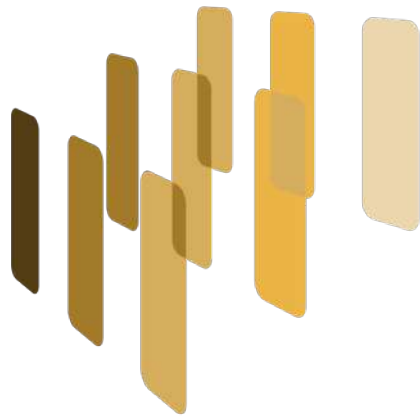


EUROPEAN LANGUAGE GRID

Introducing the European Language Grid: Deep Dive 2 – ELG Platform

Stelios Piperidis (ILSP/Athena RC), Andrejs Vasiljevs (TILDE), Florian Kintzel (DFKI)
spip@athenarc.gr, andrejs@tilde.lv, florian.kintzel@dfki.de

08/09-10-2019, Brussels – META-FORUM 2019
<http://www.european-language-grid.eu>

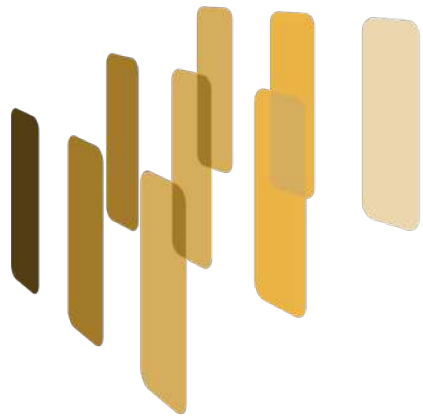


Introducing the European Language Grid

Deep Dive 2 – ELG Platform

08/09-10-2019, Brussels – META-FORUM 2019 – Stelios Piperidis, Andrejs Vasiljevs, Florian Kintzel

- ELG Platform Overview
- Architecture and Technology Choices
- Platform Backend
- User Interface



Introducing the European Language Grid

Deep Dive 2 – ELG Platform

08/09-10-2019, Brussels – META-FORUM 2019 – Stelios Piperidis, Andrejs Vasiljevs, Florian Kintzel

ELG Platform: Overview



META FORUM 2010



META SHARE

META FORUM 2015



META FORUM 2017

The picture can't be displayed.

T4ME
EU-funded project (Seventh Framework Programme) working on technologies for the Multilingual European Information Society

META-NET
Established in 2010, META-NET is a network of Excellence consisting of 60 research centres from 34 countries, building the technological foundations of a multilingual European information society

META-FORUM 2010
"Challenges for Multilingual Europe"
(November 17/18, 2010)

META FORUM 2012

META-FORUM 2012
Conference "A Strategy for Multilingual Europe"
(Brussels – June 20/21, 2012)

META-NET White Papers
Release of 32 volumes on 31 languages, revealing that there is a severe threat of digital extinction for at least 21 European languages
(December 2012)

CRACKER
EU-funded project CRACKER (Horizon2020) pushing towards an improvement of MT research in terms of efficiency and effectiveness (2015 – 2017)

Cracking the Language Barrier Federation
Founded in 2015, the federation has been assembling European research and innovation projects as well as all related community organisations working on multilingual technologies

META-FORUM 2015
Conference "Technologies for the Multilingual Digital Single Market" (Riga – April 27, 2015)

Strategic Agenda for the Multilingual Digital Single Market (Version 0.5)
Launch of the Strategic Agenda for the Multilingual Digital Single Market titled "Technologies for Overcoming Language Barriers towards a truly integrated European Online Market" (April 2015)

Riga Summit on the Multilingual Digital Single Market
Summit "Shape the future of the multilingual digital single market" (April 27–29, 2015)

"Language Equality in the Digital Age"
Workshop on "Language Equality in the Digital Age", commissioned by the EU Parliament's Science and Technology Options Assessment Committee (STOA) (January 2017)

"Language equality in the digital age: Towards a Human Language Project"
Launch of the study on "Language equality in the digital age: Towards a Human Language Project", commissioned by the EU Parliament (March 2017)

META-FORUM 2017
Conference "Towards a Human Language Project"
(Brussels – November 13/14, 2017)

Strategic Research and Innovation Agenda (V1.0)
Launch of the Strategic Research and Innovation Agenda titled "Language Technologies for Multilingual Europe – Towards a Human Language Project" (December 2017)



META-NET

2010

2011

2013

2015

2017

2012

2014

2016

2018

EUROPEAN LANGUAGE GRID

2019-2021

META NORD

META-NORD
EU-funded (ICT PSP) Project to establish an open linguistic infrastructure in the Baltic and Nordic countries (2011 – 2013)

METANET4U

METANET4U
EU-funded project (ICT PSP) to enhance the European Linguistic Infrastructure (2011 – 2013)

CESAR

CESAR
EU-funded project (ICT PSP) functioning as a part of META-NET to standardise language resources and tools (2011 – 2013)

META-FORUM 2011
Conference "Solutions for Multilingual Europe"
(Budapest – June 27/28, 2011)

META FORUM 2011

"State of the Art of Machine Translation – Current Challenges and Future Opportunities"
Workshop on "State of the Art of Machine Translation", commissioned by the EU Parliament (December 2013)

META-SHARE
Initiated in 2013, META-SHARE has functioned as an open and secure network of repositories for sharing and exchanging language data, tools and services

Strategic Research Agenda for Multilingual Europe 2020
Launch of the Strategic Research Agenda for Multilingual Europe 2020 (January 2013)

META-FORUM 2013
Conference "Connecting Europe for New Horizons"
(Berlin – September 19/20, 2013)

META FORUM 2013

Strategic Research and Innovation Agenda (Version 0.9)
Launch of the Strategic Research and Innovation Agenda titled "Language as a Data Type and Key Challenge for Big Data" (July 2016)

META-FORUM 2016
Conference "Beyond Multilingual Europe"
(Lisbon – July 4/5, 2016)



META FORUM 2016

ELG Final Proposal Submission
Final submission on Feb. 20, 2018



Introducing the European Language Grid: Deep Dive 2 – ELG Platform



A bit of background

META-SHARE LEARN ▾ DISCOVER ▾ PARTICIPATE ▾ CONNECT ▾ LOGIN

Search & exchange language resources

META-SHARE is an open and secure network of repositories for sharing and exchanging language data, tools and related web services

Share your own resources!

JOIN OUR NETWORK NOW

Already a member? [Log in](#)

Search the META-SHARE inventory

OR LEARN MORE

4,481 users

2,887 language resources

32% text corpora

27,630 number of downloads

Virtual Language Observatory Search Contributors Help CLARIN

CLARIN Virtual Language Observatory

Welcome to the VLO!

Use the **search bar** below to start searching through hundreds of thousands of language resources, or [continue](#) to browse everything and use **facets** to narrow down to your area of interest or discover new resources.

[See all records](#) [Take a quick tour](#)

Search through 1,030,321 records

Home Browse Resources Help About Register Login

European Language Resource Coordination Connecting Europe Faculty ELRC-SHARE

ELRC-SHARE Repository

Type in your keywords, please...

Welcome to the ELRC-SHARE repository!

The ELRC-SHARE repository is used for documenting, storing, browsing and accessing [Language Resources](#) that are collected through the [European Language Resource Coordination](#) and considered useful for feeding the [CEF Automated Translation \(CEFACT\)](#) platform.

If you want to contribute resources, all you have to do is [register](#) (new user) or [login](#) (returning user) and go on to describe and upload your data with a simple form.

Browse Resources Information Cart total View cart Register Login

EUROPEAN ASSOCIATION OF RA RESOURCES

1096 Language Resources (Page 1 of 55)

Order by: Resource Name A-Z

2006 CoNLL Shared Task – Arabic & Czech

Arabic Czech

ID: ELRA-W0087

ISLRN: 798-485-794-792.1

2006 CoNLL Shared Task – Arabic & Czech consists of dependency treebanks used as part of the CoNLL 2006 shared task on multi-lingual dependency parsing. The Conference on Computational Natural Language Learning (CoNLL) is accompanied every year by a shared task intended to promote natural lan...

