European Language Grid: One Year After

Georg Rehm[1], Stelios Piperidis[2], Dimitrios Galanis[2], Penny Labropoulou[2], Maria Giagkou[2], Miltos Deligiannis[2], Leon Voukoutis[2], Martin Courtois[1], Julian Moreno-Schneider[1] and Katrin Marheinecke[1]

[1]Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Berlin, Germany

[2]R. C. "Athena", Institute for Language and Speech Processing (ILSP), Athens, Greece

Corresponding author: georg.rehm@dfki.de

Abstract

The European Language Grid (ELG) is a cloud platform for the whole European Language Technology community. While the EU project that developed the platform successfully concluded in June 2022, the ELG initiative has continued. This article provides a description of the current state of ELG in terms of user adoption and number of language resources and technologies available in early 2024. It also provides an overview of the various activities with regard to ELG since the end of the project and since the publication of the ELG book, especially the co-authors' attempt to integrate the ELG platform into various data space initiatives. The article also provides an overview of the Digital Language Equality (DLE) dashboard and the current state of DLE in Europe.

Keywords: LR Infrastructures and Architectures, LR National/International Projects, Tools, Systems

1. Introduction

The European Language Grid (ELG) is a joint cloud platform for the whole European Language Technology community.[1] It provides access to thousands of language resources (LRs) and language technologies (LTs) as well as to related information such as academic and commercial organisations active in the European LT space. The platform is thoroughly documented in the form of several articles (e. g., Labropoulou et al., 2020; Rehm et al., 2021) and, most recently, a comprehensive book that provides detailed technical and conceptual information concerning all aspects of the EU project and the technology platform (Rehm, 2023).

In a development which occurred in parallel with and independently from ELG and its EU project, several different communities have started fleshing out the concept of *data spaces*, which have by now reached a certain level of relevance in Europe. Most importantly, the European Union (EU) kickstarted the development of a number of official, EU-mandated data spaces in various sectors to support the EU Data Strategy. These activities are supported through the EU project Data Spaces Support Centre (DSSC),[2] in which the main initiatives – Gaia-X,[3] IDSA,[4] FIWARE,[5] BDVA[6] – collaborate with other partners and other data space projects to develop the underlying technical and organisational specifications of the European data space ecosystem. In terms of data provisioning, data exchange, data sharing and data markets, it is safe to predict that the emerging DSSC blueprint will gain more and more importance and relevance in the coming years. Because they did not exist back when the ELG project was prepared, one crucial aspect could not have been foreseen to be further investigated in the ELG EU project and that was the relation of ELG to data spaces.

This article describes recent developments around the European Language Grid, especially developments started in 2023 that we have not been able to describe in the ELG book. The remainder of this article is structured as follows. First, Section 2 provides a brief general overview of the ELG platform while Section 3 presents an overview of current statistics in terms of users and resources provided. Section 4 describes the dashboard implemented for monitoring the situation towards Digital Language Equality (DLE) in Europe. Section 5 reports on our attempts to integrate ELG into various data space initiatives including the Common European Language Data Space (Section 5.1), IDSA (Section 5.2), and Gaia-X (Section 5.3); Section 5.4 discusses interoperability across data spaces at the level of metadata. Section 6 describes related work and Section 7 concludes the paper.

[1]https://www.european-language-grid.eu
[2]https://dssc.eu
[3]https://gaia-x.eu
[4]https://internationaldataspaces.org
[5]https://www.fiware.org
[6]https://www.bdva.eu

2. ELG: A Brief Overview

The European Language Grid (ELG) project was funded by the EU under the Horizon 2020 research

LREC-COLING 2024, pages 6353–6362

and innovation programme.[7] It developed the ELG cloud platform as a joint technology platform for the whole European LT/NLP community to overcome the stark fragmentation in this community. At the end of June 2022, the EU project was completed and Release 3 of ELG was made available.

The ELG platform brings together various LTs, such as machine translation systems, natural language processing tools, and other language-related services, as well as LRs such as datasets, corpora, vocabularies, terminologies, ontologies and models. The ELG currently provides approx. 17,500 resources and has more than 1,100 users. Its platform and APIs are valuable resources for researchers, developers, businesses, and the general public, promoting multilingualism and facilitating cross-border initiatives within Europe. The project aims to foster innovation, collaboration, and the sharing of knowledge, language technologies and resources among stakeholders across Europe. This is especially important given the rich linguistic diversity in the continent.

Descriptions of the various parts of the ELG project and components of the technology platform can be found in Rehm (2023). Piperidis et al. (2023) provide an overview of the basic concepts.

3. ELG in 2024

While the ELG EU project was concluded in June 2022, the platform continues to be available as a joint initiative for the European LT and NLP community. Compared to the most recent report (Arranz et al., 2023), the ELG catalogue has grown substantially. Figure 1 shows the development of LRTs available in ELG since the beginning of the project, while Table 1 indicates the growth of the ELG catalogue per different type of LRTs over this time period. Figure 2 shows the LRTs available for the 24 official EU languages. Figure 3 shows the development of users of the European Language Grid over time.

