Enabling Robust Automatic FAIRness Evaluation of Knowledge Graphs

Maryam Basereh B.Sc, M.Sc

Supervised by Dr Paul Clarke and Dr Andrew McCarren

A thesis presented for the degree of Doctor of Philosophy

School of Computing

Dublin City University

December 2024

Declaration

I hereby certify that this material, which I now submit for assessment on the program of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Maryam Basereh

ID No: 19215921

Date: 06 June 2025

Dedication

To my beloved Mom and brother, and in loving memory of my Dad.

Acknowledgments

I would like to express my gratitude to the people who have played significant roles in my PhD journey, supporting me in reaching this important milestone and enabling me to complete this thesis despite the challenges encountered along the way.

I would like to express my sincere gratitude to Dr Claire Bohan, Dean of Students at DCU, whose exceptional kindness, understanding, and unwavering support during a critical period of my PhD journey were truly life-changing. Her genuine care for students, her remarkable sense of responsibility, and her dedication to fostering a supportive environment set an inspiring example of what it means to lead with compassion. Without her belief in me and her invaluable help, both in advocacy and resources, I would not have been able to overcome the challenges I faced to complete my work.

I wish to extend my profound gratitude to my supervisors, Dr Paul Clarke and Dr Andrew McCarren, whose exceptional support, expertise, and dedication were pivotal in enabling me to complete this thesis. Despite their demanding schedules, they stepped up to guide me through the most challenging stages of my PhD journey with unwavering commitment and care. Their combined efforts, thoughtful discussions, and collaborative approach played a crucial role in shaping the thesis. Their belief in my abilities and their collaborative spirit exemplify the highest standards of mentorship. Without their support, guidance, and encouragement, completing this thesis would not have been possible. Thanks Paul for meticulously proofreading my entire thesis and offering comprehensive feedback.

I also deeply appreciate my friends Dr Golnar K. Mahani, Dr Sarah Carter, Dr Beyza Yaman, Joan Desmond, Niamh Ring, Marzieh Saidimehr, Ramin Ranjbarzadeh, Elham Mianji, Mohammad Fardad, Vivek Manmohan Bulani, Vitor Gaboardi dos Santos whose support has been a constant source of motivation. A special appreciation goes to my colleagues and peers at DCU School of Computing. Engaging in stimulating academic exchanges with all of you while maintaining a personal connection has been truly enriching.

I would like to express my deepest and most heartfelt gratitude to my family whose belief in my abilities and encouragement played an integral role in my academic journey. To my mom, dad, and brother: Thank you for everything. I dedicate this PhD thesis to you. Without your tremendous understanding in the past few years, it would have been impossible for me to complete my study. Mom, I hope you are aware of the significant role you have played in shaping my path.

Contents

List of Figures

List of Tables

Enabling Robust Automatic FAIRness Evaluation of Knowledge Graphs

Maryam Basereh

Abstract

A knowledge graph is a form of knowledge representation that provides a mechanism for describing the interrelatedness of entities in a dataset.

Large knowledge graphs have become increasingly important in AI due to their ability to formalize and classify knowledge, enabling more effective extraction, retrieval, and analysis. They are used extensively in systems such as Google's Gemini and Bard, and IBM's Watson platform, to support smarter, context-aware applications in search, recommendation, and decision-making. As AI becomes part of daily life, the ethical implications of how these systems understand, recommend, and decide are significant—making the reliability of their underlying data critical.

A key consideration for any dataset is its adherence to the FAIR (Findable, Accessible, Interoperable, Reusable) principles, which aim to ensure the provenance, persistence, and reusability of data. In this context, FAIRness has become a crucial measure for establishing the suitability of knowledge graphs not only for data reliability but also for model reliability in machine learning.

This thesis evaluates the three currently available automated tools for assessing knowledge graph FAIRness—F-UJI, FAIR Evaluator, and FAIR Checker—to determine their capabilities and consistency. These tools, while gaining adoption in academic and industrial settings, have not previously been systematically compared.

This work applies statistical analysis to assess the consistency of their outputs and finds that, while each tool has strengths, none alone offers a complete view. It proposes a novel consistency measurement to support complementary use of all three tools.

The systematic evaluation of FAIRness assessment tools, along with the introduction of a new supporting metric, contributes to more trustworthy knowledge graph assessments. This, in turn, provides a foundation for practitioners and researchers working in dataset curation and machine learning model development, where ethical and technical robustness are increasingly essential.

Publications

- **Under preparation**

  1. Basereh M, McCarren A, Clarke P. ”Enabling Robust Automatic FAIRness Evaluation of Knowledge Graphs,” To be submitted.

- **Published**

  1. Basereh M, Ward M, Geary U, Healy U, Fogarty M, Caputo A, Brennan R. ”Automated Transparency Evaluation of Knowledge Graphs for Clinical Risk Governance,” The 31st Irish Conference on Artificial Intelligence and Cognitive Science (AICS 23), 2023 Dec. 7-8, Atlantic Technological University, Letterkenny, Donegal, Ireland.
  2. Basereh M, Caputo A, and Brennan R. ”Automatic transparency evaluation for open knowledge extraction systems.” Journal of Biomedical Semantics 14.01 (2023): 12. https://doi.org/10.1186/s13326-023-00293-9.
  3. Basereh M, Caputo A, and Brennan R. ”AccTEF: A transparency and accountability evaluation framework for ontology-based systems.” International Journal of Semantic Computing 16.01 (2022): 5-27.
  4. Basereh M, Caputo A, Brennan R. FAIR Ontologies for Transparent and Accountable AI: A Hospital Adverse Incidents Vocabulary Case Study. In the third International Conference on Transdisciplinary AI (TransAI) 2021 Sep 20 (pp. 92-97). IEEE (invited paper).

5. Basereh M, Brennan R, Corrigan S, Abgaz Y. A risk governance framework for healthcare decision support systems based on socio-technical analysis. Poster presented at: The first Conference on Critical Perspectives on Artificial Intelligence Ethics (CPAIE); 2020 Oct 22-23; Held online.

6. Crotti Junior A, Basereh M, Abgaz Y, Liang J, Duda N, McDonald N, Brennan R. The ARK platform: enabling risk management through semantic web technologies. In: Hastings J, Loebe F, editors. ICBO 20. Proceedings of the 11th International Conference on Biomedical Ontologies. 2020 Sept 17; Virtual conference hosted in Bolzano, Italy.

Chapter 1

Introduction

1.1 Chapter overview

This chapter introduces the thesis. First, the motivations for this research are explained, and the research questions are outlined in Section 1.2. The technical and evaluation approaches are detailed in Section 1.3. The contributions of this work are then summarized in Section 1.4, and finally, the structure of the remaining chapters is described in Section 1.5.

1.2 Motivation and Research Questions

Knowledge graphs are a form of data representation that formalize and classify knowledge and makes it easier to process for machines. Tim Berners-Lee defines knowledge graphs as a standard method of publishing structured data by using vocabularies that can be connected and interpreted by machines (Berners-Lee 2006a). DBpedia[1] is an example of a knowledge graph, which is the structured content extracted from the Wikipedia[2], a free online encyclopedia.

Due to their structured nature, knowledge graphs simplify knowledge extraction, retrieval, and analysis (Reinanda, Meij, Rijke, et al. 2020; Dørpinghaus and Stefan

[1] https://www.dbpedia.org, date accessed: 06/12/2024
[2] https://www.wikipedia.org, date accessed: 06/12/2024

2019). Consequently, their use in AI systems is increasingly prevalent (Wu and Weld 2010). Since data transparency is an important component of AI transparency, enhancing and maintaining the transparency of knowledge graphs is increasingly important. Transparency is an important component of AI governance methodologies, which is necessary for accountability (Reddy et al. 2020; Lepri et al. 2018; A. F. Winfield et al. 2019; Diakopoulos 2016; da-Cruz et al. 2016). The transparency problem has been identified as critical by the European Commission's Ethics Guidelines for Trustworthy AI (European Commission 2019).

FAIR (Findable, Accessible, Interoperable, Reusable) Guiding Principles (Mark D Wilkinson, Dumontier, Aalbersberg, et al. 2016), emerged in 2016, is an important component of transparency (Basereh, Caputo, and Brennan 2023). The principles were designed to improve the infrastructure supporting the reuse of scholarly data and to ensure that scientific data is managed and shared in a way that maximizes its value and usability for humans and machines (Mark D Wilkinson, Dumontier, Aalbersberg, et al. 2016). Later, the focus shifted to include all digital objects rather than just scientific data (Mark D Wilkinson, Sansone, Schultes, et al. 2018; Jagodnik et al. 2017; Mark D Wilkinson, Dumontier, Sansone, et al. 2019; Hodson et al. 2018). FAIR principles have gained global adoption, endorsed by governments and organizations such as the European Commission (European Commission 2024a), G7 (G7 2023), and G20 (Leaders 2016). This widespread support has led to the emergence of numerous FAIR-related initiatives, projects, and organizations, such as the European Open Science Cloud (EOSC) (European Commission 2024b), GO FAIR (Bernard Mons 2017), and FAIR4Health (Alvarez-Romero et al. 2022).

FAIR is a set of principles and not a standard, which has led to a broad interpretation and varied implementations and self-assessments (Mark D Wilkinson, Sansone, Schultes, et al. 2018). As a result, diverse measures and tools have been proposed for FAIRness assessment. This wide range of tools and measures has introduced inconsistencies and confusion, not only in the assessment and analysis of FAIRness but also in selecting the appropriate tool for assessing different digital objects. This resulted in the development of guidelines and recommendations to improve FAIR implementation. However, despite efforts to harmonize FAIRness assessment practices (Oliveira et al. 2021; Amdouni, Bouazzouni, and Jonquet 2022; Mark D Wilkinson, Sansone, Marjan, et al. 2022; Hodson et al. 2018; Whyte et al. 2021; European Commission Directorate-General for Research and Innovation 2016; Barend Mons, Neylon, et al. 2017; Mark D Wilkinson, Sansone, Schultes, et al. 2018; Devaraju, Huber, et al. 2020; Mark D Wilkinson, Dumontier, Sansone, et al. 2019), gaps persist in achieving consistency across tools.

Accordingly, this thesis addresses the robustness of FAIRness assessment in knowledge graphs, motivated by the lack of existing research in this area, their increasing use in AI systems, and the presence of inconsistencies in FAIRness evaluations. The focus is specifically on automatic FAIRness assessment, which is essential for handling the vast volume of available data, enabling scalability while saving time. There is however a fragmented automated FAIRness assessment landscape at present, and no previous research has sought to quantitatively measure the consistency of the various assessment forms. This gap leads to two research questions, as follows.

1. How consistent are the results produced by automated knowledge graph FAIRness assessment tools?

2. How can confidence in automated FAIRness assessment for knowledge graphs be increased through specific techniques?

It is believed that addressing these research questions will provide a clear quantitative view of the consistencies and inconsistencies in FAIRness assessment methods and creates a basis for more research and investigation in the area of robust and trustworthy FAIRness assessment.

To address the research questions, a number of discrete work items are undertaken, as follows.

1. Establish the State-of-the-Art review on

   (a) FAIRness assessment measures and aligned tooling.

   (b) Open access knowledge graph repositories to identify the most comprehensive source of open access knowledge graphs for FAIRness assessment.

   The scope of the State-of-the-Art review is deliberately focused on these key areas. These areas were selected because they directly support the research questions, which aim to evaluate the consistency of automated FAIRness assessment tools and explore techniques to increase confidence in their outcomes. A thorough understanding of existing FAIRness evaluation methods and tools is essential to analyze their performance and limitations. Similarly, focusing on open-source knowledge graphs ensures the study is grounded in accessible and widely used datasets that can be seen as representative of the broader knowledge graph ecosystem. Topics such as manual FAIRness assessment methodologies, proprietary knowledge graphs, and alternative data quality frameworks were not included, as they fall outside the scope of this thesis and do not directly contribute to answering the research questions.

2. Collect Data: Assess the FAIRness of knowledge graphs in the identified comprehensive repository using the identified suitable tools.

3. FAIRness Assessment Analysis: Analyze the collected data using statistical methods and machine learning techniques to provide a statistically grounded view of the consistencies and inconsistencies of automated FAIRness assessment in knowledge graphs.

1.3 Technical approach

While earlier studies have provided descriptive analyses of the area, this research aims to deliver a quantitative and reproducible comparative analysis of various automated knowledge graph FAIRness assessment approaches. This work establishes a clear foundation for the field, enabling the community to perform consistent and quantitative comparisons and assess developments in this domain.

Accordingly, statistical techniques, including correlation analysis, Mahalanobis outlier detection (Mahalanobis 1936), and data visualization methods, are employed in this research to illustrate the consistency issues in automated knowledge graph FAIRness assessment techniques.

In addition, to ensure a more robust and consistent measure of transparency, machine learning methods are employed to identify FAIR measures that play a statistically significant role in addressing inconsistencies between automated knowledge graph FAIRness evaluation methods. This approach facilitates the development of a robust compound FAIRness measure that incorporates the most significant contributors across the tools.

1.4 Contributions

The research contributions, which are elaborated in greater detail in Chapter 7, can be summarized as follows.

- **Primary Contribution**: This thesis proposes a novel measure that is an indicator for both FAIRness and the consistency/robustness of FAIRness assessment. This has positive implications for various models trained using knowledge graphs, supporting the development of trustworthy AI systems reliant on robust training data.

- **Supplementary Contributions:** A number of important additional contributions have also been identified, as follows:

  - A thorough and systematic comparative analysis of the three automatic knowledge graph FAIRness assessment tools (See Chapter 4).

  - A complete FAIRness assessment of the LOD Cloud, made available as an open access dataset for research and analysis (See Chapter 5).

– A statistically grounded view of inconsistencies between three open access automatic knowledge graph FAIRness assessment tools (See Chapter 6).

1.5 Thesis Overview

The remainder of this thesis is structured as follows. Chapter 2 explores the definition and significance of AI and data transparency, the transparency of knowledge graphs, components of transparency—including quality and FAIR principles—and FAIRness assessment metrics, tools, and comparative studies. Chapter 4 presents a comparative analysis of open-access knowledge graph FAIRness assessment tools and measures. Chapter 3 details the methodology and research design adopted for this study. Chapter 5 provides an in-depth explanation of the data collection process. Chapter 6 focuses on data analysis and proposes a combined measure for robust knowledge graph FAIRness assessment. Finally, Chapter 7 summarizes the research objectives achieved, highlights the contributions, and suggests directions for future research.

Chapter 2

Background and Related Work

2.1 Chapter Overview

This chapter provides a background and discussion on related work, emphasizing research gaps. It covers the definition and importance of AI transparency (Section 2.2), with a focus on data transparency (Sections 2.3). Knowledge graphs and their transparency are discussed in Sections 2.4 and 2.5. The components of transparency, including quality and FAIR principles, are detailed in Sections 2.6, 2.7, and 2.8. Sections 2.9 through 2.12 explore the FAIR assessment metrics, tools, and comparative studies. Finally, the literature gaps are discussed in Section 2.13.

2.2 AI Transparency

According to the Cambridge Dictionary (Cambridge University Press 2024), transparency is the quality of being open and free of secrets. This concept has been adapted across various fields. For instance, in (da-Cruz et al. 2016), transparency in politics has been defined as providing society with accessible, reliable, and high-quality information about government actions. Similarly, in AI, transparency has been defined in comparable ways. Table 2.1 lists AI transparency definitions published in standards or by authorities.

Table 2.1: AI transparency definitions

Resource title	**Transparency definition**
ISO/IEC DTS 12791 standard (International Organization for Standardization 2023) - under development	Transparency in AI systems is the provision of clear and understandable information to stakeholders about the system's design, goals, data usage, and operational logic to ensure proper and safe use.
Workshop on ISO/IEC standards for AI transparency and explainability (AI Standards Hub 2024b)	Transparency is "the availability of meaningful, faithful, comprehensive, accessible and understandable information about relevant aspects of an AI system for stakeholders. Relevant aspects may include the system's life cycle, functionality, operation and impact on AI subjects.".
ISO/IEC 23894:2023 standard (International Organization for Standardization and International Electrotechnical Commission 2023a)	Transparency is the provision of comprehensive and clear information on the performance, risks, and operational details of high-risk AI systems to ensure transparency and to facilitate proper and safe usage.
The EU AI Act (European Parliament and Council 2024) - The first legal framework addressing AI risks	Article 13 mandates that high-risk AI systems include documentation and information for users, specifying that instructions must be provided in a digital format that is concise, complete, correct, clear, and easily accessible and comprehensible.
The analysis of the preliminary AI standardization work plan in support of the AI Act (Soler Garrido et al. 2023)	Transparency involves providing clear information about a system's performance, risks, and the data used in its training, which directly impacts the quality of AI systems.
The Algorithmic Transparency Standard (ATRS) (AI Standards Hub 2024a) - A part of the UK government's National Data Strategy	Algorithmic transparency means being open about how algorithmic tools support decisions. This includes providing information on algorithmic tools and algorithm-assisted decisions in a complete, open, understandable, easily-accessible, and free format.

Continued on next page

Table 2.1: AI transparency definitions (Continued)

Resource title	Transparency definition
Information Commissioner's Office (ICO)'s draft guidance on the AI auditing framework (Information Commissioner's Office 2020)	Transparency is the clear documentation and communication of AI system capabilities and behaviors to ensure they are understandable by different stakeholders, including regulators, users, and the general public.

Among the AI transparency definitions presented in Table 2.1, European Parliament and Council 2024 and AI Standards Hub 2024a focus primarily on the quality of the information to be provided about AI systems, such as clarity, accessibility, and understandability, rather than on the specific aspects of the systems that the information should cover. In contrast, the remaining five definitions address both the quality of the information and the specific aspects of AI systems that should be disclosed, such as functionality and operation. Figure 2.1 illustrates the commonalities and differences in these aspects across the various definitions.