2006 CoNLL Shared Task - Ten Languages

Bulgarian Danish Dutch Flemish German Japanese Portuguese Slovenian Spanish Catalan Swedish Turkish

ID: ELRA-W0086

ISLRN: 578-227-532-044.0

ELG Platform: Overview

- Platform with running functional services, downloadable data sets, tools, models etc.
- Catalogue of LT companies, research organisations, service and application types, languages etc.
- Built with robust, scalable, reliable, widely used technologies that are constantly developed further.
- Ability to scale with the growing demand and supply of resources.
- Interactive modern web interface with different views for users, LT providers, potential buyers, and administrators.
- Information about conference and training events, open calls, pilot projects, and other pieces of content.

Goal and Vision

Thousands of resources



corpora, language models and source code which users can search for and download.

Hundreds of LT services



can be used directly through the grid – from dependency parsing and NER to MT, speech recognition and synthesis



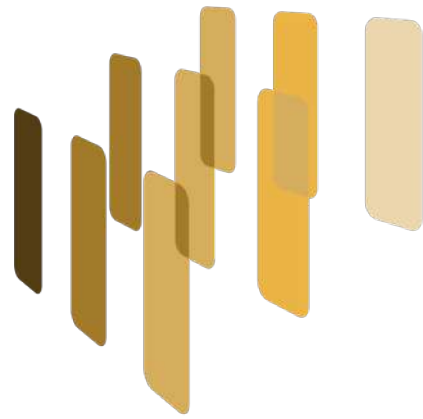
Scalable LT grid

applying containerisation through Kubernetes, with all required components and libraries included in a single container.



The **yellow pages** for European LT

listing all relevant stakeholders from industry, innovation, and research. Stakeholders will be able to register themselves in the ELG's freely accessible catalogue, ensuring vastly increased reach and visibility. Users will be able to filter and search for organisations, services, data sets and more, by languages, service types, domains, sectors, regions, countries

