Automating FAIRness: A FAIRification Tool within the Language Resources Infrastructure

Daniele Melaccio, Monica Monachini
Istituto di Linguistica Computazionale "A. Zampolli" CNR-ILC
daniele.melaccio@cnr.it, monica.monachini@cnr.it

Abstract

In addition to technical interoperability, FAIRness encompasses governance, policy, and ethical aspects, reflecting how language data are produced, represented, and managed within research infrastructures. Ensuring FAIR compliance of language resources is essential for transparent and sustainable research in the social sciences and humanities, enabling data accessibility, quality, and long-term community reuse. The FAIRification Tool — created by CLARIN IT as part of the Humanities and Heritage Italian Open Science Cloud (H2IOSC) — is a modular system that automates and enhances FAIR compliance for language resources. The tool builds upon and extends existing FAIR data assessment frameworks by combining automatic and human validation, a feedback dashboard, certification thresholds, and domain-specific extensions aligned with linguistic metadata standards. It supports FAIR-by-design practices by operationalizing FAIR concepts and embedding them into repository workflows, thereby promoting interoperability across CLARIN, H2IOSC, and EOSC. The tool's effectiveness has been demonstrated through an initial evaluation conducted on a representative set of linguistic datasets, which revealed notable improvements (30–40%) in FAIR scores, particularly in the Findable and Reusable dimensions, contributing to responsible, policy-aware, and transparent language data management within the European Open Science landscape.

Keywords: FAIR principles, CLARIN, H2IOSC, language resources, interoperability, governance, open science, data stewardship

1. Fairness and Policy Context

Ensuring FAIR compliance of language resources is essential for sustainable research in the social sciences and humanities, as it promotes data accessibility, quality, and long-term community reuse. In recent years, research infrastructures have played an important role in supporting the sustainable management and interoperability of language data within open science ecosystems. This paper contributes to current discussions on sustainable research data management by examining how language resources can be integrated and reused within a FAIR-compliant infrastructure. In particular, it addresses the FAIRness of language data within the H2IOSC infrastructure, a national research initiative aiming to contribute to the European Open Science Cloud (EOSC) (Degl'Innocenti et al., 2023; European Commission, 2022). H2IOSC integrates data and services from the humanities and social sciences under a FAIR-by-design approach, promoting open, sustainable, and interoperable data practices. The EOSC initiative aims to create a federated environment for sharing and reusing research data across disciplines and national infrastructures, promoting interoperability, trust, and responsible data governance at the European level. Within this framework, the FAIR principles have become a central policy instrument guiding the design and operation of infrastructures, repositories, and data services.

Over the past decade, the FAIR principles (Wilkinson et al., 2016) have become a cornerstone of Open Science, fostering the discoverability, accessibility, interoperability, and reusability of research data. Within the community of language resources, these principles are essential to ensure that linguistic datasets, corpora, and tools can be reused effectively across domains, infrastructures, and disciplines. By promoting machine-actionable metadata, persistent identifiers, and interoperable standards, FAIR principles support the creation of sustainable research ecosystems in which data can circulate across infrastructures and communities.

However, achieving FAIRness extends beyond technical interoperability and metadata compliance. It also involves broader governance and ethical dimensions related to how language data are produced, represented, and managed within research infrastructures. Recent discussions on research data governance suggest that FAIRness is not only a technical feature of datasets but also a sociotechnical practice shaped by institutional workflows, community standards, and policies. Research infrastructures play a key role in implementing FAIR principles by turning general guidelines into practical tools, services, and evaluation mechanisms.

In this sense, infrastructures such as H2IOSC and CLARIN can be understood not only as technical systems but as policy-driven ecosystems, where design choices influence transparency, accountability, and the long-term sustainability of data stewardship. As language repositories continue to expand, the need for systematic and reproducible

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 1197–1204
11–16 May 2026.