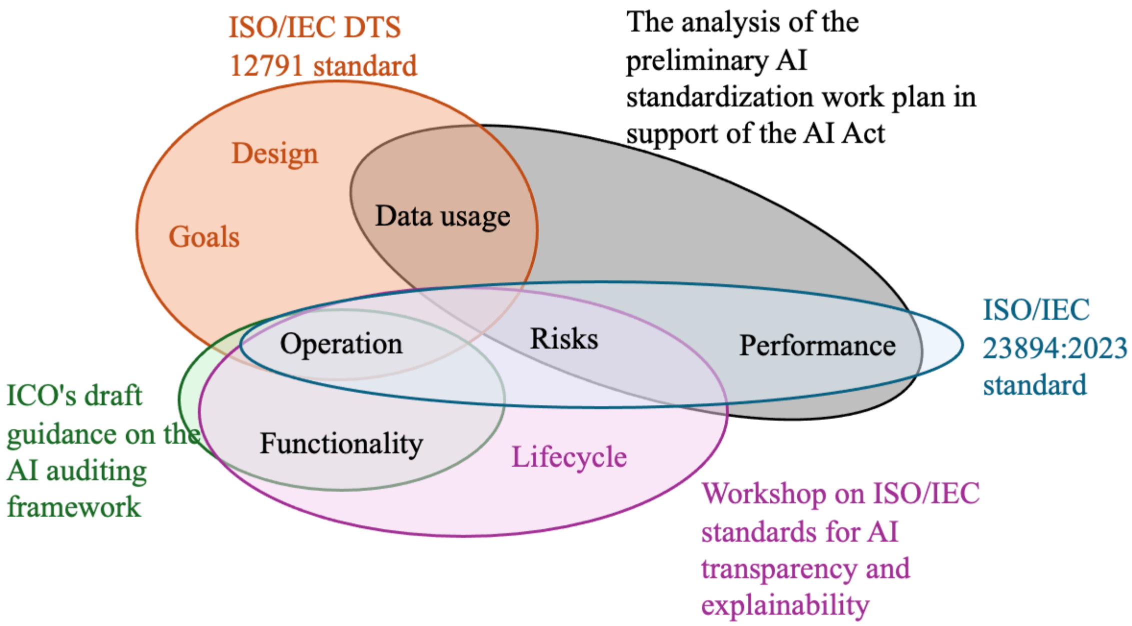


Figure 2.1: Venn diagram illustrating the commonalities and differences among AI transparency definitions listed in Table 2.1

While the various AI Transparency definitions identified in Table 2.1 are not identical, they do convey largely similar or congruent themes. In essence, the core considerations of AI Transparency revolves around the provision of clear, comprehensive, and accessible information about AI systems to ensure that stakeholders—including users, developers, regulators, and the general public-can understand

- The design, goals, and functionality of the AI system;
- The data used in training and operation, including data sources, processing methods, and usage;
- The operational logic, decision-making processes, and algorithmic support behind the AI's actions and outputs;
- The performance, potential risks, and impacts associated with the AI system throughout its life-cycle; and
- The documentation and guidelines provided for safe and proper use, including compliance with legal and ethical standards.

By enhancing transparency, AI systems become more understandable, accountable, and trustworthy, fostering confidence among stakeholders and facilitating the safe and effective use of AI technologies.

Repeated calls for improved transparency in AI systems have been made, with 'black-box' algorithms often regarded with concern due to their opacity and ethical implications.

For example, in the healthcare domain, Reddy et al. highlight issues of bias, liability, and lack of transparency in AI diagnostics, advocating for governance frameworks to ensure safety and trust (Reddy et al. 2020). Lepri et al. emphasize the dangers of algorithmic discrimination in access to public resources like employment and finance, advocating for transparent and accountable algorithmic design (Lepri et al. 2018). Winfield et al. propose ethical governance for AI and autonomous systems, especially in sensitive areas such as medicine and robotics (A. F. Winfield et al. 2019).

Gasser and Almeida stress layered governance models for regulating complex systems like AI in contexts including autonomous vehicles and surveillance (Gasser and

Almeida 2017). Wirtz et al. further identify transparency, fairness, and privacy as major concerns in AI applications in public administration, urging more integrative policy approaches (Wirtz, Weyerer, and Geyer 2019).

In algorithmic management, Lee et al. show how rideshare drivers felt monitored and unfairly judged by opaque systems that lacked human oversight (M. K. Lee et al. 2015). Lastly, Guidotti et al. survey various methods for interpreting black-box models, particularly in high-stakes contexts such as financial credit scoring and criminal sentencing, where decisions must be both accurate and explainable (Guidotti et al. 2018).

Transparency is a central issue in AI ethics (Wolf 2020). As a core component of AI governance, transparency is essential for ensuring accountability in systems that increasingly shape decisions in high-stakes domains.

Reddy et al. emphasize that in healthcare, opaque AI systems risk bias, liability issues, and eroding public trust unless robust governance models are implemented to enforce transparency and oversight mechanisms (Reddy et al. 2020).

Lepri et al. argue that without transparency, algorithmic decision-making can deepen social inequities, especially in public services such as criminal sentencing, credit scoring, and welfare allocation (Lepri et al. 2018). Winfield et al. highlight the need for ethical governance to ensure transparency in autonomous systems used in contexts like driverless cars and robotic medical assistants, where human lives may be at stake (A. F. Winfield et al. 2019).

Diakopoulos discusses transparency in algorithmic journalism, showing how automated news-writing systems can lead to misinformation if errors are not traceable and corrected, raising questions about democratic accountability (Diakopoulos 2016). Wirtz et al. note that in public administration, transparency is vital for public trust but remains underdeveloped due to weak regulatory alignment and unclear standards (Wirtz, Weyerer, and Geyer 2019).

Mozilla's report on AI transparency in practice highlights a gap between technical transparency tools and actual stakeholder understanding, emphasizing that transparency must be meaningful and tailored to diverse audiences including policymakers and end-users (Molavi-Vasse'i and McCrosky 2023). Zerilli et al. show that transparency plays a critical role in shaping human trust in AI systems, particularly in dynamic, team-based environments like healthcare or defense, where interpretability and confidence metrics may be more useful than simplistic explanations (Zerilli, Bhatt, and Weller 2022).

The ISO/IEC standards (e.g., ISO 23894) and regulatory frameworks such as those outlined by the EU's Scientific Foresight Unit call for both technical and governance-based approaches to transparency that scale with complexity and impact (International Organization for Standardization and International Electrotechnical Commission 2023a; European Parliamentary Research Service 2019).

The UK Information Commissioner's Office emphasizes a risk-based approach to AI transparency, focusing on compliance with data protection laws and the need for meaningful human oversight when AI systems are used in areas such as healthcare and recruitment (Information Commissioner's Office 2020).

Arnold et al. propose "FactSheets" for AI services—akin to product safety labels—to disclose performance, safety, and ethical concerns, thereby fostering trust in domains like speech recognition and social media monitoring (Arnold et al. 2019). Abdollahi and Nasraoui explore transparency in explainable recommender systems, underscoring how fairness and bias mitigation are tightly coupled with interpretability, especially in systems that influence personal preferences and societal behaviors (Abdollahi and Nasraoui 2018).

Shin's case studies on algorithm governance in Korea and China reveal the critical need for transparency in government-run algorithmic systems such as social credit scoring, where opacity can lead to public distrust and ethical violations (Shin 2019). The ISO/IEC 23894:2023 standard provides a comprehensive framework for managing AI risks through transparent design, emphasizing proactive risk assessment and traceability in industrial and regulatory settings (International Organization for Standardization and International Electrotechnical Commission 2023a).

Felzmann et al. introduce "Transparency by Design" as a principled approach for embedding transparency throughout the AI system lifecycle, from healthcare diagnostics to recommender systems, ensuring stakeholder trust and legal compliance from the outset (Felzmann et al. 2020). Lastly, the ISO's global AI ethics guidance stresses the importance of transparency not just as a technical requirement, but as a foundational ethical value for responsible AI (International Organization for Standardization 2024).

Enhancing transparency promotes fairness, scrutability, trust, reproducibility, effectiveness, and efficiency in AI systems by addressing critical risks across high-impact domains. Numerous studies emphasize that transparency must be tailored to stakeholders and integrated throughout the AI lifecycle.

For example, Haibe-Kains et al. demonstrate how insufficient transparency and reproducibility in AI models for breast cancer screening can undermine both scientific credibility and clinical trust, despite promising diagnostic results (Haibe-Kains et al. 2020). Similarly, Wynants et al. reveal that most early COVID-19 prediction models were poorly reported and at high risk of bias, making them unreliable for clinical use and underscoring how a lack of transparency can have direct health consequences (Wynants et al. 2020).

Reddy et al. highlight the need for transparent governance in healthcare AI to avoid compromising patient safety and institutional accountability (Reddy et al. 2020), while Morley et al. argue that transparency supports not only accountability but also ethical and legal compliance across healthcare settings (Morley et al. 2019).

In domains such as defense and clinical radiology, Zerilli et al. and Ho et al. stress that transparency mechanisms—whether through interpretable models or supplementary confidence estimates—are critical for fostering trust in human-AI collaboration and ensuring responsible integration into clinical workflows (Zerilli, Bhatt, and Weller 2022; Ho et al. 2019).

Koene et al. and Wachter et al. both explore transparency in governance frameworks, showing that a lack of interpretability in systems like autonomous vehicles and robotic assistants can pose legal and ethical risks (European Parliamentary Research Service 2019; Wachter, Mittelstadt, and Floridi 2017).

In public sector applications, Sun and Medaglia and Muralidhar et al. identify challenges in aligning transparency expectations across stakeholders and stress that understandable, interactive AI interfaces are crucial for user trust and acceptance (T. Q. Sun and Medaglia 2019; Muralidhar et al. 2023).

Various scholars and institutions propose concrete models to operationalize transparency. Arnold et al. advocate for "FactSheets" that document an AI system's purpose, performance, and ethical considerations—much like nutrition labels—and recommend their use in high-risk domains such as speech services, healthcare, and finance (Arnold et al. 2019).

The ISO/IEC 23894:2023 standard builds on this by recommending risk management principles that embed transparency from design to deployment (International Organization for Standardization and International Electrotechnical Commission 2023a), and ISO's broader AI ethics guidance frames transparency as a foundational principle for responsible AI (International Organization for Standardization 2024).

Felzmann et al. propose "Transparency by Design," a model inspired by Privacy by Design, to integrate transparency as a core design value across use cases ranging from recommender systems to medical diagnostics (Felzmann et al. 2020).

Liefgreen et al., focusing specifically on healthcare AI, emphasize that technical fixes alone are insufficient. Instead, sustained behavioral change—supported by motivational strategies rooted in psychology—is essential to embed fairness and transparency values into development practices (Liefgreen et al. 2023).

Finally, Wolf argues that in enterprise knowledge graph systems, transparency about data sources and update cycles is vital for trustworthy use in strategic decision-making and business operations (Wolf 2020).

Furthermore, transparency is key to mitigating risks and fostering innovation in AI technologies (W. Wang and Siau 2018; Wolf 2020), making it a necessary component for effective regulation and understanding of AI technologies (Gasser and Almeida 2017; Wolf 2020). Transparency also allows supporting dialogue and participation and furthering principles of democracy (Wolf 2020).

Transparency is the most frequently cited principle in the 84 AI policy documents reviewed by Jobin, Ienca, and Vayena 2019. ISO/IEC DTS 12791 (International Organization for Standardization 2023) and ISO/IEC 42001:2023 (International Organization for Standardization and International Electrotechnical Commission 2023a) both emphasize transparency as crucial for addressing bias in AI systems and ensuring trustworthy AI management through clear documentation and decision-making explanations. Transparency is also one of the eight principles in IEEE Ethically Aligned Design (A. Winfield et al. 2022), supported by the General Data Protection Regulation (GDPR)'s "right to explanation," which mandates accountability in automated decision-making (Goodman and Flaxman 2017). In February 2024, the AI white paper consultation response announced that ATRS will become mandatory for all central government departments, with plans to expand it to the broader public sector (AI Standards Hub 2024a).

2.3 Data Transparency

Data, encompassing both inputs and outputs of an AI system, is a key factor in AI transparency (Muralidhar et al. 2023; Bertino, S. Merrill, et al. 2019; Haibe-Kains et al. 2020). Ensuring transparency in data collection, storage, processing, and usage is vital for ethical AI use, avoiding biases, and protecting user privacy, especially in high-risk fields like healthcare (Soler Garrido et al. 2023; Panch, Mattie, and Celi 2019). Transparency also supports accountability and trustworthiness in data use (Bertino, S. Merrill, et al. 2019; Pushkarna, Zaldivar, and Kjartansson 2022). The UK government's National Data Strategy highlights this in its Algorithmic Transparency Standard (AI Standards Hub 2024a).

Data transparency enhances the quality of AI systems and supports the explainability of data-driven decisions (Soler Garrido et al. 2023; Bertino, S. Merrill, et al.

2019). It aids in data breach investigations, regulatory compliance auditing, and optimizing data processes. Clear and comprehensive dataset information is crucial for the robustness and generalizability of AI models, particularly in clinical settings (Daneshjou et al. 2021). Transparency ensures data trustworthiness, privacy, and quality, fostering trust and compliance in systems that use or manage the data (Diakopoulos 2016; Bertino, Kundu, and Sura 2019; International Organization for Standardization and International Electrotechnical Commission 2023b).

Similar to AI transparency, data transparency has been defined in various yet comparable ways. Table 2.2 presents definitions of data transparency as published in standards or by authoritative bodies.

Table 2.2: Data transparency definitions

Resource title	**Data transparency definition**
ISO 42001 Annex A Control A.7 – Data for AI Systems (International Organization for Standardization and International Electrotechnical Commission 2023b)	Data transparency is maintained by meticulously documenting • The selection process of data used in AI systems, including data source characteristics, data subject demographics, and any previous uses of the data, • Data provenance information, i.e., data origins and transformations, which helps providing clear explanations of AI decisions and outputs, • Data preparation methods, including the techniques used, such as data cleaning, normalization, labeling, and encoding, along with the rationale behind their selection. It also highlights the importance of ensuring conformity with privacy and security requirements from the outset.

Continued on next page

Table 2.2: Data transparency definitions (Continued)

Resource title	**Data transparency definition**
Recommendations on shaping technology according to GDPR provisions (European Union Agency For Network and Information Security (ENISA) 2018)	Transparency involves clarity about data collection purposes and ensuring that individuals understand how their data is being used.
GDPR European Parliament and Council of the European Union 2016	Transparency of data processing requires clear communication about how personal data is used and providing data subjects with access to their data.

Figure 2.2 illustrates the commonalities and differences in these aspects across the various definitions.

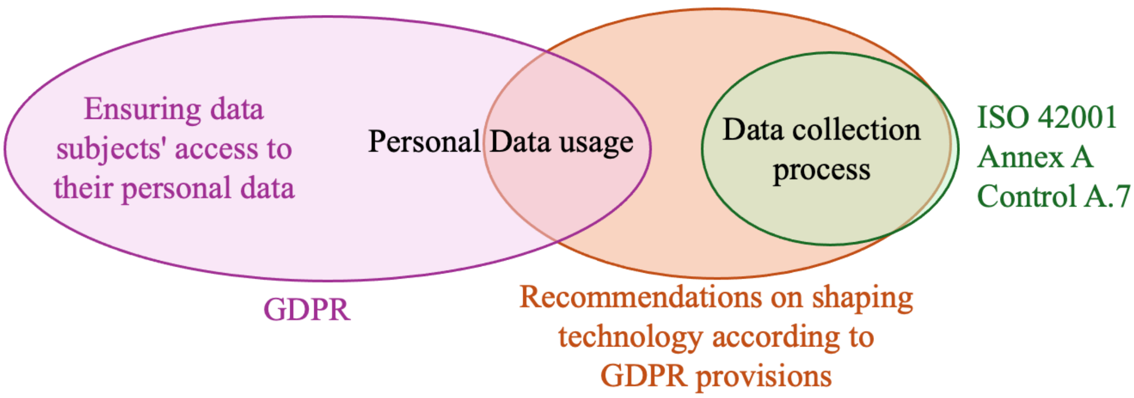


Figure 2.2: Venn diagram illustrating the commonalities and differences among AI transparency definitions listed in Table 2.2

While the various data Transparency definitions identified in Table 2.2 are not identical, they do convey largely similar or congruent themes. At its core, data transparency centers on the clear, thorough, and easily accessible documentation and communication of

- The selection and preparation of data used in AI systems, including the characteristics of data sources, demographics of data subjects, any previous usage of the data, data preparation methods, such as data cleaning, normalization, labeling, and encoding, along with justifications for these choices;

- Data provenance and processing, including data origins, transformations, and flow;

- Adherence to Privacy and Security Standards, such as the GDPR;

- The purposes of data collection, the specific ways in which personal data will be used, and ensuring that individuals are fully informed about and understand these processes. This includes granting data subjects access to their personal data and control over how it is used.

By integrating these elements, data transparency promotes trust, accountability, and compliance with legal and ethical standards in data management practices.

Bertino et al. provide a foundational multidimensional definition of data transparency, identifying the needs of various stakeholders—data participants, victims, users, and curators—and highlighting its growing importance in domains such as healthcare and sociology where personal data use is most sensitive (Bertino, S. Merrill, et al. 2019).

In computational journalism, Diakopoulos underscores how algorithmically generated news content—used at scale in financial reporting and political communication—can result in errors that undermine democratic accountability and public trust, illustrating the urgent need for transparency in automated content systems (Diakopoulos 2016). Muralidhar et al. stress the importance of transparency in AI system interfaces, arguing that clear explanations of algorithmic decisions—particularly in domains like recidivism risk scoring and job application filtering—are essential for user trust and perceived fairness (Muralidhar et al. 2023).

Similarly, Bertino et al. (in a separate paper) advocate for blockchain-based frameworks to achieve decentralized, auditable transparency in data-intensive systems, linking transparency directly to AI ethics, especially where personally identifiable information is used (Bertino, Kundu, and Sura 2019). Gebru et al. propose "Datasheets for Datasets" to standardize documentation about dataset provenance, composition, and intended use, thereby enabling more responsible data practices in critical sectors such as hiring, criminal justice, and infrastructure (Gebru et al. 2021).

Pushkarna et al. extend this idea through "Data Cards," tailored for real-world industrial deployment, capturing not just metadata but also contextual, ethical, and usability considerations across the AI lifecycle (Pushkarna, Zaldivar, and Kjartansson 2022). Coleti et al.'s TR-Model presents a metadata schema designed for improving personal data transparency, especially in consumer-facing digital platforms, supporting GDPR compliance and user empowerment (Coleti et al. 2020).

Daneshjou et al. expose serious transparency and bias issues in dermatology-related AI datasets, where poor demographic documentation and lack of data availability have limited the generalizability and trustworthiness of clinical models (Daneshjou et al. 2021). These studies demonstrate that although many promising transparency frameworks have been proposed, they are still inconsistently applied and rarely externally validated—underscoring the need for standardization and broader adoption.