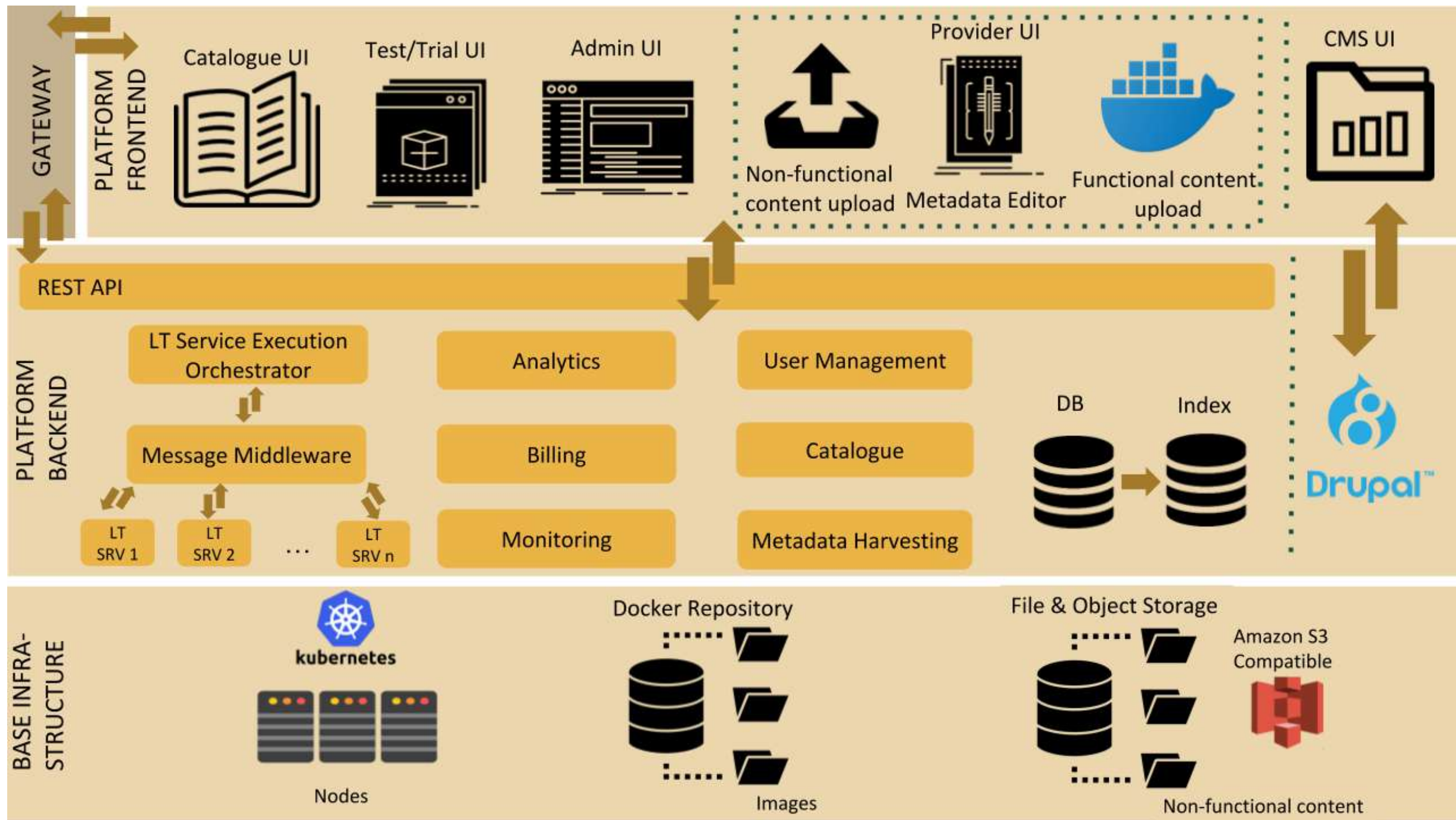


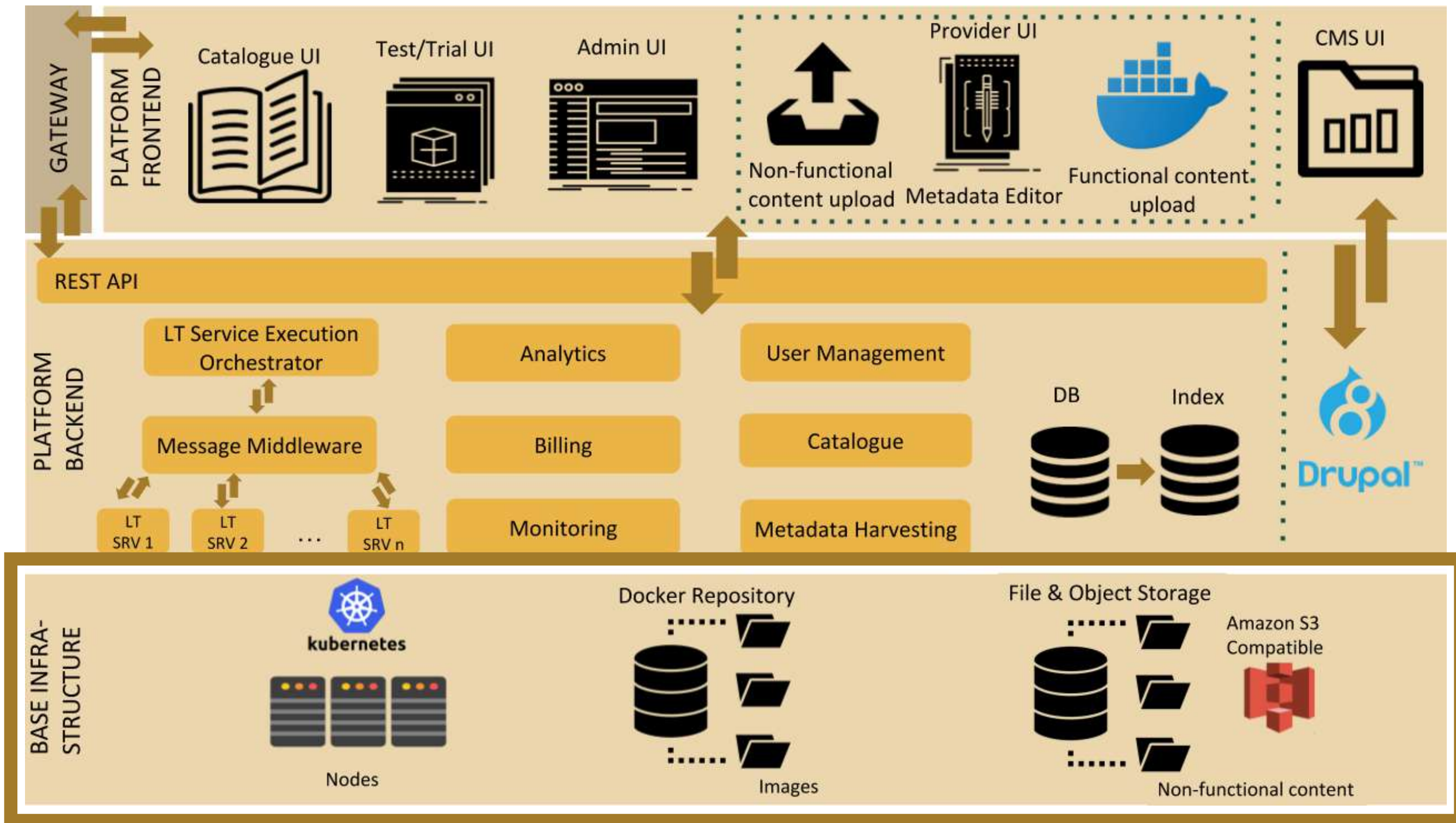
Introducing the European Language Grid

Deep Dive 2 – ELG Platform

08/09-10-2019, Brussels – META-FORUM 2019 – Stelios Piperidis, Andrejs Vasiljevs, Florian Kintzel

Architecture and Technology Choices

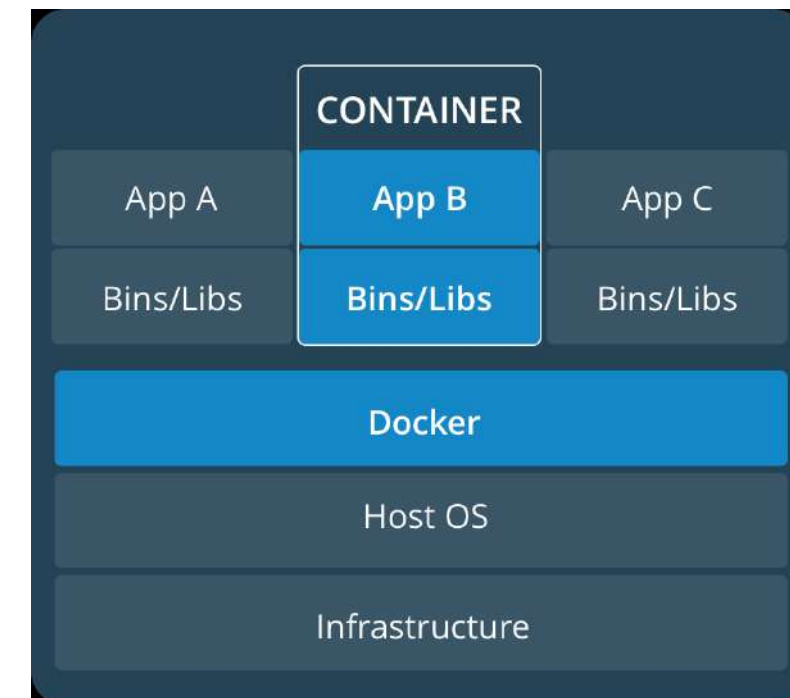




Docker

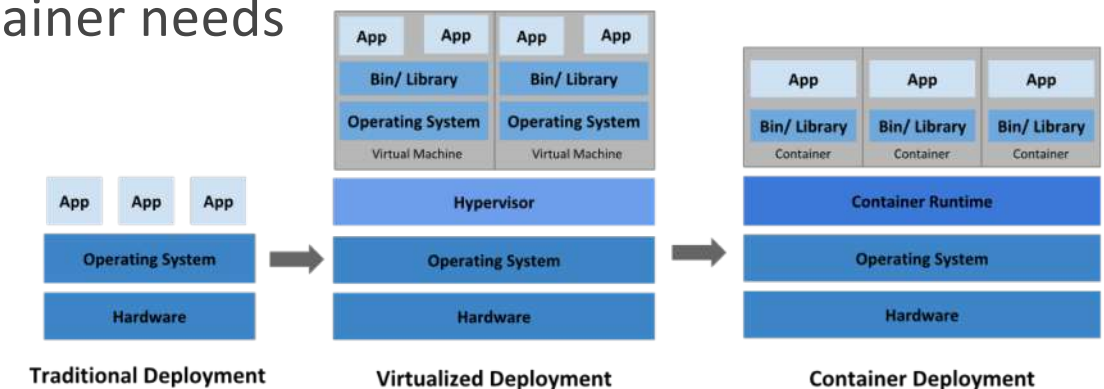


- **Flexible:** Even the most complex applications can be containerized.
- **Lightweight:** Containers leverage and share the host kernel.
- **Interchangeable:** You can deploy updates and upgrades on-the-fly.
- **Portable:** You can build locally, deploy to the cloud, and run anywhere.
- **Scalable:** You can increase and automatically distribute container replicas.
- **Stackable:** You can stack services vertically and on-the-fly.



