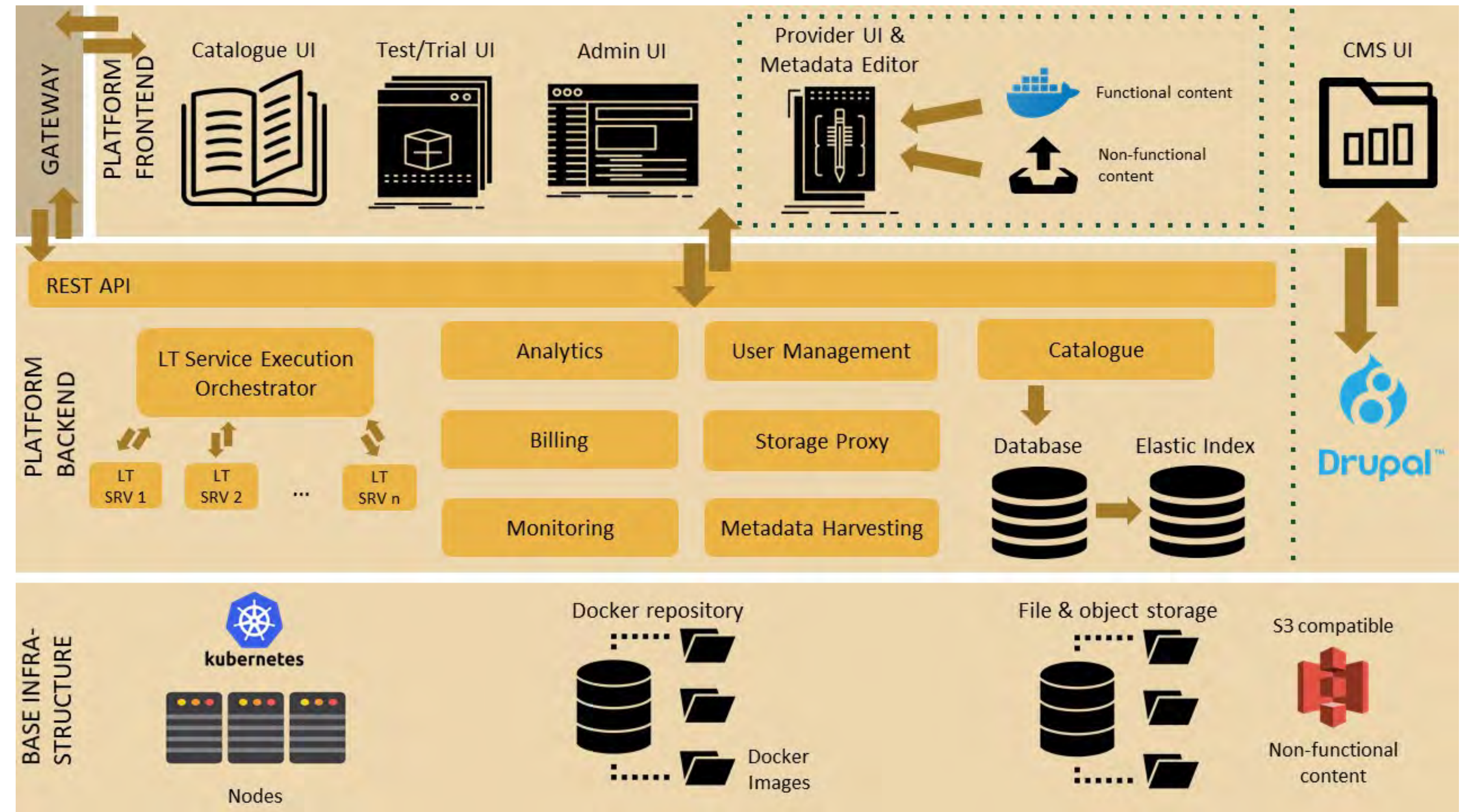


European Language Grid – META-FORUM 2021
Stelios Piperidis: European Language Grid – Platform

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

ELG Platform Architecture

- All layers built with robust scalable, reliable, widely used technologies
- Docker containers for all services and applications which comprise the ELG platform
- Kubernetes for container orchestration
- Ability to scale with the growing demand and supply of resources
- Laying the foundations for interoperable data and services spaces



- Consumers can 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, intended application, conditions of use (for data resources), license, related entities
- View detailed information (metadata)
- View statistics of number of views and downloads (for ELG hosted resources)
- Download data (depending on access conditions)
- Export metadata records
- Check what is forthcoming in terms of data and services

On the data consumer side

The screenshot displays the European Language Grid (ELG) website interface. The top navigation bar includes links for Technologies, Resources, Community, Events, Documentation, and About ELG. A search bar is prominently featured at the top right.