Given the significant impact of data transparency on AI transparency, this thesis focuses on data transparency, specifically, Knowledge graph transparency (See Section 2.4 for full introduction).

2.4 Knowledge Graphs

Knowledge graphs, as defined by Tim Berners-Lee, are a standard method for publishing structured data using vocabularies that machines can interpret and connect (Berners-Lee 2006a). In essence, knowledge graphs represent information as a set of interconnected triples, where each triple consists of a subject (the entity being described), a predicate (the relationship or property), and an object (the related value or entity). For example, in the statement 'Earth is a planet', 'Earth' serves as the subject, 'is a' acts as the predicate, and 'planet' is the object. These triples form the basic units of knowledge graphs, enabling both humans and machines to understand the relationships they express (Heath and Bizer 2011). Figure 2.3 visually illustrates this example.

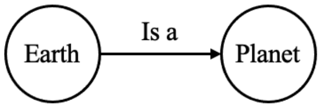


Figure 2.3: An example of a triple.

To effectively publish structured data in knowledge graphs, vocabularies, known as ontologies, are relied upon. Ontologies provide a standardized way to describe the entities, properties, and relationships within a domain. Tim Berners-Lee's concept of the Semantic Web ties directly into the use of ontologies to support the structure of knowledge graphs. Gruber defines an ontology as 'a formal, explicit specification of a shared conceptualization,' establishing a framework through which knowledge can be represented consistently across different contexts (Gruber 1993). Essentially, ontologies are the backbone of knowledge graphs, guiding how triples (subjects, predicates, and objects) are organized and interpreted. Popular ontologies include FOAF[1], Dublin Core[2], Schema.org[3], and PROV-O[4].

The Resource Description Framework (RDF) is used to represent these structured relationships on the web. RDF is the formal model that underpins knowledge graphs on the Semantic Web. It leverages triples (subject, predicate, object) and URIs[5] to uniquely identify and link resources across the web (World Wide Web Consortium 2014b). RDF enables seamless data interchange and integration, making it possible to connect and expand diverse datasets into a unified graph model. Figure 2.4 demonstrates the RDF representation of the above-mentioned example, i.e., 'Earth is a planet'.

In this example, http://dbpedia.org/ontology/, http://www.w3.org/1999/02/22-rdf-syntax-ns#, and http://www.w3.org/1999/02/22-rdf-syntax-ns# are the ontolo-

[1]Friend of a Friend: An ontology for describing relationships between people, their activities, and interests.
[2]A vocabulary for describing resources like books, web pages, and digital objects.
[3]A collection of schema for structured data on the internet, aiding search engines and applications in understanding web content.
[4]An ontology for representing provenance information, facilitating the interchange of provenance data on the web.
[5]Uniform Resource Identifier is a string of characters used to identify a resource on the Internet either by location, name, or both Berners-Lee, Fielding, and Masinter 2005.

```
@prefix dbo: <http://dbpedia.org/ontology/> .
@prefix dbr: <http://dbpedia.org/resource/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

dbr:Earth rdf:type dbo:Planet .
```

Figure 2.4: An example of an RDF code.

gies that have been used to create the knowledge graph. 'dbr:Earth' is the subject of the triple. 'rdf:type' is the predicate, and 'dbo:Planet' is the object.

Thus, knowledge graphs, ontologies, and RDF are tightly intertwined: ontologies define the structure and meaning of entities, RDF serves as the framework for encoding and linking the entities, and the knowledge graph itself is the product of these connections, enabling both human and machine interpretation of complex datasets.

Knowledge graphs are important because they formalize and classify knowledge and make it easier to process for machines. This simplifies knowledge extraction, retrieval, and analysis (Reinanda, Meij, Rijke, et al. 2020; Dørpinghaus and Stefan 2019). Also, the use of knowledge graphs is increasing in AI systems (Wu and Weld 2010) and the transparency of knowledge graphs has a prominent share in the transparency of the systems that create and use them.

2.5 Knowledge Graph Transparency

Knowledge graph is a type of structured data and the data transparency definition can be adapted to knowledge graphs. However, some researchers have specifically defined knowledge graph transparency in various yet comparable ways. Table 2.3 presents definitions of knowledge graph transparency.

Table 2.3: Knowledge graph transparency definitions

Research	Knowledge graph transparency definition
Zaveri et al. 2016	Transparency is the provision of clear and accessible information about the quality and reliability of knowledge graphs.
Wolf 2020	Transparency in knowledge graphs encompasses efforts to enhance explainability and interpretability by providing clear and understandable information about the sources, construction, refinement, and maintenance of the knowledge contained within the graphs, that isn't usually visible in end-use applications.
Andersen, Cazalens, and Lamarre 2021	knowledge graph transparency involves making the information in the knowledge graph accessible, understandable, and interpretable, including details about the data origin (provenance), how it is processed, and how it is integrated within the knowledge graph, and ensuring that users can understand how data are connected and how conclusions are drawn from these connections.

While the various data Transparency definitions identified in Table 2.3 are not identical, they do convey largely similar or congruent themes. At its heart, data transparency focuses on offering clear, accessible, and understandable information about the quality, sources, construction, processing, and integration of data within the knowledge graph, ensuring that users can interpret the data's origin, connections, and the conclusions drawn from these connections.

Recent literature highlights the critical need for transparency in knowledge graphs, particularly in enterprise applications, to ensure their effective and trustworthy use (Wolf 2020). Despite its importance, research on knowledge graph transparency remains limited. Enhancing transparency requires automated evaluation methods, as outlined by the IEEE Standard for Transparency of Autonomous Systems (A. Winfield et al. 2022). Automated transparency evaluation improves scalability, saves time, and helps stakeholders validate data sources and human contributors, thereby increasing confidence in ML systems (Barclay et al. 2019). According to Datta, Sen, and Zick 2016, transparency quantification serves several key purposes: accountability by identifying harms and ensuring fair decisions; error detection by correcting input data errors; guidance for improvement by offering insights into decision-making; and oversight by detecting algorithmic biases or errors through testing. Additionally, measuring data transparency directly impacts the accuracy of ML-based systems. It ensures that ML results are derived from reliable and comprehensive information, reducing the risk of significant deviations between predicted and actual outcomes, thereby supporting better decision-making (Gatti et al. 2024).

2.6 The Components of Transparency in Knowledge Graphs

There are currently no formal automated tools available for evaluating the transparency of knowledge graphs. However, transparency is a multidimensional concept. In this section, the main components of transparency in knowledge graphs are introduced to find out if there are state-of-the-art automatic tools that are capable of evaluating some transparency dimensions in knowledge graphs. Wolf 2020 introduces three fundamental transparency requirements that are essential for knowledge graphs in enterprise settings. These include

1. Source of knowledge: Emphasizes the need for transparency regarding the origins of information in the knowledge graph, allowing users to assess data reliability and relevance.

2. Currency of knowledge: Highlights the importance of knowing whether the data is up-to-date, as this impacts the validity of decisions based on the knowledge graph.

3. Evidence supporting associations: Stresses the necessity of providing details on how relationships within the knowledge graph are derived, including the datasets or algorithms used to extract the relationships, to help users understand the confidence level of associations.

Building on these requirements, Andersen Andersen, Cazalens, and Lamarre 2021 propose a formalized approach to defining transparency in knowledge graphs, outlining key dimensions such as:

1. Data provenance[6], which involves understanding the origin and method by which data in a knowledge graph is obtained;

2. Data quality, which focuses on ensuring that the data is accurate and reliable; and

3. Data accessibility, which ensures that users can easily access and comprehend the data within the knowledge graph.

In exploring transparency components for knowledge graphs, the relevance of FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Mark D Wilkinson, Dumontier, Aalbersberg, et al. 2016) becomes evident. Originally proposed to improve the usability of scholarly digital resources for both humans and machines (Miranda Azevedo and Dumontier 2020; Poveda-Villalón et al. 2020), FAIR principles are now recognized for enhancing transparency by ensuring well-documented and accessible data and metadata management practices. Adherence to these principles boosts transparency, making data easier to find, understand, and reuse, thereby contributing to trustworthy and reliable datasets (Bahim, Casorrán-Amilburu, et al. 2020). Additionally, the intersection of transparency and quality is emphasized in multiple studies, with the implementation of FAIR principles shown to enhance data quality. Bishop and Hank 2018 argue that transparency provided by FAIR principles improves data usability, supporting diverse scientific applications and ensuring reliable data management. Similarly, Iturbide et al. 2022 demonstrate that applying FAIR principles in the IPCC's Atlas repository enhanced transparency and improved climate data quality, making it more dependable for policymakers and researchers.

Jacobsen et al. 2020 emphasize that structuring data and metadata according to FAIR principles enhances transparency and data quality by improving accessibil-

[6]Data provenance has been identified as one of data quality dimensions by Zaveri et al. 2016.

ity and openness. Petrosyan et al. 2023 and Lia et al. 2023 further connect these principles to transparency and quality, particularly in agricultural datasets and data reusability. They argue that structured assessments and adherence to FAIR principles boost accessibility, reusability, and overall data quality, which are essential for advanced analytics and decision-making.

In specific applications like machine learning pipelines and bio-simulation models, integrating FAIR principles enhances transparency and data quality. Samuel, Löffler, and König-Ries 2020 and Welsh et al. 2021 show that ensuring data provenance and applying semantic annotations—key aspects of FAIR—lead to more transparent, reproducible, and higher-quality models. This aligns with the broader consensus that transparency through FAIR principles directly improves data quality across various domains.

The literature highlights the role of FAIR compliance tools and frameworks in evaluating the transparency of knowledge graphs and other data-intensive resources. Amdouni and Jonquet 2021 suggest that FAIRness assessments for semantic resources and ontologies enhance transparency and improve quality, making these resources more useful and reliable. Barend Mons, Schultes, et al. 2020 support this, arguing that structuring, standardization, and comprehensive metadata inclusion—central to FAIR principles—enhance transparency and data management quality across scientific communities.

Based on the literature, it can be inferred that knowledge graph quality and FAIR principles may serve as indicators of knowledge graph transparency. The following sections will review these two indicators in more detail.

2.7 Quality Evaluation in Knowledge Graphs

Quality and transparency are closely linked concepts (Diakopoulos 2016; Wirtz, Weyerer, and Geyer 2019; Sofi-Mahmudi and Raittio 2022), with overlapping dimensions. This overlap suggests that tools used for quality evaluation may also be useful for transparency evaluation. Wang and Strong (R. Y. Wang and Strong

1996) proposed one of the first frameworks for hierarchical data quality assessment in 1996, identifying four key dimensions: intrinsic, contextual, representational, and accessibility.

In 2011, Duque-Ramos et al. 2011 proposed a framework for ontology quality evaluation, based on the SQuaRE standard for software quality, to help users make informed decisions on which ontology to use. Since ontologies are foundational to knowledge graphs, their quality directly impacts the quality of the knowledge graphs. The framework introduces a series of quality dimensions, characteristics, and sub-characteristics, each accompanied by specific quality metrics for evaluation.

In 2012, Zaveri et al. 2016 proposed a comprehensive knowledge graph quality evaluation framework with six quality categories and 23 dimensions, each with associated metrics identified in the literature. The Data Quality Vocabulary[7] defines a category as a group of quality dimensions using a common type of information as a quality indicator, and a dimension as criteria relevant for assessing quality. This framework, widely accepted and cited, is used in state-of-the-art quality evaluation tools and methods (Mihindukulasooriya 2020; Debattista, Auer, and Lange 2016).

Mihindukulasooriya 2020 proposed a mechanism for knowledge graph quality assessment using a profiling technique tailored for quality evaluation, which feeds an ML-based RDF shape induction mechanism to extract data shape constraints for quality assessments. However, this framework lacks open access tools. Kontokostas et al. 2014 introduced a methodology inspired by test-driven software development, using SPARQL[8] (World Wide Web Consortium 2013) query templates to create domain-specific quality test queries, allowing for the discovery of data quality issues beyond conventional methods. The open-access tool RDFUnit[9] was developed based on this method. However, while RDFUnit offers flexibility for context-based quality tests, it does not provide standardized quality evaluation tests aligned with widely accepted frameworks like (Zaveri et al. 2016).

[7] https://www.w3.org/TR/vocab-dqv/
[8] SPARQL is a query language used to retrieve and manipulate data stored in RDF format.
[9] https://github.com/AKSW/RDFUnit, http://rdfunit.aksw.org/

Szarkowska et al. 2021 used knowledge graph-BERT for assessing the quality of hierarchical structures in knowledge graphs through binary classification, achieving high performance across four different graphs. Jia et al. 2019 proposed a neural network-based model for measuring the trustworthiness of knowledge graph triples by quantifying semantic correctness and the truthfulness of expressed facts. This method synthesizes semantic information and global inference to evaluate trustworthiness at the entity, relationship, and global levels. However, these approaches are complex and their evaluated dimensions are limited, lacking flexibility to add new quality dimensions.

Debattista, Auer, and Lange 2016 propose Luzzu, a conceptual methodology for assessing knowledge graphs and a framework for knowledge graph quality assessment. Luzzu allows defining new quality metrics, creating RDF quality metadata and quality problem reports, provides scalable dataset processors for data dumps, SPARQL endpoints, and big data infrastructures, and a customisable ranking algorithm for user-defined weights. Luzzu scales linearly against the number of triples in a dataset. It is open source and includes 37 pre-implemented quality evaluation metrics from intrinsic, accessibility, contextual, and representational dimensions, based on Zaveri et al. 2016 work. It has also been used to evaluate the quality of Linked Open Data Cloud (Debattista, Lange, et al. 2018), which is one of the largest collections of knowledge graphs on the web (Assaf, Troncy, and Senart 2015; Debattista, Attard, et al. 2019).

2.8 FAIR Principles

In January 2014, the FORCE11 Community[10] held a workshop in Leiden, Netherlands, which led to the concept that a minimal set of community-agreed principles could enable both machines and humans to more easily discover, access, interoperate, and reuse scientific data (FORCE11 2024). This idea prompted the creation

[10]FORCE11 (Future Of Research Communications and E-Scholarship) is a community of scholars, librarians, archivists, publishers, and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing (FORCE11 2011).

of the FAIR (Findable, Accessible, Interoperable, Reusable) Guiding Principles for Scientific Data Management and Stewardship (Mark D Wilkinson, Dumontier, Aalbersberg, et al. 2016). FAIR principles emphasize the need for data to be FAIR. Table 2.4 outlines FAIR principles.

Table 2.4: FAIR principles (Mark D Wilkinson, Dumontier, Aalbersberg, et al. 2016)

FAIR	**Principles**
Findable	F1. (meta)data are assigned a globally unique and eternally persistent identifier. F2. data are described with rich metadata. F3. (meta)data are registered or indexed in a searchable resource. F4. metadata specify the data identifier.
Accessible	A1. (meta)data are retrievable by their identifier using a standardized communications protocol. A1.1 the protocol is open, free, and universally implementable. A1.2 the protocol allows for an authentication and authorization procedure, where necessary. A2. metadata are accessible, even when the data are no longer available.
Interoperable	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (meta)data use vocabularies that follow FAIR principles. I3. (meta)data include qualified references to other (meta)data.
Reusable	R1. meta(data) have a plurality of accurate and relevant attributes. R1.1. (meta)data are released with a clear and accessible data usage license. R1.2. (meta)data are associated with their provenance. R1.3. (meta)data meet domain-relevant community standards.

FAIR principles were originally designed to improve the infrastructure supporting the reuse of scholarly data and to ensure that scientific data is managed and shared in a way that maximizes its value and usability for humans and machines (Mark D

Wilkinson, Dumontier, Aalbersberg, et al. 2016). However, the focus then shifted to include all digital objects rather than just scientific data (Mark D Wilkinson, Sansone, Schultes, et al. 2018; Hodson et al. 2018). A digital object is a structured set of data stored in a digital format, typically in memory or storage devices, that holds informational value, meaning it represents an entity such as a document, dataset, image, or software (Bonino-da-Silva-Santos 2022).

FAIR principles have gained global adoption, endorsed by governments and organizations like the European Commission (European Commission 2024a), G7 (G7 2023), and G20 (Leaders 2016). This widespread support has led to more data resources striving to be FAIR (Barend Mons, Neylon, et al. 2017) and the emergence of numerous FAIR-related initiatives, projects, and organizations, such as

- The European Open Science Cloud (EOSC), which aims to build a metadata-rich infrastructure to enhance data reuse (European Commission 2024b);
- GO FAIR, which is an international initiative that supports the EOSC and the global Internet of FAIR Data and Services (IFDS) (Bernard Mons 2017);
- The Horizon 2020 program, which offers guidelines on implementing FAIR data principles to help beneficiaries manage their research data, maximizing reuse and advancing scientific progress (European Commission Directorate-General for Research and Innovation 2016);
- FAIRsFAIR, a Horizon 2020 project, which was established to share knowledge, expertise, guidelines, and educational resources on FAIR practices across the European Union (FAIRsFAIR 2019-2022); and
- The Dutch Techcentre for Life Sciences (DTL) [11] (Eijssen et al. 2015), which is a partnership of over 50 life science organizations in the Netherlands, focuses on advancing FAIR data stewardship. The DTL FAIR Data team has developed tools like FAIRifier, Metadata Editor, FAIR Data Point, FAIR Search Engine,

[11] `https://www.dtls.nl/`

and ORKA, which together form the 'Data FAIRport' system, supporting data interoperability.

The European Commission funds numerous FAIR-related projects, including EOSC Future (Athens Research Center for Innovation in Information and Technologies 2021), FAIR4Health (Alvarez-Romero et al. 2022), FAIRplus (FAIRplus 2019-2022), and others. In addition, initiatives like BioSharing (BioSharing 2024) and FAIRdom (FAIRdom 2024) focus on supporting machine-friendly, high-quality, reproducible science. Moreover, funders like ZonMw (Dutch Organisation for Health Research and Development 2024), SNSF (Swiss National Science Foundation 2024), Science Europe (Science Europe 2024), and HRB (Health Research Board 2024) support FAIR research. There are also national initiatives such as EOSC-Nordic (EOSC-Nordic 2024), EOSC-Pillar (EOSC-Pillar 2024), EOSC-synergy (EOSC-synergy 2024), and ExPaNDS (ExPaNDS 2024), which work to expand the research and adoption of FAIR principles.

