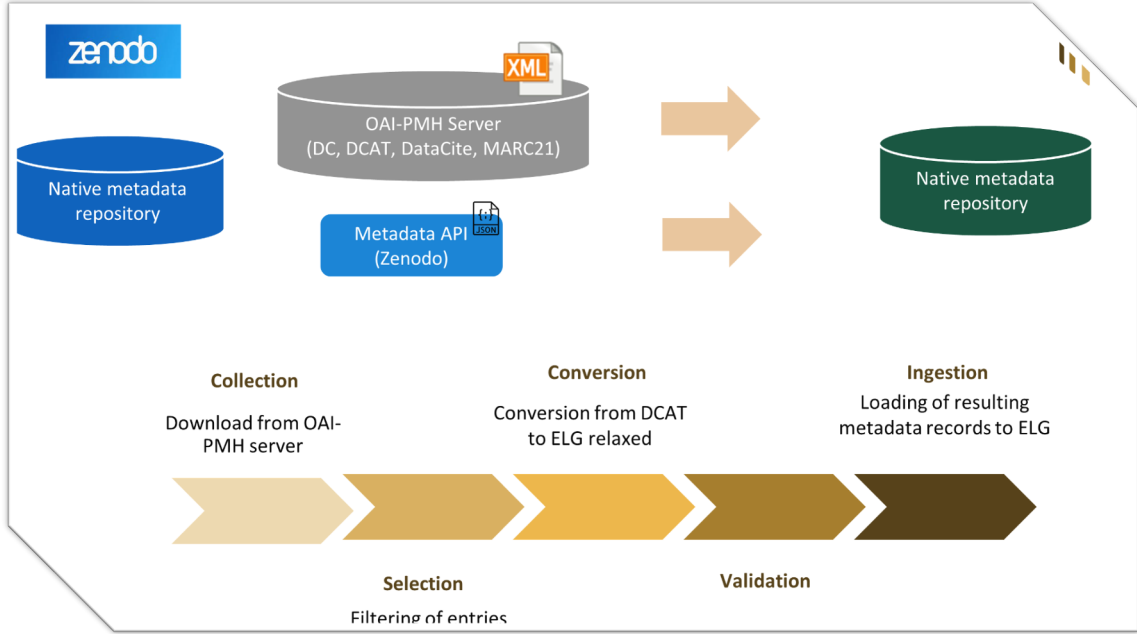
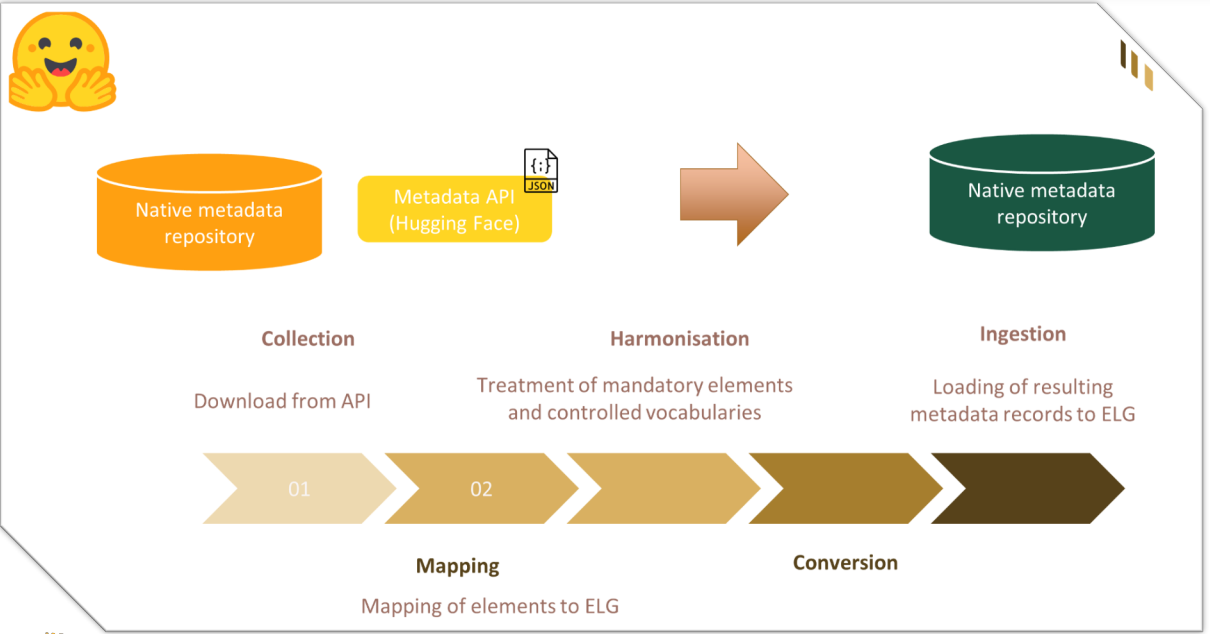


