



# EUROPEAN LANGUAGE EQUALITY



## European Language Equality: Preliminary Results Defining Digital Language Equality

Federico Gaspari (DCU, ADAPT Centre)

15/16/17-11-2021 META-FORUM 2021 – Using the European Language Grid (virtual conference)

<http://www.european-language-grid.eu>

# Defining Digital Language Equality

- Work conducted as part of WP1 on the **current situation of language equality in Europe**
- Preliminary definition of **Digital Language Equality (DLE)** produced early **to guide the whole project**
  - Input received from **all consortium partners** with broad **consensus**
  - Ongoing exchange with a wide **range of external stakeholders**
  - Preliminary working definition to be refined and **finalized by early 2022**

# Defining Digital Language Equality

- **Aims**
  - Provide a **solid theoretical basis to achieve DLE** in Europe by 2030
  - A key **ingredient** of the evidence-based **SRIA** and **roadmap**
  - Enable **comparisons among Europe's languages**
  - Identify **gaps in LR and LT provision** to help prioritize **future interventions**
- Work is in progress to refine the **final definition of DLE** and the associated **metric by early 2022**

# Preliminary Working Definition: Key Features

- **Modular** and **flexible** design
  - Well-defined (separate and independent, but tightly integrated) **quantifiers, measures** and **indicators**
- **Compatibility** and **interoperability** with the metadata schema of the **European Language Grid (ELG)**
  - The DLE definition is **fully aligned with ELG**
- Basic Language Resource Kit (BLARK) (Krauwer, 2003)
- Firmly rooted in the **state-of-the-art**, but intended to guide and prioritize much-needed **future LT development** and **LR creation, collection** and **curation** activities

# Digital Language Equality: Preliminary Working Definition



Digital Language Equality (DLE) is the state of affairs in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age.

# Digital Language Equality *Metric*

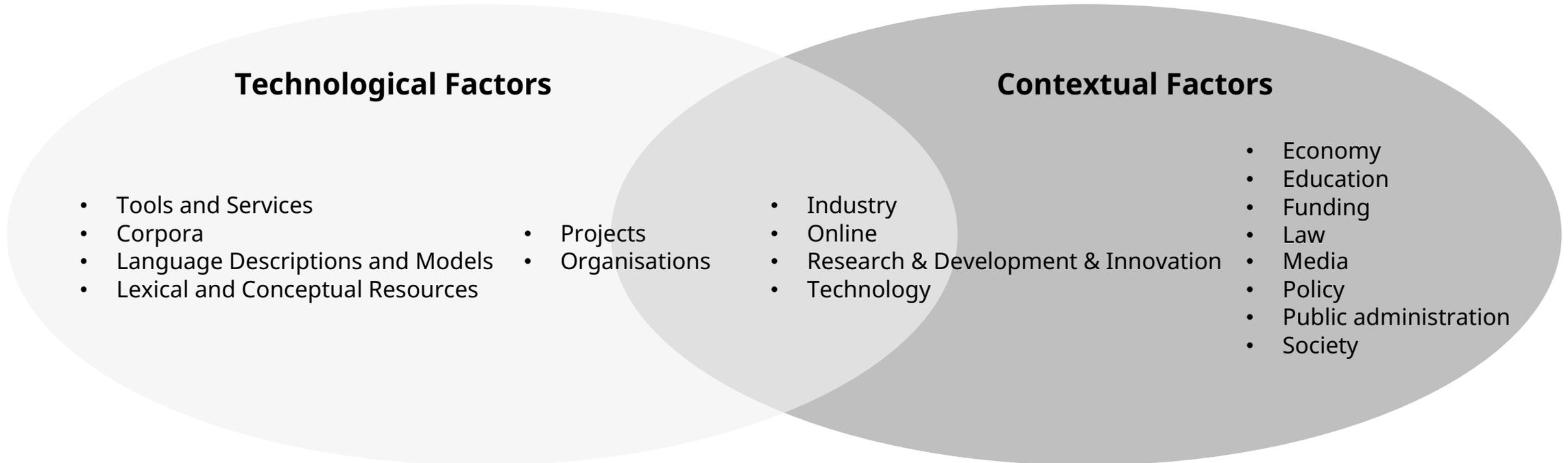


The DLE metric is a measure that reflects the digital readiness of a language and its contribution to the state of technology-enabled multilingualism, tracking its progress towards the goal of DLE.