	Number	
Resource Type	**08/2022**	**03/2024**
Corpus	6937	8074
Language Description	479	512
Lexical Conceptual Resource	2699	2825
Organisation	1776	1779
Project	37	514
Tool or Service	3451	3854
Total Resources	15379	17558

Table 1: Number of LRTs available in ELG

[7] https://cordis.europa.eu/project/id/825627

4. ELG and Digital Language Equality

The ELG catalogue is Europe's most comprehensive list of LRTs. On top of the LRTs hosted or deployed in the ELG cloud, the ELG catalogue is continuously populated and updated through targeted activities, such as harvesting several major LR/LT repositories[8] and crowd-sourcing endeavours (Labropoulou et al., 2023). The catalogue comprises more than 17,500 metadata records, encompassing both data and tools/services, covering almost all European languages – both official and regional/minority ones (see Section 3). We consider the present status of the ELG repository representative with regard to the current situation of technology support for Europe's languages.

This is why ELG was selected as the underlying data source of a mechanism for dynamically exposing and monitoring the support that Europe's languages receive through technology. This mechanism was implemented by ELG and its sister project, European Language Equality (ELE, Rehm and Way, 2023), as an interactive dashboard.[9]

The dashboard provides an overview of the Digital Language Equality (DLE) metric (Gaspari et al., 2023), and the two main components contributing to it, i. e., the technological (TFs, Gaspari et al., 2022) and contextual factors (CFs, Grützner-Zahn and Rehm, 2022). It exposes the TFs (based on the contents of the ELG catalogue) and the CFs as interactive visuals dynamically created in response to user queries. With regard to the TFs, as the ELG catalogue organically grows over time, at any given point in time the dashboard provides an up-to-date overview of LT support that each language enjoys, also showing where the status is less than ideal or not at the expected level.

The dashboard also enables users to dive into more detailed comparisons of subsets of the TFs, both across languages and within a language. Datasets can be compared per resource types (corpora, tools, lexical resources, etc.) and resource features (e. g., media types, access rights, etc.). Tools and services are compared as per the basic functions they perform (e. g., text or speech processing, NLG, translation technologies etc.). Cross-language comparisons are available in both table- and chart-based comparative visualisations, in particular as heatmaps and tables or radial bars (Figure 4 shows an example). The generated tabular data can be downloaded by the user as CSV files for further analysis, while the visuals can be downloaded as SVG files.

[8] ELRC-SHARE, LINDAT CLARIAH-CZ, CLARIN.SI, CLARIN-PL, Hugging Face and Zenodo.

[9] https://www.european-language-grid.eu/catalogue/dashboard

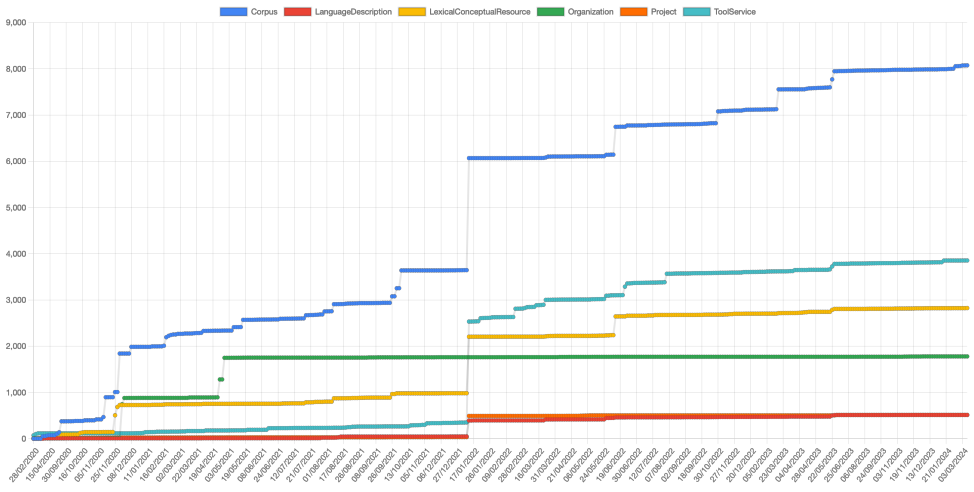

Figure 1: Growth of the ELG catalogue over time

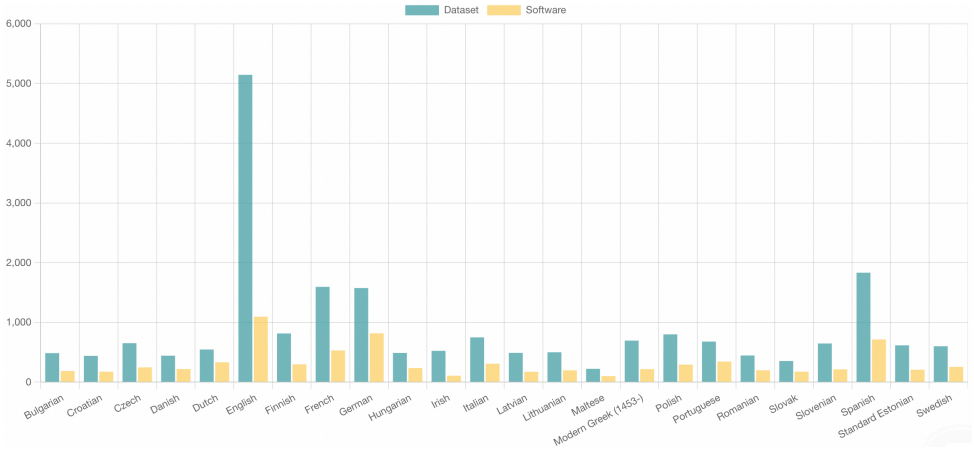

Figure 2: Number of LRTs available in ELG (24 EU languages only)

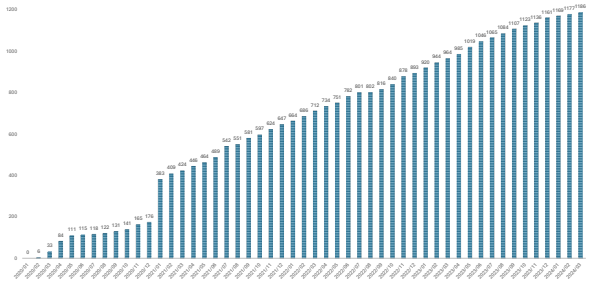

Figure 3: Number of ELG users (since 01/2020)

One of the most significant recent features of the dashboard is the consideration of the perspective of time and the possibility to compare languages as per the evolution of the number of resources over time and to monitor this development.