Researchers have applied FAIR principles across various domains. Oliveira et al. 2021 developed a structured workflow to enhance data management and reuse, applying it to the VODAN BR (Virus Outbreak Data Network Brazil) [12] pilot for COVID-19 research. Queralt-Rosinach et al. 2022 focused on making COVID-19 patient data within hospitals FAIR, using ontological models and semantic web technologies to create machine-actionable digital objects for medical research. Jagodnik et al. 2017 created a framework within the BD2K (Margolis et al. 2014) Commons (Bourne et al. 2015) to improve the sharing and reuse of biomedical research data, emphasizing the need for deep metadata, community participation, and collaborative efforts to develop FAIRness metrics and tools.

The emergence of FAIRness enhancement projects has led to guidelines and recommendations aimed at improving the implementation of the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Initiatives such as the 2018 FAIR Data Action Plan (Hodson et al. 2018) outline 34 key recommendations, including

[12] `https://portal.fiocruz.br/en/vodan-brazil`

validating FAIR data (Rec. 9), certifying supporting services (Rec. 11), and aligning FAIR data policies to reduce inconsistencies (Rec. 15). Similarly, the FAIRsFAIR project's D3.2 report (Whyte et al. 2021) underscores the need for robust metrics, a self-assessment framework for research infrastructures, and stronger stakeholder engagement to harmonize policies and enhance interoperability.

Despite this progress, FAIR remains a set of guiding principles rather than a formal standard. This openness allows for flexibility but also introduces significant challenges. The lack of standardization leads to varying interpretations and implementations across domains, complicating efforts to assess FAIRness consistently. Moreover, conceptual overlaps among the FAIR principles further hinder clarity in assessment and execution. For example, the Reusability sub-principles, which emphasize rich metadata and provenance, often overlap with aspects of Findability, such as persistent identifiers and metadata indexing. This blurring of conceptual boundaries makes it difficult to distinctly evaluate compliance with each individual principle.

As expert groups continue to refine FAIRness metrics, there is a growing recognition that both the interpretative flexibility and the internal overlaps within the FAIR framework pose barriers to consistent and objective assessment.

2.9 FAIRness Assessment Metrics

The broad interpretation of FAIR principles has resulted in varied implementations and self-assessments (Mark D Wilkinson, Sansone, Schultes, et al. 2018). To address this, the Research Data Alliance(RDA) FAIR Data Maturity Model Working Group, including co-authors of the original FAIR principles, was established to develop universal FAIRness assessment metrics (Hodson et al. 2018). They created a framework with semi-quantitative metrics focused on machine-actionability, introducing 14 Generation 1 (Gen1) metrics, which were used in a questionnaire for resource owners and users (Mark D Wilkinson, Sansone, Schultes, et al. 2018). Feedback led to the development of Generation 2 (Gen2) Maturity Indicators (MIs) and compliance tests, offering a more objective, automated evaluation of FAIRness with a focus on machine-readability.

Table 2.5: Gen2 FAIRness assessment metrics (Mark D Wilkinson, Sansone, Schultes, et al. 2018)

#	Metric	Description
1	F1A-Identifier Uniqueness	Ensures that each digital object has a unique identifier.
2	F1B-Identifier Persistence	Checks if the identifiers for digital resources remain consistent over time.
3	F2A-Structured Metadata	Assesses the presence of structured metadata.
4	F2B-Grounded Metadata	Evaluates whether metadata is based on established standards.
5	F3-Use of Globally Unique Identifiers (GUIDs) in Metadata	Ensures that metadata includes GUIDs.
6	F4-Metadata Indexing	Verifies if the metadata is indexed in a searchable resource.
7	A1.1-Open Protocol for Data Retrieval	Checks if an open protocol is used for retrieving data.
8	A1.2-Support for Authentication and Authorization	Assesses the support for authentication and authorization protocols.
9	A2-Metadata Persistence	Ensures that metadata remains accessible over time.
10	I1A-Use of Knowledge Representation Language (Weak)	Evaluates the use of a knowledge representation language in a less stringent manner.
11	I1B-Use of Knowledge Representation Language (Strict)	Evaluates the use of a knowledge representation language in a more stringent manner.
12	I2A-Use of FAIR Vocabularies (Weak)	Checks for the use of FAIR vocabularies in a less stringent manner.
13	I2B-Use of FAIR Vocabularies (Strict)	Checks for the use of FAIR vocabularies in a more stringent manner.

Continued on next page

Table 2.5: Gen2 FAIRness assessment metrics (Mark D Wilkinson, Sansone, Schultes, et al. 2018) (Continued)

#	Metric	Description
14	I3A-Qualified Outward Links	Assesses the presence of qualified outward links in the metadata.
15	R1.1A-Metadata Includes License (Weak)	Ensures that metadata includes a license statement.
16	R1.1B-Metadata Includes License (Strict)	Ensures that metadata includes a detailed license statement.

In 2020, Devaraju, Huber, et al. 2020 introduced the FAIRsFAIR Data Object Assessment Metrics_v0.5 to address key challenges in FAIRness assessments, such as subjectivity in manual evaluations, difficulty in meeting domain-specific needs, resource-intensive processes, the need for continuous updates, and technical barriers for some repositories. These metrics aim to create a common framework for assessing the FAIRness of data objects, which is essential for improving data management and enhancing data sharing and reuse across various domains. The proposed metrics include detailed descriptions of rationale, methodology, and expected outcomes (see Table 2.6, following the anatomy of metric identifiers in Figure 2.5). These metrics were adapted from the RDA FAIR Data Maturity Model developed by Mark D Wilkinson, Sansone, Schultes, et al. 2018.

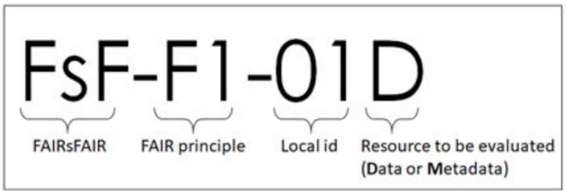


Figure 2.5: Anatomy of FAIRsFAIR metric identifier (Devaraju, Huber, et al. 2020).

Table 2.6: List of FAIRsFAIR Data Object Assessment Metrics_v0.5 (Devaraju, Huber, et al. 2020)

#	Identifier	Description
1	FsF-F1-01D	Data is assigned a globally unique identifier.
2	FsF-F1-02D	Data is assigned a persistent identifier.

Continued on next page

Table 2.6: List of FAIRsFAIR Data Object Assessment Metrics_v0.5 (Devaraju, Huber, et al. 2020) (Continued)

#	**Identifier**	**Description**
3	FsF-F2-01M	Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.
4	FsF-F3-01M	Metadata includes the identifier of the data it describes.
5	FsF-F4-01M	Metadata is offered in such a way that it can be retrieved by machines.
6	FsF-A1-01M	Metadata contains access level and access conditions of the data.
7	FsF-A1-02M	Metadata is accessible through a standardized communication protocol.
8	FsF-A1-03D	Data is accessible through a standardized communication protocol.
9	FsF-A2-01M	Metadata remains available, even if the data is no longer available.
10	FsF-I1-01M	Metadata is represented using a formal knowledge representation language.
11	FsF-I2-01M	Metadata uses semantic resources.
12	FsF-I3-01M	Metadata includes links between the data and its related entities.
13	FsF-R1-01MD	Metadata specifies the content of the data.
14	FsF-R1.1-01M	Metadata includes license information under which data can be reused.
15	FsF-R1.2-01M	Metadata includes provenance information about data creation or generation.
16	FsF-R1.3-01M	Metadata follows a standard recommended by the target research community of the data.
17	FsF-R1.3-02D	Data is available in a file format recommended by the target research community.

In 2021, Devaraju, Mokrane, et al. 2021 developed 15 core FAIRness assessment metrics (v0.3) within the FAIRsFAIR project, aiming to bridge the gap between conceptualizing and implementing FAIR metrics in trustworthy data repositories. These metrics, building on the RDA FAIR Data Maturity Model (Mark D Wilkinson, Sansone, Schultes, et al. 2018), were refined through community feedback and testing, focusing on unique identifiers, metadata quality, accessibility, and interoperability. Using the F-UJI tool (Devaraju and Huber 2020), the authors validated these metrics by evaluating 500 datasets, showing improvements in accessibility and reusability. The study concludes that developing FAIR metrics should be an ongoing process with iterative testing and feedback to support broader adoption of FAIR principles and enhance data availability and reuse.

It can be concluded that, although different sets of FAIRness assessment metrics (Mark D Wilkinson, Sansone, Schultes, et al. 2018; Devaraju, Huber, et al. 2020; Devaraju, Mokrane, et al. 2021) have been developed, inconsistencies in FAIRness assessments still exist. This implies that various tools may produce different results when evaluating the same digital objects due to varied interpretations of FAIR principles. If FAIRness is to be consistently applied and interpreted, it is fundamental that existing measurement techniques are robustly evaluated. A standardized approach to FAIRness measurement may ultimately be established. Currently, various implementations and self-assessments are being utilized, but their strengths, weaknesses, and adherence to fundamental FAIR principles remain insufficiently understood.

2.10 FAIR Metrics Standardization

In 2020, the FAIR Metrics and Data Quality Task Force (TF) was established to implement and oversee FAIR metrics and data quality within the EOSC (EOSC Task Force FAIR Metrics and Data Quality 2020). The TF aims to test and validate existing FAIR metrics across various research communities, ensuring they are practical and beneficial for EOSC stakeholders. As of June 13, 2024, the TF has published five documents, with four focusing on FAIR-related research (EOSC 2024).

On December 1, 2022, the TF published a document addressing challenges in governing FAIRness assessments for digital research objects (Mark D. Wilkinson et al. 2022). The main challenges include a lack of standardization, leading to inconsistent interpretations of FAIR principles, and diverse implementations across communities, resulting in varied assessment outcomes. The TF emphasizes the need for a standardized approach to measure FAIRness and stresses the importance of establishing a governance model for transparent and objective assessments. The document reviews existing models from organizations like the Internet Engineering Task Force (Internet Engineering Task Force 2024), World Wide Web Consortium (W3C) (World Wide Web Consortium 2024), EOSC (European Commission 2024b), and GO FAIR (Bernard Mons 2017), which offer practical approaches for governing FAIRness. The TF calls for a trusted and sustainable governance mechanism to harmonize evaluations, with the goal of professionalizing assessments and providing consistent, reliable tools and guidelines across all domains.

On December 20, 2022, the TF published a document detailing their efforts to harmonize FAIRness assessments by developing a standardized approach called 'FAIR Signposting' for metadata provision and ensuring consistency across assessment tools (Mark D Wilkinson, Sansone, Marjan, et al. 2022). This method uses web standards to guide automated agents in discovering and retrieving metadata and data from digital objects, aiming to reduce inconsistencies in assessments by standardizing metadata provision. The document also discusses the outcomes of workshops and hackathons that brought developers together to address these inconsistencies. Further details are provided in Section 2.12.

On January 11, 2024, the TF reported on its efforts to develop and promote FAIRness assessment tools and methodologies (Mark D Wilkinson, Sansone, Grootveld, et al. 2024). The report highlights challenges, methodologies, and outcomes from workshops and hackathons aimed at improving consistency and reliability in FAIRness assessments. Key concepts include FAIR Signposting—a method for organizing and publishing metadata that makes digital resources easier to find and access, Community-Driven Governance—emphasizing a governance model shaped by community input, and FAIRness Assessment Tools—developing standardized tools to reduce inconsistencies and improve comparability. The TF conducted six workshops and hackathons, focusing on findability and reusability, leading to the following outcomes.

- Development of reference environments: Created benchmarks to evaluate metadata harvesters' compliance with FAIR Signposting;
- Community uptake: Engaged repositories and tool-builders, with early implementations by platforms like Dataverse (Dataverse Project 2024) and Zenodo (Zenodo 2024); and
- FAIRness assessment governance: Highlighted the need for a governance model to ensure consistency across tools and platforms.

The document identifies major challenges, including inconsistencies in FAIRness assessments due to varying interpretations of metrics and high implementation costs of FAIR Signposting. To address these, the authors recommend

- Adoption of common standards for FAIR data exchange,
- Emphasizing comprehensive documentation and metadata,
- Ensuring data errors are corrected at the source,
- Regular user engagement to meet data quality expectations,
- Developing standardized FAIRness assessment tests, and
- Establishing a governance body to oversee assessment standards.

The document concludes by endorsing FAIR Signposting as a standards-compliant method for metadata publication and stresses the need for a unified governance structure to manage data quality and FAIRness assessments, aiming to ensure consistent application of FAIR principles across the research community.

On March 8, 2024, the TF published a report on a community survey evaluating FAIRness (Papadopoulou et al. 2024). Conducted between late 2022 and early 2023, the survey gathered insights from 78 respondents, mostly from academia (94%) across 62 organizations, including developers and users of FAIRness assessment tools. The survey aimed to support harmonization and explore community-driven governance. Key findings include:

- Most respondents applied FAIR principles to assess data, software, or research outputs, with self-assessments being the most common method.
- Challenges were noted in interpreting FAIR principles, understanding criteria, and validating assessment results.
- Trust in assessment results is moderate, with higher trust among technical respondents.
- There is strong support for establishing a FAIRness assessment Governance Body.
- Respondents emphasized the need for community involvement, transparency, and developing best practices and infrastructures.
- Confidence in interpreting FAIR principles varies, with suggestions to increase researcher awareness.

The survey highlights that while the research community is generally familiar with FAIR principles, challenges remain in tool interpretation and assessment criteria. There is a clear need for better governance, training, and community engagement.

In conclusion, the FAIR Metrics and Data Quality Task Force has underscored the need for standardized and consistent FAIRness assessments across research communities. Despite progress in developing tools like FAIR Signposting and gathering feedback, challenges persist in interpreting FAIR principles and ensuring consistent assessments. The TF stresses the importance of creating a unified governance structure, improving training, and increasing community involvement to better implement FAIR principles.

2.11 FAIRness Assessment Tools

In the years following the introduction of FAIR principles, initiatives and researchers,

such as EOSC Task Force FAIR Metrics and Data Quality Charter (EOSC Task Force FAIR Metrics and Data Quality 2020), the European Commission Expert Group on FAIR data (Hodson et al. 2018), the Digital Curation Centre (Whyte et al. 2021), the NIH Big Data to Knowledge (BD2K) initiative (Jagodnik et al. 2017), Barend Mons, Neylon, et al. 2017, and Mark D Wilkinson, Sansone, Schultes, et al. 2018 have focused on assessing the FAIRness of digital objects.

Various tools have been developed to evaluate compliance with FAIR principles using specific metrics (See Section 2.9). These tools fall into three categories: automatic, manual, and hybrid (Petrosyan et al. 2023). Each type has its own strengths and weaknesses; manual tools consider subjective details, while automatic tools follow stricter criteria, offering scalability but with greater challenges (Petrosyan et al. 2023).

FAIR-IMPACT, an EOSC initiative, focuses on developing tools to assess FAIRness across various digital objects and disciplines, promoting cross-collaboration and alignment (FAIR-IMPACT 2024). As of May 10, 2024, the FAIR-IMPACT website lists three automated tools: F-UJI (Devaraju and Huber 2020), O'FAIRe (Amdouni, Bouazzouni, and Jonquet 2022), and FOOPS! (Garijo, Corcho, and Poveda-Villalón 2021). The FAIRsFAIR project (FAIRsFAIR 2019-2022) also supports FAIRness assessment within EOSC and mentions F-UJI as a FAIRness assessment tool. Additionally, FAIRassist (FAIRsharing Team Data Readiness Group 2024), an educational component of FAIRsharing, tracks and lists 28 tools (14 manual, 2 hybrid, and 12 automated) for assessing digital objects against FAIR principles. However, different tools may yield varying results due to flexible interpretations of FAIR principles. Among the automated tools, registered on FAIRassist, F-UJI (Devaraju and

Huber 2020), FAIR-Checker (Gaignard et al. 2023), and FAIR Evaluator (Mark D Wilkinson, Dumontier, Sansone, et al. 2019) are suitable to assess the FAIRness of knowledge graphs. The following paragraphs describe these automated tools in chronological order.

In 2018, Mark D Wilkinson, Dumontier, Sansone, et al. 2019 introduced FAIR Evaluator, building on their earlier framework for automated FAIRness assessment (Mark D Wilkinson, Sansone, Schultes, et al. 2018). FAIR Evaluator is a framework with measurable indicators, an open-source tool, and community participation guidelines for domain-specific FAIRness assessments of digital objects. The framework includes several key components.

1. Maturity Indicators (MIs): Community-authored specifications that define automatically-measurable FAIR behaviors.

2. Compliance tests: Small web applications that test digital resources against individual MIs.

3. The evaluator: A web application that assembles and applies relevant compliance tests to digital resources, providing detailed FAIRness reports.

The framework supports various approaches to FAIRness assessment, including self-assessment, task forces, crowd-sourcing, and automation, allowing scalability across numerous digital objects. Initially tested with Gen1 metrics on 11 biotechnology data resources, the FAIR Evaluator helps data stewards improve their resources by generating detailed FAIRness reports. Built on Ruby on Rails, the application interacts with components via JSON interfaces and includes a metadata harvester for extracting metadata from various GUIDs. The public interface is a JavaScript-based Single Page Application.

The authors compare their proposed framework, FAIR Evaluator, with other initiatives such as the Research Data Alliance SHARC IG (David et al. 2018), CSIRO/OzNome 5-star System (Yu and Cox 2017), DANS 'FAIR enough?' Questionnaire (Dutch Data Archiving and Networked Services 2020), ARDC FAIR Self-

Assessment Tool (Australian Research Data Commons 2020), and ELIXIR Data Stewardship Wizard (Suchánek and Pergl 2018). Unlike these initiatives, which often rely on questionnaire-based approaches that measure intentions, FAIR Evaluator focuses on detecting machine-readable and reusable behaviors of digital objects, providing a more outcome-oriented assessment of FAIRness.

F-UJI, introduced by the FAIR-IMPACT, FAIRsFAIR, and FAIRassist resources, is an automated FAIRness assessment tool developed by Devaraju and Huber 2020. The name "F-UJI" combines "F" for "FAIR" and "UJI," meaning "Test" in Malay. It is a REST API service using the OpenAPI specification and is open-source under the MIT License. F-UJI assesses digital objects based on aggregated metadata, including metadata from landing pages and Persistent Identifiers (PIDs) (C. Sun, Emonet, and Dumontier 2022). It supports PIDs like Handle, DOI, and others, and considers metadata standards such as DublinCore, DCAT, DataCite, and Schema.org. F-UJI evaluates the FAIRness of digital objects using 16 metrics distributed across the FAIR principles: five for findability, three for accessibility, three for interoperability, and five for reusability (Devaraju, Huber, et al. 2020). These metrics, known as "FAIRsFAIR Data Object Assessment Metrics_v0.5," were introduced by Devaraju, Huber, et al. 2020 (See Section 2.9).

