EUROPEAN LANGUAGE GRID

META-NET
META-FORUM 2021
EUROPEAN LANGUAGE EQUALITY

European Language Grid:
Broad Overview – Services and Resources – Platform

Georg Rehm (DFKI), Stelios Piperidis (ILSP/R.C. Athena), Kalina Bontcheva (University of Sheffield),
georg.rehm@dfki.de, spip@athenarc.gr, k.bontcheva@sheffield.ac.uk
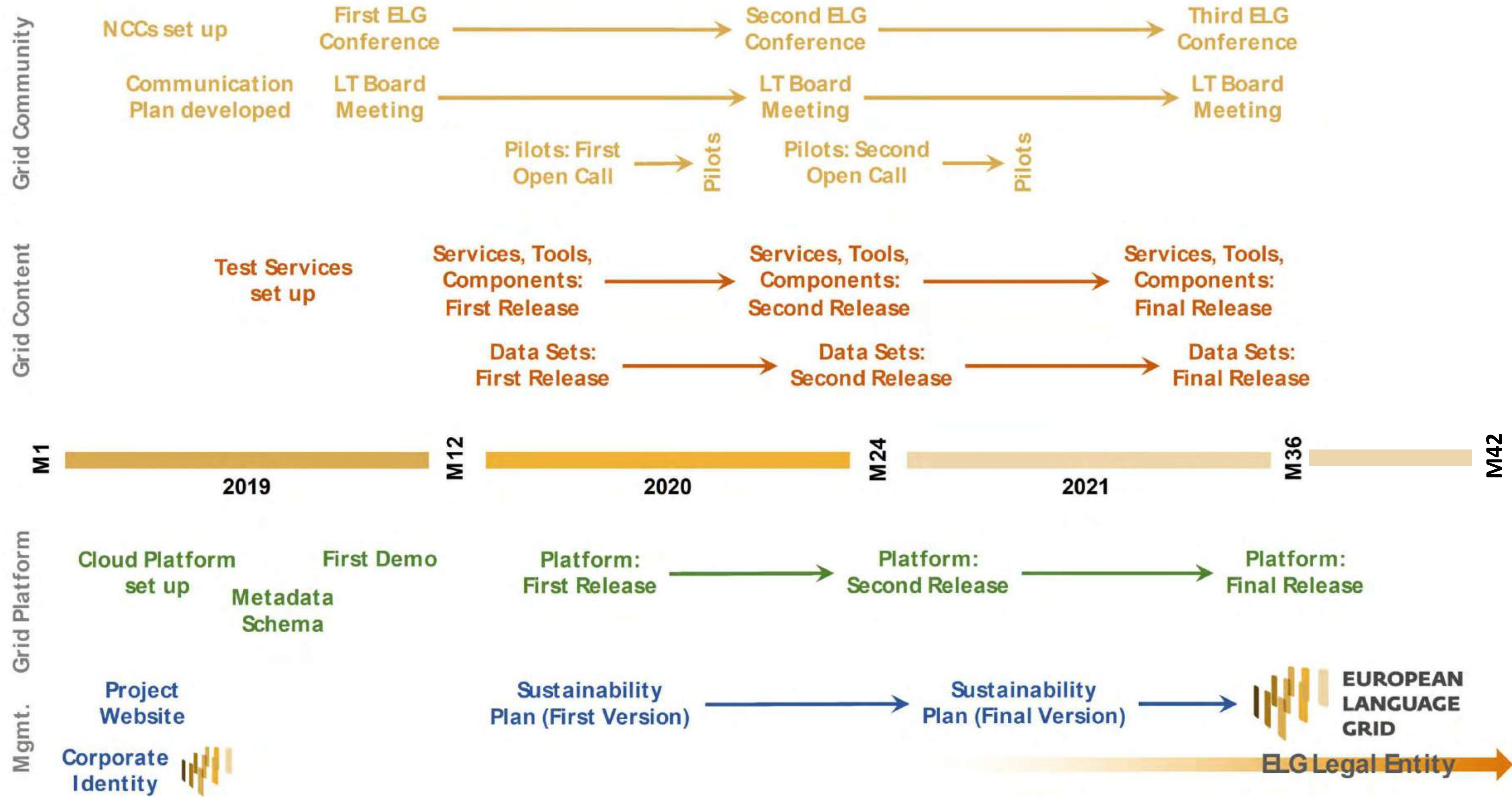
Objectives (Selection)

1. **Establish the ELG as the primary Language Technology platform and market place in Europe to tackle the fragmentation of the European LT landscape.**

2. ELG as a platform for commercial and non-commercial, industry-related LTs (functional and non-functional).

3. Enable the European LT community to upload services and data sets, to deploy them and to connect with, and make use of those resources made available by others.

4. Enable businesses to grow and benefit from scaling up.

5. Unleash enormous potential for innovation.

Kick-off meeting, 22/23 January 2019

Grid Community

NCCs set up

First ELG Conference → Second ELG Conference → Third ELG Conference

Communication Plan developed

LT Board Meeting → LT Board Meeting → LT Board Meeting

Pilots: First Open Call → Pilots

Pilots: Second Open Call → Pilots

Grid Content

Test Services set up

Services, Tools, Components: First Release → Services, Tools, Components: Second Release → Services, Tools, Components: Final Release

Data Sets: First Release → Data Sets: Second Release → Data Sets: Final Release

M1 — 2019 — M12 — 2020 — M24 — 2021 — M36 — M42

Grid Platform

Cloud Platform set up

First Demo

Metadata Schema

Platform: First Release → Platform: Second Release → Platform: Final Release

Mgmt.