FAIR assessment becomes increasingly evident. Manual and semi-structured evaluation procedures, still prevalent in many infrastructures, are time-consuming, difficult to reproduce, and ultimately unsustainable given the growing volume and diversity of deposited resources.

In the language domain, the complexity of metadata formats further complicates the evaluation of FAIR compliance. Language resources often include layered annotations, multimodal content, and domain-specific metadata schemas, which make their evaluation more complex than in many other scientific domains. While repositories hosted by CLARIN centres have implemented community standards and persistent identifier systems (de Jong et al., 2022), mechanisms for systematically assessing and improving FAIRness remain limited.

Existing FAIR evaluation practices in many repositories rely on manual checks or loosely structured guidelines, which makes systematic monitoring difficult and limits the reproducibility of FAIR assessments. As the volume of digital resources continues to grow, infrastructures increasingly require automated and transparent mechanisms capable of supporting both the evaluation and improvement of FAIR compliance.

This paper tries to address this gap by presenting the *FAIRification Tool*, an integrated environment designed to support the automated evaluation, monitoring, and improvement of FAIRness within linguistic repositories. The tool extends the F-UJI assessment framework with domain-specific metrics tailored to linguistic metadata standards and integrates the evaluation workflow directly into repository operations. By embedding FAIR-compliance mechanisms within the H2IOSC service architecture, the tool contributes to a model of *responsible FAIRness*, in which technological innovation and policy awareness are mutually reinforcing.

2. State of the Art and Related Works

The FAIR principles introduced by (Wilkinson et al., 2016) established a conceptual foundation for making data Findable, Accessible, Interoperable, and Reusable. Since their introduction, a growing body of work has focused on translating these high-level principles into operational frameworks capable of supporting practical data management and automated evaluation.

Several initiatives have attempted to formalize FAIR assessment through measurable indicators and evaluation models. Among the most influential contributions, the Research Data Alliance (RDA) proposed the *FAIR Data Maturity Model* (Bahim et al., 2020), which defines a set of community-driven indicators aimed at assessing the FAIRness of datasets across different levels of maturity. In parallel, the FAIRsFAIR project developed a comprehensive set of *FAIR Data Object Assessment Metrics* designed to support reproducible and machine-actionable FAIR evaluation across repositories and infrastructures (FAIRsFAIR Project, 2021).

Despite these advances, translating the FAIR principles into measurable and machine-actionable indicators remains a significant challenge. As noted by (Devaraju and Huber, 2021), many FAIR metrics remain qualitative or domain-agnostic, making it difficult to evaluate compliance automatically and consistently across disciplines. In particular, the lack of domain-specific evaluation criteria often results in ambiguous or incomplete assessments when applied to complex datasets.

In response to these challenges, several automated FAIR assessment tools have been proposed. Early initiatives such as *FAIRshake* (Clarke et al., 2019) and the *FAIR Evaluation Services* (Wilkinson et al., 2019) introduced community-driven scoring frameworks and semi-automated evaluation mechanisms aimed at promoting FAIR awareness and benchmarking datasets. However, these tools were primarily developed within the life sciences and biomedical domains, where data structures and metadata practices differ significantly from those found in linguistic infrastructures.

2.1. The F-UJI Framework and FAIRsFAIR Assessment Metrics

Among the most relevant initiatives addressing automated FAIR assessment is the *F-UJI* framework, developed within the FAIRsFAIR project as a reference implementation of the FAIR Data Object Assessment Metrics (Devaraju et al., 2020; FAIRsFAIR Project, 2021; Devaraju and Huber, 2021). F-UJI provides an automated service capable of evaluating the FAIRness of digital objects by executing a series of machine-actionable tests derived from the FAIRsFAIR metric framework.

The FAIRsFAIR metrics translate the abstract FAIR principles into a set of measurable indicators designed to assess whether digital objects satisfy the requirements for being Findable, Accessible, Interoperable, and Reusable. In total, the framework defines seventeen metrics distributed across the four FAIR dimensions. These metrics evaluate several aspects of data publication and management, including the presence of globally unique and persistent identifiers, the availability and machine-readability of metadata, the use of formal knowledge representation languages and controlled vocabularies, the specification of access protocols, and the clarity of licensing and provenance information.