F-UJI is a continuously evolving tool designed to evaluate large numbers of datasets without requiring individual entry of each identifier. Petrosyan et al. 2023 used F-UJI to assess the FAIRness of 6,288 agricultural datasets, revealing varying levels of compliance with FAIR principles across different repositories. Findability scored the highest, while reusability scored the lowest. The authors recommend identifying the causes of low reusability and finding ways to improve it.

FAIR-Checker (Gaignard et al. 2023) is a web-based tool aimed at assessing and improving the FAIRness of metadata, particularly in the life sciences. It utilizes Knowledge Graphs and Semantic Web technologies to create, evaluate, and enhance machine-actionable metadata. Designed for various users, including data producers, software developers, and repository developers, FAIR-Checker helps make resources

FAIRer by evaluating repositories, software, and enriched metadata for registries. FAIR-Checker simplifies automated FAIRness assessment by helping non-experts evaluate and improve metadata quality. It offers two main modules: the check module, which thoroughly evaluates metadata and suggests improvements for FAIRness, and the inspect module, which helps users enhance metadata quality. The tool uses SPARQL queries and SHACL[13] (World Wide Web Consortium 2017) constraints to automatically assess FAIR metrics and notify users about missing or recommended metadata.

FAIR-Checker was developed in Python using the Flask web framework and employs various libraries, including Requests (Reitz 2024), Selenium (Selenium Project 2024), Extruct (Scrapinghub 2024), RDFlib (RDFlib Developers 2024), and pySHACL (pySHACL Developers 2024), to handle RDF data and perform evaluations. It generates SHACL shapes from Bioschemas profiles (Bioschemas 2024) to check metadata annotations' completeness and correctness. Caching techniques optimize performance and reduce the load on external SPARQL endpoints. FAIR-Checker was tested on over 25,000 bioinformatics software descriptions, identifying compliance with mandatory and recommended properties. The tool helps identify metadata gaps and provides actionable recommendations to improve adherence to FAIR principles.

The three automated FAIRness assessment tools described above may be used with knowledge graphs. However, other automated tools registered on the FAIRassist webpage are not suitable for assessing knowledge graphs. SciScore (Menke et al. 2022) evaluates the FAIRness of research manuscripts, PresQT (PresQT 2024) offers FAIR maturity metrics and a RESTful API for assessing research datasets with DOIs, and OpenAIRE (OpenAIRE 2018) provides a dashboard that includes a FAIRness assessment service for data repositories.

Key takeaways from reviewing FAIRness assessment tools include the need for harmonizing methodologies to ensure consistent evaluations, developing unified cri-

[13] SHACL is a language for validating RDF data against a set of conditions or rules to ensure it conforms to a specified structure.

teria and guidelines, and fostering cross-disciplinary collaboration to create universally applicable tools. Ongoing development and user feedback are also essential for improving the accuracy of these assessments. In conclusion, consistent and comprehensive FAIRness assessments require harmonized criteria, cross-disciplinary efforts, and addressing gaps in existing tools.

2.12 Comparative Studies of FAIRness Assessment Tools

As discussed earlier, several automated FAIRness assessment tools are available, raising the question: "What are the differences between these tools?" Krans et al. 2022 evaluated ten tools, including online self-assessment survey-based tools, online (semi-)automated tools, offline self-assessment tools, and other tools, using datasets from the nanomaterials and microplastics risk assessment domain. They compared the tools based on nine criteria grouped into prerequisite knowledge, ease of use, and the type and detail of output. Based on the results, the authors recommend using online survey tools for initial assessments and (semi-)automated tools for detailed evaluations, suggesting a combination for comprehensive results. They call for improved guidance, consistency in scoring, and better harmonization of tools to ensure more consistent and actionable FAIRness assessments.

C. Sun, Emonet, and Dumontier 2022 systematically compared three automated FAIRness assessment tools—F-UJI, FAIR Evaluator, and FAIR-Checker—by testing datasets from different repositories, including GeoData (GeoData 2020), CORD-19 (L. L. Wang et al. 2020), and NL-Covid-19 from RIVM (Dutch Institute for Public Health and Environment 2024). The study aimed to highlight differences in the tools' design, implementation, and results, emphasizing the need for consistent FAIRness assessment tools to improve data management practices. The authors found significant variations in evaluation results, driven by differences in the tools' design, implementation, and metric documentation.

The comparison revealed that F-UJI and FAIR Evaluator both support APIs and provide JSON output with detailed logs, while FAIR-Checker uses the FAIR Evaluator API, offering a more visually appealing presentation but lacking detailed logs and metric test selection. F-UJI focuses on reusability, whereas FAIR Evaluator emphasizes interoperability, leading to conflicting results due to differing definitions and metric implementations. In dataset testing, GeoData scored perfectly with F-UJI but failed five tests with FAIR Evaluator, CORD-19 failed more tests with FAIR Evaluator than F-UJI due to poor metadata quality, and NL-Covid-19 scored lower overall, struggling with license information in both tools.

The authors identified several issues with FAIRness assessment tools, including conflicting results due to varying interpretations of data and metadata identifiers, differences in the level of detail extracted (especially for license information), and inconsistent standards for evaluating relationships between local and third-party data. These discrepancies highlight the need for standardized benchmarks and clearer metric definitions to ensure consistent evaluations. Additionally, improving transparency in metric test implementation and documentation is crucial for helping users enhance their data's FAIRness. The study's limitations include the selection of datasets, the lack of detailed examination of metric implementations, and the ongoing development of the tools, which may lead to changes over time.

Peters-von Gehlen et al. 2022 emphasize the need for a standardized, globally accepted procedure for FAIRness evaluation, particularly regarding domain-specific dataset requirements. The study applied five different FAIRness evaluation tools to the World Data Center for Climate (WDCC) (World Data Center for Climate 2024) to offer recommendations for improving these tools and methodologies, ensuring they are suitable for comprehensive assessments of research data repositories. The tools used include the Checklist for Evaluation of Dataset Fitness for Use (Austin et al. 2019), which assesses data fitness comprehensively; FAIR Evaluator, which provides detailed automated assessments; FAIRshake (Clarke et al. 2019), a hybrid tool combining automated and manual evaluations with domain-specific flexibility;

F-UJI, a mature automated tool for assessing various FAIR metrics; and a Self-Assessment approach (Bahim, Dekkers, and Wyns 2020) using self-developed metrics to provide a holistic view of WDCC's (meta)data curation practices.

The WDCC-archived data had an overall mean FAIR score of 0.67 out of 1, ranging from 0.5 to 0.88. Manual approaches generally yielded higher scores than automated ones, with hybrid methods showing the highest scores. Data collections with DOIs and ample metadata scored better. Manual and hybrid approaches aligned closely, while automated tools like FAIR Evaluator and F-UJI were useful for machine-actionable aspects but struggled with contextual reusability. Manual tools captured more contextual details but were subject to interpretation bias. However, manual evaluation is impracticable across the very wide range of data presently available and utilized in AI systems. The hybrid approach of FAIRshake combined the strengths of both. Based on these findings, the authors recommend adopting hybrid approaches for future tools to capture both technical and contextual reusability, involving the community in designing discipline-specific FAIRness metrics, improving automated tools to better handle contextual information, and fostering collaboration between tool developers and evaluators to refine tools and ensure they meet user needs.

In December 2022, the FAIR Metrics and Data Quality Task Force published a document detailing their efforts to harmonize FAIRness assessments across different tools (Mark D Wilkinson, Sansone, Marjan, et al. 2022). A benchmark repository was created to test and compare tool performance uniformly. Tools such as FAIR Evaluator, F-UJI, FAIR Enough?/SATIFYD, FAIR Checker, ENVRI-FAIR, and FAIRshake were included in the study. The authors stress the ongoing need to harmonize FAIRness assessment tools, encourage new tools to adopt benchmark environments, and emphasize the role of data publishers and the EOSC Association in promoting these practices. They also highlight the sustainability of FAIR Signposting and benchmark environments, supported by FAIRsharing and governance initiatives, and underline the importance of community involvement and continuous updates for long-term success.

In March 2024, the FAIR Metrics and Data Quality Task Force published a report on a community survey conducted in late 2022 and early 2023 aimed to support the harmonization and community-driven governance of FAIRness assessments (Papadopoulou et al. 2024). Among 29 respondents who provided tool information, F-UJI was the most mentioned tool (16 times), followed by the FAIR Evaluator (7 times). In total, fourteen different tools were mentioned, with F-UJI widely used across domains.

Vogt et al. 2024 address the limitations in current research data management, where data is scattered across repositories with varying structures and terminologies, hindering efficient integration and reuse. They propose FAIR 2.0, an extension of the original FAIR principles, to enhance semantic interoperability, which they define as consisting of two key aspects, terminological interoperability, ensuring consistent interpretation of terms (both ontological and referential), and propositional interoperability, ensuring uniform application of data schemata and logical frameworks across datasets.

To manage data overload and improve reusability, the authors propose FAIR services: a terminology service for controlled vocabularies and ontologies, a schema service for managing data schemata and ensuring uniform representation, and an operations service for providing executable functions to enhance machine-actionability. They also suggest a framework extending the original FAIR principles to address semantic aspects, ensuring data interoperability and meaningfulness across contexts. The authors emphasize that machine-actionability is crucial for achieving FAIR data, where data must be not only machine-readable but also actionable by machines.

The study critiques current FAIRness assessment tools like FAIR-Checker, F-UJI, and FAIR Evaluator for their limited focus on basic metadata and licensing, and their lack of support for domain-specific data. It highlights the challenge of establishing uniform data schemata and logical frameworks across diverse datasets, acknowledging the significant effort needed to create and maintain mappings and crosswalks for interoperability. The study addresses gaps such as the lack of focus on semantic interoperability, the limitations of existing assessment tools, and the need for standardized services to manage terminological, schema, and operational aspects of data interoperability.

While efforts to implement FAIR principles have made progress, harmonizing FAIRness assessment tools across domains remains a persistent challenge. Initiatives like the FAIR Data Action Plan and FAIRsFAIR highlight the critical need for standardized metrics and policies, and task forces such as the RDA FAIR Data Maturity Model and the FAIR Metrics and Data Quality Task Force have developed semi-quantitative metrics to promote consistency in evaluations (Hodson et al. 2018; Whyte et al. 2021; Mark D Wilkinson, Sansone, Schultes, et al. 2018; EOSC 2024). However, comparative studies continue to reveal inconsistent results, largely due to disparities in tool design and metric interpretation, underscoring the urgent need for clearer definitions and standardized benchmarks (C. Sun, Emonet, and Dumontier 2022; Devaraju, Huber, et al. 2020; Mark D Wilkinson, Sansone, Marjan, et al. 2022). Furthermore, the lack of studies comparing automated FAIRness assessment tools for knowledge graphs reveals critical gaps in these harmonization efforts (Krans et al. 2022). Ongoing efforts emphasize the necessity of a unified governance structure and improved tool functionality to ensure reliable and consistent assessments across diverse research domains (Papadopoulou et al. 2024; Mark D Wilkinson, Sansone, Grootveld, et al. 2024).

In conclusion, while there are compelling reasons for adopting FAIR principles in areas such as data trustworthiness and AI systems development, research into the consistent automated measurement of FAIRness in knowledge graphs is lacking. This significantly undermines the robust application of FAIRness principles in practice, as the volume of data held in knowledge graphs, together with the growing number of knowledge graphs, is too large to be suited to manual FAIRness assessment. This research will examine and reduce this gap, undertaking a systematic and deep evaluation of the currently available automated FAIRness assessment tools using machine learning techniques.

2.13 Highlighted Gaps

This chapter provides an overview of AI transparency (Section 2.2), data transparency (Sections 2.3), knowledge graphs (Section 2.4), their transparency (Sections 2.5 and 2.6), and its key components, including quality and FAIR (Sections 2.7 to 2.12). Here are the key takeaways from this chapter:

1. Data transparency is central to AI transparency.
2. The growing use of knowledge graphs in AI systems makes their transparency crucial for ensuring system integrity and trustworthiness.
3. There is limited published research on knowledge graph transparency.
4. Transparency in knowledge graphs requires evaluating data provenance, currency, and the evidence supporting associations (Section 2.6).
5. Key dimensions of transparency include data provenance, quality, and accessibility, with knowledge graph quality and FAIR principles as indicators of transparency (Sections 2.6 and 2.7).
6. Existing tools for evaluating knowledge graph quality and FAIRness could be adapted for transparency evaluation (Sections 2.7 and 2.8).

According to this literature review, as discussed in Sections 2.9, 2.11, and 2.12, the following gaps related to FAIRness assessment are identified.

1. There are inconsistencies in interpreting and implementing FAIR principles, metrics, and FAIRness assessment. However, different initiatives have tried to harmonize the FAIRness assessment across various tools, the inconsistencies still persist.

2. Second, no systematic research has focused on a broad analysis of automated FAIRness measurement across a large variety of knowledge graphs, and therefore the true scale of measurement volatility is unknown. This lack of quantification raises considerable challenges for the community as it grapples with the FAIRness measurement at scale.

Chapter 3

Methodology and Research Design

3.1 Chapter Overview

This chapter outlines the methodology and research design employed in the study. The research approach used in this study is a quantitative approach. This allows utilizing statistical methods and Machine Learning techniques for the first time to provide a tangible view of the inconsistencies between three automated FAIRness assessment tools, created by European Union-funded projects (Mark D Wilkinson, Sansone, Schultes, et al. 2018; Devaraju, Mokrane, et al. 2021; FAIRplus 2019-2022), as opposed to subjective claims about these inconsistencies made in the literature.

These tools may be used for assessing the FAIRness of Knowledge Graphs (KGs). Furthermore, potential strategies to minimize these inconsistencies are explored.

Quantitative research, as described by J. W. Creswell and J. D. Creswell 2018, serves as a method for testing objective theories by examining relationships among variables. This approach relies on numerical data to investigate these relationships and establish correlations between variables and outcomes (J. W. Creswell and J. D. Creswell 2018; Choy 2014). Quantitative research methods offer several notable advantages, including their suitability for testing theories and hypotheses, the ability to collect large-scale responses through surveys, and their highly systematic and structured nature.

Quantitative methods are deemed suitable in this context, allowing for an in-depth analysis of inconsistencies between the tools' results, exploring the factors influencing these inconsistencies, and identifying a reliable and comprehensive approach for assessing FAIRness in KGs. The quantitative approach aligns with the research questions and aims to provide rich insights into the subject matter. The research questions which form the basis for this study are:

1. How consistent are the results produced by automated knowledge graph FAIRness assessment tools?
2. How can confidence in automated FAIRness assessment for knowledge graphs be increased through specific techniques?

The structure of this chapter is organized as follows. Section 3.2 outlines the research philosophy. Section 3.3 details the research design. Finally, Section 3.4 presents the concluding remarks.

3.2 Research Philosophy

Two core branches of philosophy, ontology and epistemology, play a significant role in determining the research paradigm for a scientific study. Ontology focuses on understanding the nature of "reality"—the essence of things or concepts that researchers seek to describe. On the other hand, epistemology examines what qualifies as "knowledge" about that reality, distinguishing it from belief (Žukauskas, Vveinhardt, and Andriukaitienė 2018). These philosophical perspectives underpin three primary methodological paradigms commonly used in IT innovation research, which have evolved from positivism to interpretivism and critical realism.

3.2.1 Ontology

Ontology, a branch of philosophy, explores the fundamental nature of reality, examining the types and arrangements of objects, events, processes, and concepts (Viinikkala 2004). It seeks to classify entities across various domains, with this study focusing on conceptual entities like resistance to innovation, coping responses, and motivating factors. Ontology is divided into two main perspectives: objectivism (or realism) and subjectivism. Objectivism asserts that reality exists independently of human perception, making it universally applicable, while subjectivism emphasizes the influence of individual or collective interpretations (Al-Saadi 2014; Bryman 2016).

This study adopts a realist ontological perspective, consistent with critical realism, which acknowledges the objective existence of phenomena—specifically, inconsistencies and unreliability in FAIRness assessments—while recognizing that our understanding of these phenomena is inherently subjective and imperfect.

3.2.2 Epistemology

Epistemology, as defined by Richards 2003, pertains to beliefs about the nature of knowledge and what constitutes acceptable knowledge within a discipline. According to Al-Saadi 2014, researchers' epistemological assumptions significantly influence their methodological choices. A key distinction in epistemology lies between objective knowledge, seen as tangible and independent, and subjective knowledge, regarded as personal and context-dependent.

In this study, an objective, positivist view of knowledge is adopted, aligning with a realist ontological stance consistent with critical realism. As noted by Ritchie et al. 2013, assuming knowledge is objective often necessitates a detached observer role and the use of quantitative methods. Accordingly, this study employs a quantitative approach to provide a statistically grounded analysis of the inconsistencies and reliability of automated FAIRness assessments.

Epistemological positions are often categorized as either positivist or interpretive/constructivist. Positivism prioritizes objectivity, treating the world as unaffected by the researcher. In contrast, interpretivism and constructivism view knowledge as constructed through human perceptions and social interactions (Tennis 2008). In the fields of technology and computer science, the positivist approach is widely adopted due to its emphasis on objectivity and quantifiable results.

3.2.3 Positivism

Positivism, as a research paradigm, is rooted in realist ontology and objective epistemology (Dudovskiy 2020), emphasizing that science is the sole means of discovering truth. It asserts that valid knowledge can only be obtained through systematic observation and measurement. In this paradigm, researchers are expected to focus on the objective collection of data and its limited interpretation, akin to methodologies employed in the hard sciences, such as physics (Gorski 2018). Positivist research requires researchers to remain independent (Ray 2017), minimizing interaction with participants and maintaining objectivity throughout the research process.

