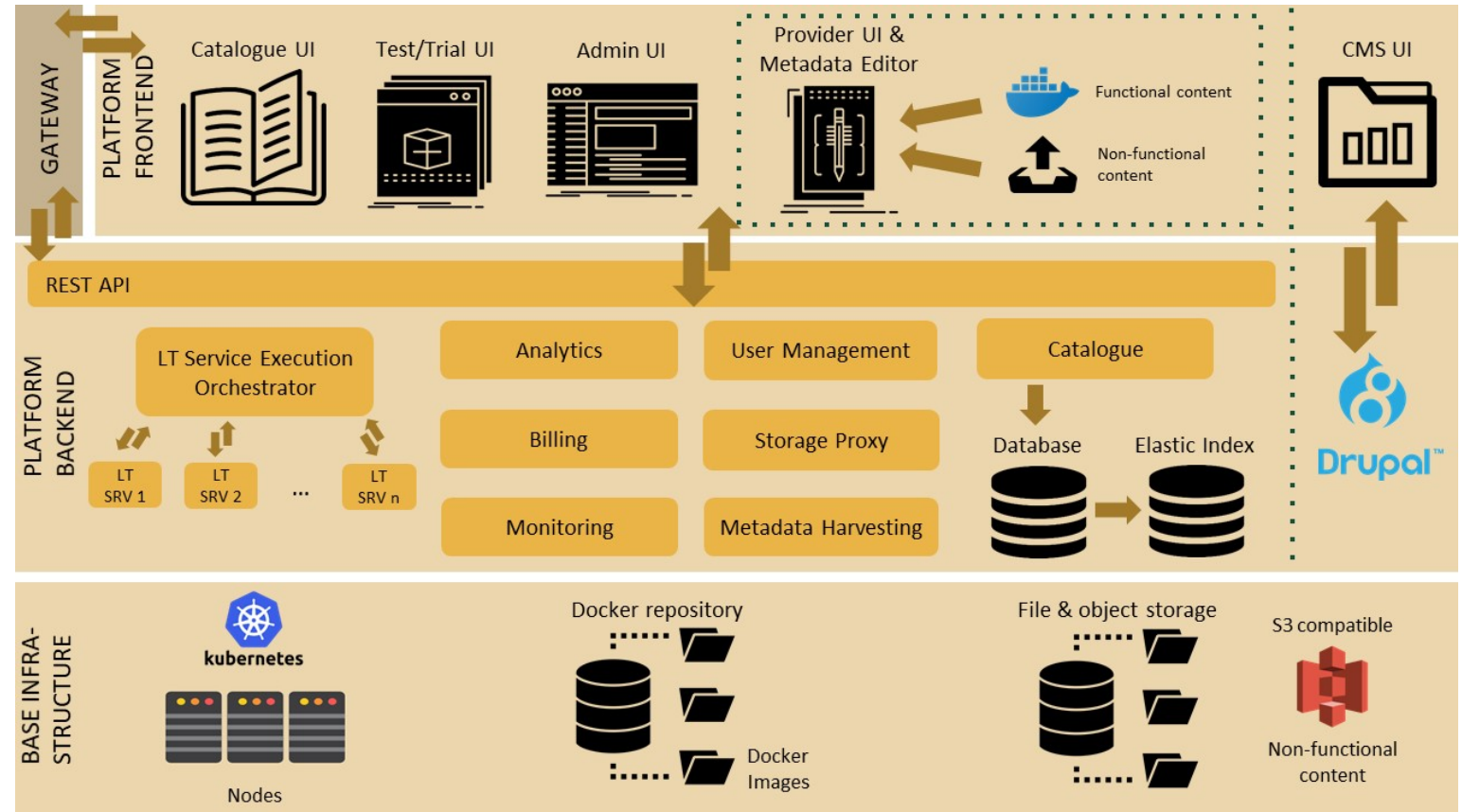


European Language Grid – META-FORUM 2020
Stelios Piperidis: European Language Grid – Platform