The F-UJI service implements these metrics

through automated tests that analyze dataset landing pages, repository APIs, and embedded metadata formats such as JSON-LD or RDFa. For each evaluated resource, the system generates a structured report summarizing the compliance level across the FAIR dimensions and providing diagnostic feedback on missing or incomplete FAIR indicators. This approach enables scalable FAIR assessments that can be applied to large collections of datasets across heterogeneous repositories.

While F-UJI represents one of the most advanced frameworks for automated FAIR evaluation, its design remains intentionally domain-agnostic. As a consequence, the framework does not explicitly account for domain-specific metadata schemas and standards used in particular research infrastructures. In the context of linguistic data, repositories frequently rely on specialized metadata frameworks such as CMDI profiles and controlled vocabularies maintained within the CLARIN infrastructure. These elements are essential for evaluating interoperability and reusability but are only partially captured by generic FAIR metrics.

While F-UJI represents one of the most advanced frameworks for automated FAIR evaluation, its design remains intentionally domain-agnostic. As a consequence, the framework does not explicitly account for domain-specific metadata schemas and standards used in particular research infrastructures. In the context of linguistic data, repositories frequently rely on specialized metadata frameworks such as CMDI profiles and controlled vocabularies maintained within the CLARIN infrastructure. These elements are essential for evaluating interoperability and reusability but are only partially captured by generic FAIR metrics.

This limitation highlights the need for domain-adapted FAIR evaluation environments capable of integrating automated metrics with infrastructure-specific validation mechanisms. The FAIRification Tool developed within the H2IOSC project addresses this gap by extending the F-UJI assessment framework with domain-specific indicators tailored to linguistic metadata standards and repository practices.

The FAIRification Tool developed within the H2IOSC project addresses this gap by extending the F-UJI assessment framework with domain-specific indicators tailored to linguistic metadata standards and repository practices. In addition to supporting automated FAIR evaluation, the tool integrates the assessment process into repository operations, enabling iterative FAIRification cycles that combine automated evaluation with curator validation and structured feedback.

Through its modular architecture and its integration with existing infrastructures, the system bridges FAIR evaluation and repository governance, enabling continuous monitoring, quality improvement, and certification of linguistic resources within interconnected research infrastructures (Rehm et al., 2024).

3. The FAIRification Tool: Design and Implementation

The FAIRification Tool[1] is grounded in the idea that FAIR assessment should not be treated as a one-off certification step, but as a continuous and participatory process integrated into the daily workflows of linguistic repositories. Within this perspective, FAIRness becomes both a technical and an organizational practice that supports transparency, accountability, and long-term sustainability in data governance. The tool operationalizes these principles by translating the FAIR-by-design philosophy into a working environment where automation, human validation, and feedback coexist to promote community engagement and quality improvement.

At the conceptual level, the design of the FAIRification Tool builds on three guiding principles: *automation*, *transparency*, and *feedback*. Automation ensures scalable evaluation of datasets through machine-actionable tests; transparency guarantees that assessment criteria and results remain interpretable by repository managers and data providers; and feedback mechanisms allow users to iteratively improve metadata quality and FAIR compliance.

The FAIRification workflow is articulated in four phases: *pre-assessment*, *automated evaluation*, *human validation*, and *certification and feedback*. These phases can be executed independently or repeated as iterative curation cycles. In practice, the workflow begins when a dataset identifier or landing page URL is submitted to the system. The tool retrieves available metadata from the repository and performs an automated FAIR evaluation using the F-UJI assessment service. The resulting report highlights strengths and weaknesses across the four FAIR dimensions. Curators can subsequently validate the automated results, refine metadata descriptions, and trigger a new evaluation cycle, progressively improving FAIR compliance before final certification.