Project Website

Sustainability Plan (First Version) → Sustainability Plan (Final Version) → EUROPEAN LANGUAGE GRID

Corporate Identity

ELG Legal Entity

European Language Grid – Current State of Play (November 2021)

- **Current state of play ELG Release 2 (Nov. 2021):**
  - 450 functional services and tools
  - 3563 corpora and data sets
  - 948 lexical/conceptual resources
  - 27 language descriptions, grammars, models
  - 1789 organisations (research orgs., companies)
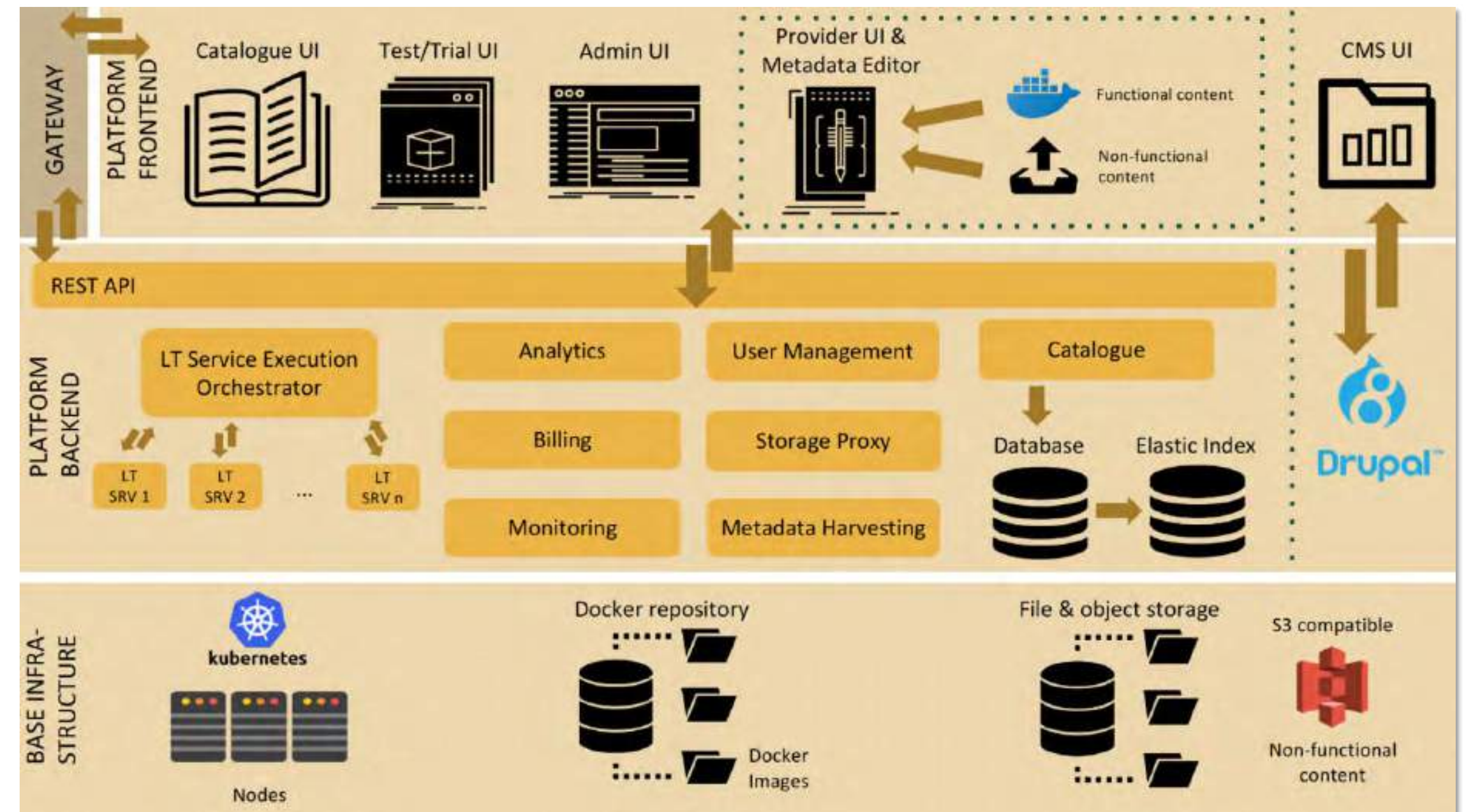
Users can connect to the ELG cloud platform via ELG APIs, remote APIs, ELG GUI, Python SDK, download of containers or source code.

➔ Kalina Bontcheva: ELG LT services and resources

➔ Stelios Piperidis: ELG platform and demonstration

Brows

urce

Te

parser

Using ELG Services – it's really very simple!

- In the project PANQURA we perform automated credibility assessment of online content.

- The corresponding tool is hosted in ELG and can be called through the public ELG API.

- The service's response is exposed to the user – in this specific example clearly indicating the source of the service and also where the server is located, which is important to users.