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

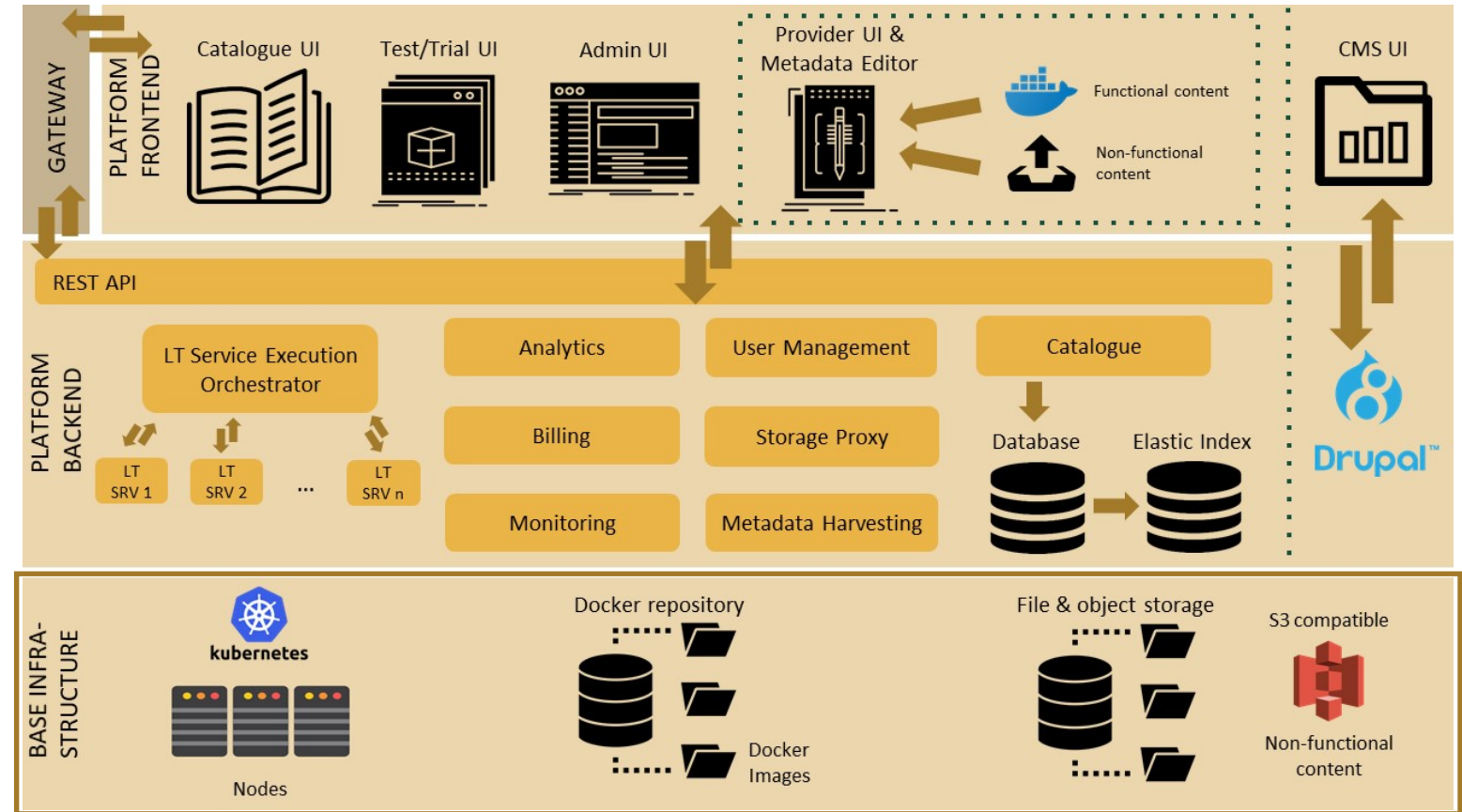
ELG Platform Architecture

- Three-layer architecture
- All layers built with robust scalable, reliable, widely used technologies
- Ability to scale with the growing demand and supply of resources
- Laying the foundations for interoperable data and services spaces