3.1. Integration with the F-UJI FAIR Assessment Framework

The automated evaluation component of the FAIRification Tool builds upon the *F-UJI* framework (De-

[1]The FAIRification tool described in this paper is currently under testing in a staging environment and will be made publicly available upon completion of the validation phase.

varaju et al., 2020; Devaraju and Huber, 2021), which implements the FAIRsFAIR Data Object Assessment Metrics (FAIRsFAIR Project, 2021). These metrics translate the FAIR principles into machine-actionable tests that can be automatically executed on digital objects and their associated metadata.

F-UJI evaluates datasets by retrieving metadata from landing pages, repository APIs, and structured metadata formats such as JSON-LD or RDFa. The system verifies the presence and quality of key FAIR indicators including persistent identifiers, machine-readable metadata, access protocols, semantic vocabularies, and licensing information. Each evaluated dataset receives a structured report summarizing compliance across the four FAIR dimensions.

The FAIRification Tool adopts the F-UJI framework as its evaluation backbone while extending it with additional validation procedures tailored to linguistic data repositories. In particular, the tool incorporates checks related to the presence of CMDI metadata elements, alignment with the CLARIN Concept Registry (CCR), and the correct use of community vocabularies relevant for language resources.

Table 1 summarizes the seventeen FAIRsFAIR metrics implemented by the F-UJI framework and adopted within the FAIRification workflow.

3.2. System Architecture

From a technical perspective, the FAIRification Tool has been implemented as a distributed system based on a microservice architecture. The architecture separates three main layers: a *presentation layer*, a *service layer*, and a *persistence layer*. This design ensures modularity, scalability, and interoperability with external repository systems.

The presentation layer provides the user interface through which repository curators and data providers interact with the system. It offers dashboards for submitting datasets, visualizing FAIRness scores, and accessing diagnostic reports.

The service layer orchestrates the FAIRification workflow. It manages dataset ingestion, metadata retrieval, communication with the F-UJI assessment engine through a dedicated API wrapper, and the transformation of evaluation results into structured FAIR indicators.

The persistence layer maintains a registry of evaluated datasets, metadata records, FAIRness scores, and historical evaluation results. This layer enables longitudinal monitoring of FAIR improvements and supports comparative analyses across datasets and repository collections.

Figure 1 illustrates the simplified architecture of the FAIRification Tool and its integration with repository systems and the F-UJI evaluation service.

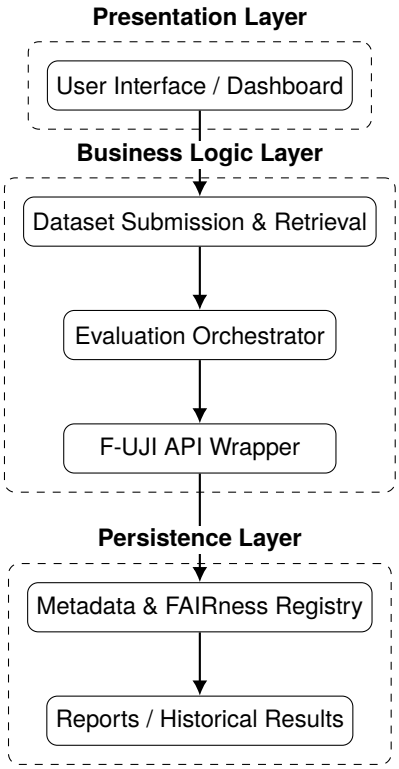

Figure 1: Layered architecture of the FAIRification Tool.

Security and compliance are ensured through encrypted communications, GDPR-aligned data management policies, and role-based access control mechanisms. Accessibility requirements follow the WCAG 2.1 guidelines, ensuring that the platform can be used by a wide range of users across the research community.