The empiricist perspective is reflected in positivism, as it holds that knowledge is derived from human experience. However, this does not equate ordinary human common sense with scientific knowledge. Instead, positivism prioritizes precise observation and logical reasoning. It adopts a deductive, or ”top-down,” approach, where general theories are formulated to generate predictions (hypotheses) that are then subjected to scientific testing. The role of research, according to this paradigm, is to collect data that either supports or refutes these hypotheses (Gorski 2018). The present study aligns with this perspective by testing hypotheses about the consistencies and unreliability in FAIRness assessment methods derived from existing research (Oliveira et al. 2021; Hasnain and Rebholz-Schuhmann 2018; Amdouni, Bouazzouni, and Jonquet 2022; Mark D Wilkinson, Sansone, Marjan, et al. 2022; Hodson et al. 2018; Whyte et al. 2021; European Commission Directorate-General for Research and Innovation 2016; Barend Mons, Neylon, et al. 2017; Mark D Wilkinson, Sansone, Schultes, et al. 2018; RDA FAIR Data Maturity Model Working Group 2020; Devaraju, Huber, et al. 2020; Mark D Wilkinson, Dumontier, Sansone, et al. 2019), using statistical methods and machine learning techniques on a collected dataset (See Chapter 5).

3.2.4 Interpretivism

Interpretivism aligns with subjective, non-realist perspectives of ontology and epistemology and emphasizes qualitative analysis, making it suitable for studies in the human and social sciences (Ryan 2018). It employs naturalistic methods, such as observations and interviews, to examine phenomena in real-life contexts, often incorporating the researcher's experiences and following an inductive "bottom-up" approach (O'Donoghue 2018). While it produces authentic and valid data, its subjectivity limits generalizability (Glesne 2016).

Interpretivism is not adopted in this study due to its inherent subjectivity and limited generalizability. The study prioritizes the need for statistically grounded and broadly applicable results, which are not aligned with the interpretivist approach.

3.2.5 Critical Realism

Critical realism (CR) is a philosophical paradigm that combines positivist and interpretivist elements, offering a comprehensive framework for understanding reality. As Archer et al. 2016 explain, CR distinguishes between the "real" world, which exists independently of human perceptions, and the "observable" world, shaped by human experiences. It asserts that unobservable structures cause observable events, emphasizing the complexity of reality and rejecting the reduction of ontology (reality) to epistemology (knowledge).

Developed by K. Bhaskar 1979 and elaborated by scholars such as Lawson 1997 and Sayer 1992, CR resolves tensions between positivism and interpretivism, providing an inclusive approach to scientific inquiry (Denzin and Lincoln 2011). Danermark 2002 describe CR's stratified view of reality, comprising the empirical (human experience), actual (independent events), and real (causal mechanisms) levels. This stratification highlights the limitations of positivism and constructivism, both of which oversimplify the nature of reality (R. Bhaskar et al. 1998).

CR also acknowledges the openness and complexity of the world, where entities possess causal powers that influence events (Psillos 2007). Knowledge in CR evolves over time as human understanding improves, reflecting the transitive nature of knowledge (Danermark 2002).

3.3 Research Design

Research design refers to the overarching strategy employed to logically and coherently integrate the various components of a study, ensuring that research questions are effectively addressed J. W. Creswell and J. D. Creswell 2013. A research design connects broader decisions, such as selecting the overall methodology (qualitative or quantitative) and the philosophical framework (research paradigm), with detailed choices about how data will be collected.

A quantitative research design is adopted in this study, enabling the application of statistical methods and Machine Learning techniques to conduct a detailed analysis of the consistencies and inconsistencies evident in various sets of measures used for automated FAIRness assessment in KGs. Additionally, the study explores more reliable and comprehensive methods for assessing FAIRness.

The process begins by identifying the state-of-the-art automated FAIRness assessment tools applicable to evaluating the FAIRness of KGs (detailed in Chapter 2). Next, a comparative analysis of the selected tools is conducted—for the first time—to provide a structured view of the differences in the metrics they use to assess FAIRness (detailed in Chapter 4). Subsequently, a diverse set of openly accessible knowledge graphs is selected, and their FAIRness is evaluated using the identified tools (detailed in Chapter 5). The results are then analyzed using statistical methods to quantify inconsistencies, offering an objective perspective that contrasts with the largely subjective interpretations found in the literature (detailed in Chapter 6).

Using a newly introduced statistical measure of consistency range, Machine Learning methods are employed to predict the range between the overall FAIRness scores. A smaller range is interpreted as indicative of a more reliable FAIRness assessment. Accordingly, knowledge graphs with higher FAIRness scores and lower ranges are considered more suitable for reuse and interoperation, particularly for

ML training purposes. This predictive analysis also identifies the most influential FAIRness measures affecting the range, providing insights for developing a potentially combined method for FAIRness assessment that is both comprehensive and reliable.

Figure 3.1 illustrates the overall research design including the methodology and steps undertaken in this study.

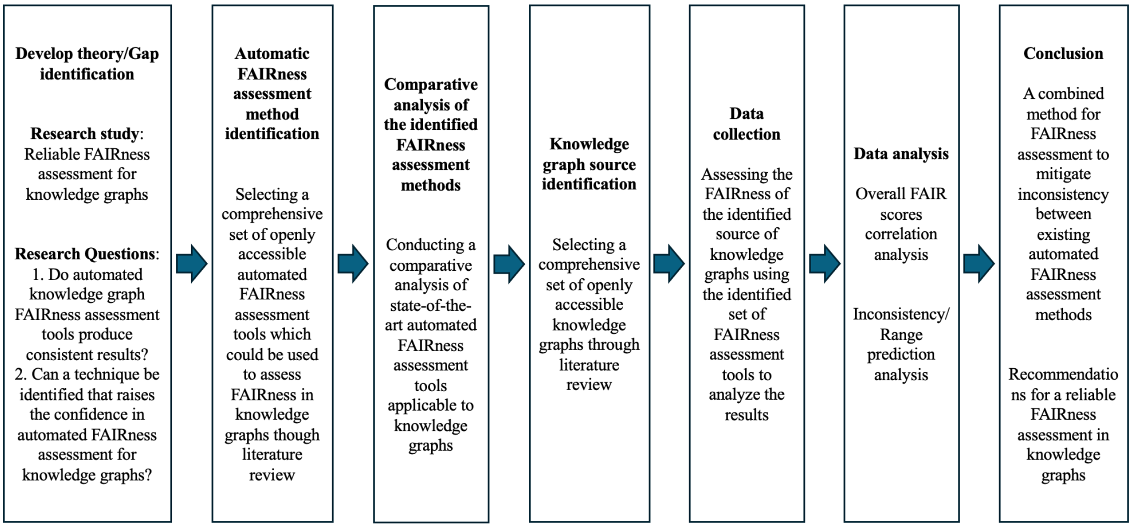


Figure 3.1: Overview of the research design adopted in this study.

The first two steps are done through the literature review (Chapter 2). Section 3.3.1 to Section 3.3.5 briefly describe the next research steps.

3.3.1 Comparative Analysis of the Identified FAIRness Assessment Tools

The comparative analysis of FAIRness assessment tools is conducted to understand the scope, depth, and consistency of measurement provided by each tool when applied to knowledge graphs. This step is undertaken by examining the publicly available documentation and source code implementations of the three tools considered suitable for assessing KG FAIRness: FAIR-Checker, F-UJI, and FAIR Evaluator. The analysis is carried out by first identifying the set of metrics used by each tool, then mapping these metrics to the FAIR principles to uncover overlaps, gaps, and conceptual interpretations. Special attention is given to how each tool operationalizes FAIR sub-principles through specific tests, how these are grouped into composite scores, and the expected metadata inputs required for each. This structured comparison enables a grounded understanding of each tool's design logic and interpretability, laying the foundation for subsequent evaluation and validation.

Alternative approaches—such as expert interviews or black-box testing based solely on tool outputs—were considered but ultimately rejected due to their limitations in transparency and reproducibility. The chosen method provides a more rigorous and systematic analysis by leveraging direct access to the tools' source code and technical documentation. Validation is achieved by cross-checking the documented descriptions of each test or metric against its actual implementation in code, ensuring consistency between intended functionality and practical behavior. This not only confirms the reliability of the comparative mapping but also supports reproducibility for future research. By using openly available materials, the analysis ensures transparency and supports FAIR principles in the evaluation process itself.

3.3.2 Knowledge Graph Source Identification

To evaluate automated FAIRness assessment tools, this study first identifies a suitable and representative source of open-access knowledge graphs. Several repositories—such as KG-Hub[1] (Caufield et al. 2023), OpenAIRE Graph[2] (Manghi et al. 2022), and the EU Open Data Portal[3] (European Union 2024)—offer valuable domain-specific datasets, but their limited topical coverage makes them less suitable for evaluating general FAIRness performance. To ensure broader domain representation, the Linked Open Data (LOD) Cloud[4] (Jentzsch, Cyganiak, and Bizer 2011; Schmachtenberg, Bizer, and Paulheim 2014) is selected. The LOD Cloud is the largest curated collection of interlinked datasets on the Web of Linked Data, spanning diverse domains including life sciences, media, and social networks (Assaf, Troncy, and Senart 2015; Hitzler 2021), and is widely used in academic and

[1] `https://kghub.org`, date accessed: 18th November 2024
[2] `https://graph.openaire.eu/`, date accessed: 18th November 2024
[3] `https://data.europa.eu/en`, date accessed: 18th November 2024
[4] `https://lod-cloud.net`, date accessed: 25th June 2024

industrial research (Debattista, Attard, et al. 2019; Debattista, Lange, et al. 2018; Nogales, Angel Sicilia-Urban, and García-Barriocanal 2017; Debattista, Auer, and Lange 2016).

The LOD Cloud's adherence to Linked Data principles and its alignment with FAIR publishing practices (Haller et al. 2020) make it a suitable and scalable resource for this study. Its inclusion criteria—such as RDF formatting, minimum triple count, and linkage to other datasets—support structured access and objective FAIRness evaluation (McCrae et al. 2024; Open Knowledge Foundation 2015; Berners-Lee 2006b). Compared to other sources, it offers a balanced combination of diversity, maturity, and accessibility, making it a methodologically sound choice for evaluating automated FAIRness tools. Chapter 5 includes detailed information about this step.

3.3.3 Data Collection

To conduct a robust and scalable evaluation of automated FAIRness assessment tools, this study systematically collects metadata links for knowledge graphs (KGs) listed in the Linked Open Data (LOD) Cloud[5]. As FAIRness assessment tools—F-UJI (Devaraju and Huber 2020), FAIR-Checker (Gaignard et al. 2023), and FAIR Evaluator (Mark D Wilkinson, Dumontier, Sansone, et al. 2019)—are all metadata-dependent, the identification of suitable metadata entry points is essential. A decision tree (Figure 5.1) is developed to formalize and automate the metadata source selection process. The tree prioritizes links from FAIRsharing[6] (Sansone et al. 2019) for KGs likely to receive updates, followed by Datahub[7] (Open Knowledge International 2024), the Mannheim Linked Data Catalog (University of Mannheim, Data and Web Science Group 2014a), the KG's own webpage, and, if none are available, the LOD Cloud HTML page. This structured and repeatable process ensures the metadata selected is as current and representative as possible. Ultimately, meta-

[5] `https://lod-cloud.net`, date accessed: 25th June 2024
[6] `https://fairsharing.org/`, date accessed: 18th November 2024
[7] `https://datahub.io`, date accessed: 18th November 2024

data for 1,230 KGs is collected using this method: 1116 from Datahub, 115 from FAIRsharing, 62 from KG websites, and 15 from LOD Cloud HTML pages.

The chosen data collection method balances practicality and comprehensiveness, capturing a wide coverage of the LOD Cloud while remaining feasible for automated evaluation. Manual metadata curation was considered but dismissed due to its subjectivity, scalability limitations, and potential for inconsistency. The decision tree approach offers a transparent and replicable process for selecting appropriate links while addressing metadata freshness—particularly important given that many public catalogs are no longer actively maintained. Validation was built into the process through iterative refinement: if a chosen metadata source failed to produce a valid result for a given tool, the next source in the hierarchy was used. Adjustments were made, for example, when FAIR-Checker or FAIR Evaluator servers failed to respond or returned errors. The process ensures that each KG is assessed using the most suitable and technically compatible metadata link available, supporting the objective of evaluating FAIRness tools at scale. See Chapter 5 for detailed information.

3.3.4 Data Analysis

Following the collection of FAIRness assessment results for over a thousand knowledge graphs using three automated tools—F-UJI, FAIR-Checker, and FAIR Evaluator—the data analysis step involves systematically examining the consistency, divergence, and patterns among the outputs. Statistical techniques such as descriptive analysis, correlation assessment, and dimensionality reduction (e.g., PCA) are applied to summarize and visualize trends across tools and FAIR principles. The analysis also incorporates techniques for inconsistency detection and score aggregation to construct a holistic view of each tool's behavior across different metrics and datasets. This step is conducted using Python, with particular focus on standardizing input formats, handling missing or failed assessments, and ensuring the comparability of metric values across tools. Throughout the process, the emphasis remains on quantifying patterns objectively, rather than relying on subjective judgments about tool behavior.

Alternative methods such as manual benchmarking or purely qualitative comparison were considered but rejected due to limitations in scalability, reproducibility, and bias control. The chosen approach is preferred for its alignment with FAIRness assessment goals—particularly transparency and reusability. Validation is embedded throughout the analysis pipeline by cross-checking output consistency, testing the robustness of results under different assumptions (e.g., tool output completeness), and ensuring reproducibility through scripted workflows. While the full details of the statistical analysis appear in Chapter 6, this methodological step is designed to be generalizable and adaptable to other FAIRness evaluation contexts, thereby strengthening the study's contribution to future research in this space. See Chapter 6 for detailed information.

3.3.5 Conclusion Derivation

The final step in this research involves synthesizing findings from the data analysis, comparative evaluation, and literature review to draw structured, evidence-based conclusions about the consistency and reliability of automated FAIRness assessment tools. This process is grounded in a critical interpretation of the observed inconsistencies, statistical patterns, and limitations identified in earlier steps. Conclusions are derived through an iterative, triangulation-based approach: insights from the quantitative evaluation of tool outputs are cross-checked with methodological assumptions, and reflections on the broader implications for FAIR implementation and assessment practice are contextualized using the FAIR principles and community standards.

Rather than applying a separate formal method for conclusion derivation (e.g., expert panels or Delphi studies), this research follows a results-driven analytical narrative supported by empirical evidence. This approach is consistent with common practices in computational and applied data science research, where conclusions are justified through the rigor of the analysis pipeline. To validate the robustness of the conclusions, particular attention is paid to ensuring consistency across multiple tools, fairness in interpretation (e.g., avoiding bias toward any single tool), and traceability of findings to raw results. The conclusions are further informed by the researcher's critical evaluation of the tools' limitations, highlighting areas for improvement and future work. This method ensures that the conclusions not only reflect the data but also meaningfully contribute to ongoing discussions about standardizing FAIRness assessments. See Chapter 7 for detailed information.

3.4 Concluding Remarks

In conclusion, this study employs a quantitative research design, utilizing statistical methods and Machine Learning techniques to examine existing automated FAIRness assessment tools. By providing a comparative analysis of the state-of-the-art tools and assessing the FAIRness of evaluating a comprehensive source of Knowledge Graphs using state-of-the-art tools and analyzing results, a clear understanding of these inconsistencies is achieved. Machine Learning enhances this analysis by identifying key measures and predicting inconsistency, enabling the development of a more robust FAIRness assessment method. This approach ensures rigor and objectivity while advancing FAIRness practices and the development of reusable, interoperable Knowledge Graphs for model training.

Chapter 4

Comparative Analysis of Automated Knowledge Graph FAIRness Assessment

4.1 Chapter Overview

In this chapter, an overview of the FAIRness assessment process in F-UJI, FAIR Evaluator, and FAIR-Checker is provided. These three assessment tools were selected for this study as they represent the only freely available automated approaches suitable for the FAIRness assessment of Knowledge Graphs. Additionally, they are widely adopted in related research, further supporting their relevance for this analysis. Subsequently, a comparative analysis of the measures employed by these three FAIRness assessment tools is conducted. The purpose is to elucidate which measures are utilized by these tools, the meaning of each measure, how they operate (i.e., which features or properties in the KGs they consider to evaluate each metric), their score values, and the results they produce. The structure of the chapter is as follows: Section 4.2 outlines and explains the FAIRness assessment measures used in F-UJI. Section 4.3 provides an overview and explanation of the measures employed by FAIR Evaluator. Section 4.4 discusses the FAIRness assessment process in FAIR-

Checker. This is followed by Section 4.5, which presents a comparative analysis of the measures utilized by the three tools. Finally, concluding remarks are provided in Section 4.6.

4.2 FAIRness Assessment in F-UJI

F-UJI employs a hierarchical structure to provide numerical assessments for each FAIR principle. Specifically, for each FAIR principle, there is a set of metrics, and for each metric, there is a corresponding set of practical tests. Table 4.1 presents the metrics and practical tests associated with FAIR sub-principles, along with their descriptions and score values. F-UJI's output includes scores obtained for each practical test, an aggregate score for each corresponding sub-principle (F1, F2, F3, F4, A1, I1, I2, I3, R1, R1.1, R1.2, R1.3), aggregate scores for each principle (F, A, I, R), and an overall aggregate FAIR score.

Table 4.1: FAIRness assessment metrics, their corresponding practical tests, and explanations of their operation in F-UJI.

FAIR sub-principle	Metric	Practical test	Score value
F1. (Meta)data are assigned a globally unique and persistent identifier	**FsF-F1-01D**: Data is assigned a Globally Unique Identifier (GUID)[1]. Possible score: 0, 0.5, 1	**FsF-F1-01D-1**: Identifier is resolvable and follows a defined unique identifier syntax (IRI, URL)	0, 1
		FsF-F1-01D-2[2]: Identifier is not resolvable but follows an UUID[3] or HASH type syntax[4]	0, 0.5

Continued on next page

[1]It is associated with only one resource at any given time, e.g., Uniform Resource Identifiers (URI), Internationalized Resource Identifiers (IRI)-extend the functionality of URIs by allowing non-ASCII characters from Unicode, e.g., 'https://español.ejemplo.com/', and Digital Object Identifiers (DOI)-A unique alphanumeric string assigned to a digital object to provide a permanent and reliable link to its location on the internet, e.g., '10.5281/zenodo.6505846' (FAIRsFAIR Project 2024).

[2]As a personal view, the inability of a dataset link to be resolvable represents a significant issue that fundamentally undermines its findability. Assigning a score of 0.5 solely based on the link following a defined pattern does not effectively align with the principles and objectives of FAIR.