The main content area shows search results for 'Bilingual documents Bulgarian-English in the field of transport (Processed)'. The results are organized into facets: Keyword (parallel corpus, machine translation, elrc data), Domain (TRANSPORT, INTERNATIONAL RELATIONS), and Corpus subclass (raw corpus). The 'Corpus part' section shows a 'TEXT' resource with a 'Linguality type' of 'bilingual' and a 'Multilinguality type' of 'multilingual single text'. The 'Language' facet shows 'English' and 'Bulgarian', and the 'Text genre' facet shows 'official'.

Below the search results, there is a section for 'Ethics' with the following information: Personal data included: false, Sensitive data included: false, Anonymized: false.

The 'Licence' section shows 'Creative Commons Attribution 4.0 International'.

The 'Universal Semantic Annotator' section shows a 'work in progress' status and a 'Functional service' label. The description states: 'Universal Semantic Annotator produces annotations for Word Sense Disambiguation, Semantic Role Labelings and Semantic Parsing (Abstract Meaning Representation) in multiple languages.' The keywords are 'Semantic Parsing', 'Semantic Role Labeling', and 'Word Sense Disambiguation'. The licence is 'CC-BY-NC-SA + I.DC99T42'.

The 'Semantic Systems S.A.' section shows a '0 views' status.

The 'Le Monde Diplomatique' section shows three corpora: 'Arabic tagged corpus' (117 views), 'Text corpus in Arabic' (201 views), and 'Text corpus in English' (39 views). Each corpus entry includes a description, keyword, language, and license.

The 'Creative Commons Attribution Non Commercial Share Alike 4.0 International' license is shown with the URL <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode> and <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

The 'Cost' section shows '0 euro'.

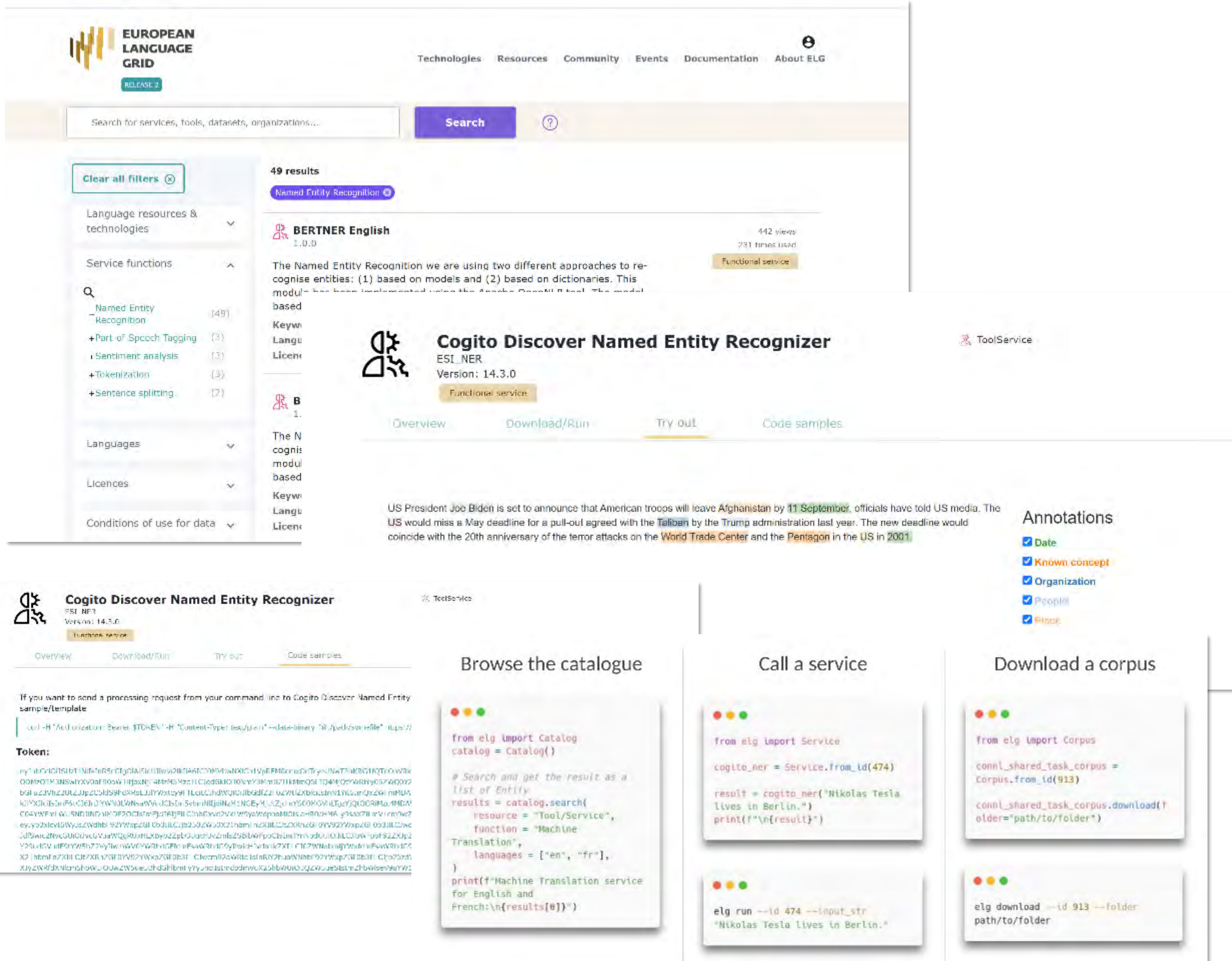
The 'Membership institution' section shows 'ELRA'.