The "Evolution over time" section of the dashboard creates charts displaying either the overall data evolution over time, or the intensity at which data evolved in each time period. The overall evolution of the number of resources shows data as a series of points connected by a line. The data is plotted on a coordinate system, with the horizontal axis representing time and the vertical axis representing the number of resources. Multiple languages can be selected for simultaneous comparison and the time periods can also be viewed per semester or quarter, depending on the granularity of the visualisation and analysis that is required by the user. The charts also include additional features such as labels, axes, grid-lines, and legends to enhance readability and facilitate data interpretation in a user-friendly manner.

The "Intensity evolution" chart provides an overview of the rate at which data for each language has been created or imported into ELG over time. Again, multiple languages can be simultaneously selected for comparison and the time periods can also be viewed per semester or quarter. Figure 6 shows an example. In this chart, the data for each language is plotted on a vertical axis, with time on the horizontal axis. Each language is represented by a separate line drawn as a filled-in area. The areas are arranged in such a way that

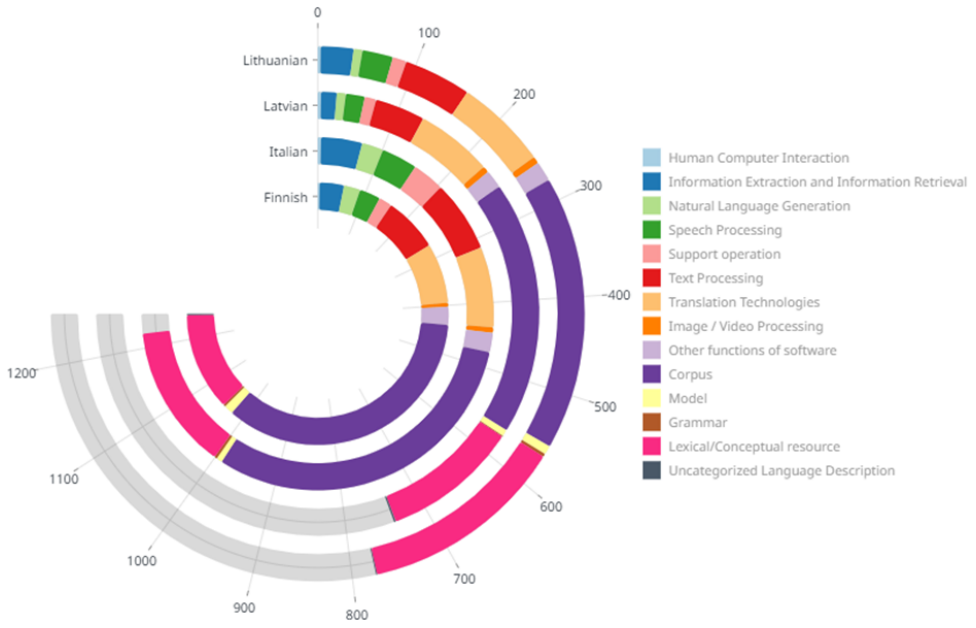

Figure 4: Example of a radial bar comparing Lithuanian, Latvian, Italian and Finnish

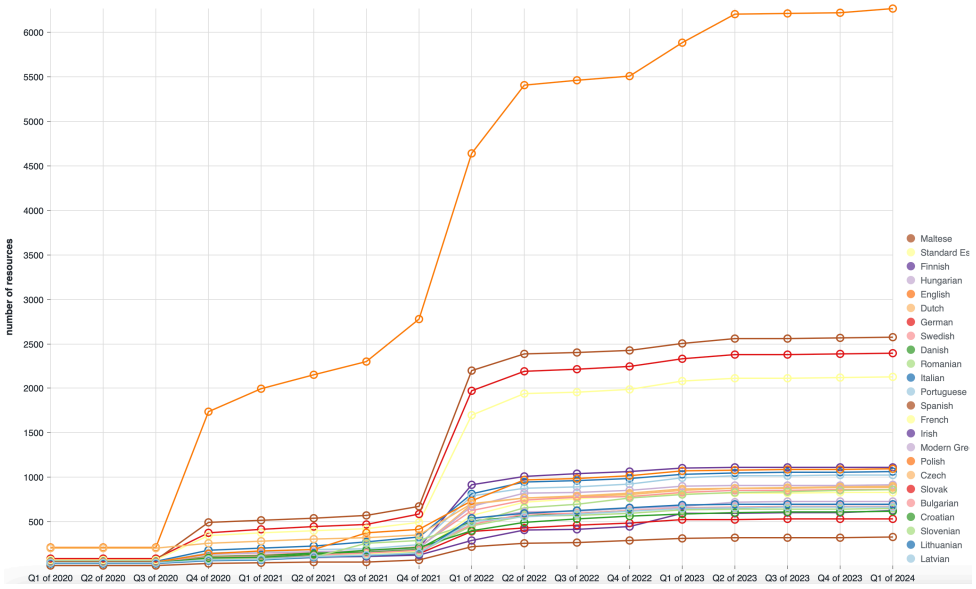

Figure 5: Evolution over time (24 official EU languages only)

the language that has the highest number of resources appears wider. As time progresses, the areas become wider or thinner, reflecting changes in the relative positions of the language's contribution. This type of chart can be particularly useful when comparing multiple languages, tracking relative changes and positioning over time, e. g., as a result of dedicated funding and policy interventions (or, rather, in case of lack thereof).