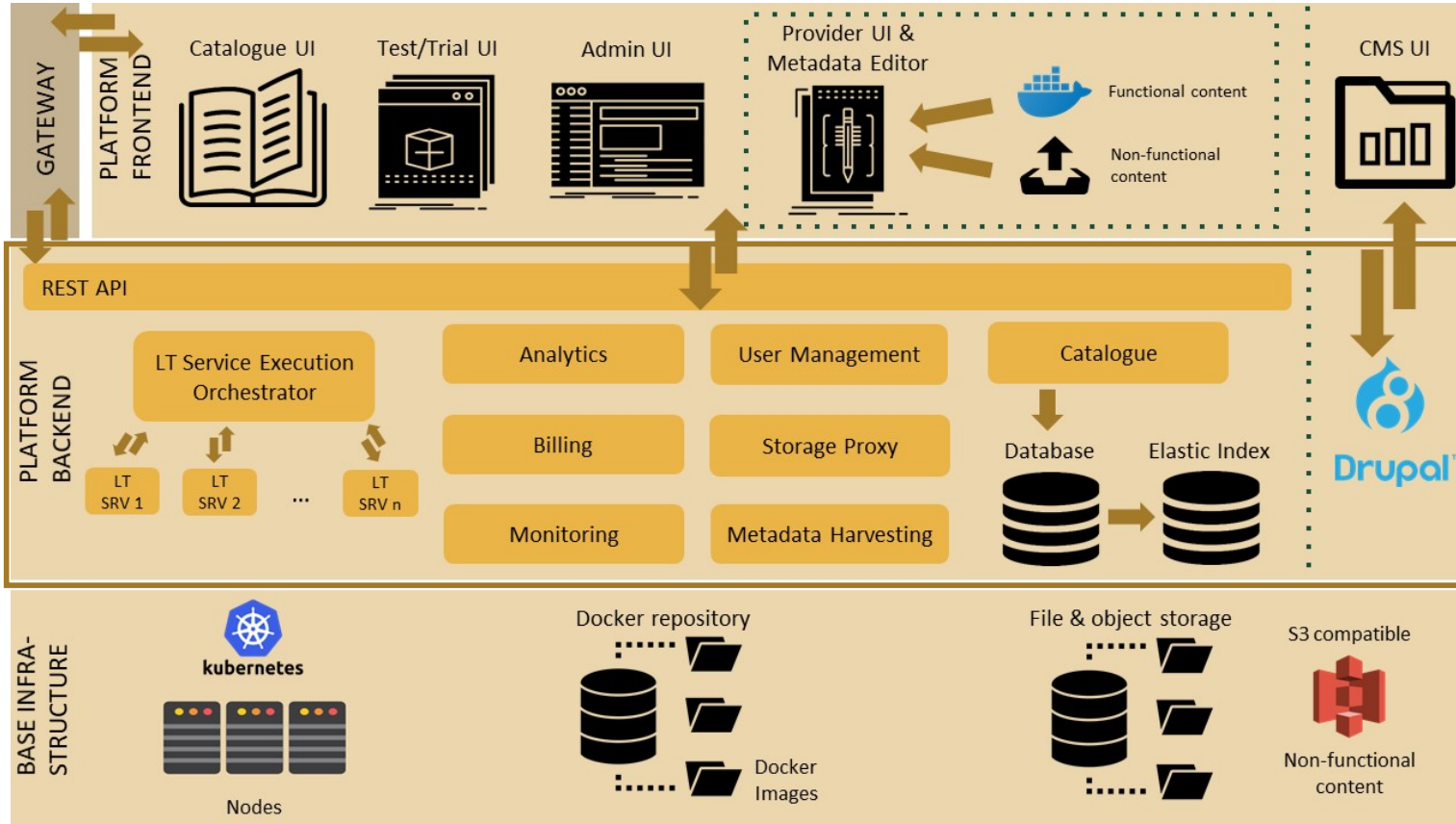


ELG Platform Base Infrastructure

- Docker containers for all services and applications which comprise the ELG platform
- Kubernetes for container orchestration
- Storage solution, S3 compatible, for language data/resources and other related content
- Supporting tools for development and management of ELG software