The 'Availability' section shows '05/07/2019 -'.

The 'Distribution rights holder' section shows 'ELRA' and a 'Website' link.

A 'Download' button is visible at the bottom right.

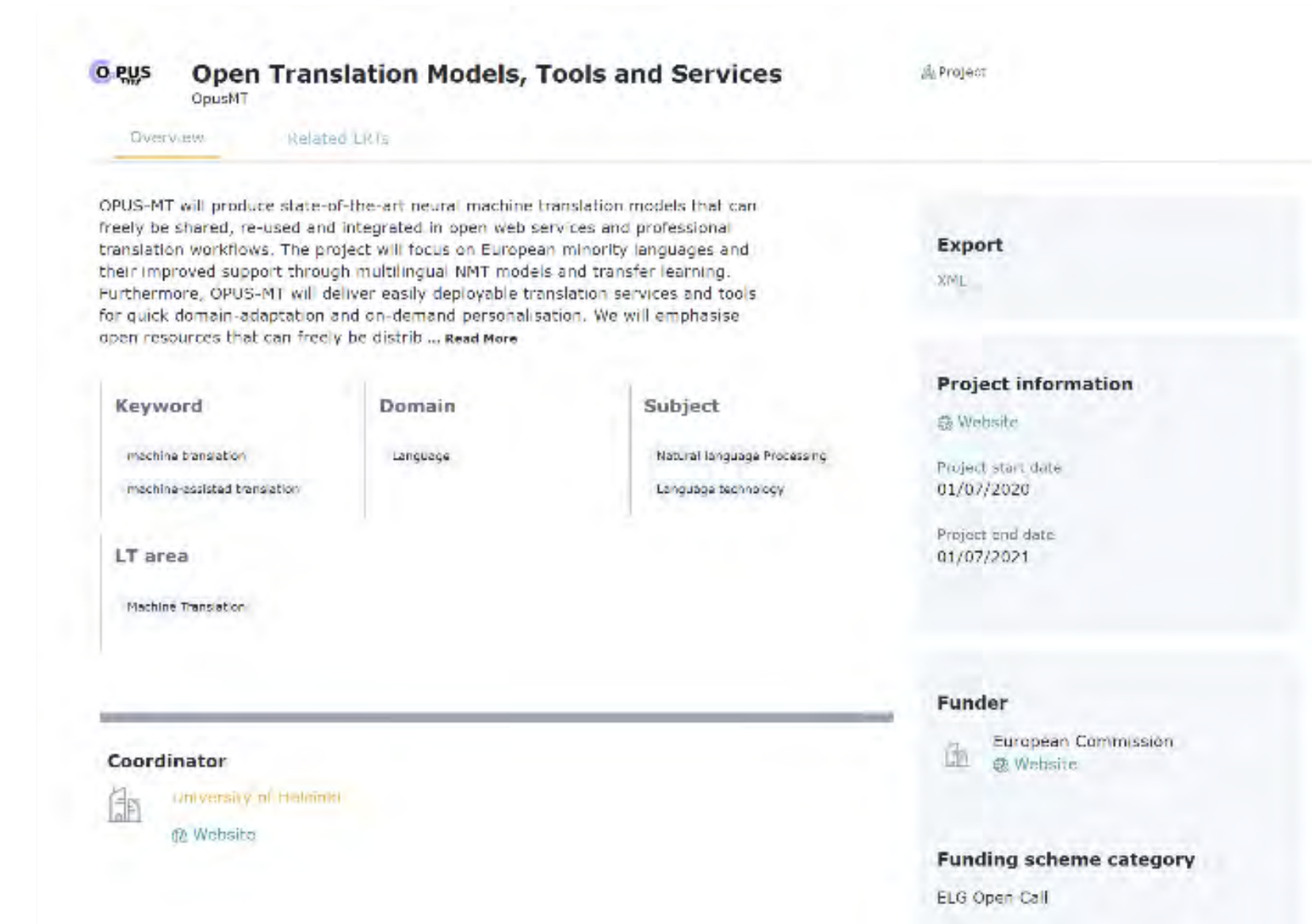
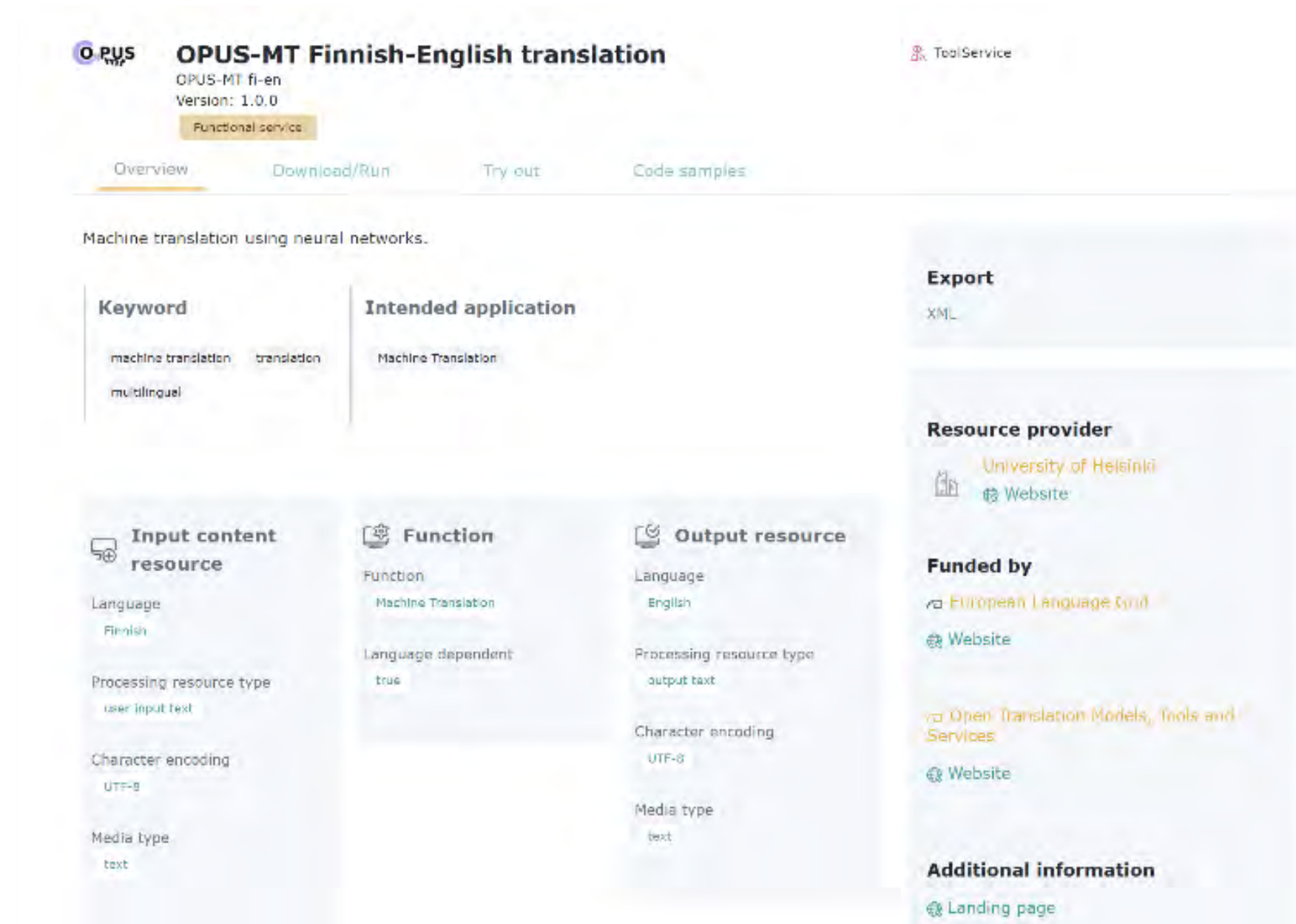