Stakeholders and Users

Companies that

- … *develop* Language Technologies
- … *integrate* Language Technologies
- … *purchase* Language Technologies

Universities and research centres that

- … *develop* Language Technologies
- … *use* Language Technologies

META-FORUM 2019 (8/9 October) – Brussels, Belgium

Public administrations that *purchase* or *use* Language Technologies

Other organisations (e.g., NGOs) that *purchase* or *use* Language Technologies

Funding agencies that support the development of Language Technologies

Stakeholders and Users and Collaborators

ELG Open Calls: 15 pilot projects

ICT-29b research and innovation projects

ELE – European Language Equality

AI4EU – European AI on demand platform

Other related projects: Elexis, Lynx,
   MeMAD, CEF INEA projects (MAPA, NTEU
   etc.), OpenGPT-X, NFDI4DataScience etc.

Core networks: META-NET, LT Innovate

Additional initiatives: CLAIRE, AI PPP, ELRC,
   ECSPM, EFNIL, BDVA, NPLD, CLARIN, W3C,
   NGI, RDA, EOSC, OpenAIRE etc.

Community: NCCs and LTC

32 National Competence Centres (NCCs)

- ◦ Strong international network of national networks to broaden ELG's reach, identify content for the ELG and interest companies in using the ELG.

- ◦ Main goal: *support the mission of the ELG project*.

European LT Council (LTC) (work in progress)

- ◦ A pan-European body, in which strategic LT-related matters can be discussed and coordinated

- ◦ Main goal: *represent and support European LT community*.

Austria — DAGMAR GROMANN — Centre for Translation Studies, University of Vienna

Belgium — WALTER DAELEMANS — Computational Linguistics and Psycholinguistics, University of Antwerp

Bulgaria — SVETLA KOEVA — Institute for Bulgarian Language, Bulgarian Academy of Sciences

Croatia — MARKO TADIĆ — Department of Linguistics, University of Zagreb

Cyprus — DORA LOIZIDOU — University of Cyprus

Czech Republic — JAN HAJIČ — Institute of Formal and Applied Linguistics, Charles University Prague

Denmark — BOLETTE SANDFORD PEDERSEN — Centre for Language Technology, Department of Nordic Studies and Linguistics, University of Copenhagen

European Language Grid: Sustainable Operational Model & Legal Entity

- ELG is supposed to be a long-term, sustainable initiative – a legal entity is needed.

- The technical and operational requirements – high availability and performance, SLAs, billing, support etc. – create non-trivial costs: hosting; bandwidth; ELG team; legal; etc.

- We've identified several ways and approaches of covering the costs on a long-term basis.

- Establish consensus for a sustainable operational model.

- Options: a) for-profit or b) *not-for-profit company*, c) association, d) foundation

- Business Model Canvas prepared

- First set of approx. 20 products defined

- Next up: market validation of the product ideas

- Q1/Q2 2022: Establish legal entity

Open Calls for Pilot Projects

Two open calls for pilot projects

- **Open Call #1:** 03/04 2020
- **Open Call #2:** 10/11 2020

Pilot projects shall

- **Type A:** broaden ELG's portfolio or
- **Type B:** demonstrate usefulness of ELG

Up to €200,000 per project

Approx. €2,000,000 FSTP in total

Available in Open Call #1: 1.3M€

Available in Open Call #2: 585k€

Project duration: **9-12 months**

Eligibility: **SMEs, research organisations**

Open Call #1 – Statistics

110 project proposals evaluated !
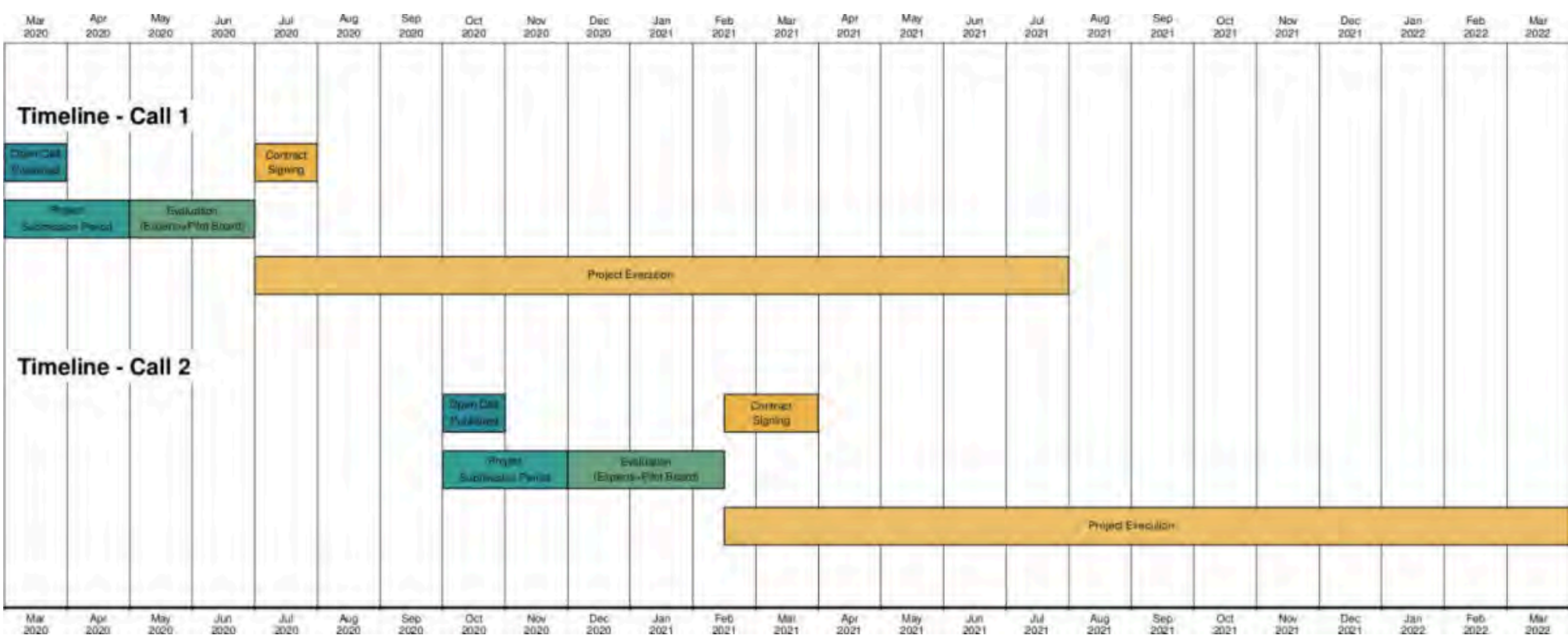
- **Type A:** 79 – **Type B:** 31 proposals