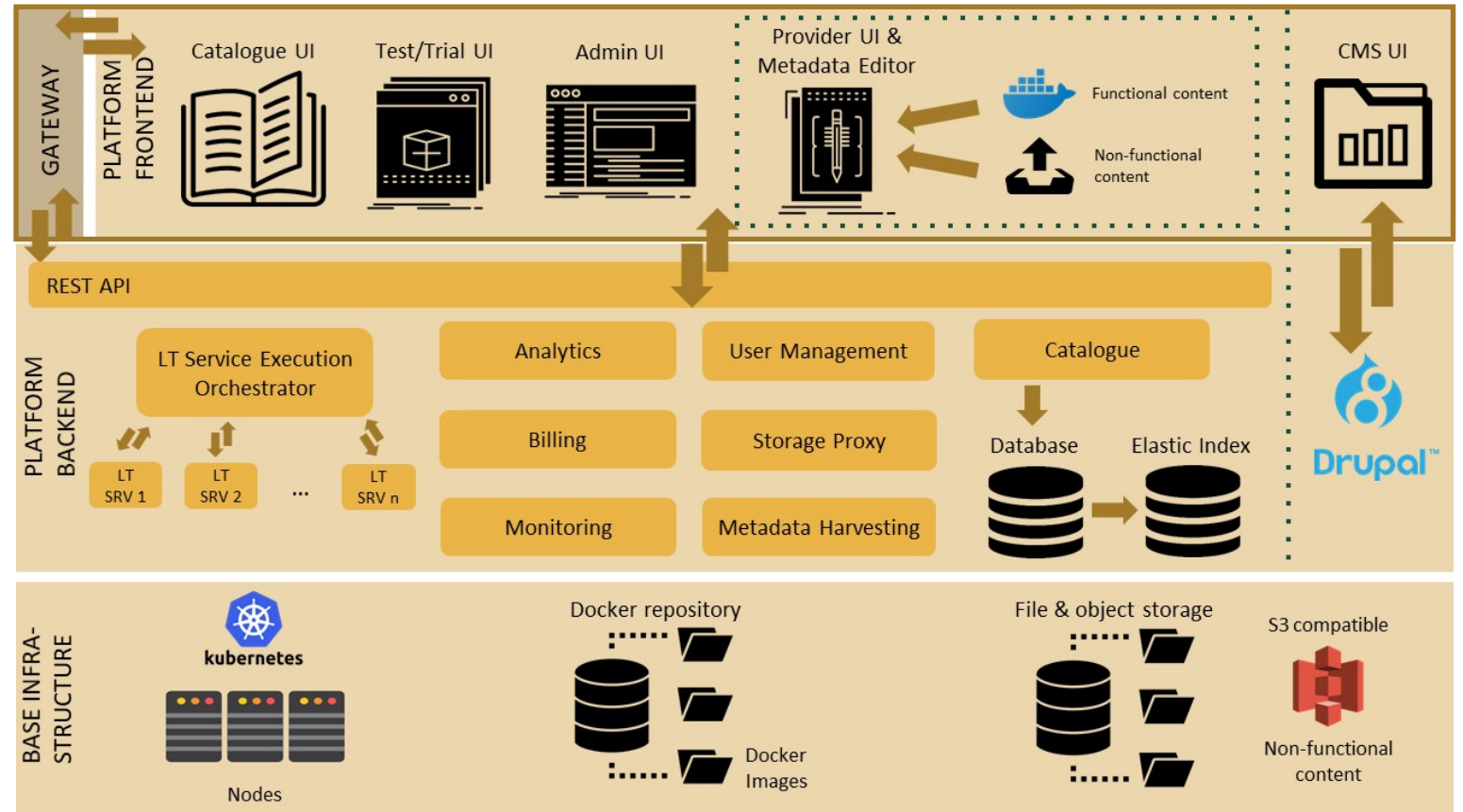
ELG Platform Backend



- ELG platform repository and catalogue
 - Application for managing metadata records according to the ELG publication lifecycle policies
 - Database and indexing mechanisms
- LT execution server
 - Invokes functional services
 - Flexible ways of integration
- User management
 - Different user categories and roles
 - Support and monitoring mechanisms