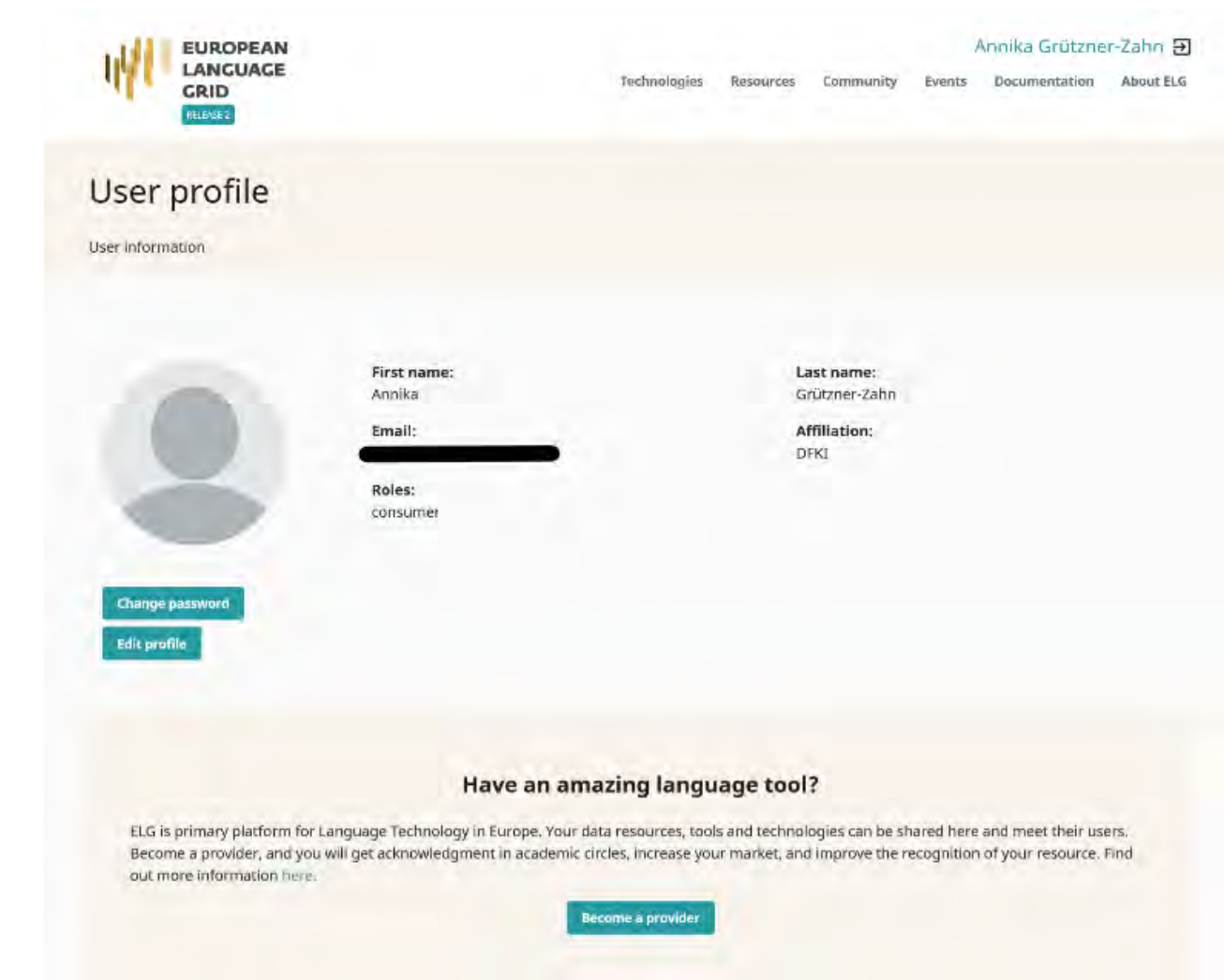
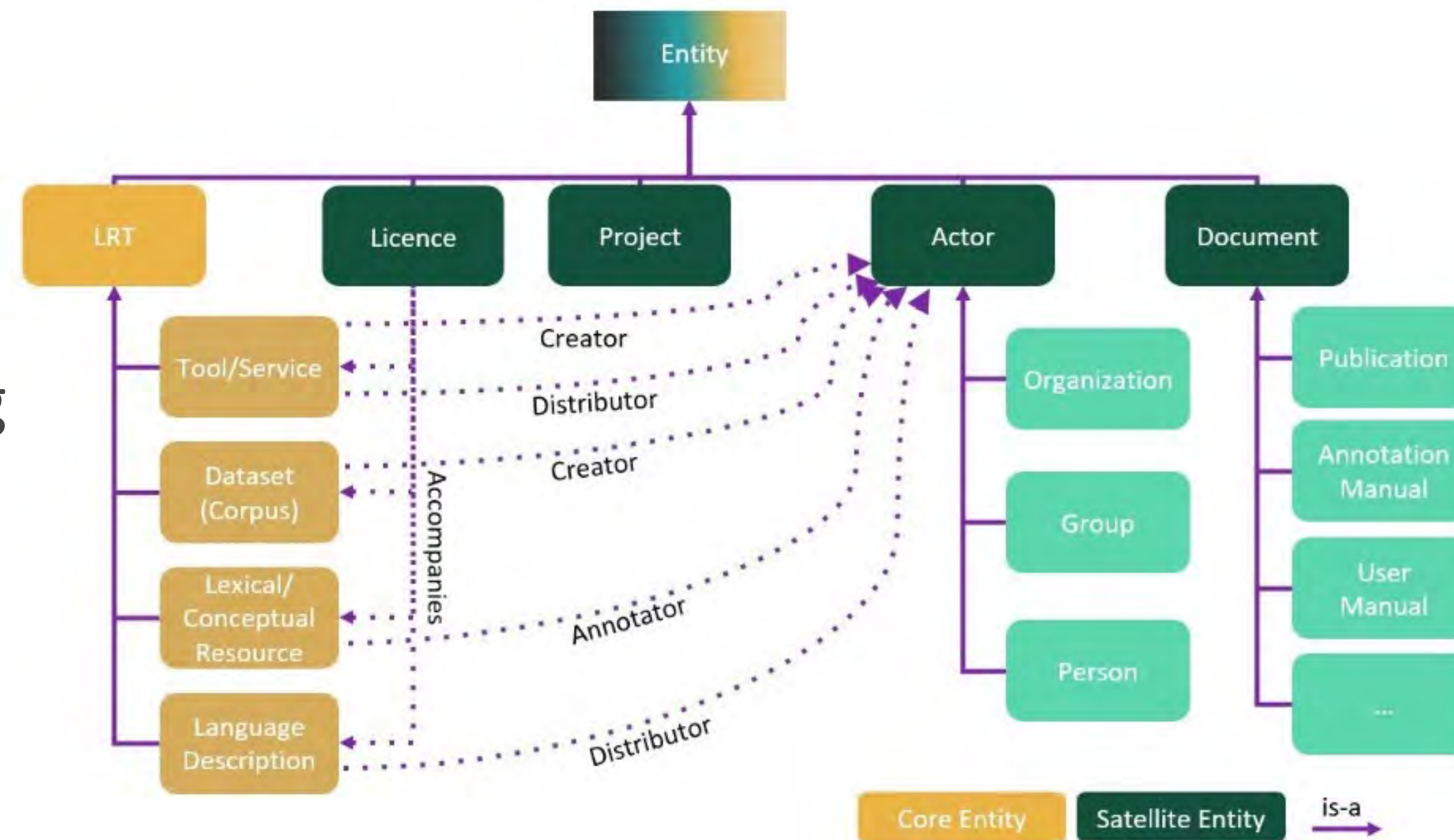
On the services consumer side