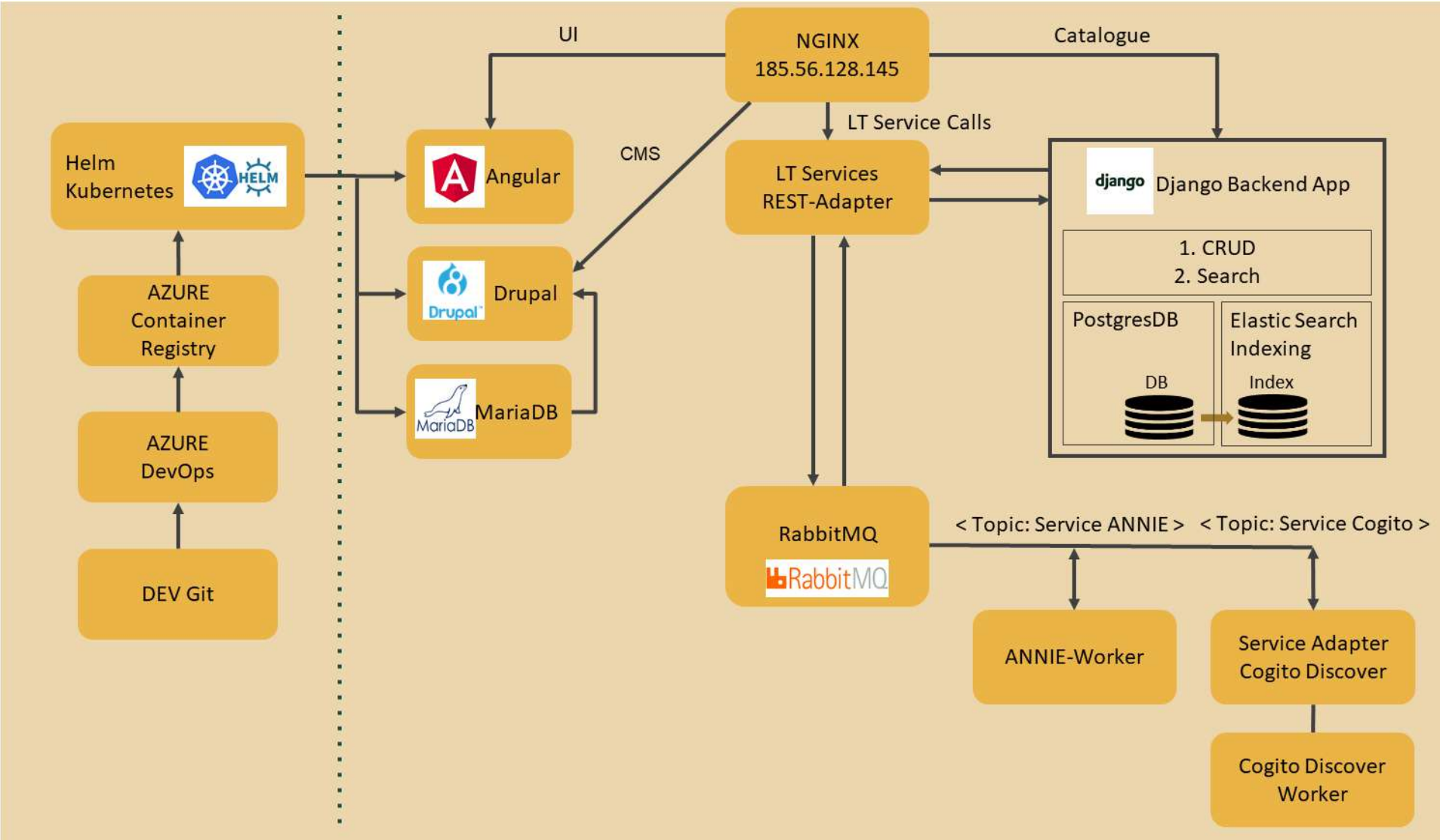
Kubernetes

- **Service discovery and load balancing**
 - expose a container using the DNS name or own IP address
 - load balance and distribute the network traffic so that the deployment is stable.
- **Storage orchestration**
 - allows you to automatically mount a storage system of choice (local storages, public cloud providers, etc.)
- **Automated rollouts and rollbacks**
 - describe the desired state for your deployed containers
 - can change the actual state to the desired state (e.g. create new containers, remove existing containers, etc.)
- **Automatic bin packing**
 - allows you to specify how much CPU and memory (RAM) each container needs
 - can make better decisions to manage the resources for containers.
- **Self-healing**
 - restarts containers that fail, replaces containers, kills containers that don't respond to your user-defined health check, etc.
- **Secret and configuration management**
 - lets you store and manage sensitive information, such as passwords, OAuth tokens, and ssh keys.
 - lets you deploy and update secrets and application configuration without rebuilding your container images, and without exposing secrets in your stack configuration.



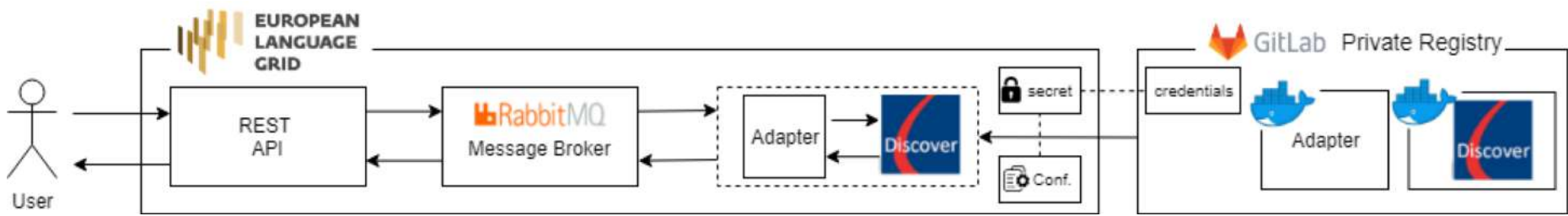
Service Integration Architecture

- ELG as a scalable LT platform, based on encapsulating services in containers.
- Solving issues of interoperability.
- Enabling commercial and academic parties to provide their services through the ELG.



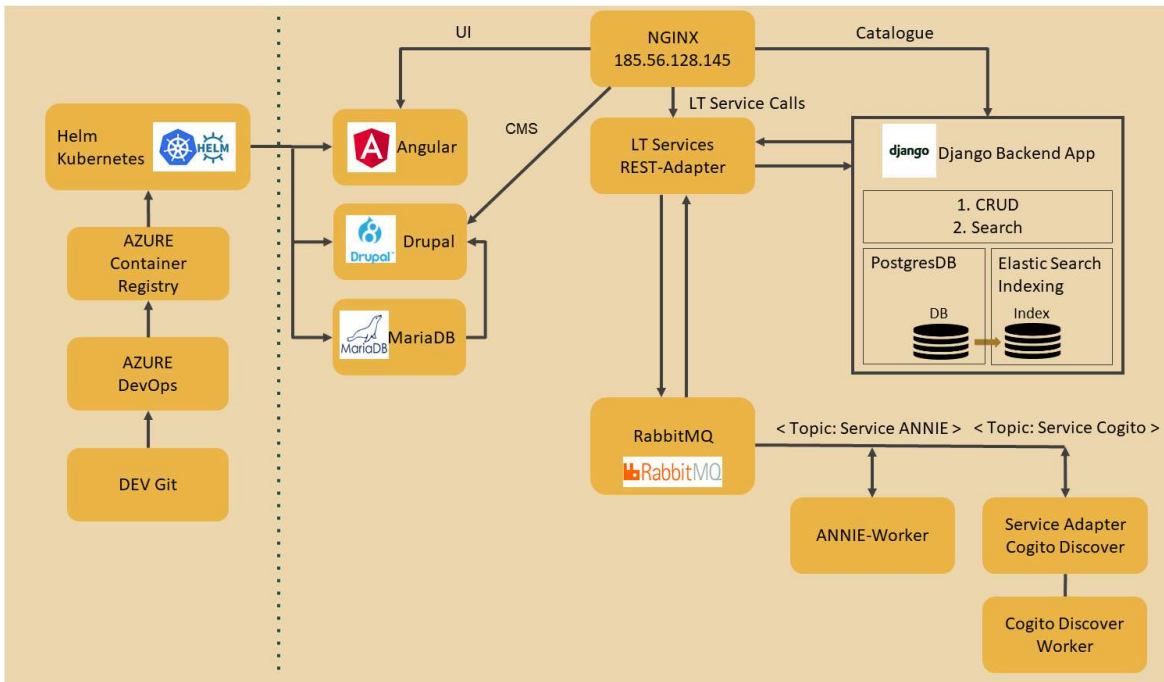
Service Integration: Six Steps

1. **From script to service:**
how to fit the ELG API (RabbitMQ)
2. **From service to container:**
how to create a docker image
3. **Push docker images into a registry**
(e.g., ELG Gitlab)
4. **Request a Kubernetes namespace,**
in case of proprietary software with
restricted access
5. **Deploy service**
6. **Add service to the ELG catalogue:**
contact ELG admins and provide
metadata for your service



ELG service integration success stories:

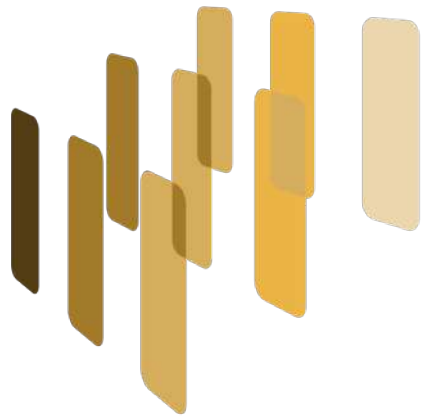
<https://gitlab.com/european-language-grid/platform/infra/wikis/home>