Total amount requested: 16.9M€

→ **10 projects selected**

Open Call #2 – Statistics

106 project proposals evaluated !

→ **5 projects selected**

Open Call #1	Fondazione Bruno Kessler	European Clinical Case Corpus ☑	IT	139,370€
	Lingsoft, Inc.	Lingsoft Solutions as Distributable Containers ☑	FI	140,625€
	Coreon GmbH	MKS as Linguistic Linked Open Data ☑	DE	167,375€
	Elhuyar Fundazioa	Basque-speaking smart speaker based on Mycroft AI ☑	ES	117,117€
	Universita' Degli Studi di Torino	Italian EVALITA Benchmark Linguistic Resources, NLP Services and Tools ☑	IT	126,125€
	University of Helsinki	Open Translation Models, Tools and Services ☑	FI	154,636€
	University of Vienna	Extracting Terminological Concept Systems from Natural Language Text ☑	AT	132,977€
	University of Turku	Textual paraphrase dataset for deep language modelling ☑	FI	166,085€
	Weber Consulting KG	Virtual Personal Assistant Prototype ☑	AT	87,445€
	FZI Research Center for IT	Streaming Language Processing in Manufacturing ☑	DE	132,160€
Open Call #2	Institute for Bulgarian Language	Multilingual Image Corpus 2021	BG	110,960€
	EDIA BV	CEFR Labelling and Assessment Services	NL	137,560€
	University of West Bohemia	Motion Capture 3D Sign Language Resources	CZ	85,421€
	Sapienza University of Rome	Universal Semantic Annotator	IT	113,228€
	Sign Time GmbH	Sign language explanations for terms in a text	AT	137,227€

European Language Equality (ELE)
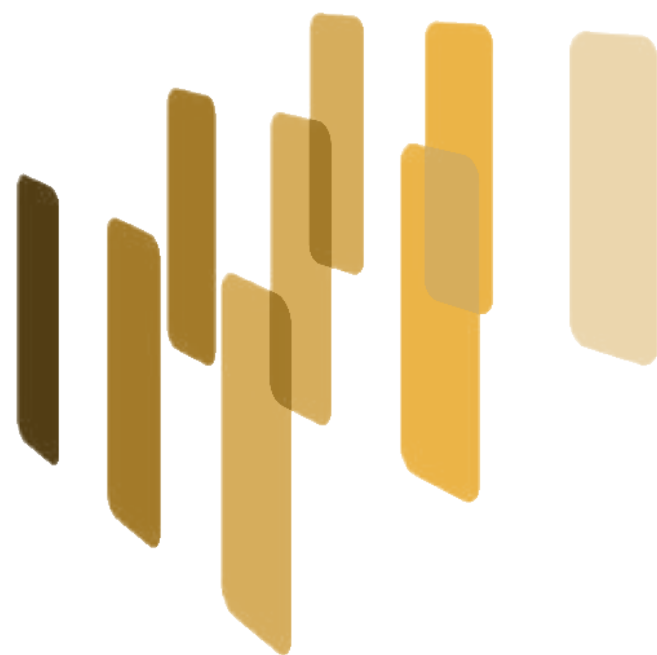
Consortium: 52 partners from all over Europe.

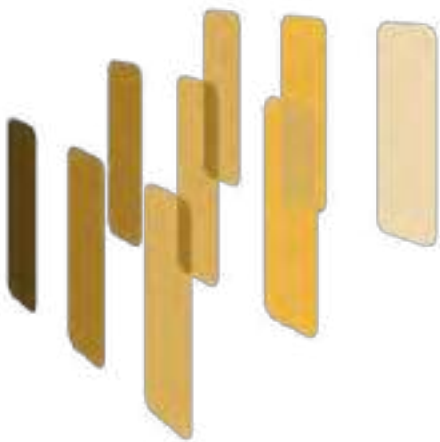**Coordinator:** ADAPT Centre (Dublin City University).

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

Runtime: 18 months – ELE and ELG will both end in June 2022.