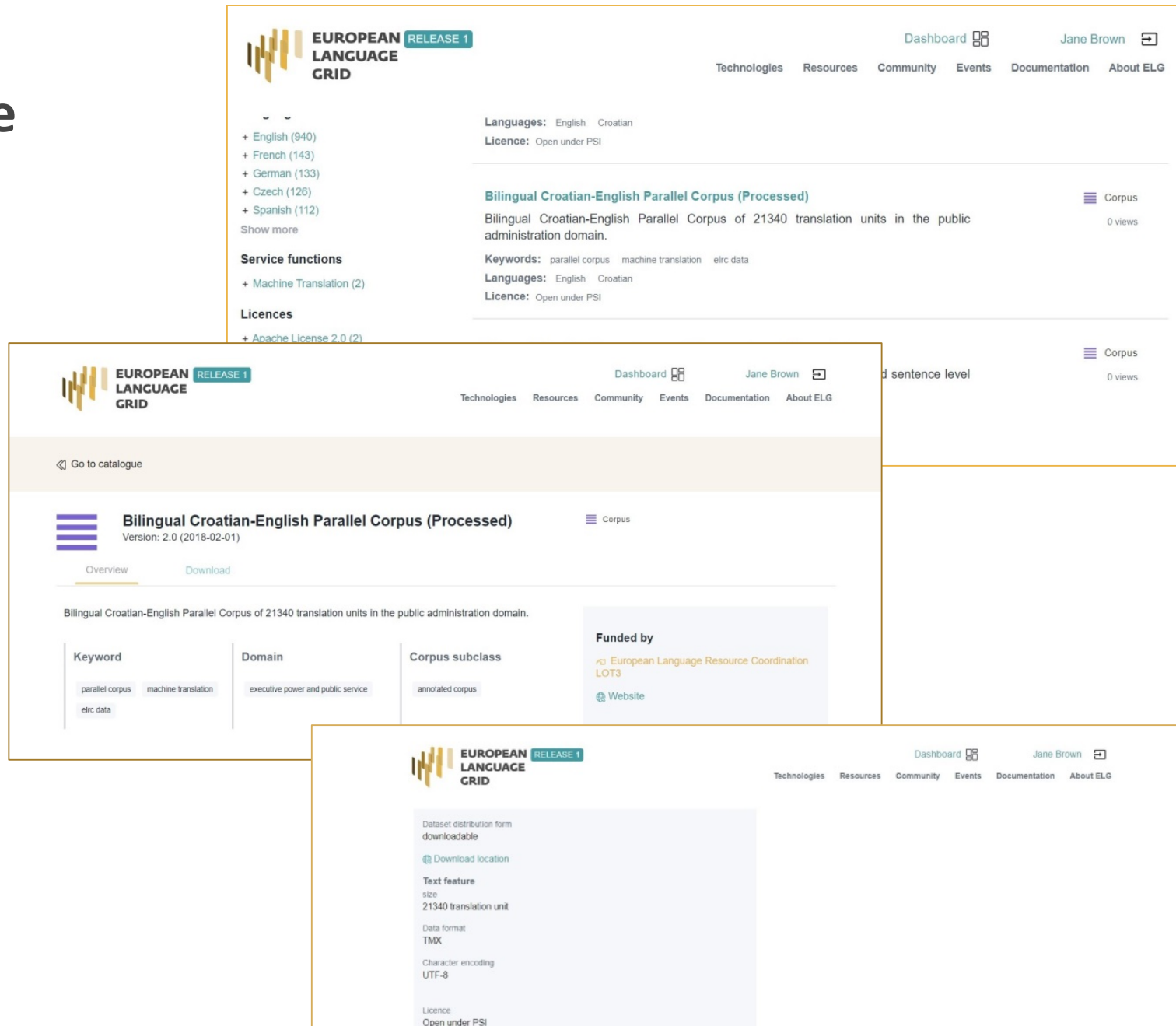
ELG Platform Frontend

- GUIs for catalogue application
 - User interactions, e.g., browsing, search
 - Providing and editing descriptions (metadata)
 - Admin interfaces, e.g., for validating submitted resources
- User interfaces for trying and testing functional services
- Code samples
- Content management system for ELG related content and information



On the data consumer side

- Consumers can **search and browse** the ELG catalogue
 - for **different types of data, language processing services, related projects and organisations** in Europe
 - using simple and advanced search: **facets for resource type, language, service function, license, related entities**
- Download data (depending on access conditions)
- Check number of views/downloads
- Check what is forthcoming in terms of data and services



The screenshot displays the European Language Grid (ELG) website interface. The top navigation bar includes the ELG logo, a 'RELEASE 1' badge, and a user profile 'Jane Brown' with a dashboard icon. The main menu consists of 'Technologies', 'Resources', 'Community', 'Events', 'Documentation', and 'About ELG'. The page title is 'Bilingual Croatian-English Parallel Corpus (Processed)', with a version of '2.0 (2018-02-01)'. The description states it is a 'Bilingual Croatian-English Parallel Corpus of 21340 translation units in the public administration domain'. The 'Keywords' section lists 'parallel corpus', 'machine translation', and 'elrc data'. The 'Domain' section lists 'executive power and public service'. The 'Corpus subclass' section lists 'annotated corpus'. The 'Funded by' section mentions 'European Language Resource Coordination LOT3' and includes a 'Website' link. The 'Licence' section states 'Open under PSI'. The 'Dataset distribution form' section is labeled 'downloadable' and includes a 'Download location' link. The 'Text feature' section lists 'size: 21340 translation unit', 'Data format: TMX', and 'Character encoding: UTF-8'. The 'Licence' section states 'Open under PSI'. The page also shows a 'Corpus' icon and '0 views'.