[3]A Universally Unique Identifier is a 36-character alphanumeric string that can be used to identify information, e.g., 550e8400-e29b-41d4-a716-446655440000.

[4]A string containing a '#' followed by the fragment identifier of the URL, e.g., http://example.com/file#!md5!b3187253c1667fac7d20bb762ad53967.

Table 4.1 – continued from previous page

FAIR sub-principle	Metric	Practical test	Score value
	FsF-F1-02D: Data is assigned a persistent identifier (PID). Possible score: 0, 0.5, 1	**FsF-F1-02D-1**: Identifier follows a defined PID syntax	0, 0.5
		FsF-F1-02D-2: The identifier is resolvable to a valid URI	0, 1
F2. Data are described with rich metadata (defined by R1 below)	**FsF-F2-01M**: Metadata includes descriptive core elements to support data findability. Possible score: 0, 0.5, 1, 1.5, 2	**FsF-F2-01M-1**: Metadata has been made available via common web methods	0, 0.5
		FsF-F2-01M-2: Core data citation metadata is available (creator, title, publisher, publication_date, object_identifier, object_type)	0, 0.5
		FsF-F2-01M-3: Core descriptive metadata is available (creator, title, object_identifier, publication_date, publisher, object_type, summary, keywords)	0, 1
F3. Metadata clearly and explicitly include the identifier of the data they describe	**FsF-F3-01M**: Metadata includes the identifier of the data it describes. Possible score: 0, 0.5, 1	**FsF-F3-01M-1**: Metadata contains data content related information (file name, size, type)	0, 0.5
		FsF-F3-01M-2: Metadata contains a PID or URL which indicates the location of the downloadable data content	0, 0.5
F4. (Meta)data are registered or indexed in a searchable resource	**FsF-F4-01M**: Metadata is offered in such a way that it can be retrieved by machines. Possible score: 0, 1, 2	**FsF-F4-01M-1**: Metadata is given in a way major search engines can ingest it for their catalogs.	0, 1
		FsF-F4-01M-2: Metadata is registered in major research data registries (DataCite (DataCite 2022))	0, 1

Continued on next page

Table 4.1 – continued from previous page

FAIR sub-principle	Metric	Practical test	Score value
A1. (Meta)data are retrievable by their identifier using a standardized communications protocol	**FsF-A1-01M**: Metadata contains access level and access conditions of the data. Possible score: 0, 1	**FsF-A1-01M-1**: Information about access restrictions or rights can be identified in metadata	0, 0.5
		FsF-A1-01M-2: Data access information is machine readable. This is verified against controlled vocabularies, e.g., COAR[5].	0, 1
		FsF-A1-01M-3: Data access information is indicated by (not machine readable) standard terms.	0, 1
	FsF-A1-02M: Metadata is accessible through a standardized communication protocol	**FsF-A1-02M-1**: Landing page link is based on standardized web communication protocols.	0, 1
	FsF-A1-03D: Data is accessible through a standardized communication protocol. Possible score: 0, 1	**FsF-A1-03D-1**: Metadata includes a resolvable link to data based on standardized web communication protocols.	0, 1
I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	**FsF-I1-01M**: Metadata is represented using a formal knowledge representation language. Possible score: 0, 1, 2	**FsF-I1-01M-1**: Parsable, structured metadata is embedded in the landing page code.	0, 1
		FsF-I1-01M-2: Parsable, graph data is accessible through content negotiation, typed links, or SPARQL endpoint	0, 1
I2. (Meta)data use vocabularies that follow FAIR principles	**FsF-I2-01M**: Metadata uses semantic resources. Possible score: 0, 1	**FsF-I2-01M-2**: Namespaces of known semantic resources can be identified in metadata	0, 1

Continued on next page

[5]Controlled Vocabularies for Repositories: https://vocabularies.coar-repositories.org, date accessed: 03/12/2024

Table 4.1 – continued from previous page

FAIR sub-principle	Metric	Practical test	Score value
I3. (Meta)data include qualified references to other (meta)data	**FsF-I3-01M**: Metadata includes links between the data and its related entities. Possible score: 0, 1	**FsF-I3-01M-1**: Related resources are explicitly mentioned in metadata.	0, 1
		FsF-I3-01M-2: Related resources are indicated by machine readable links or identifiers.	0, 1
R1. (Meta)data are richly described with a plurality of accurate and relevant attributes	**FsF-R1-01MD**: Metadata specifies the content of the data. Possible score: 0, 1, 2, 3, 4	**FsF-R1-01MD-1**: Minimal information about available data content is given in metadata	0, 1
		FsF-R1-01MD-2: Verifiable data descriptors (file info, measured variables or observation types) are specified in metadata	0, 1
		FsF-R1-01MD-3: Data content matches file type and size specified in metadata	0, 1
		FsF-R1-01MD-4: Data content matches measured variables or observation types specified in metadata	0, 1
R1.1. (Meta)data are released with a clear and accessible data usage license	**FsF-R1.1-01M**: Metadata includes license information under which data can be reused. Possible score: 0, 1, 2	**FsF-R1.1-01M-1**: License information is given in an appropriate metadata element	0, 1
		FsF-R1.1-01M-2: Recognized license is valid (community specific or registered at SPDX)	0, 1

Continued on next page

Table 4.1 – continued from previous page

FAIR sub-principle	Metric	Practical test	Score value
R1.2. (Meta)data are associated with detailed provenance[6]	**FsF-R1.2-01M**: Metadata includes provenance information about data creation or generation[7]. Possible score: 0, 1, 2	**FsF-R1.2-01M-1**: Metadata contains elements which hold provenance information and can be mapped to PROV	0, 1
		FsF-R1.2-01M-2: Metadata contains provenance information using formal provenance ontologies	0, 1
R1.3. (Meta)data meet domain-relevant community standards	**FsF-R1.3-01M**: Metadata follows a standard recommended by the target research community of the data. Possible score: 0, 1	**FsF-R1.3-01M-1**: Community specific metadata standard is detected using namespaces or schema found in provided metadata or metadata services outputs.	0, 1
		FsF-R1.3-01M-2: Community specific metadata standard is listed in the re3data[8] record of the responsible repository.	0, 1
		FsF-R1.3-01M-3: Multidisciplinary but community endorsed metadata, such as fairsharing (Sansone et al. 2019) standard is listed in the re3data record or detected by namespace.	0, 1
	FsF-R1.3-02D: Data is available in a file format recommended by the target research community. Possible scores: 0, 1	**FsF-R1.3-02D-1**: The format of a data file given in the metadata is listed in the long term file formats, open file formats, or scientific file formats controlled list	0, 1

[6]Refers to the individuals, entities, and processes involved in data creation, management, and long-term curation.

[7]Includes creator, contributors, creation and modification dates, version, source, and relationships indicating data creation activities (Devaraju and Huber 2020).

[8]REgistry of REsearch Data REpositories: https://www.re3data.org/#, date accessed: 03/12/2024.

4.3 FAIRness Assessment in FAIR Evaluator

FAIR Evaluator uses 22 maturity indicator compliance tests to assess FAIRness of digital resources (8 for Findable, including 3 for F1, 2 for F2, 2 for F3, 1 for F4; 5 for Assessable, including 2 for A1.1, 2 for A1.2, 1 for A2; 5 for Interoperable, including 4 for I1, 2 for I2, 1 for I3; 2 for Reusable, including 2 for R1.1). Table 4.2 includes FAIR Evaluator's maturity indicators compliance tests, their descriptions, score values, and how measured, organized based on their corresponding FAIR principles. The output of each test performed by FAIR Evaluator is either failure (0) or success (1) and when running the tests on the online user interface, it also gives the number of successes and failures out of the 22 tests.

Table 4.2: FAIRness assessment indicator compliance tests and explanations of their operation in FAIR Evaluator.

FAIR Sub-principle	Maturity indicator compliance test	Score value
F1	**FAIR Metrics Gen2- Unique Identifier**: Metric to test if the metadata resource has a unique identifier. This is done by comparing the GUID to the patterns (by regexp) of known GUID schema such as URLs and DOIs. Known schema are registered in FAIRsharing[9]	0, 1
	FAIR Metrics Gen2 - Identifier Persistence: Metric to test if the unique identifier of the metadata resource is likely to be persistent. Known schema are registered in FAIRsharing. For URLs that don't follow a schema in FAIRsharing we test known URL persistence schema, such as purl[10].	0, 1
	FAIR Metrics Gen2 - Data Identifier Persistence: Metric to test if the unique identifier of the data resource is likely to be persistent. Known schema are registered in FAIRsharing. For URLs that don't follow a schema in FAIRsharing we test known URL persistence schema.	0, 1
F2	**FAIR Metrics Gen2 - Structured Metadata**: Tests whether a machine is able to find structured metadata, such as RDF Turtle[11]	0, 1

Continued on next page

[9]https://fairsharing.org/standards/?q=&selected_facets=type_exact:identifier%20schema, date accessed: 25/11/2024.

[10]Persistent URL: https://www.opengis.net/docs/index.html, date accessed: 04/12/2024.

[11]A textual syntax for RDF: https://www.w3.org/TR/turtle/, date accessed: 04/12/2024.

Table 4.2 – continued from previous page

FAIR Sub-principle	Maturity indicator compliance test	Score value
	FAIR Metrics Gen2 - Grounded Metadata: Tests whether a machine is able to find 'grounded' metadata. i.e. metadata terms that are in a resolvable namespace, where resolution leads to a definition of the meaning of the term, such as any form of RDF.	0, 1
F3	**FAIR Metrics Gen2 - Data Identifier Explicitly In Metadata**: Metric to test if the metadata contains the unique identifier to the data. This is done by searching for a variety of properties.	0, 1
	FAIR Metrics Gen2- Metadata Identifier Explicitly In Metadata: Metric to test if the metadata contains the unique identifier to the metadata itself.	0, 1
F4	**FAIR Metrics Gen2 - Searchable in major search engine**: Tests whether a machine is able to discover the resource by search, using Microsoft Bing.	0, 1
A1.1	**FAIR Metrics Gen2 - Uses open free protocol for data retrieval**: Data may be retrieved by an open and free protocol. Tests data GUID for its resolution protocol.	0, 1
	FAIR Metrics Gen2 - Uses open free protocol for metadata retrieval: Metadata may be retrieved by an open and free protocol. Tests metadata GUID for its resolution protocol.	0, 1
A1.2	**FAIR Metrics Gen2 - Data authentication and authorization**: Test a discovered data GUID for the ability to implement authentication and authorization in its resolution protocol. It also searches the metadata for the Dublin Core 'accessRights' property, which may point to a document describing the data access process. Recognition of other identifiers will be added upon request by the community.	0, 1
	FAIR Metrics Gen2 - Metadata authentication and authorization: Tests metadata GUID for the ability to implement authentication and authorization in its resolution protocol.	0 or 1
A2	**FAIR Metrics Gen2 - Metadata Persistence**: Metric to test if the metadata contains a persistence policy, explicitly identified by a http://www.w3.org/2000/10/swap/pim/doc#persistencePolicy predicate in Linked Data.	0, 1

Continued on next page

Table 4.2 – continued from previous page

FAIR Sub-principle	Maturity indicator compliance test	Score value
I1	**FAIR Metrics Gen2 - Metadata Knowledge Representation Language (weak)**: Maturity Indicator to test if the metadata uses a formal language broadly applicable for knowledge representation. This particular test takes a broad view of what defines a 'knowledge representation language'; in this evaluation, anything that can be represented as structured data will be accepted	0, 1
	FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong): Maturity Indicator to test if the metadata uses a formal language broadly applicable for knowledge representation. This particular test takes a broad view of what defines a knowledge representation language. In this evaluation, a knowledge representation language is interpreted as one in which terms are semantically-grounded in ontologies. Any form of RDF will pass this test.	0, 1
	Gen2 Data Knowledge Representation Language (Weak): Maturity Indicator to test if the data uses a formal language broadly applicable for knowledge representation. In this evaluation, a knowledge representation language is interpreted as one in which terms are semantically-grounded in ontologies. Any form of structured data will pass this test.	0, 1
	Gen2 Data Knowledge Representation Language (Strong): Maturity Indicator to test if the data uses a formal language broadly applicable for knowledge representation. In this evaluation, a knowledge representation language is interpreted as one in which terms are semantically-grounded in ontologies. Any form of ontologically-grounded linked data will pass this test.	0, 1
I2	**FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)**: Maturity Indicator to test if the linked data metadata uses terms that resolve. This tests only if they resolve, not if they resolve to FAIR data, therefore is a somewhat weak test.	0, 1
	FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (strong): Maturity Indicator to test if the linked data metadata uses terms that resolve to linked (FAIR) data.	0, 1
I3	**FAIR Metrics Gen2 - Metadata contains qualified outward references**: Maturity Indicator to test if the metadata links outward to third-party resources. It only tests metadata that can be represented as Linked Data.	0, 1
R1.1	**FAIR Metrics Gen2 - Metadata Includes License (strong)**: Maturity Indicator to test if the linked data metadata contains an explicit pointer to the license.	0, 1

Continued on next page

Table 4.2 – continued from previous page

FAIR Sub-principle	Maturity indicator compliance test	Score value
	FAIR Metrics Gen2 - Metadata Includes License (weak): Maturity Indicator to test if the metadata contains an explicit pointer to the license. This 'weak' test will use a case-insensitive regular expression, and scan both key/value style metadata, as well as linked data metadata.	0, 1

4.4 FAIRness Assessment in FAIR-Checker

FAIR-Checker uses 12 metrics to assess the FAIRness of digital resources (4 for Findable, including 2 for F1, 2 for F2; 2 for Assessable, including 1 for A1.1, 1 for A1.2; 3 for Interoperable, including 1 for I1, 1 for I2, 1 for I3; 3 for Reusable, including 1 for R1.1, 1 for R1.2, 1 for R1.3). Table 4.3 includes FAIR-Checker's metrics, their descriptions, score values and how measured, and is organized based on their corresponding FAIR principles. FAIR-Checker gives the result of each metric (0, 1, or 2) and provides a whole percentage score which is obtained by dividing the whole score obtained by 24 (the maximum possible score).

Table 4.3: FAIRness assessment Metrics and explanations of their operation in FAIR-Checker.

FAIR Sub-principle	Metric	Score value
F1	**F1A - Unique IDs**: Checks if the identifier can be reached with an HTTP or HTTPs request. It's better if the URL is persistent (PURL or DOI).	0, 1, 2
	F1B - Persistent IDs *Strong: FAIR-Checker verifies that the identifier property from DCTerms (dct:identifier) or Schema.org (schema:identifier) vocabularies is present in metadata. *Weak: Checking that at least one namespace from identifiers.org (life-science oriented registry) is in metadata. This identifier can be either the URL itself or encoded in the metadata as a dct:identifier or schema:identifier property.	0, 1, 2

Continued on next page

Table 4.3 – continued from previous page

FAIR Sub-principle	Metric	Score value
F2	**F2A - Structured metadata** *Strong: Checking that at least one of the access policy properties (dct:title, dct:description, dcat:accessURL, dcat:downloadURL, dcat:endpointDescription, dcat:endpointURL) is found in metadata. *Weak: Verifies that at least one RDF triple can be found in metadata. Structured metadata should be embedded as machine readable content into the HTML file. A variety of RDF-compliant options are available, such as RDFa[12], HTML Microdata[13], and JSON-LD[14].	0, 1, 2
	F2B - Shared vocabularies for metadata Strong: Checking if all classes used in RDF are known in OLS[15], LOV[16], or BioPortal[17] Weak: Checking if at least one class used in RDF is known in OLS, LOV, or BioPortal	0, 1, 2
A1.1	**A1.1 - Open resolution protocol**: Checking if the the resource is accessible via an open protocol, for instance the protocol needs to be HTTP.	0, 1, 2
A1.2	**A1.2 - Authorization procedure or access rights**: Checking that at least one of the access policy properties, i.e., odrl:hasPolicy, dct:rights, dct:accessRights, dct:license, or schema:license is found in metadata.	0, 1, 2
I1	**I1 - Machine readable format** *Strong: Checking that at least one of the discoverability properties, i.e., dct:title, dct:description, dcat:accessURL, dcat:downloadURL, dcat:endpointDescription, dcat:endpointURL is found in metadata. *Weak: Checking if data is structured, looking for at least one RDF triple.	0, 1, 2

Continued on next page

[12] https://www.w3.org/MarkUp/2009/rdfa-for-html-authors, date accessed: 26/11/2024
[13]https://www.w3.org/TR/2021/NOTE-microdata-20210128/, date accessed: 04/12/2024.
[14]https://www.w3.org/TR/json-ld11/#relationship-to-rdf, date accessed: 04/12/2024.
[15]Ontology Lookup Service: https://www.ebi.ac.uk/ols4, date accessed: 04/12/2024.
[16]Linked Open Vocabularies: https://lov.linkeddata.es/dataset/lov/, date accessed: 04/12/2024.
[17]A comprehensive repository of biomedical ontologies: https://bioportal.bioontology.org, date accessed: 04/12/2024.

Table 4.3 – continued from previous page

FAIR Sub-principle	Metric	Score value
I2	**I2 - Use shared ontologies** Weak: FAIR-Checker verifies that at least one used ontology class or property are known in major ontology registries (OLS, BioPortal, LOV) Strong: FAIR-Checker verifies that all used ontology classes or properties are known in major ontology registries (OLS, BioPortal, LOV)	0, 1, 2
I3	**I3 - External links**: Checking that at least 3 different URL authorities are used in the URIs of RDF metadata.	0, 1, 2
R1.1	**R1.1 - Metadata includes license**: Checking that at least one of the following license properties is found in metadata: schema:license, dct:license, doap:license, dbpedia-owl:license, cc:license, xhv:license, sto:license, nie:license.	0, 1, 2
R1.2	**R1.2 - Metadata includes provenance**: Checking that at least one of the following provenance properties is found in metadata: prov:wasGeneratedBy, prov:wasDerivedFrom, prov:wasAttributedTo, prov:used, prov:wasInformedBy, prov:wasAssociatedWith, prov:startedAtTime, prov:endedAtTime, dct:hasVersion, dct:isVersionOf, dct:creator, dct:contributor, dct:publisher, pav:hasVersion, pav:version, pav:hasCurrentVersion, pav:createdBy, pav:authoredBy, pav:retrievedFrom, pav:importedFrom, pav:createdWith, pav:retrievedBy, pav:importedBy, pav:curatedBy, pav:createdAt pav:previousVersion, schema:creator, schema:author, schema:publisher, schema:provider, schema:funder.	0, 1, 2
R1.3	**R1.3 - Community standards** *Weak: Checking that at least one used ontology class or property are known in major ontology registries (OLS, BioPortal, LOV) *Strong: FAIR-Checker verifies that all used ontology classes or properties are known in major ontology registries (OLS, BioPortal, LOV)	0, 1, 2

4.5 Comparative Analysis of FAIRness Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

As discussed in Sections 4.2, 4.3, and 4.4, F-UJI, FAIR Evaluator, and FAIR-Checker employ different measures for assessing FAIRness. Currently, there are no established criteria for selecting the most suitable FAIRness assessment tool. As a result, analyzing the FAIRness of a knowledge graph (KG) can be challenging and may lead to confusion when interpreting the results provided by different tools. In this section, a comparative analysis of FAIRness assessment methodologies in F-UJI, FAIR Evaluator, and FAIR-Checker is presented. This analysis aims to provide a clearer understanding of the assessment procedures and the differences among these tools. The comparison begins with an examination of the inputs required by each tool, followed by an evaluation of their respective FAIRness assessment measures.