These features of the dashboard include evolution over time and correlations between various factors for each language. The updated visuals unveil supplementary data attributes that can be leveraged for monitoring the advancement of technology support across different languages. They have the potential to spark inquiries, stimulate further research, and offer insightful ideas to individuals, domain experts, as well as policy- and decision-makers, at the European, national, regional and local levels.

5. ELG in the Data Spaces Landscape

A data space (DS) is a decentralised infrastructure for the trustworthy sharing and exchange of resources (e. g., datasets, services, LLMs) between its participants (e. g., companies, academic

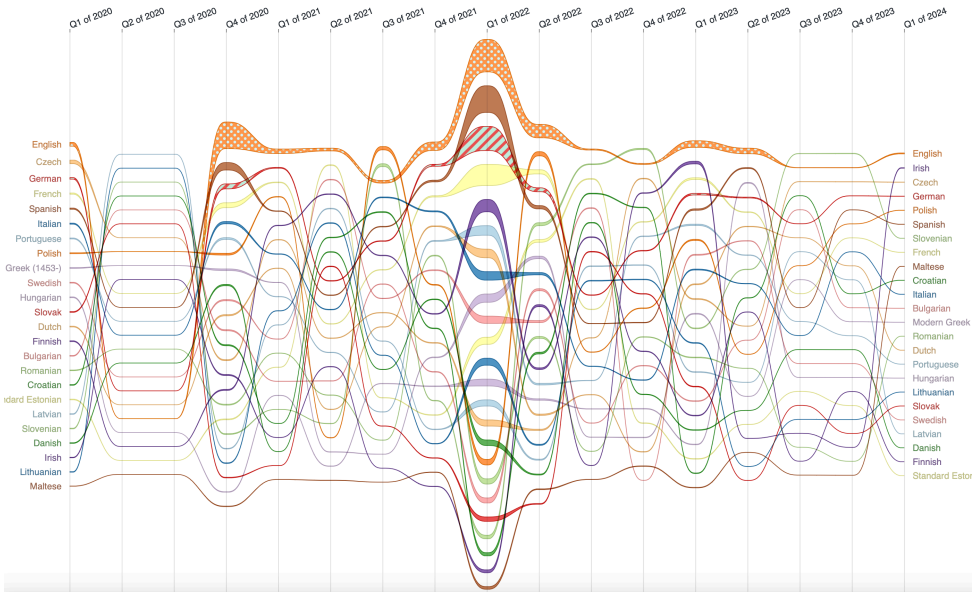

Figure 6: Intensity evolution over time (24 official EU languages only)

and public organisations, etc.) based on commonly agreed upon principles. Important reference documents based on which data spaces are designed and built include the International Data Spaces Association (IDSA) Reference Architecture Model (RAM)[10], the OpenDEI Design Principles (Nagel and Lycklama, 2021) and, most recently, the DSSC blueprint Version 1.0 (Guilloud et al., 2024).

Organisations that wish to join a data space follow a registration and onboarding process based on governance policies and implement the appropriate technical process. According to this, participants get an access token that verifies their membership/identity. Data space participants interact with one another through a well-defined protocol and the respective agent, a software system called *data space connector* configured with the access token. The connector provides (to its users) all required functionalities for 1. hosting resources and documenting them with appropriate descriptions, 2. defining usage and access policies and contract offers for these resources, 3. querying other connectors for their offerings, 4. negotiating (i. e., controlling) access to the offerings on a case-by-case basis and 5. exchanging or transferring the actual resources when an agreement has been reached between two parties.

There is currently a large number of data spaces (e. g., Mobility Data Space,[11] Catena-X,[12], for more examples see the Data Space Radar[13]) and each of them uses a different connector implementation tailored to the specific domain and project requirements. There are two main basic connector technologies to use and extend: the Data Space Connector[14] and the Eclipse Dataspace Components connector (EDC),[15] an open source project hosted by the Eclipse Foundation. The latter is the reference implementation of the Dataspace Protocol (DSP),[16] which is designed to be the foundation for technical interoperability in data spaces over the next years; this is the reason why its adoption is increasing at a significant rate.

In addition to connectors, the architecture of data spaces foresees a central catalogue (metadata broker), in which all participating connectors publish their resources. However, the use and deployment of a central catalogue in a data space is typically considered optional. Other technical components perform required functionalities (e. g., identity provision and management, logging of transactions, etc.).