The DLE metric is computed for each language on the basis of various factors, grouped into **technological factors** (e.g., the available language resources, tools and technologies) and a range of situational **contextual factors** (e.g., societal, economic, educational, industrial).

*dynamic nature, score can be (regularly) updated*

# DLE Metric: Technological and Contextual Factors



# DLE Metric: Six Technological Factors (TFs)

- *Corpora*
- *Tools and services*
- Language models and computational grammars (a.k.a. language descriptions)
- Lexical and conceptual resources
- Projects
- Organizations

*Various relevant studies, e.g., most recently **Joshi et al. (2020)** and **Blasi et al. (2021)***

*TFs documented for all of Europe's languages, fully **aligned with metadata schema of ELG***

*The six TFs correspond to as many **indicators of linguistic digital readiness** (comparable across languages)*

# Example of Technological Factors: *Corpora*

*Corpora or datasets, including collections of text documents, text segments, audio transcripts, audio and video recordings, etc., monolingual or bi- or multi-lingual, raw or annotated.*

- Language(s)
- Domain(s)
- Licence
- Type of access
- Corpus subclass
- Media type(s) of parts
- Multilinguality type
- Corpus size



# Key Sources of Inspiration for the Contextual Factors (CFs)

- **FLaReNet** report (Calzolari et al., 2011)
- META-NET White Paper series *Europe's Languages in the Digital Age* (Rehm and Uszkoreit, 2012)
- META-NET's *Strategic Agenda for Multilingual Europe 2020* (Rehm and Uszkoreit, 2013)
- **STOA** report (2017)
- European Parliament resolution of September 2018 on *Language Equality in the Digital Age*
- EFNIL's **European Language Monitor (ELM)**, especially **ELM4** from **2019**
- Project on **Modeling the Linguistic Consequences of Digital Language Contact**
- **Digital Language Diversity Project**
- *etc.*

# DLE Metric: Contextual Factors

- **72 CFs** identified as part of the **preliminary working definition of DLE**, currently being refined
- Grouped into **12 categories**:
  - ***Economy***
  - ***Education***
  - Funding
  - Industry
  - Law
  - Media
  - Online
  - Policy
  - Public Administration
  - Research & Development & Innovation
  - Society
  - Technology

# Example of Contextual Factors: *Economy*

- Size of the economy of the respective country, countries, region(s)
  - Size of the LT/NLP market in the respective country, countries, region(s)
  - Size of the language service and translation / interpreting market in the respective country, countries or region(s)
  - Percentage of the IT/ICT sector relative to the whole economy of the respective country, countries or region(s)
  - Investment instruments or accelerator programs targeting AI/LT/NLP start-ups
  - Regional or national LT/NLP/LSP etc. market (including forecast)
  - Average socio-economic status of members of the language community
- Currently defining **relative weighting of technological factors and contextual factors**
  - Open issue: identify reliable and consistent **sources of up-to-date data** for as many CFs as possible

# Conclusion and Next Steps

- DLE definition useful to the **LT community** for future inclusive and sustainable **progress**
- Defining **relative weighting between TFs and CFs**, so that the **DLE metric** can provide **intuitive and transparent scores**
- Plans for **comparative visualization** of DLE metric scores **across languages** (dashboard)
- **Final definition of DLE and specification of DLE metric** in early 2022, reflecting **consensus in the consortium**, also considering the **whole community**, including **external stakeholders** such as:
  - industry players, developers, vendors, academics, researchers, scientists, users, consumers, decision- and policy-makers, language (and rights) advocates and activists, etc.



# Thank you!



The European Language Equality project has received funding from the EU under grant agreement No LC-01641480 – 101018166 (ELE)

Federico Gaspari (DCU, ADAPT Centre)

15/16/17-11-2021 META-FORUM 2021 – Using the European Language Grid (virtual conference)

<http://www.european-language-grid.eu>