Hugging Face





ELG Bridging Strategies

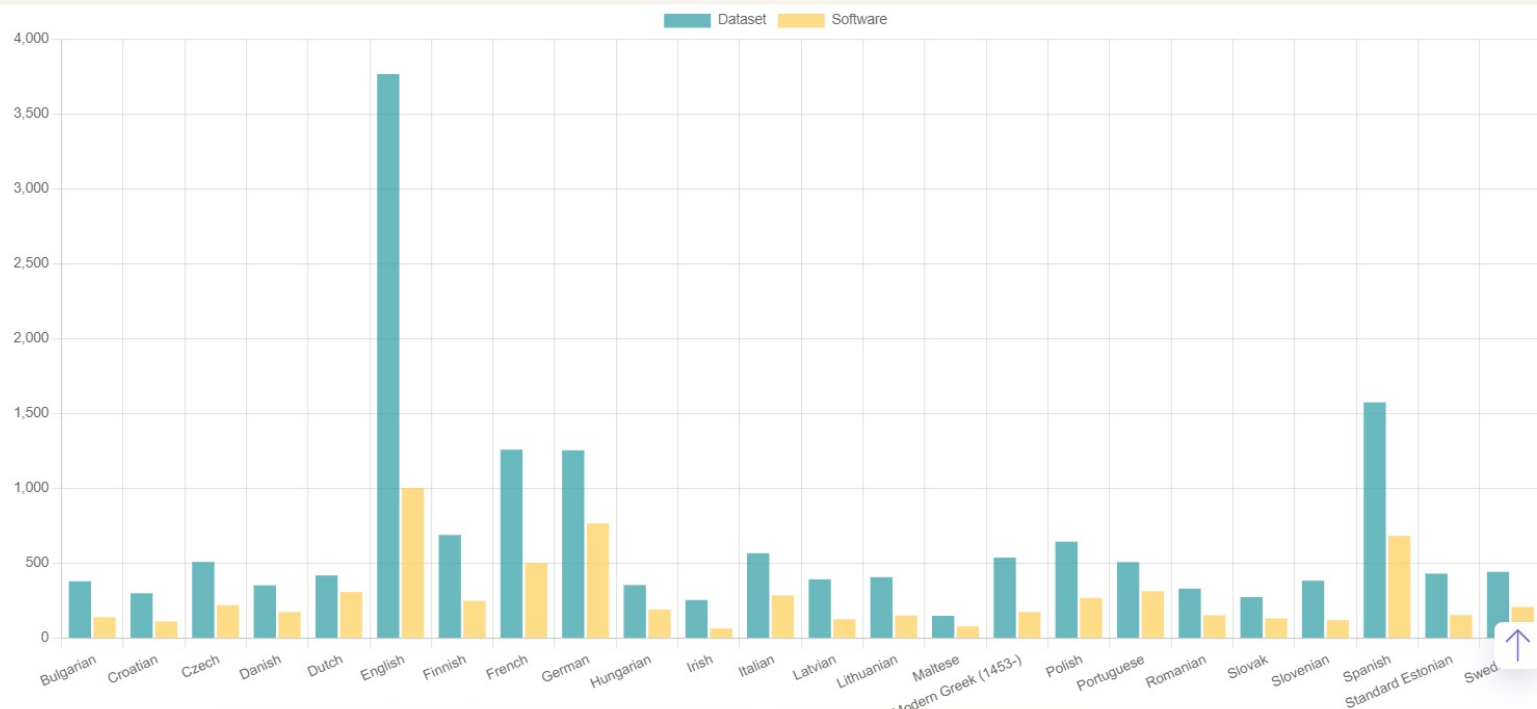
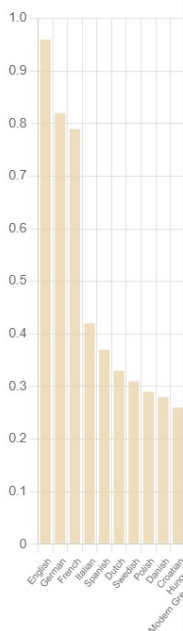
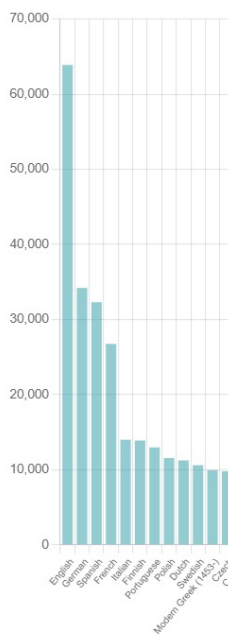


A living dynamic LT Observatory

live.european-language-grid.eu/catalogue/dashboard



- ☒ Bulgarian
- ☒ Croatian
- ☒ Czech
- ☒ Danish
- ☒ Dutch
- ☒ English
- [Show more](#)
- Other European languages**
- ☐ Select all other
- ☒ Albanian
- ☒ Apulia-Calabrian Greek
- ☒ Aragonese
- ☒ Arbëreshë Albanian
- ☒ Asturian



➔ More about the ELE dashboard tomorrow