Started on **1 January 2021**.
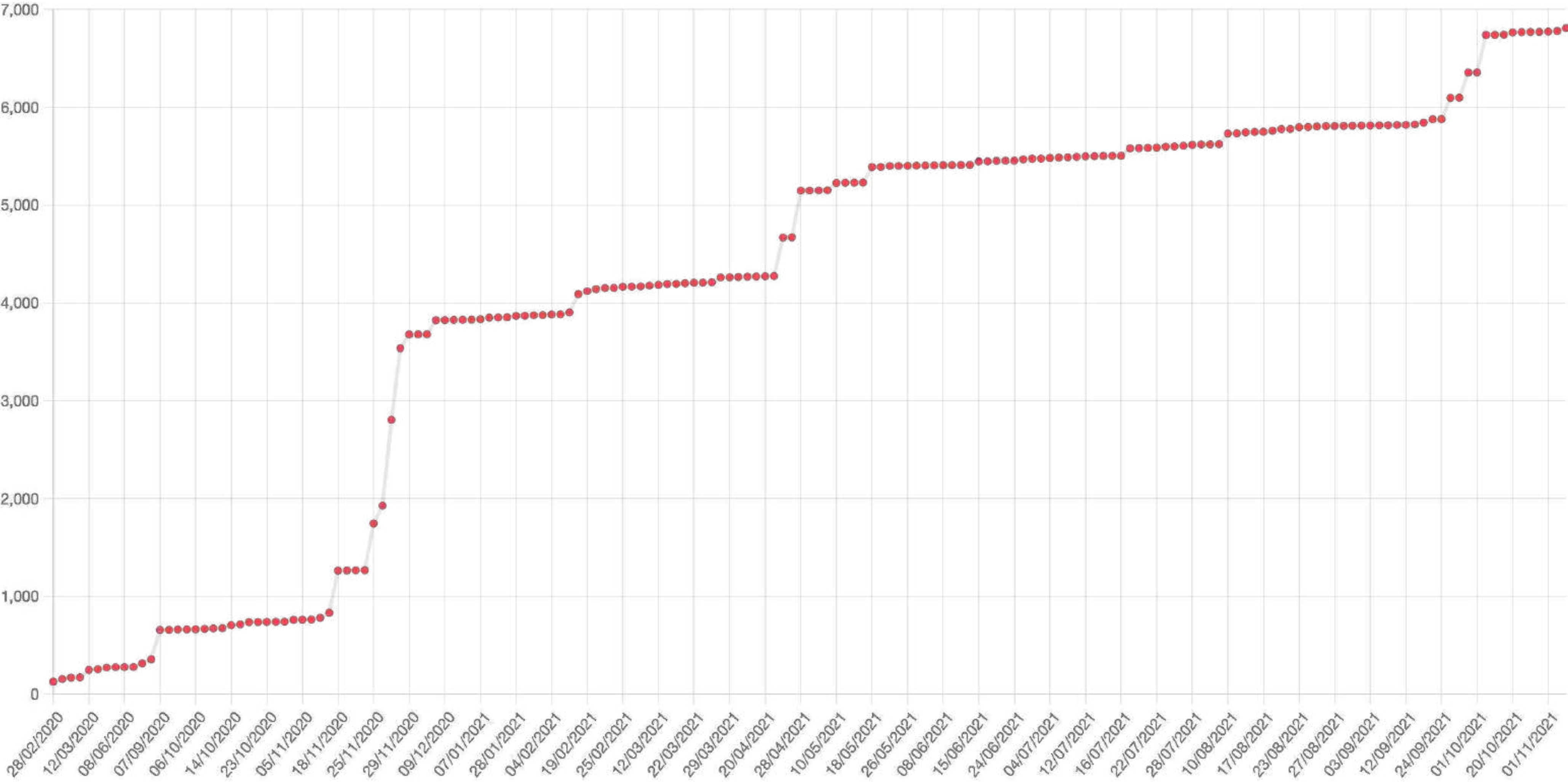
➔ ELE session (Wednesday, 09:00-10:30)

Resource Types

	01/12/2020 META-FORUM 2020	12/11/2021 META-FORUM 2021
Corpora	1851	3563
Tools/Services	173	450
Language Descriptions	7	37
Lexical Conceptual Resources	728	948
Projects	24	32
Total	**3679**	**5030**
Resources collected by ELE	–	approx. 6000
Total (online soon)	**3679**	**11030**