5.1. ELG and LDS

The Language Data Space (Rehm et al., 2024)[17] (LDS) will support the collection, creation, sharing and re-use of multilingual and multimodal language resources (e. g., datasets, LLMs, lexica, on-

[10] https://docs.internationaldataspaces.org/knowledge-base/ids-ram-4.0

[11] https://mobility-dataspace.eu

[12] https://catena-x.net

[13] https://internationaldataspaces.org/adopt/data-space-radar/

[14] https://github.com/International-Data-Spaces-Association/DataspaceConnector

[15] https://github.com/eclipse-edc

[16] https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/overview/readme

[17] https://language-data-space.ec.europa.eu

tologies, etc.). It is one of the officially mandated Common European Data Spaces[18] and funded through a procurement contract in the Digital Europe Programme (DEP).[19] The LDS follows the specifications for data spaces and especially the recommendations from the Data Spaces Support Centre[20] (Guilloud et al., 2024), a project initiated by the European Commission to achieve interoperability among the Common European Data Spaces (see Section 6).

The LDS intends to bring together all existing and emerging stakeholders involved in language data provision and consumption with a focus on new industry stakeholders, and support their activities. ELG is planned to participate in LDS and make available its (hosted) resources. ELG will follow the required procedure for joining LDS, according to its governance framework, and implement the necessary technical modules, such as the setup and configuration of the LDS connector, once fully implemented. In addition, the ELG metadata model will be mapped to the LDS one and converters will be developed.

The LDS project started in January 2023, and the governance and technical frameworks are still under discussion. At this stage, initial plans for the technical modules include the design and development of the LDS connector on the basis of the EDC technology; further extensions will be added for supporting multilinguality and other required features. In addition, the LDS metadata model is under development based on the most popular vocabularies in data spaces and the community (see also Section 5.4).

5.2. ELG and DataBri-X (IDSA)

The DataBri-X (DBX) project[21] is funded by the European Union under Horizon Europe. It started in October 2022 and will last for three years. The project develops a toolbox governed by a modern process, enabling users in data-intensive projects of various domains in conducting their data analysis activities as well as customising the process according to their objectives and needs supported by automated tool chain execution and monitoring. The DBX toolbox is designed to be easily interoperable with existing and emerging data spaces in multiple domains. Interoperability will be achieved through the respective connectors that will be developed for this purpose. For example, one dedicated connector will enable DBX to access ELG data and use it as input to tool execution workflows. The proposed solution will be compliant with the IDSA RAM architecture and the IDS Information Model (see Section 5.4).

5.3. ELG and Gaia-X

Gaia-X is a federated infrastructure that has many similarities with data spaces, but the two initiatives also have differences (IDSA, January 2021). Both infrastructures are federations that consist of participants that are accepted following a registration process, however, in Gaia-X a federated catalogue (along with other central services) is obligatory. In this catalogue, a Gaia-X participant can register metadata (known as self-descriptions, SDs) for resources (datasets, services, etc.) based on its model, the Gaia-X core ontology.[22] However, the use of a connector is not required. The participant (e. g., an administrator) can directly use the respective REST services of the Gaia-X federated catalogue.[23]

As part of the OpenGPT-X project[24] we are interfacing ELG with the Gaia-X federated catalogue set up by OpenGPT-X. To achieve this, ELG metadata records that describe resources are converted to Gaia-X compliant SDs. In addition, an SD is created for ELG itself which is identified as a Gaia-X legal entity; this happens because all SDs/resources should be associated with a participant. Then the resources' SDs are signed using the mechanism described in the Gaia-X architecture and they are submitted to the federated catalogue, making the respective ELG resources discoverable and searchable on the Gaia-X federation through their SDs. The resources are published in the federated catalogue only if their signatures can be verified. This signature and verification procedure is crucial for Gaia-X, contributing to the establishment of trust. The verification is performed in a decentralised manner, i. e., with Decentralized Identities (DIDs) that are inserted into the SDs during signing. DIDs resolve to DID Documents containing the required information (i. e., a public key) used for validation/verification. The DID documents are stored on the web servers managed by the participants and for that reason are considered trusted and secure.

[18] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

[19] https://digital-strategy.ec.europa.eu/en/activities/digital-programme

[20] https://dssc.eu

[21] https://databri-x.eu

[22] https://gaia-x.gitlab.io/gaia-x-community/gaia-x-self-descriptions/core/core.html

[23] The EDC Connector is designed to be interoperable with Gaia-X and can be used for exchanging resources.

[24] https://opengpt-x.de

5.4. Interoperability across Data Spaces at the Metadata Level

Data interoperability is a key enabling factor for data spaces and refers to technical and semantic interoperability.[25] The latter, ensuring that data is correctly interpreted, is attained by sharing semantic vocabularies, such as ontologies, data models, schema mappings, API specifications, etc. These are used by providers to describe and annotate their assets (data or data-related services), and by consumers to understand their contents, structure, ways of accessing and processing them, etc. (Guilloud et al., 2024; Nagel and Lycklama, 2021).

The use of a common model for descriptions (i. e., metadata records or self-descriptions)[26] of products offered in a data space, possibly extended with domain-specific vocabularies that cater for domain-specific needs and idiosyncratic features of these products, lays the foundations for interoperability at the metadata level and facilitates discoverability and accessibility. However, despite ongoing discussions, there is no consensus yet on a single core model used across data space initiatives. In addition, other vocabularies and models may be used by providers and, in this case, they should publish them through the vocabulary hub, a central component in data spaces, supporting semantic interoperability.