The development and initial deployment of the tool were accompanied by a preliminary evaluation conducted on twelve representative datasets from the ILC4CLARIN repository[2]. The results showed a mean improvement of approximately 38% in FAIRness scores, with the most significant progress observed in the *Findable* and *Reusable* dimensions. These outcomes demonstrate that the combination of automated FAIR evaluation and curator-driven validation can effectively enhance metadata quality and support FAIR-by-design practices in real repository contexts.

4. Evaluation and Case Studies

The first validation phase of the FAIRification Tool aimed to assess both its technical performance and its practical usefulness within real repository workflows. The evaluation was conducted on a representative sample of twelve linguistic resources

[2]ILC4CLARIN CLARIN-IT repository: https://dspace-clarin-it.ilc.cnr.it/home

Metric	FAIR Principle	Description
F1	Findable	Persistent identifier assigned to the data object
F2	Findable	Data described with rich metadata
F3	Findable	Metadata include the identifier of the data object
F4	Findable	Metadata registered or indexed in searchable resources
A1	Accessible	Data retrievable via standardized communication protocol
A1.1	Accessible	Protocol is open, free, and universally implementable
A1.2	Accessible	Protocol allows authentication and authorization if required
A2	Accessible	Metadata remain accessible even if the data are no longer available
I1	Interoperable	Data use a formal, accessible, shared language for knowledge representation
I2	Interoperable	Metadata use vocabularies that follow FAIR principles
I3	Interoperable	Metadata include qualified references to other data
R1	Reusable	Metadata include accurate and relevant attributes describing the data
R1.1	Reusable	Data are released with a clear and accessible usage license
R1.2	Reusable	Metadata include detailed provenance information
R1.3	Reusable	Data comply with domain-relevant community standards
R1.3.1	Reusable	Metadata follow community-endorsed metadata standards
R1.3.2	Reusable	Data formats conform to disciplinary best practices

Table 1: FAIRsFAIR Data Object Assessment Metrics implemented in the F-UJI evaluation framework.

from the ILC4CLARIN repository, which serves as the CLARIN-IT certified centre for the preservation and dissemination of language resources.

The selected sample was designed to reflect the heterogeneity of resources typically hosted within linguistic repositories. The evaluated materials included textual corpora, lexical resources, linguistic processing tools, language description datasets, and multimedia resources. This diversity allowed the evaluation to cover different metadata structures, documentation levels, and repository ingestion practices.

Table 2 presents a selection of representative resources included in the evaluation sample.

Category	Representative Resources
Text Corpora	MQDQ, Rosmini Corpus, HT-Archilochus
Lexical Resources	ItalWordNet, PAROLE-SIMPLE-CLIPS
Linguistic Tools	LexO-server, LMF ML Merger
Language Resources	AraMorph Data Plus
Audio Resources	Concerto di Caterina Bueno
External References	Historia Mongalorum, PMLAN dataset

Table 2: Categories of language resources used in the evaluation.

The evaluation protocol followed the FAIRification workflow implemented in the system and included two components: automated FAIR assessment and validation carried out by expert curators. Baseline FAIRness scores were first obtained through automated evaluation of the original metadata records using the F-UJI assessment service. Subsequently, datasets underwent the FAIRification process, which included metadata refinement, validation of persistent identifiers, enrichment of licensing information, and alignment with controlled vocabularies.

Each dataset therefore underwent the complete FAIRification workflow, which consisted of the following steps:

1. Retrieval of dataset metadata from repository landing pages and APIs.
2. Automated FAIR assessment using the F-UJI evaluation service.
3. Identification of missing or incomplete FAIR indicators.
4. Metadata refinement performed by repository curators through the FAIRification interface.
5. Re-evaluation of the updated metadata records.

This evaluation pipeline allowed the integration of machine-actionable FAIR indicators with curator interpretation and domain expertise.

The results confirmed the potential of automated FAIR evaluation to significantly improve FAIR compliance. Across the twelve evaluated resources, the average composite score increased from 0.54 before FAIRification to 0.75 after the FAIRification process, corresponding to an overall improvement of approximately 38%. The distribution of improvements across the four FAIR dimensions is summarized in Table 3.