Number of Resources over Time

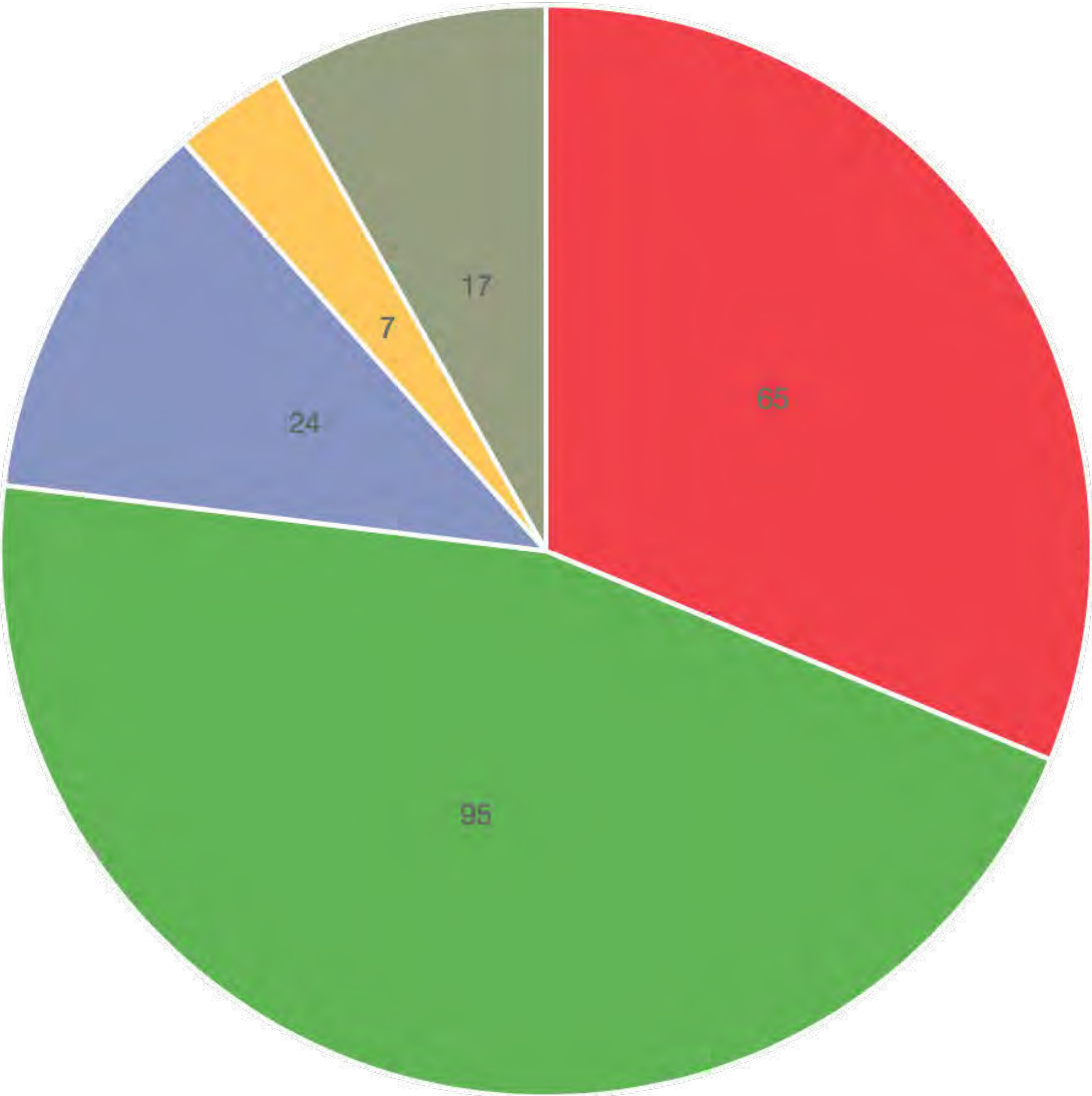# Number of Resources and Resource Types
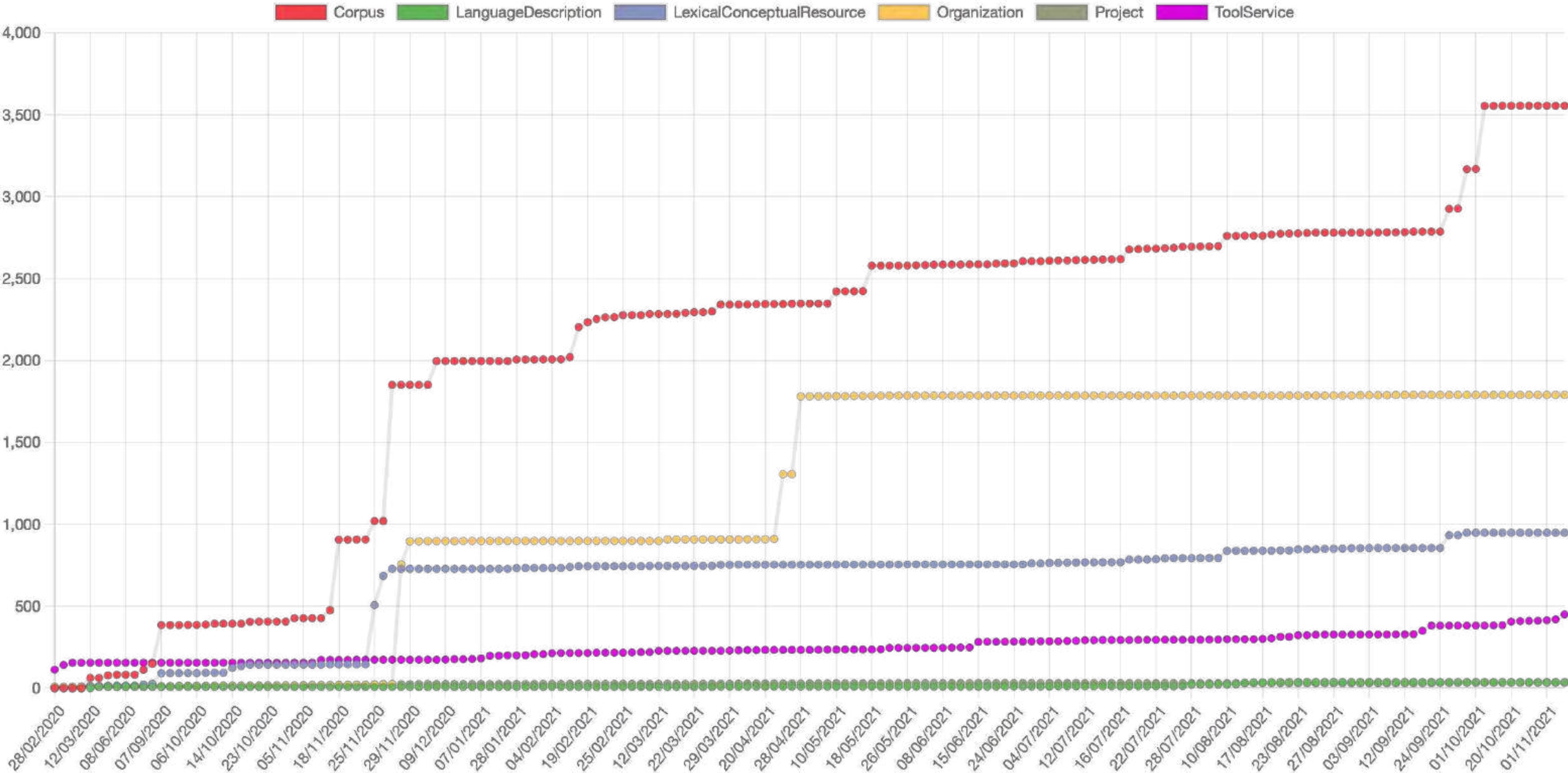
Number of Users