The Dataspace protocol (Guilloud et al., 2024) recommends DCAT[27] for the description of catalogues, products and services. The same recommendation is made by the DSSC blueprint which also refers to DCAT-AP,[28] a DCAT application profile designed for use between data catalogues and portals especially in the European Union. The metadata model to be used in LDS builds upon DCAT-AP and extends it so as to be able to describe language resources and models as well as language processing services. As mentioned in Section 5.2, DataBri-X resources will be compliant with the IDSA recommendations and, therefore, the IDSA Information Model.[29] The IDSA model, re-using DCAT and other popular vocabularies, and complemented with a large set of new classes and properties, is a rich and well-organised model covering all aspects of information required for the description and operation of data products and services. Finally, to interface with OpenGPT-X, ELG resources will be imported in the federated catalogue (see Section 5.3) and must therefore comply with the Gaia-X core ontology[30] and its extension used in the Gaia-X Trust Framework (Gaia-X, 2022) also re-uses DCAT and other vocabularies, yet to a lesser extent, introducing new classes and properties for the description of the same data products from their own perspective.

The ELG metadata model (Labropoulou et al., 2020) caters for the description of multilingual and multimodal data resources, categorised as corpora, lexical/conceptual resources and language descriptions, and language processing services and language technology applications (see Section 4). It includes a set of common properties used for all resource types and sets of features specific to each resource type (Figure 7). The optionality status for the use of properties, set to *mandatory*, *recommended* and *optional*, facilitates flexibility while guaranteeing that the essential features of a data or service are included in a metadata record.

Interfacing ELG to LDS, DataBri-X and OpenGPT-X relies on the mapping of the ELG metadata model to their own models and the conversion of the ELG metadata records into records compliant with these models. This is a non-trivial task as it involves first the mappings of classes and then their properties, often with differences in the way these may be conceptualised (Labropoulou et al., 2023). Thus, mappings can be categorised into direct one-to-one mappings, overlaps, mappings necessitating the fulfilment of specific conditions, mappings of properties with different data types (free text values vs. controlled vocabularies), etc. To address also the need for mappings and conversions into multiple models a generic semi-automatic workflow has been devised. The design of generic templates for the encoding of matches organised in the above categories supports the creation of customised scripts that implement the appropriate conversions. Using this workflow, a subset of the ELG model has already been mapped to the classes and their properties in the IDSA and Gaia-X models and, in anticipation of the LDS model, to the DCAT-AP profile. Table 2 shows a few examples.

6. Related Work

As mentioned in Section 1, a few years ago several communities have started developing the concept of *data spaces*, each with their own empha-

[25] See Rehm et al. (2020) for a first attempt at describing the challenges with regard to establishing interoperability between AI and NLP platforms.

[26] Both Gaia-X and IDSA use the term "self-description" for metadata records, albeit with differences as detailed by IDSA (January 2021).

[27] https://www.w3.org/TR/vocab-dcat-3/

[28] https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe/release/300

[29] https://international-data-spaces-association.github.io/InformationModel/docs/index.html

[30] https://gaia-x.gitlab.io/gaia-x-community/gaia-x-self-descriptions/core/core.html

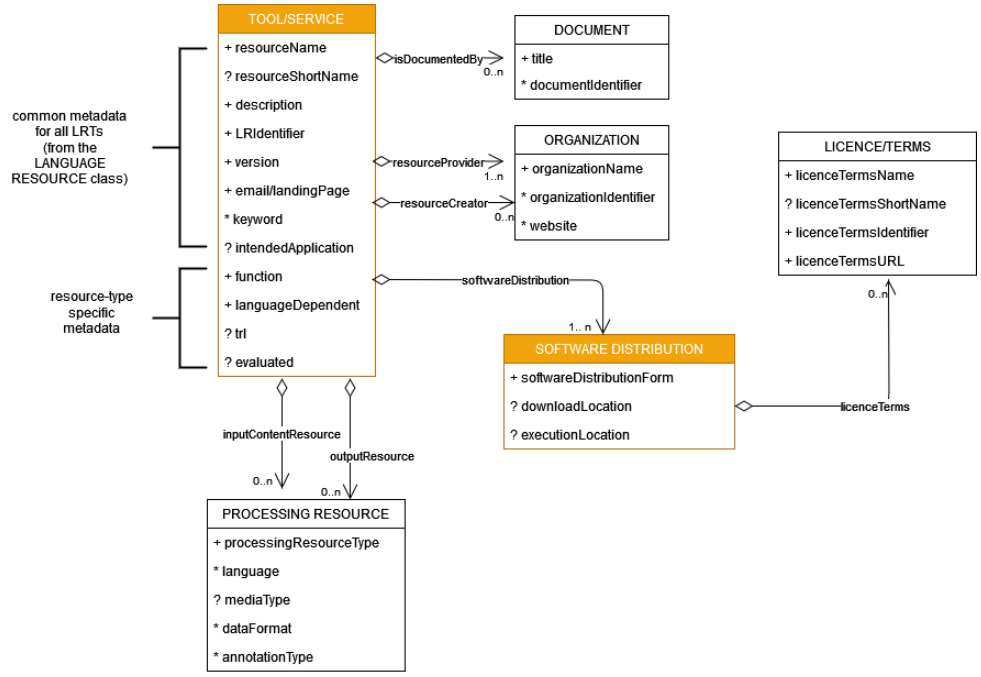

Figure 7: Minimal ELG schema with emphasis on the category “tool/service”