Global Under Resourced Media Translation



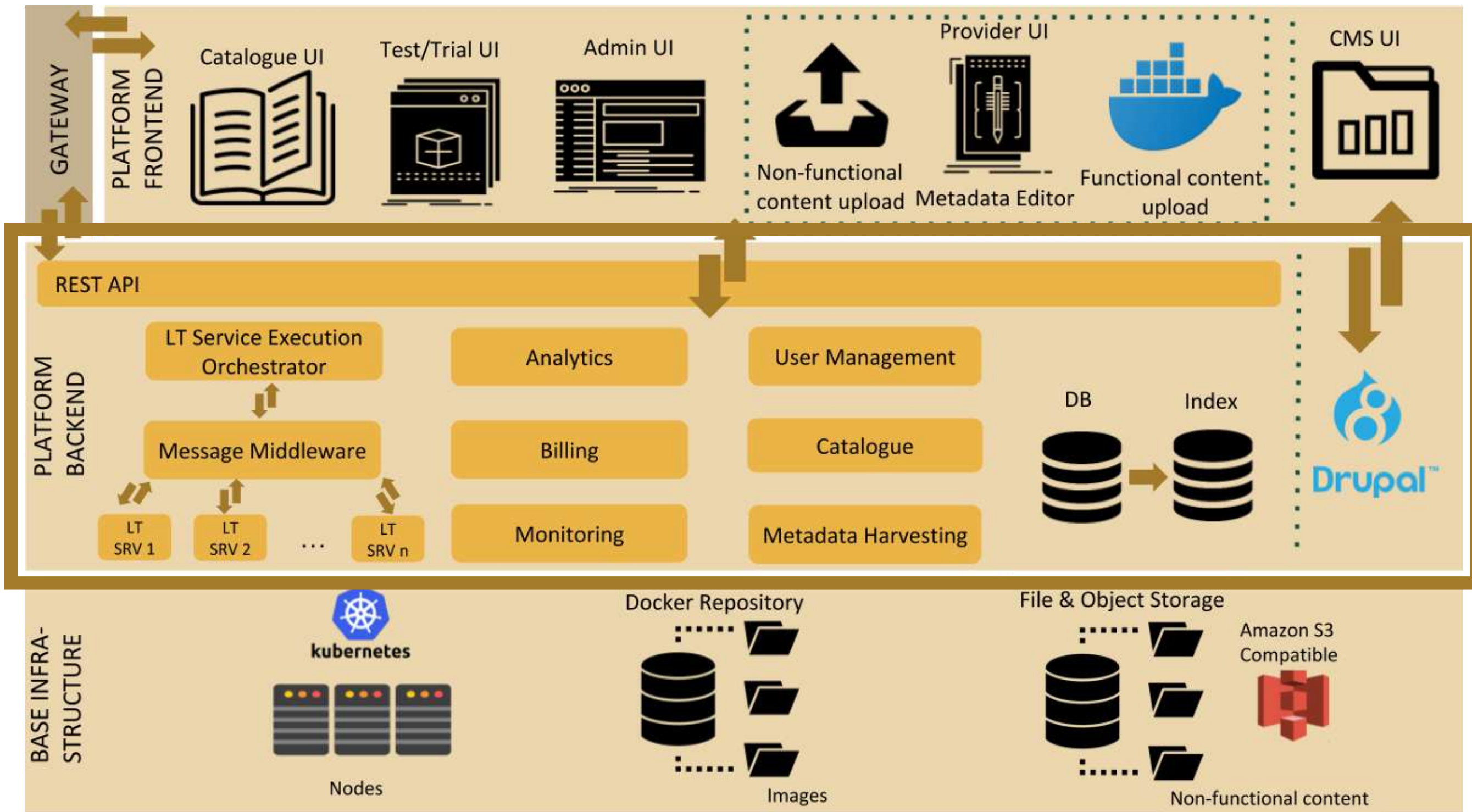


Introducing the European Language Grid

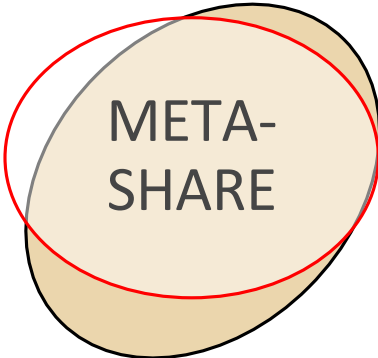
Deep Dive 2 – ELG Platform

08/09-10-2019, Brussels – META-FORUM 2019 – Stelios Piperidis, Andrejs Vasiljevs, Florian Kintzel

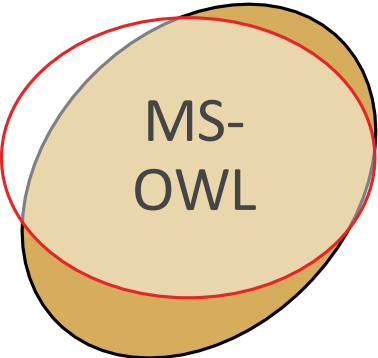
Platform Backend



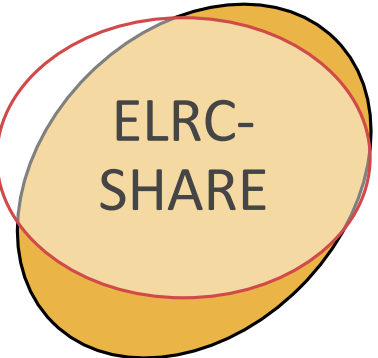
Starting Point and Evolution



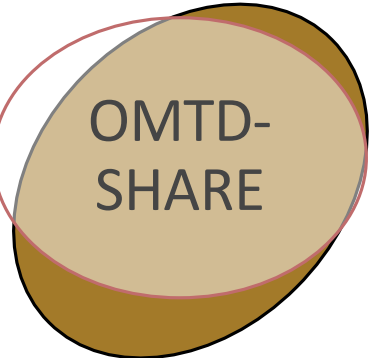
- Resource type: Corpora, Lexical/conceptual resources, Language descriptions, Tools/Services
- Media type: text, audio, video, image
- Updated in CRACKER
- XSD <https://github.com/metashare/META-SHARE/tree/master/misc/schema>



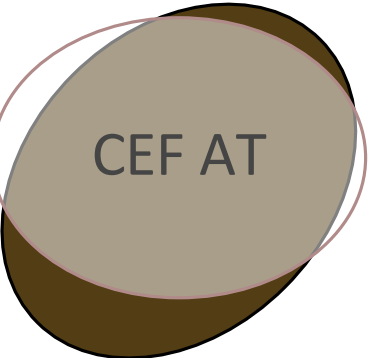
- RDF/OWL implementation of META-SHARE with some changes
- Improvements (e.g. legal metadata, distribution)
- <http://purl.org/net/def/meta-share>
- Currently in the process of update



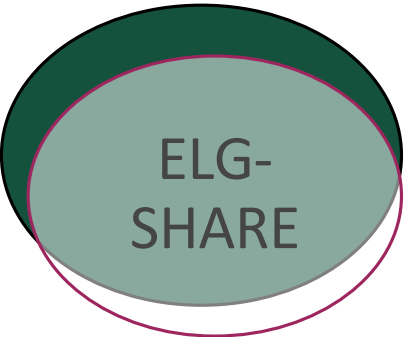
- Same entities but only for text
- Emphasis on public sector
- Additions/Improvements (e.g. GDPR)
- XSD <https://gitlab.com/ilsp-nlpi-elrc-share/elrc-share-repository/tree/master/misc/schema/ELRC2>



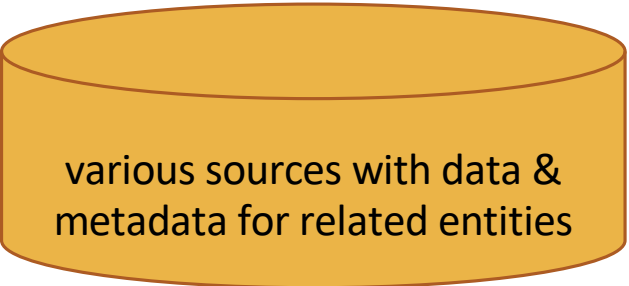
- Focus on TDM
- Addition of “publication”
- Division of software (components & apps)
- Focus on interoperability
- XSD <https://github.com/openminted/omtd-share-schema>
- Partial RDF/OWL (application type/function, annotation type, format & TDM method) <http://w3id.org/meta-share/omtd-share/>