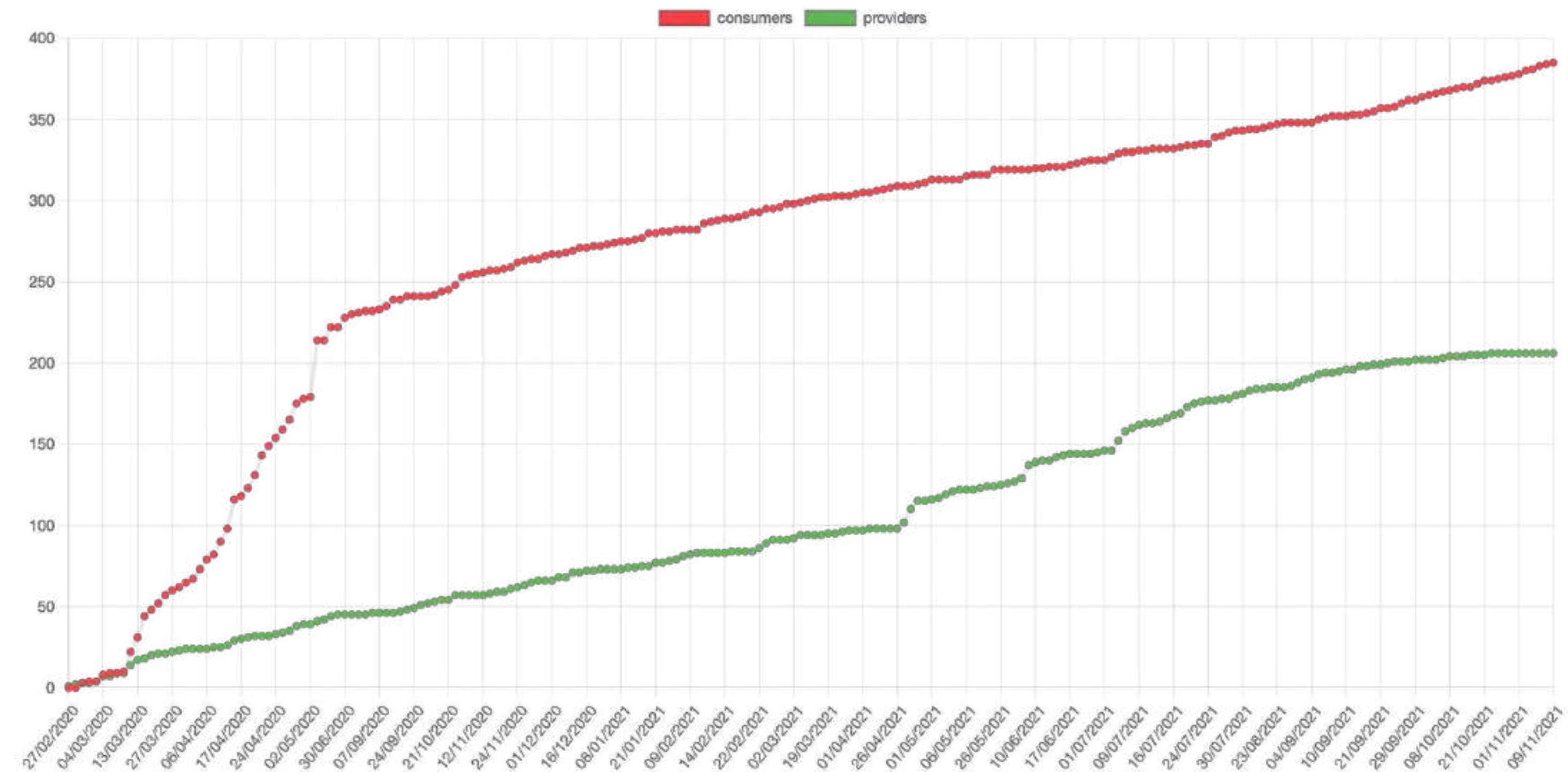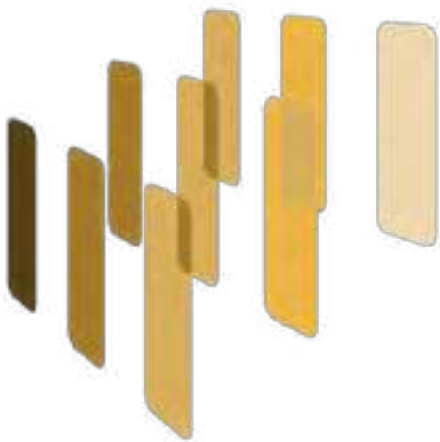
	01/12/2020 META-FORUM 2020	28/10/2021 META-FORUM 2021
Consumers	283	385
Providers	50	206
Total	**333**	**591**