ELG	DCAT-AP	IDSA	Gaia-X
ms:Corpus	dcat:dataset (to be further typed)	ids:Resource or ids:TextResource or ... (depending on ms:mediaType value)	gax-t-f:Resource
ms:resourceName	dct:title	dct:title	gax-t-f:name
ms:keyword	dcat:keyword	dcat:keyword	dcat:keyword
ms:format	dct:format or dcat:mediaType	dcat:mediaType (with ids:IANAMediaType)	gax-t-f:formatType
ms:licenceTerms	dct:license	ids:transferContract	gax-t-f:license
ms:language	dct:language (with values from EU vocabulary)	dct:language (with ids:Language values)	n/a (dct:language will be used)

Table 2: Sample of the semantic mappings between the ELG, DCAT-AP, IDSA and Gaia-X models

sis on certain technical or conceptual aspects of their own approach towards data spaces. These initiatives are IDSA (International Data Spaces Association), Gaia-X, FIWARE and BDVA (Big Data Value Association). These four initiatives joined forces and established the Data Space Business Alliance (DSBA),[31] which already published a first “technical convergence discussion document” (DSBA, 2023), describing their “common vision on Data Spaces” including selected technical aspects. Most recently, in March 2024, the EU project DSSC has published Version 1.0 of the Data Spaces Blueprint (Guilloud et al., 2024), an emerging set of technical, operational and legal building blocks that has been developed together with the DSSC partners (including IDSA, Gaia-X, FIWARE and BDVA) and representatives of other data space initiatives. While the field is quite dynamic, there appears to be a general understanding that the DSSC blueprint will be the most relevant document for specifying the emerging European data space ecosystem, especially with regard to the officially mandated official EU data spaces that will operate in various sectors and domains, including the Common European Language Data Space (LDS).[32] The authors of this paper are actively engaged in the above mentioned initiatives and will ensure that the upcoming development of ELG, integrating the platform into various data space initiatives (especially in terms of

[31] https://data-spaces-business-alliance.eu

[32] https://language-data-space.ec.europa.eu

IDSA and Gaia-X) will be performed according to the emerging specifications, to which we actively contribute.

In yet another parallel development, the EU established the European Digital Infrastructure Consortium (EDIC) as a new legal entity type.[33] EDICs enable EU Member States to set up and implement multi-country projects with regard to the Digital Decade Policy Programme 2030. One of the first is the Alliance for Language Technologies (ALT-EDIC), which seeks to establish a Europe-wide technology centre to develop a common infrastructure for NLP and LLMs. While the EDIC will be governed and financed by the EU Member States, industry, academia, NGOs and other stakeholders will also play a role. Several co-authors of this paper are actively engaged in this development and will attempt to synchronise the emerging plans of the ALT-EDIC with the existing ELG platform.

7. Summary and Future Work

This paper provides an overview of recent developments regarding the European Language Grid cloud platform. We highlight the substantial growth of the ELG platform in terms of available resources and services, we describe the Digital Language Equality dashboard, which can be used to compare and visualise the technology support of Europe's languages and we emphasise our efforts with regard to integrating ELG into a number of European data space initiatives.

With regard to future work, we (still) plan to establish a legal entity for the ELG initiative to make it more sustainable (see Rehm et al., 2023). Furthermore, we will align the ELG platform as much as we can with current European initiatives, especially LDS but also other data spaces, and also with the ALT-EDIC.

8. Acknowledgements

The European Language Grid EU project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 825627. This work has also been partially funded by the project DataBri-X, which has received funding from the EU's Horizon Europe research and innovation programme under grant agreement no. 101070069, see https://databri-x.eu. Additionally, this work has been partially funded by the project SciLake, which has received funding from the EU's Horizon Europe research and innovation programme under grant agreement no. 101058573, see https://scilake.eu. The European Language Equality project has received funding from the European Union under the grant agreements no. LC-01641480 – 101018166 (ELE) and LC-01884166 – 101075356 (ELE2), see https://european-language-equality.eu. Finally, the work presented in this paper has received funding from the German Federal Ministry for Economic Affairs and Climate Action (BMWK) through the project OpenGPT-X (project no. 68GX21007D), see https://opengpt-x.de.

[33] https://digital-strategy.ec.europa.eu/en/policies/edic

9. Bibliographical References

Victoria Arranz, Khalid Choukri, Valérie Mapelli, Mickaël Rigault, Penny Labropoulou, Miltos Deligiannis, Leon Voukoutis, and Stelios Piperidis. 2023. Datasets, Corpora and other Language Resources. In Georg Rehm, editor, *European Language Grid: A Language Technology Platform for Multilingual Europe*, Cognitive Technologies, pages 151–169. Springer, Cham, Switzerland.

DSBA. 2023. Technical convergence discussion document. Data Spaces Business Alliance, April 2023.

Gaia-X. 2022. Trust framework - gaia-x trust framework - 22.04 release.

Federico Gaspari, Owen Gallagher, Georg Rehm, Maria Giagkou, Stelios Piperidis, Jane Dunne, and Andy Way. 2022. Introducing the Digital Language Equality Metric: Technological Factors. In *Proc. of the Workshop Towards Digital Language Equality (TDLE 2022; co-located with LREC 2022)*, pages 1–12, Marseille, France. 20 June 2022.

Federico Gaspari, Annika Grützner-Zahn, Georg Rehm, Owen Gallagher, Maria Giagkou, Stelios Piperidis, and Andy Way. 2023. Digital Language Equality: Definition, Metric, Dashboard. In Georg Rehm and Andy Way, editors, *European Language Equality: A Strategic Agenda for Digital Language Equality*, Cognitive Technologies, pages 39–73. Springer, Cham, Switzerland.