On the services consumer side

The image shows two screenshots of the European Language Grid (ELG) website. The top screenshot shows the search results for 'organizations places', displaying 7 results. The bottom screenshot shows a detailed view of the 'Cogito Discover Named Entity Recognizer' service, including its description, version (14.3.0), and a sample text snippet with annotations for Organization, Percentage, Place, and Time.

European Language Grid (ELG) Website Screenshot 1: Search Results

Search results for 'organizations places':

- Language resources & technologies
- Tool/Service (7)
- Languages
 - English (7)
 - French (3)
 - German (3)
 - Dutch (1)
 - Italian (1)

7 search results for organizations places

Cogito Discover Named Entity Recognizer

Annotation of entities: People, Organizations, Places, Known concepts, Unknown concepts. And also tags: urls, mail addresses, phone numbers, addresses, dates, time, measures, money, percentage, file folder.

Keywords: multilingual, English, Spanish, German

149 views
291 times used

Functional service

European Language Grid (ELG) Website Screenshot 2: Service Details

Cogito Discover Named Entity Recognizer

ESI_NER

Version: 14.3.0

Functional service

Overview Download/Run Test/Try out Code samples

The total number of corona infections proven so far in Germany has exceeded the mark of one million. Since the beginning of the pandemic, the Robert Koch Institute (RKI) in Berlin counted a total of 1,006,394 proven infections with the SARS-CoV-2 pathogen in Germany. More than 67,000 people are now genius according to the RKI's estimates. The number of Corona dead gives the federal authority 15,586. Within 24 hours, 426 people died in Germany in connection with the virus. This is the highest number of deaths within one day since the beginning of the pandemic. That four percent more people died last month than on average of the four previous years, however, cannot be directly linked to the Corona pandemic.