Active Providers
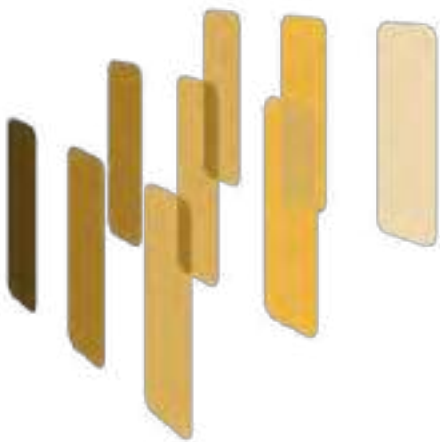
Legend: ■ 0 resource ■ 1 resource ■ 2-5 resources ■ 6-10 resources ■ +10 resources

Pie chart values: 65, 95, 24, 7, 17

Languages of the Resources
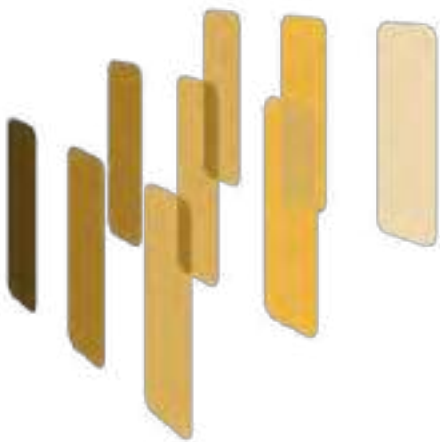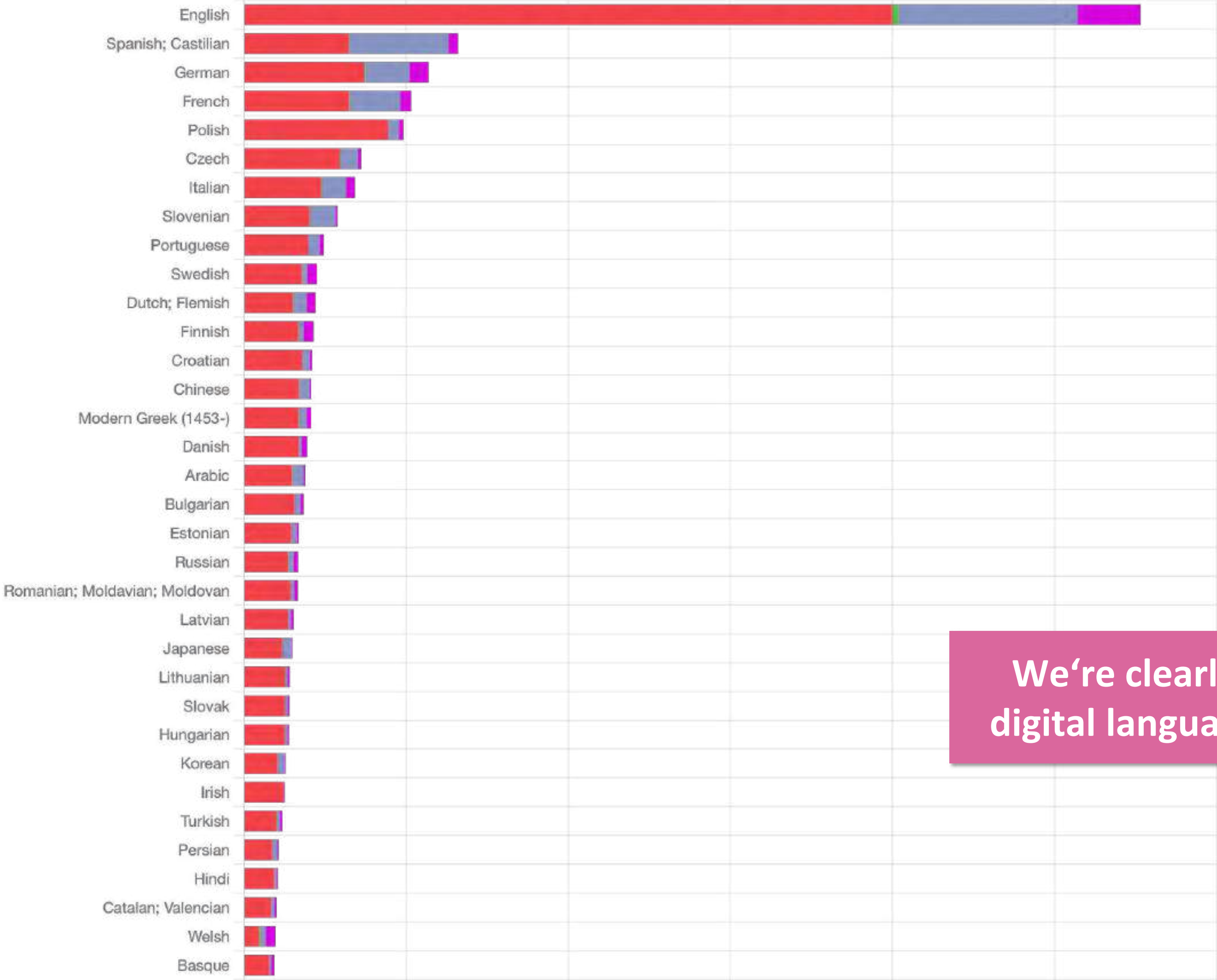
We're clearly very far away from digital language equality in Europe.