- Consumers can **try out and test** language processing services
 - registration/authentication is required
 - daily quotas apply
- **Call a service** from the command line directly (via its common REST API) and integrate it in their own workflows
- Current APIs support
 - **machine translation**
 - **information extraction**
 - **text classification**
 - **speech recognition**
 - **speech synthesis**
- **Python-based API** for accessing the ELG catalogue, searching and directly fetching datasets to feed them into, e.g., their model training pipeline, running services, combining services in a pipeline, etc.

On the provider side

- LT community members
 - can **contribute** by providing formal descriptions compliant with a dedicated metadata schema
 - describe **organisations** (parents or divisions) & **projects**
 - describe & provide access (remote or integrated in ELG) **LRTs**
 - **linked with each other**
 - must register and get authenticated as providers



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**

The screenshot shows the 'Describe Organization' form in the ELG interface. It includes a sidebar with 'IDENTITY', 'ACTIVITIES', and 'CONTACT' sections. The main form area has tabs for 'DESCRIBE ORGANIZATION' and 'DESCRIBE PARENT'. The 'DESCRIBE ORGANIZATION' tab is active, showing fields for 'Organization name', 'Language', 'Organization identifier', 'Organization short name', 'Organization description', and 'Logo'. There are also buttons for 'Save draft' and 'Save'.



The screenshot shows the 'My Grid' page in the ELG interface. It displays a list of items and their status. The items are categorized by type (Corpus, Organization, Tool/Service) and status (syntactically valid, draft). The table includes columns for item name, version, and status.