Annotations

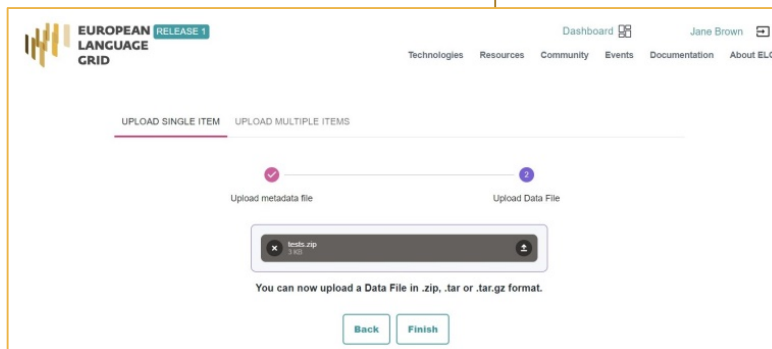
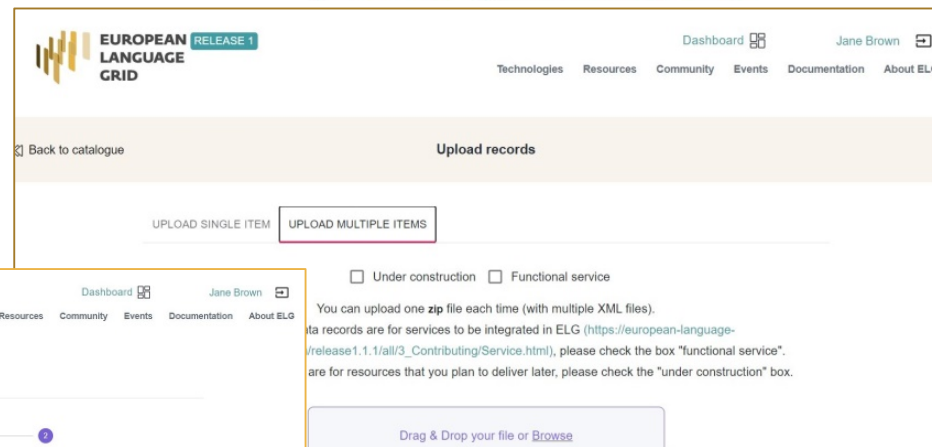
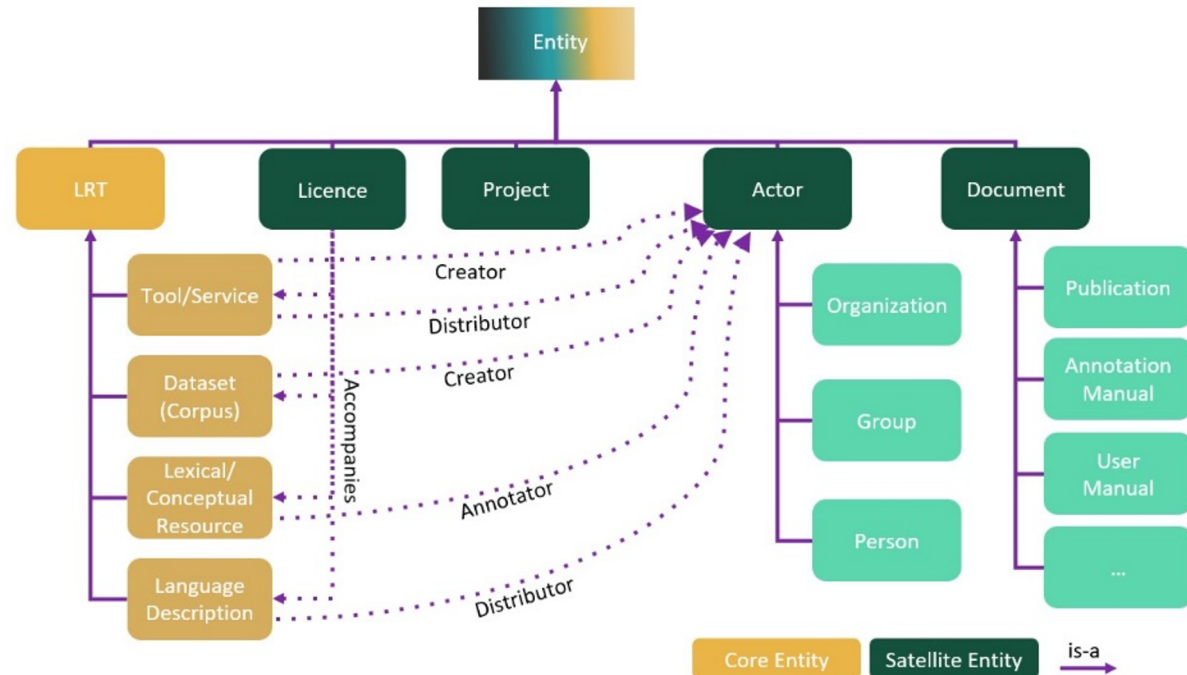
- Organization
- Percentage
- Place
- Time

BACK

- 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
- View code samples
- Current APIs support
 - **machine translation (MT)**
 - **information extraction (IE)**
 - **speech recognition (ASR)**
 - **speech synthesis (TTS)**
 - **text classification**
- Consumers can use a Python-based API for accessing the ELG catalogue, searching and directly fetching datasets to feed them into, e.g., their model training pipeline

On the provider side

- Language data/resources/services providers can register metadata descriptions by
 - registering and getting authenticated as providers
 - using a **formal metadata schema**
 - uploading a **schema compliant XML file**
 - using an **interactive metadata editor**
- Get support through the **online documentation**
- View their resources in their **dashboard**
 - including their status according to the ELG **publication lifecycle**
 - lifecycle: draft → syntactically valid → submitted → published
- Upload data files
- Batch uploading also supported
- **“Claim”** metadata records and enrich them



On the service provider side

The image displays three overlapping screenshots of the European Language Grid (ELG) service provider interface. The top screenshot shows the 'Create a new Service or Tool' page with tabs for 'LANGUAGE RESOURCE/TECHNOLOGY', 'TOOL/SERVICE', and 'DISTRIBUTION'. The 'LANGUAGE RESOURCE/TECHNOLOGY' tab is active, showing fields for 'LRT name' (tagger), 'LRT identifier', and 'LRT short name'. The middle screenshot shows the 'DISTRIBUTION' tab, which includes a 'Software distribution 1' section for describing distributable forms of the tool. The bottom screenshot shows a 'Contribute a service' guide, which provides an overview and steps for integrating a service into the ELG. The guide includes a sidebar with links to 'Contribute downloadable software', 'Contribute a corpus/dataset', 'Contribute a model', and 'Contribute a grammar'.