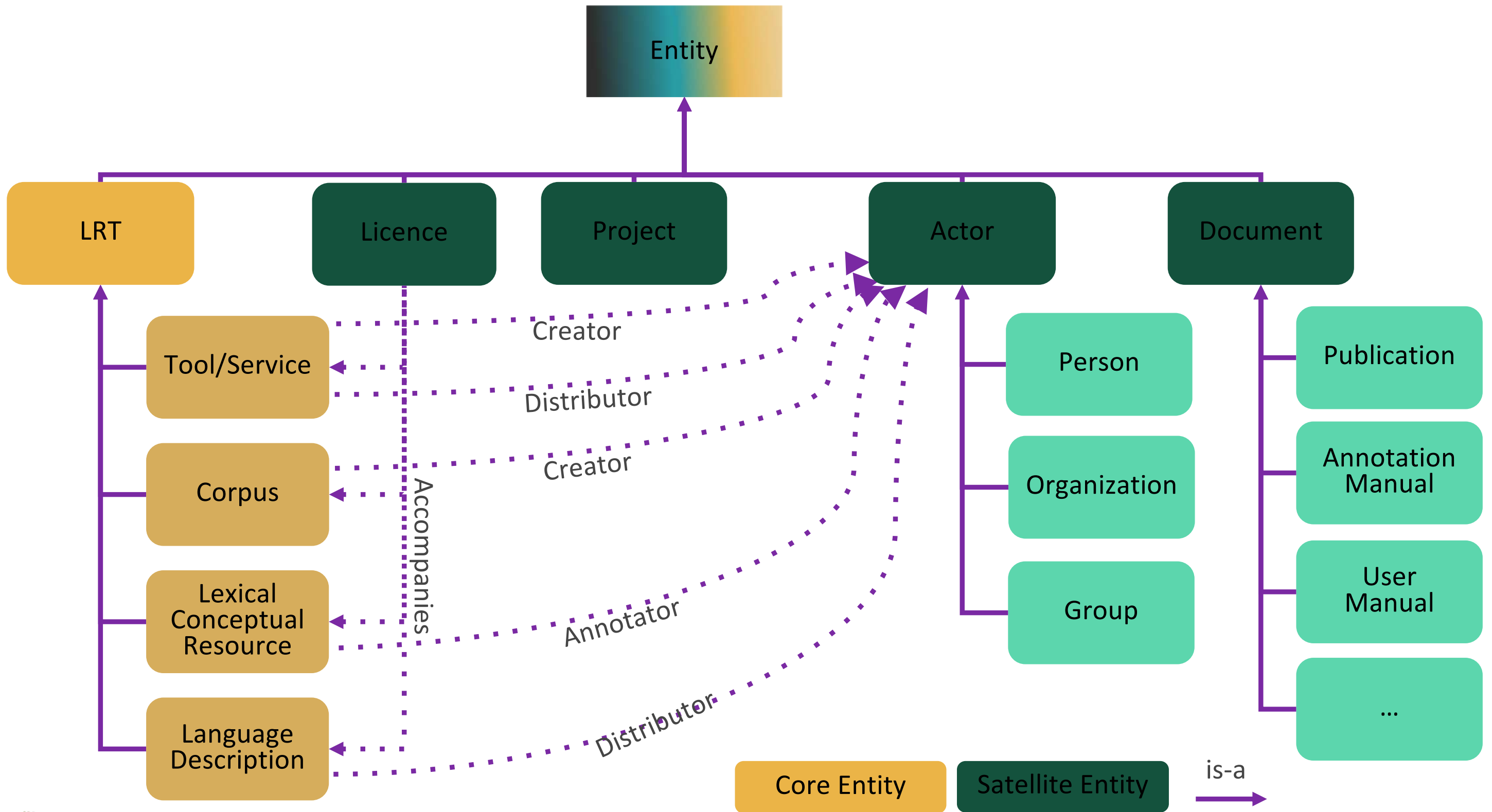


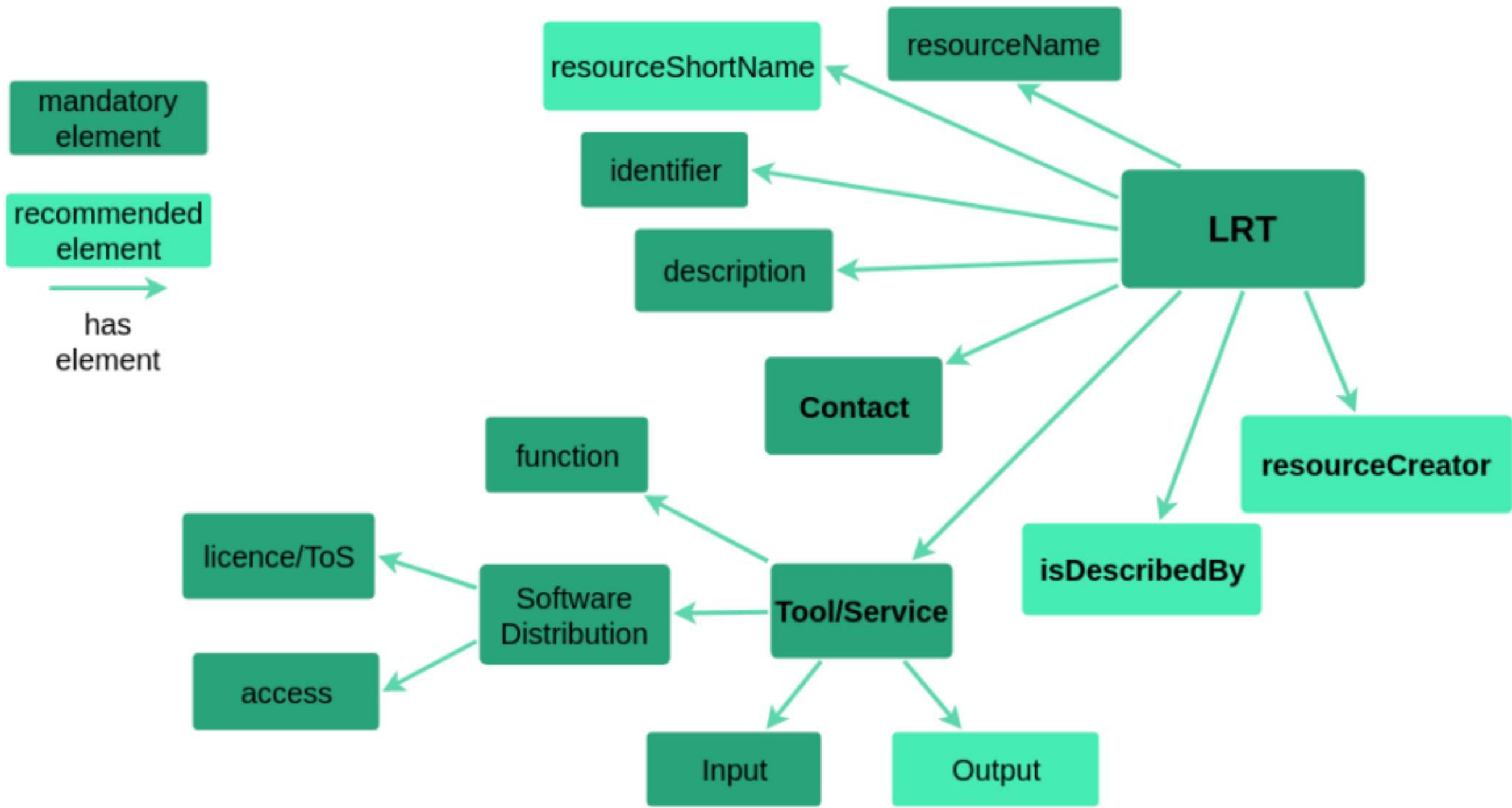
- Addition of “service”
- Geared to EU services
- Only software
- XSD not yet published



- Consolidation & update
- Linking with LT agents, activities, etc.
- FAIR principles
- XSD with RDF/OWL concepts
- JSON-LD export

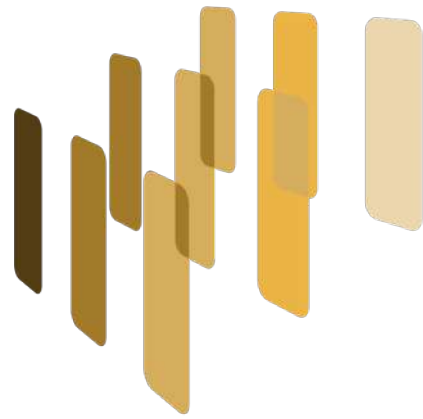






Four different Types of LT Assets

Type	Description	ELG Consortium brings in
1) Functional grid content: LT Tools and Services	Containerised services; can be uploaded and deployed as well as integrated through ELG	GATE, GATECloud, DKT, UDPipe, Tilde's and UEDIN's MT services, SAIL's ASR, KWS, sentiment (polarity) detection, age & gender detection tools, etc.
2) Functional grid content (remotely invoked): LT Tools and Services	Remote APIs can be registered, described, searched and integrated with the help of ELG	Research and commercial services (e.g. GATE Cloud with 65+ NLP services, UDPipe), TILDE's and UEDIN's MT services, SAIL's and EXPSYS's services, etc.
3) Non-functional grid content: LRs, data sets etc.	Upload, describe, search, download data sets, source code, models etc.	Data sets from META-SHARE, ELRC-SHARE, ELRA, resources, models, tools etc.
4) Yellow Pages of European LT	ELG catalogue entries, e.g., of an LT provider company	CRACKER, ELRA, ELRC (NAPs etc.), META-SHARE, LingHub, LT World etc.