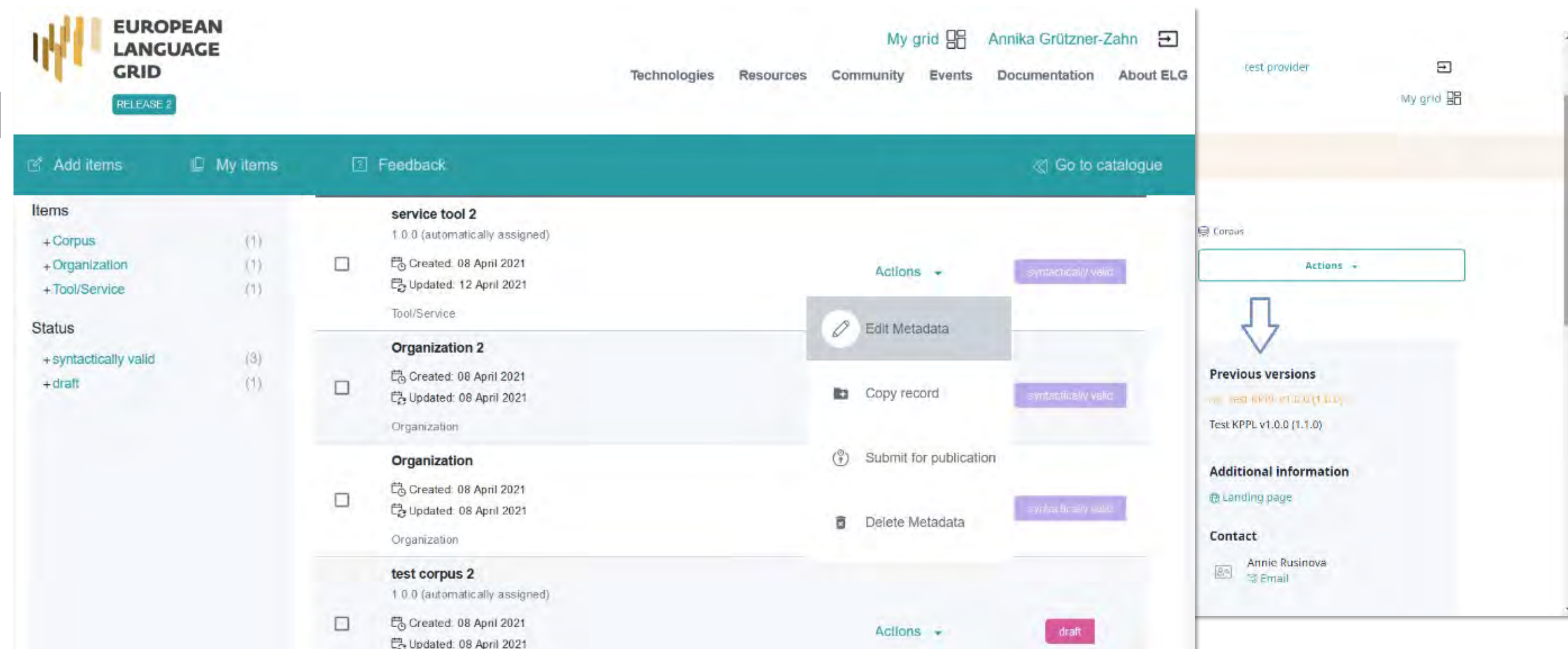
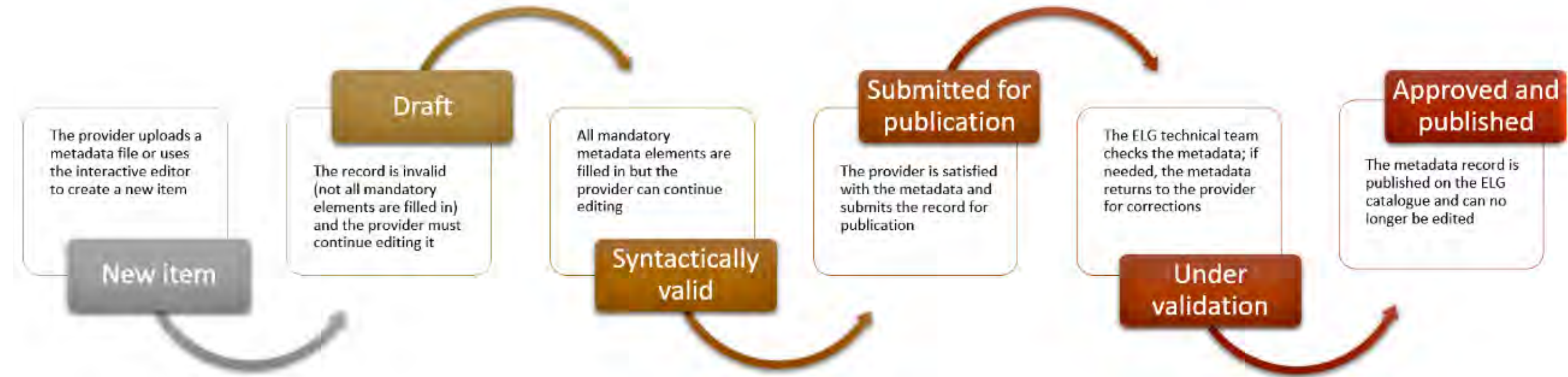
Items	Status
+ Corpus (1)	
+ Organization (1)	
+ Tool/Service (1)	
+ syntactically valid (3)	
+ draft (1)	

Item Name	Version	Created	Updated	Status
service tool 2	1.0.0 (automatically assigned)	08 April 2021	12 April 2021	
Organization 2		08 April 2021	08 April 2021	
Organization		08 April 2021	08 April 2021	
test corpus 2	1.0.0 (automatically assigned)	08 April 2021	08 April 2021	

The screenshot shows the 'Add items' dropdown menu in the ELG interface. The menu includes options for 'Interactive editor', 'Validate XML', and 'Upload XML'. Below the menu, the 'My Grid' page is visible, showing a list of items and their status. The items are categorized by type (Corpus, Organization, Tool/Service) and status (syntactically valid, draft). The table includes columns for item name, version, and status.