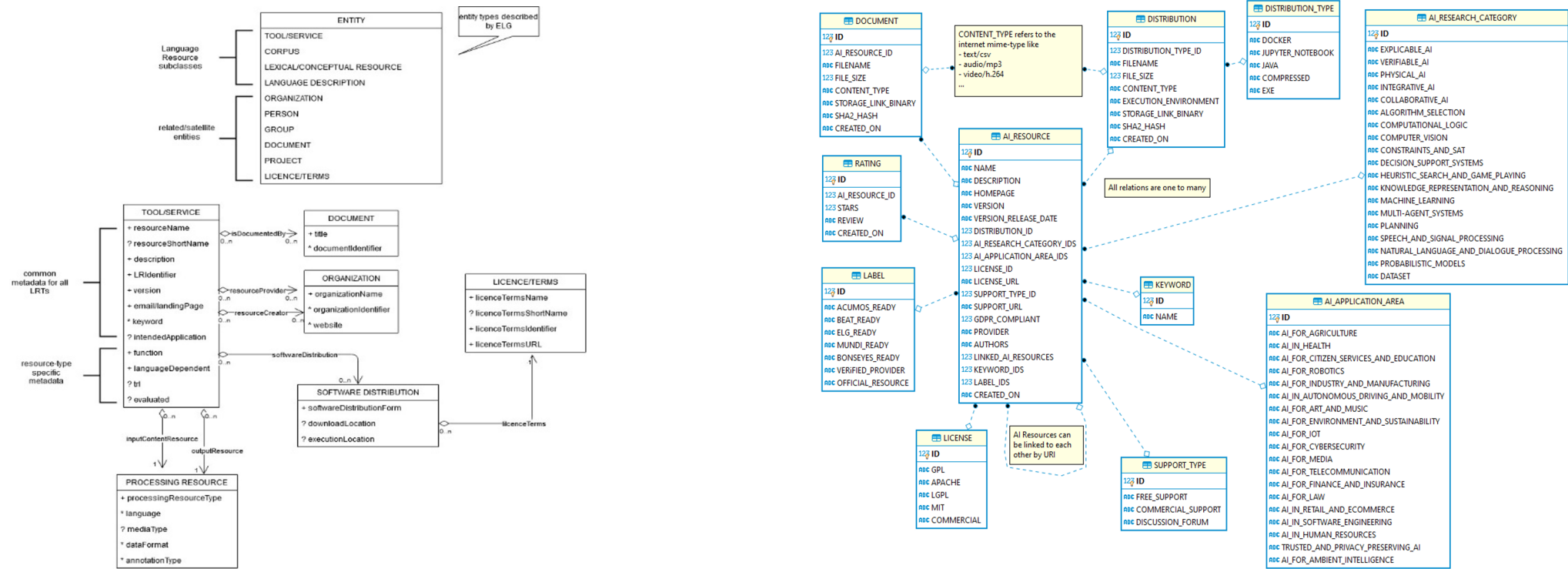
- LT service providers need to provide a **Docker image** with their LT tool or service
- Docker images have to be uploaded to a **Docker registry** (e.g., GitLab, DockerHub)
- **Three different options:**
 - LT tool packaged in one Docker image exposing an ELG-compliant endpoint
 - LT tool running outside 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 in the wider LT and AI ecosystem

- First step has been to **build bridges** to existing platforms/infrastructures
 - Mainly in terms of metadata-based descriptions
- First bridges with
 - **ELRC-SHARE**, of the European Language Resource Coordination initiative
 - **LINDAT-CLARIAH**
- Both repositories are automatically harvested once a week
 - Based on **open protocols** (OAI-PMH)
 - Respecting their own policies



ELG ⇔ AI4EU bridge



- Mapping of the underlying ELG & AI4EU ontologies
- In order to support cross-platform search



Come and visit us at the ELG booth!