The most significant improvements were observed in the *Findable* and *Reusable* dimensions. These improvements were mainly driven by the

	Baseline	After FAIR	Gain (%)
Findable	0.52	0.78	+50
Accessible	0.61	0.72	+18
Interoperable	0.49	0.69	+41
Reusable	0.54	0.76	+40

Table 3: Average FAIRness improvement across the evaluation sample.

introduction of persistent identifiers, the clarification of licensing information, and the enrichment of metadata descriptions. Improvements in interoperability were also observed due to better alignment of metadata elements with controlled vocabularies and with the CLARIN Concept Registry (CCR).

Qualitative feedback from repository curators highlighted the usability and pedagogical value of the tool. The visual dashboards and diagnostic reports allowed non-technical users to identify missing FAIR indicators and understand how specific metadata improvements influenced the final FAIRness score. The automated validation of identifiers and licenses reduced manual curation time by approximately 40%.

Performance testing indicated stable system behaviour under realistic workloads. On average, each dataset evaluation required approximately twelve seconds, while API response times remained below 250 milliseconds. System uptime during testing exceeded 99.6%, confirming the robustness of the architecture and its readiness for integration into production repository environments.

5. Discussion

The preliminary evaluation demonstrates that automating the assessment of FAIR principles can have a tangible impact on the quality, consistency, and sustainability of language resource management. The results presented in Section 4 show that even relatively limited metadata interventions—such as the introduction of persistent identifiers, the clarification of licensing information, and the alignment of metadata with controlled vocabularies—can substantially improve FAIRness scores across multiple dimensions.

Beyond the improvement of individual datasets, the FAIRification Tool introduces a new operational model for the management of linguistic resources within research infrastructures. By integrating automation, human validation, and structured feedback, the system transforms FAIR assessment from a reactive verification procedure into an iterative curation process embedded within repository workflows. This approach allows repositories to continuously monitor FAIR compliance, identify recurring metadata issues, and progressively improve the quality of their collections.

An important outcome of this experimentation is the recognition of FAIRification as a shared responsibility among different actors involved in data stewardship. The interaction between data depositors, repository curators, and infrastructure administrators—facilitated through the tool's dashboards and evaluation reports—creates a distributed model of quality assurance. This collaborative process strengthens trust in repository infrastructures and promotes a common understanding of FAIR principles within the linguistic research community.

From an infrastructural perspective, the FAIRification Tool contributes to the harmonization of FAIR assessment practices across CLARIN, H2IOSC, and the broader European Open Science Cloud ecosystem. Building upon standardized FAIR metrics implemented in the F-UJI framework while extending them with domain-specific validation mechanisms, the system demonstrates how generic FAIR evaluation models can be adapted to specialized research domains, such as language resources. In particular, the inclusion of checks related to CMDI metadata structures and references to the CLARIN Concept Registry ensures that linguistic datasets are evaluated according to the standards and practices adopted within the CLARIN infrastructure.

More broadly, the development of the FAIRification Tool highlights the strong interdependence between technological solutions and governance frameworks in the implementation of FAIR principles. Automated evaluation systems can significantly support the operationalization of FAIR guidelines, but they must remain embedded within institutional practices that involve human expertise and community-driven standards. In this sense, the system embodies an approach of *responsible automation*, where algorithmic evaluation complements rather than replaces curator judgement and institutional decision-making.

6. Conclusions and Future Work

The FAIRification Tool presented in this paper represents a concrete step toward the operationalization of the FAIR principles for language resources within research infrastructures. By combining automated evaluation with curator validation and policy-aware design, the system provides a practical environment for assessing and progressively improving the FAIR compliance of linguistic datasets.