Introducing the European Language Grid

Deep Dive 2 – ELG Platform

08/09-10-2019, Brussels – META-FORUM 2019 – Stelios Piperidis, Andrejs Vasiljevs, Florian Kintzel

User Interface

User Requirements Gathering

User Survey

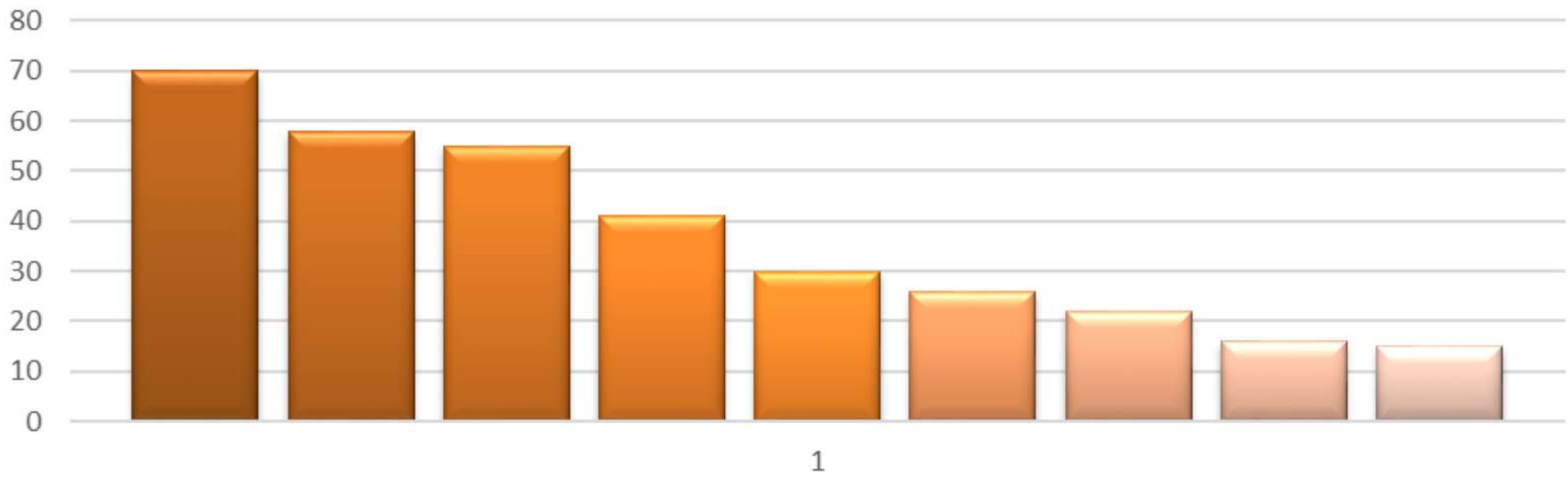
- Online survey to identify essential features and prioritize functional requirements
- 43 questions, 158 respondents

User Interviews

- In-depth interviews with user group representatives to uncover usage context, needs, motivations and success factors



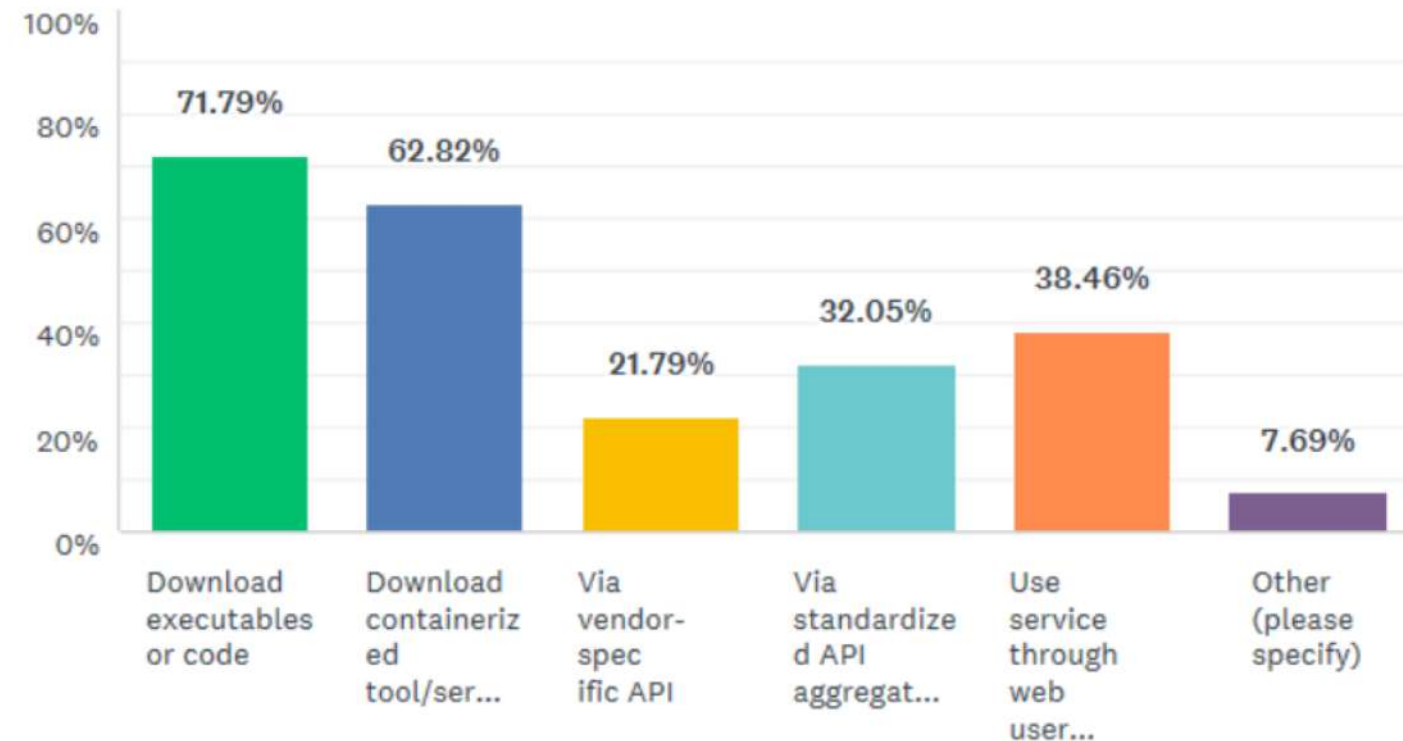
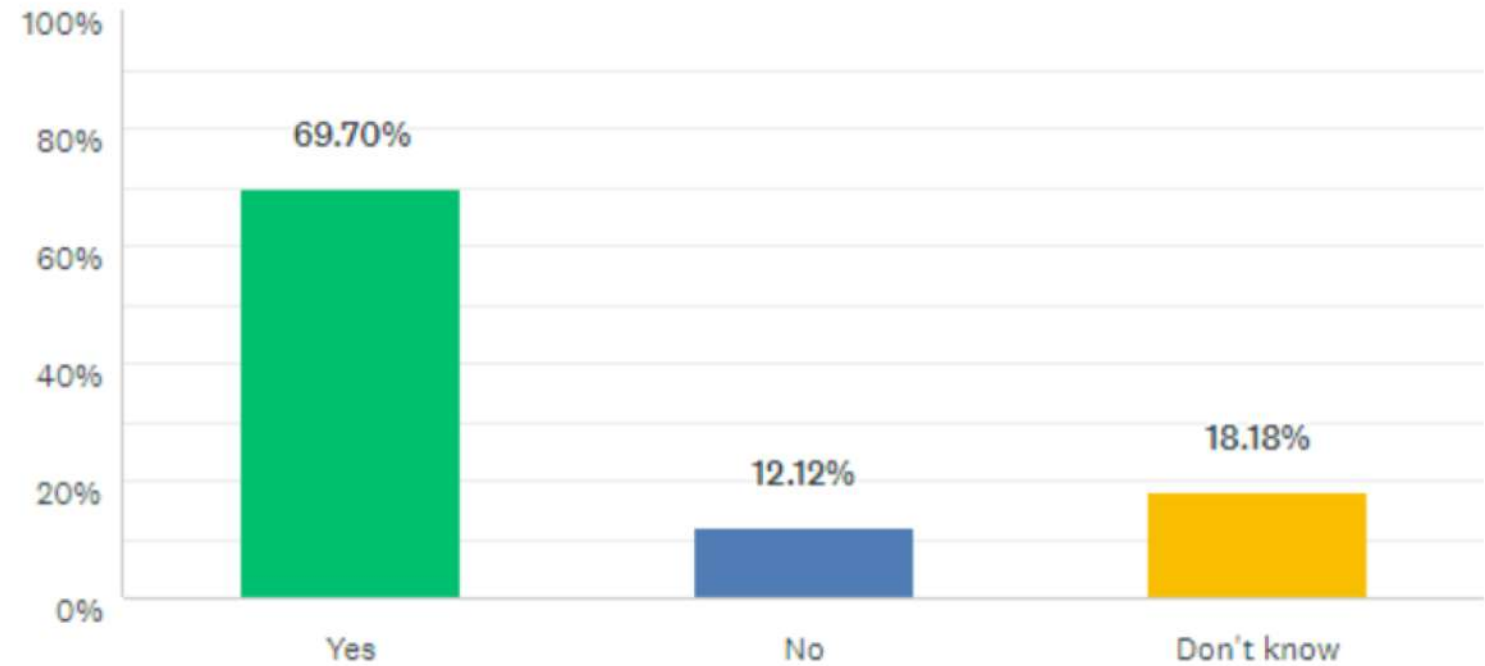
Most important criteria when searching for language data