Annika Grützner-Zahn and Georg Rehm. 2022. Introducing the Digital Language Equality Metric: Contextual Factors. In *Proceedings of the Workshop Towards Digital Language Equality (TDLE 2022; co-located with LREC 2022)*, pages 13–26, Marseille, France. 20 June 2022.

Geraud Guilloud, Mirjam Huisintveld, Tonia Sapia, and Clara Pezuela, editors. 2024. *Data Spaces*

Blueprint 1.0, March 2024. Data Spaces Support Centre (DSSC).

IDSA. January 2021. IDSA and Gaia-X: Position Paper.

Penny Labropoulou, Katerina Gkirtzou, Maria Gavriilidou, Miltos Deligiannis, Dimitris Galanis, Stelios Piperidis, Georg Rehm, Maria Berger, Valérie Mapelli, Michael Rigault, Victoria Arranz, Khalid Choukri, Gerhard Backfried, José Manuel Gómez Pérez, and Andres Garcia-Silva. 2020. Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid. In *Proc. of the 12th Language Resources and Evaluation Conf. (LREC 2020)*, pages 3421–3430, Marseille, France. ELRA.

Penny Labropoulou, Stelios Piperidis, Miltos Deligiannis, Leon Voukoutis, Maria Giagkou, Ondřej Košarko, Jan Hajič, and Georg Rehm. 2023. Interoperable Metadata Bridges to the wider Language Technology Ecosystem. In Georg Rehm, editor, *European Language Grid: A Language Technology Platform for Multilingual Europe*, Cognitive Technologies, pages 107–127. Springer, Cham, Switzerland.

Lars Nagel and Douwe Lycklama. 2021. Design Principles for Data Spaces. Position Paper. Technical report, Berlin.

Stelios Piperidis, Penny Labropoulou, Dimitris Galanis, Miltos Deligiannis, and Georg Rehm. 2023. The European Language Grid Platform: Basic Concepts. In Georg Rehm, editor, *European Language Grid: A Language Technology Platform for Multilingual Europe*, Cognitive Technologies, pages 13–36. Springer, Cham, Switzerland.

Georg Rehm, editor. 2023. *European Language Grid: A Language Technology Platform for Multilingual Europe*. Cognitive Technologies. Springer, Cham, Switzerland.

Georg Rehm, Dimitrios Galanis, Penny Labropoulou, Stelios Piperidis, Martin Welß, Ricardo Usbeck, Joachim Köhler, Miltos Deligiannis, Katerina Gkirtzou, Johannes Fischer, Christian Chiarcos, Nils Feldhus, Julián Moreno-Schneider, Florian Kintzel, Elena Montiel, Víctor Rodríguez Doncel, John P. McCrae, David Laqua, Irina Patricia Theile, Christian Dittmar, Kalina Bontcheva, Ian Roberts, Andrejs Vasiļjevs, and Andis Lagzdiņš. 2020. Towards an Interoperable Ecosystem of AI and LT Platforms: A Roadmap for the Implementation of Different Levels of Interoperability. In *Proceedings of the 1st International Workshop on Language Technology Platforms (IWLTP 2020, co-located with LREC 2020)*, pages 96–107, Marseille, France. 16 May 2020.

Georg Rehm, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajič, Khalid Choukri, Andrejs Vasiļjevs, Gerhard Backfried, Katja Prinz, Jose Manuel Gómez-Pérez, and Ulrich Germann. 2023. Sustaining the European Language Grid: Towards the ELG Legal Entity. In Georg Rehm, editor, *European Language Grid: A Language Technology Platform for Multilingual Europe*, Cognitive Technologies, pages 233–254. Springer, Cham, Switzerland.

Georg Rehm, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Victoria Arranz, Andrejs Vasiļjevs, Gerhard Backfried, José Manuel Gómez Pérez, Ulrich Germann, Rémi Calizzano, Nils Feldhus, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Julian Moreno-Schneider, Dimitris Galanis, Penny Labropoulou, Miltos Deligiannis, Katerina Gkirtzou, Athanasia Kolovou, Dimitris Gkoumas, Leon Voukoutis, Ian Roberts, Jana Hamrlová, Dusan Varis, Lukáš Kačena, Khalid Choukri, Valérie Mapelli, Mickaël Rigault, Jūlija Meļņika, Miro Janosik, Katja Prinz, Andres Garcia-Silva, Cristian Berrio, Ondrej Klejch, and Steve Renals. 2021. European Language Grid: A Joint Platform for the European Language Technology Community. In *Proc. of the 16th Conf. of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, pages 221–230, Kyiv, Ukraine. Association for Computational Linguistics (ACL).

Georg Rehm, Stelios Piperidis, Khalid Choukri, Andrejs Vasiļjevs, Katrin Marheinecke, Victoria Arranz, Aivars Bērziņš, Miltos Deligiannis, Dimitris Galanis, Maria Gavriilidou, Maria Giagkou, Katerina Gkirtzou, Dimitris Gkoumas, Annika Grützner-Zahn, Athanasia Kolovou, Penny Labropoulou, Andis Lagzdiņš, Elena Leitner, Valérie Mapelli, Hélène Mazo, Simon Ostermann, Stefania Racioppa, Mickaël Rigault, and Leon Voukoutis. 2024. Common European Language Data Space. In *Proc. of the Joint Int. Conf. on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Turino, Italy. ELRA. Accepted for publication. May 20-25, 2024.

Georg Rehm and Andy Way, editors. 2023. *European Language Equality: A Strategic Agenda for Digital Language Equality*. Cognitive Technologies. Springer.