The results of the preliminary evaluation demonstrate that the integration of automated FAIR assessment into repository workflows can significantly improve metadata quality and support the adoption of FAIR-by-design practices. Rather than treating FAIR compliance as a one-time certification pro-

cess, the tool enables an iterative FAIRification workflow in which evaluation, feedback, and improvement are embedded within the lifecycle of digital resources.

Beyond its technical functionality, the FAIRification Tool contributes to the broader objective of strengthening transparency and accountability in research data governance. By supporting systematic FAIR assessment across infrastructures such as CLARIN and H2IOSC, the system helps align linguistic data management practices with the wider European Open Science ecosystem.

Future developments will focus on consolidating the integration of the tool with repository platforms such as *DSpace 7*[3] and expanding its analytical capabilities through visual FAIR dashboards, longitudinal monitoring tools, and enhanced feedback mechanisms. Additional work will also explore the extension of the system to multimodal and multilingual datasets, where FAIR evaluation poses additional challenges due to the complexity of annotation layers and metadata structures.

At the time of the *LREC 2026 conference*, a live demonstration of the FAIRification Tool will showcase its architecture, usability, and interoperability within the broader European Open Science Cloud. The demonstration will also present updated evaluation results and insights from early adopters, contributing to a broader community dialogue on responsible and sustainable FAIR data governance.

In the longer term, the FAIRification Tool is envisioned as part of a federated ecosystem of services supporting the transparent, ethical, and reusable management of linguistic data across European research infrastructures, reinforcing the role of language resources in the evolving Open Science landscape.

7. Acknowledgements

This work was carried out within the *Humanities and Heritage NATIONAL Open Science Cloud (H2IOSC)*, funded by the European Union – NextGeneration EU (CUP B63C22000730005).

8. Bibliographical References

Charif Bahim, Robert Huber, Patricia Herterich, Andreas Petzold, Anusuriya Devaraju, and et al. 2020. Rda fair data maturity model: Specification and guidelines. Technical report, Research Data Alliance. RDA Recommendation.

Daniel J. B. Clarke, Lily Wang, Andrew Jones, Martin Wojciechowicz, Yifan Luo, et al. 2019. FAIRshake: Toolkit to evaluate the FAIRness of research digital resources. *Cell Systems*, 9(5):417–421.

Franciska de Jong, Dieter Van Uytvanck, Francesca Frontini, Antal van den Bosch, Darja Fišer, and Andreas Witt. 2022. *Language Matters*, pages 31–58. De Gruyter, Berlin, Boston. Accessed October 2025.

Emiliano Degl'Innocenti, Monica Monachini, Alberto Bucciero, Enrico Pasini, Bruno Fanini, and Francesca Frontini. 2023. H2iosc: Humanities and heritage open science cloud. In *La memoria digitale: forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD*, pages 63–64.

Anusuriya Devaraju and Robert Huber. 2021. An automated solution for measuring the progress toward fair research data. *Patterns*, 2(11):100370.

Anusuriya Devaraju, Robert Huber, Peter Herterich, Andreas Petzold, Philippe Rocca-Serra, and Peter Wittenburg. 2020. F-UJI: An automated FAIR data assessment tool. Technical report, FAIRsFAIR Project. FAIRsFAIR Project Report.

European Commission. 2022. EOSC interoperability framework. Technical report, European Commission. Version 1.0, February 2022.

FAIRsFAIR Project. 2021. FAIRsFAIR data object assessment metrics v0.8. Deliverable D5.4.

Georg Rehm, Stelios Piperidis, Khalid Choukri, Andrejs Vasiļjevs, Katrin Marheinecke, Victoria Arranz, Aivars Bērziņš, Miltos Deligiannis, Dimitris Galanis, 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 *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3579–3586, Torino, Italia. ELRA and ICCL.

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(160018).

[3] https://wiki.lyrasis.org/display/DSDOC7x/DSpace+7.x+Documentation

Mark D. Wilkinson, Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino da Silva Santos, and Michel Dumontier. 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data*, 6(1):174.