- Language of the contents
- License/access conditions
- Domain/topic/type of data (e.g. finance, politics, biology, tweets, etc.)
- Size of dataset
- Price
- Available documentation
- Data type/format (e.g. RDF, Wave, MPEG, Text/Utf-8, etc)
- Name
- Ability to preview

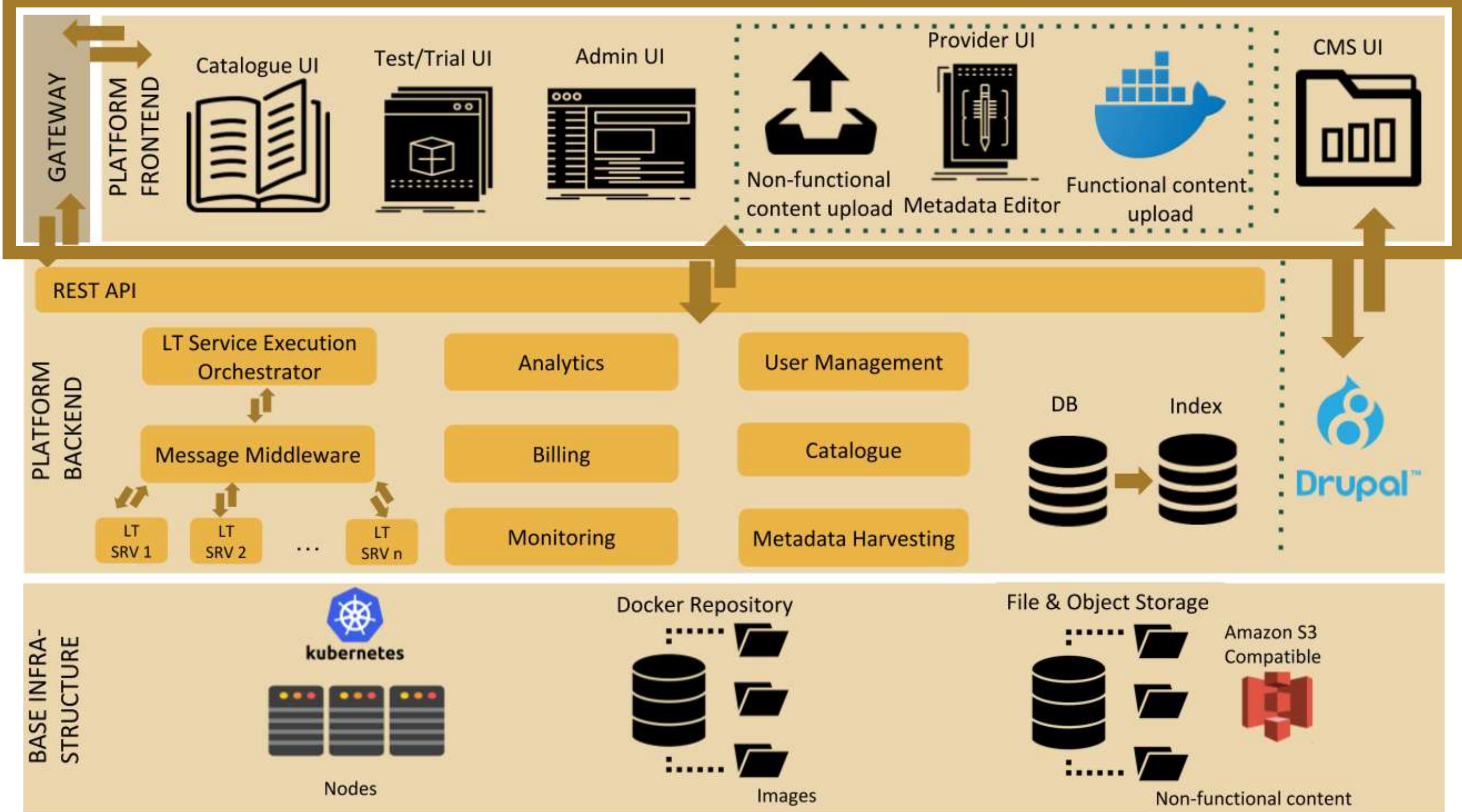
Interested in **providing** software as a container

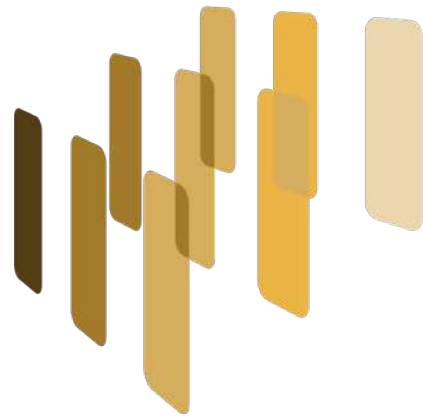
Interested in **using** software tools from the platform



Modelling User Journeys







Introducing the European Language Grid

Deep Dive 2 – ELG Platform

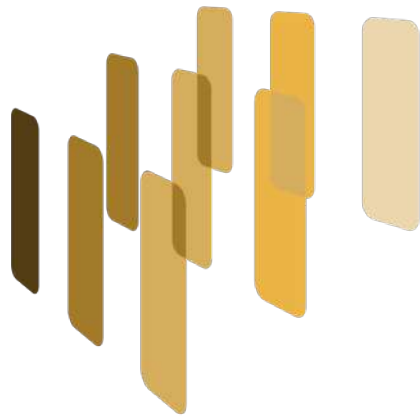
08/09-10-2019, Brussels – META-FORUM 2019 – Stelios Piperidis, Andrejs Vasiljevs, Florian Kintzel

Summary

Summary

- ELG designed as a modern platform integrating language technology services, downloadable data sets, tools, models etc.
- Enhancing visibility, discoverability and operability of language technology services and other assets.
- Tackling some of the interoperability problems.
- Facilitating industrial innovation.
- Building on past achievements and bridging to all major European, national and international initiatives in language technology and AI.





Introducing the European Language Grid Deep Dive 2 – ELG Platform

Thank you!



The European Language Grid has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825627 (ELG).

Stelios Piperidis (ILSP/Athena RC), Andrejs Vasiljevs (TILDE), Florian Kintzel (DFKI)
spip@athenarc.gr, andrejs@tilde.lv, florian.kintzel@dfki.de

08/09-10-2019, Brussels – META-FORUM 2019
<http://www.european-language-grid.eu>