Item Name	Version	Created	Updated	Status
service tool 2	1.0.0 (automatically assigned)	08 April 2021	12 April 2021	
Organization 2		08 April 2021	08 April 2021	
Organization		08 April 2021	08 April 2021	
test corpus 2	1.0.0 (automatically assigned)	08 April 2021	08 April 2021	

- Providers can **manage** their items (data, resources, services, etc.) through a dedicated page (“my items”)
 - according to their status in the **ELG publication lifecycle**: draft → syntactically valid → submitted → published
- Upload and manage files
- Create new versions and copy records

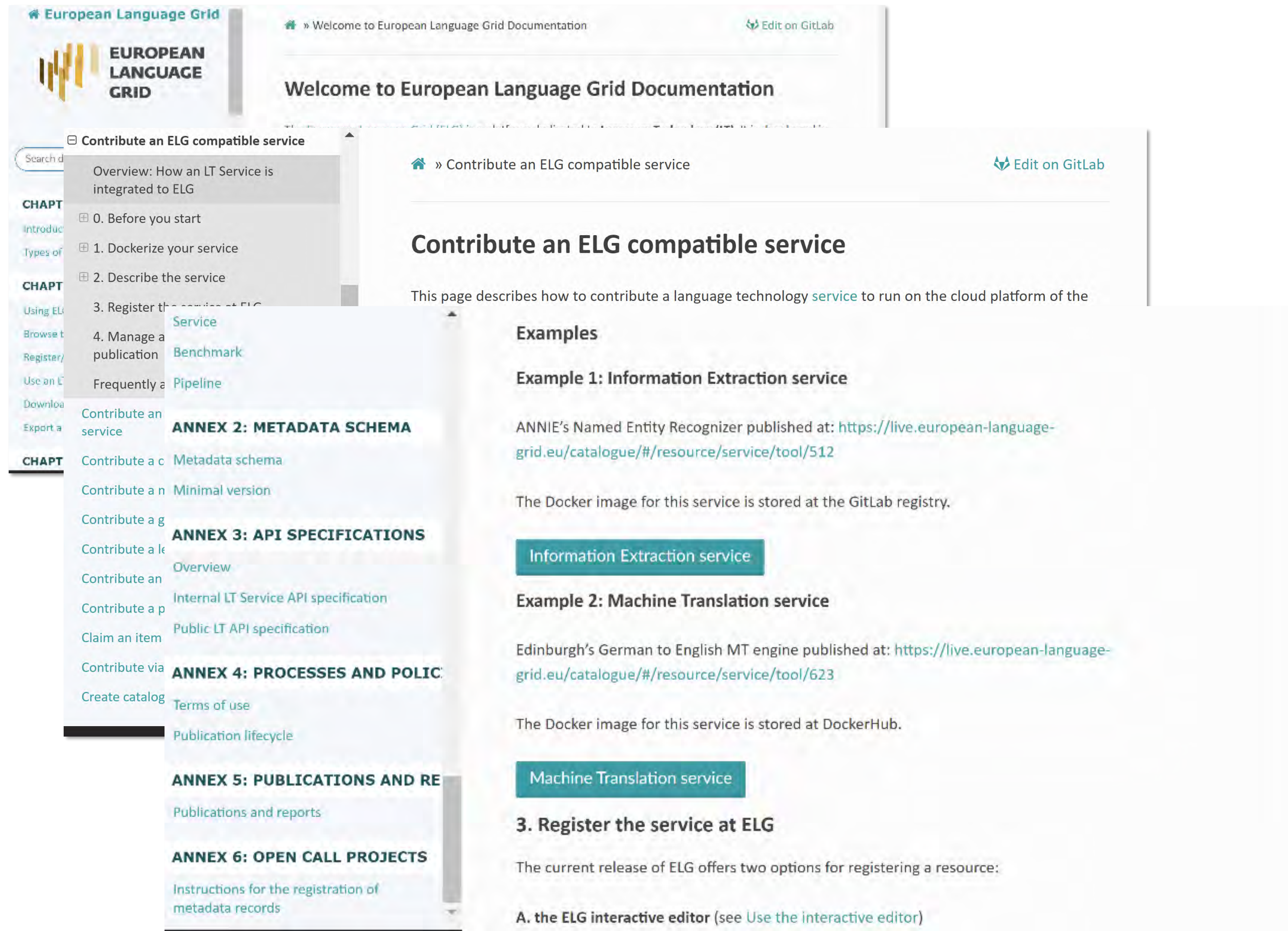


A few details on service provision

The image displays three overlapping screenshots of the European Language Grid (ELG) web interface. The top screenshot shows the 'Create a new Service or Tool' page with a teal header and a sidebar with tabs for 'LANGUAGE RESOURCE/TECHNOLOGY', 'TOOL/SERVICE', and 'DISTRIBUTION'. The middle screenshot shows the 'TOOL/SERVICE' tab with fields for 'LRT name' (tagger), 'LRT identifier', and 'LRT short name'. The bottom screenshot shows the 'DISTRIBUTION' tab with a 'Software distribution 1' section. A sidebar on the left of the bottom screenshot lists 'Contribute a service' steps and a 'Contribute a service' document link.

- 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)
- **Three different options:**
 - LT tool packaged in one Docker image exposing an ELG-compliant endpoint
 - LT tool running outside the ELG infrastructure – proxy image with one or more ELG-compliant endpoints
 - LT tool requiring an adapter – adapter image exposes an ELG-compliant endpoint as proxy

ELG Support services



- **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 is **building bridges** to existing platforms/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