Table 4.4 provides an overview of the suitable inputs for each tool. As shown in the table, all tools require a unique identifier for the resource's metadata, which is then used to evaluate the resource's FAIRness based on the metadata contents.

Table 4.4: Suitable inputs for FAIRness assessment tools: F-UJI, FAIR Evaluator, and FAIR-Checker.

Tool	Input
F-UJI	The unique identifier of the data object to be evaluated and if available, the repository's metadata provision service
FAIR Evaluator	The GUID of the metadata of the data
FAIR-Checker	A web page URL

Table 4.5 presents the percentage distribution of data-related, metadata-related, and data-and-metadata-related metrics in F-UJI, FAIR Evaluator, and FAIR-Checker. As indicated in the table, the measures of all three tools are predominantly focused on the provision of specific machine-readable metadata properties within the resource's metadata, with only a small percentage addressing the resource itself or a combination of the resource and its metadata.

Table 4.6 presents the metrics employed by F-UJI, FAIR Evaluator, and FAIR-

Table 4.5: Percentage of data, metadata, and data-metadata-related metrics in F-UJI, FAIR Evaluator, and FAIR-Checker.

Tool	**Data-related metrics (%)**	**Metadata-related metrics (%)**	**Data-and-metadata-related metrics (%)**
F-UJI	15.15%	72.73%	12.12%
FAIR-Checker	8.30%	91.70%	0%
FAIR Evaluator	25%	75%	0%

Checker, aligned with their corresponding FAIR sub-principles. The metrics are displayed side by side to facilitate comparison.

Table 4.6: FAIRness measures employed in F-UJI, FAIR Evaluator, and FAIR-Checker, aligned with their corresponding FAIR sub-principles.

FAIR Sub-principle	**F-UJI**	**FAIR Evaluator**	**FAIR-Checker**
F1	FsF-F1-01D (practical tests: (FsF-F1-01D-1, FsF-F1-01D-2)): (meta)data is assigned a GUID	FAIR Metrics Gen2- Unique Identifier	Metric: F1A - Unique IDs
	FsF-F1-02D (practical tests: FsF-F1-02D-1, FsF-F1-02D-2): (meta)data is assigned a PID	FAIR Metrics Gen2 - Identifier Persistence	Metric: F1B - Persistent IDs
	-	FAIR Metrics Gen2 - Data Identifier Persistence	-
F2	FsF-F2-01M (Practical tests: FsF-F2-01M-1, FsF-F2-01M-2, FsF-F2-01M-3): metadata includes descriptive core elements to support data findability	FAIR Metrics Gen2 - Structured Metadata	Metric: F2A - Structured metadata
	-	FAIR Metrics Gen2 - Grounded Metadata	Metric: F2B - Shared vocabularies for metadata

Continued on next page

Table 4.6 – continued from previous page

FAIR Sub-principle	**F-UJI**	**FAIR Evaluator**	**FAIR-Checker**
F3	FsF-F3-01M (practical tests: FsF-F3-01M-1, FsF-F3-01M-2): metadata includes the identifier of the data it describes	FAIR Metrics Gen2 - Data Identifier Explicitly In Metadata	N/A
	-	FAIR Metrics Gen2- Metadata Identifier Explicitly In Metadata	N/A
F4	FsF-F4-01M (practical tests: FsF-F4-01M-1, FsF-F4-01M-2): metadata is offered in such a way that it can be retrieved by machines	FAIR Metrics Gen2 - Searchable in major search engine	N/A
A1	FsF-A1-01M (practical tests: FsF-A1-01M-1, FsF-A1-01M-3, FsF-A1-01M-2): metadata contains access level and access conditions of the data	N/A	N/A
	FsF-A1-02M (practical test: FsF-A1-02M-1): metadata is accessible through a standardized communication protocol	N/A	N/A
	FsF-A1-03D (practical test: FsF-A1-03D-1): data is accessible through a standardized communication protocol	N/A	N/A
A1.1	N/A	FAIR Metrics Gen2 - Uses open free protocol for data retrieval	Metric: A1.1 - Open resolution protocol
	N/A	FAIR Metrics Gen2 - Uses open free protocol for metadata retrieval	-

Continued on next page

Table 4.6 – continued from previous page

FAIR Sub-principle	**F-UJI**	**FAIR Evaluator**	**FAIR-Checker**
A1.2	N/A	FAIR Metrics Gen2 - Data authentication and authorization	Metric: A1.2 - Authorisation procedure or access rights
	N/A	FAIR Metrics Gen2 - Metadata authentication and authorization	-
A2	N/A	FAIR Metrics Gen2 - Metadata Persistence	N/A
I1	FsF-I1-01M (practical tests: FsF-I1-01M-1, FsF-I1-01M-2): metadata is represented using a formal knowledge representation language	FAIR Metrics Gen2 - Metadata Knowledge Representation Language (weak)	Metric: I1 - Machine readable format
	-	FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong)	-
I2	FsF-I2-01M (practical tests: FsF-I2-01M-1, FsF-I2-01M-2): metadata uses semantic resources	FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)	Metric: I2 - Use shared ontologies
	-	FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (strong)	-
I3	FsF-I3-01M (practical tests: FsF-I3-01M-1, FsF-I3-01M-2): metadata includes links between the data and its related entities	FAIR Metrics Gen2 - Metadata contains qualified outward references)	Metric: I3 - External links

Continued on next page

Table 4.6 – continued from previous page

FAIR Sub-principle	F-UJI	FAIR Evaluator	FAIR-Checker
R1	FsF-R1-01MD (practical tests: FsF-R1-01MD-1, FsF-R1-01MD-2, FsF-R1-01MD-3, FsF-R1-01MD-4): metadata specifies the content of the data	N/A	N/A
R1.1	FsF-R1.1-01M (practical tests: FsF-R1.1-01M-1, FsF-R1.1-01M-2): metadata includes license information under which data can be reused	FAIR Metrics Gen2 - Metadata Includes License (strong)	Metric: R1.1 - Metadata includes license
	-	FAIR Metrics Gen2 - Metadata Includes License (weak)	-
R1.2	FsF-R1.2-01M (practical tests: FsF-R1.2-01M-1, FsF-R1.2-01M-2): metadata includes provenance information about data creation or generation	N/A	Metric: R1.2 - Metadata includes provenance
R1.3	FsF-R1.3-01M (practical tests: FsF-R1.3-01M-1, FsF-R1.3-01M-2, FsF-R1.3-01M-3): metadata follows a standard recommended by the target research community of the data	N/A	Metric: R1.3 - Community standards
	FsF-R1.3-02D (practical test: FsF-R1.3-02D-1): data is available in a file format recommended by the target research community	N/A	-

As is evident from Table 4.6, the three tools assess different sets of FAIR sub-principles. Figure 4.1 visualizes the hierarchy of FAIRness assessment measures used in each technique per FAIR sub-principles, while Figure 4.2 illustrates the number of metrics used by each tool to assess each FAIR sub-principle. According to the figures, A1[18] and R1[19] are only assessed by F-UJI, and FAIR Evaluator is the only tool among these three that assesses A2[20]. A more detailed comparative analysis has been provided in Subsection 4.5.1 to Subsection 4.5.15.

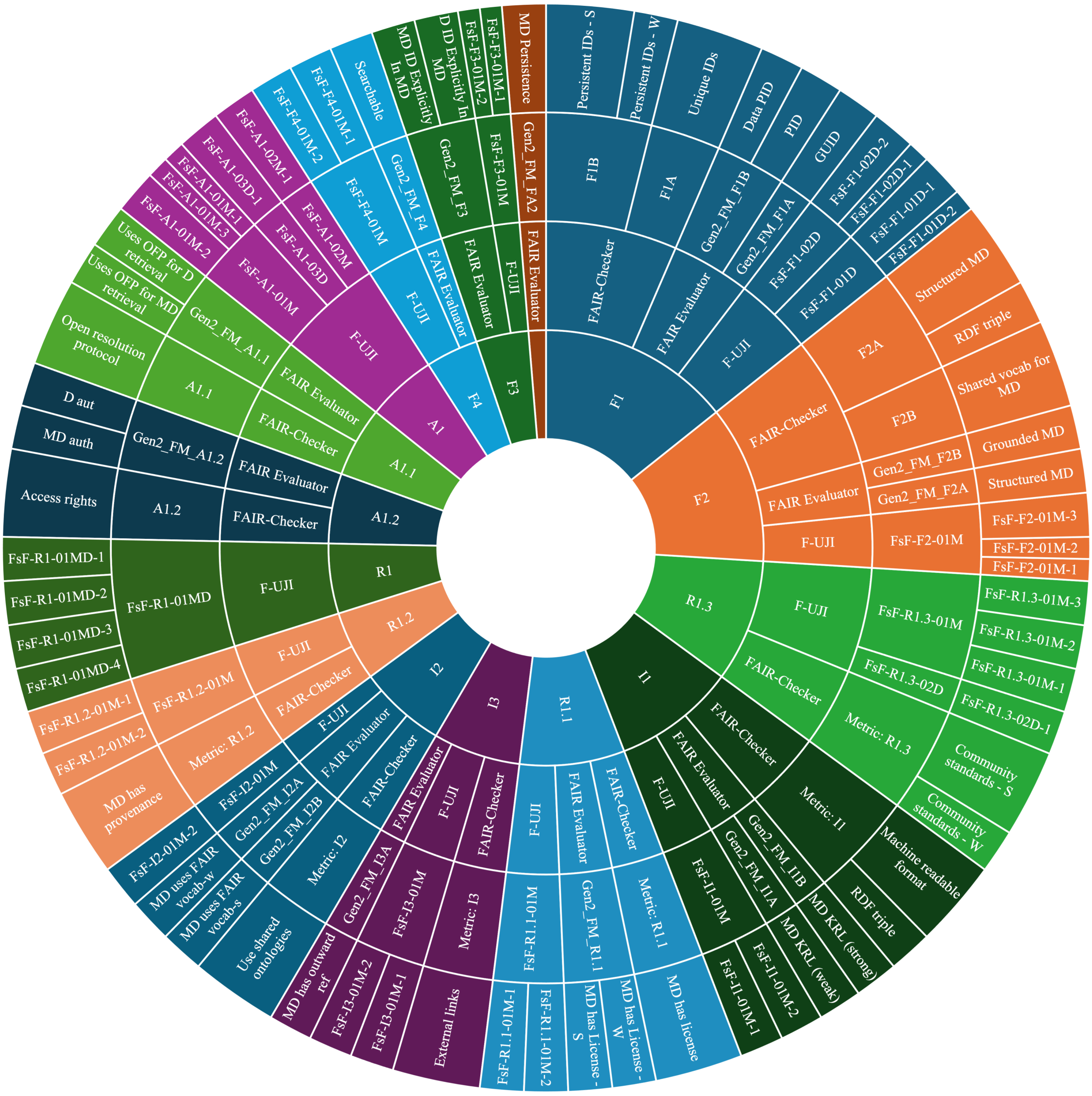


Figure 4.1: F-UJI, FAIR Evaluator, and FAIR-Checker metrics comparison – Hierarchical chart (Zoom in to see the details.).

While the three tools differ in the sub-principles they cover, there is currently no established evidence in the literature suggesting that certain FAIR sub-principles are universally more important than others. As such, a tool's omission of specific sub-principles, such as FAIR-Checker's exclusion of R1.3, is considered a limitation

[18](Meta)data are retrievable by their identifier using a standardized communications protocol.
[19](Meta)data are richly described with a plurality of accurate and relevant attributes.
[20]Metadata are accessible, even when the data are no longer available.

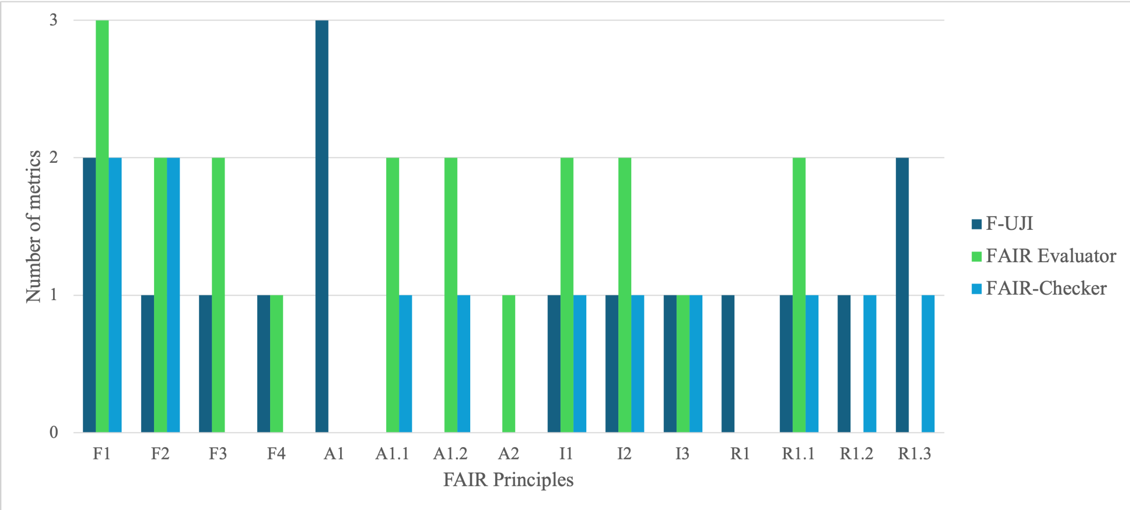


Figure 4.2: The number of metrics per principle in F-UJI, FAIR Evaluator, and FAIR-Checker.

of the tool rather than a reflection of the sub-principle's relevance. Ideally, a comprehensive FAIRness assessment should address all sub-principles to ensure balanced and complete evaluation.

4.5.1 Comparative Analysis of F1 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI, FAIR Evaluator, and FAIR-Checker share commonalities in assessing FAIR principle F1, focusing on globally unique and persistent identifiers. All three prioritize evaluating identifiers for compliance with global standards, such as DOIs and URLs, while testing their persistence and alignment with recognized schema. Each uses discrete scoring systems (e.g., 0, 0.5, or 1) to ensure consistent and comparable assessments across datasets.

Significant differences exist in the approaches and implementation of F-UJI, FAIR Evaluator, and FAIR-Checker. F-UJI adopts a granular approach, using multiple metrics to separately evaluate identifier syntax and resolvability, allowing for intermediate scores that reflect partial compliance. FAIR Evaluator uses broader metrics to assess uniqueness and persistence, aligning identifiers with schema in FAIRsharing and persistence standards like purl and w3id. However, its binary scoring (0 or 1) does not distinguish partial FAIRness compliance. FAIR-Checker integrates domain-specific contexts, i.e., life sciences and employs a wider scoring range (0, 1, or 2) to differentiate between weak and strong compliance. It also emphasizes practical tests, such as HTTP reachability, and verifies metadata properties for persistence, enhancing applicability in specific domains.

These differences highlight the unique strengths of each tool, with F-UJI excelling in detailed and nuanced assessments, FAIR Evaluator focusing on recognized standards and binary evaluations, and FAIR-Checker offering domain-specific insights and flexible scoring mechanisms.

4.5.2 Comparative Analysis of F2 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI, FAIR Evaluator, and FAIR-Checker adopt distinct methodologies for assessing F2. F-UJI employs a detailed approach, emphasizing descriptive core elements, compliance with metadata standards, and practical tests to evaluate metadata comprehensiveness. FAIR Evaluator prioritizes machine-actionable metadata and semantic grounding, ensuring metadata is interpretable through resolvable namespaces and structured formats like JSON-LD. FAIR-Checker incorporates domain-specific evaluations, distinguishing between strong and weak compliance levels for structured metadata and shared vocabularies, leveraging repositories like OLS, LOV, and BioPortal.

Despite these differences, all tools share a common focus on structured, machine-readable, and semantically rich metadata to enhance findability. The approaches vary in granularity, with F-UJI emphasizing descriptive metadata, FAIR Evaluator focusing on machine-readability, and FAIR-Checker adopting a domain-specific perspective. These distinctions highlight diverse ways to assess F2, allowing for tailored evaluations depending on the context. Additionally, they suggest that a comprehensive FAIR analysis could be achieved through a combined approach.

4.5.3 Comparative Analysis of F3 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

The measurement of F3 varies significantly among the tools. F-UJI focuses on data metadata, evaluating identifiers like file name, size, type, and resolvable PIDs or URLs through specific practical tests. FAIR Evaluator takes a broader approach, assessing both data identifiers and metadata identifiers using distinct metrics and techniques such as metadata resolution via DOI. FAIR-Checker does not measure F3, leaving this sub-principle unaddressed.

F-UJI and FAIR Evaluator share an emphasis on the importance of identifiers and rely on structured evaluations with defined metrics. However, F-UJI concentrates solely on data identifiers, while FAIR Evaluator includes metadata identifiers, offering a more comprehensive analysis.

In summary, F-UJI provides a focused assessment of data identifiers, FAIR Evaluator adopts a broader scope by including metadata identifiers, and FAIR-Checker omits this sub-principle entirely. The differences reflect the variability in FAIRness assessment methodologies and the need to select tools based on specific analytical goals.

4.5.4 Comparative Analysis of F4 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F4 is assessed differently by F-UJI and FAIR Evaluator, while FAIR-Checker does not evaluate this sub-principle. F-UJI employs a comprehensive approach, evaluating metadata compatibility with machine-readability standards (e.g., JSON-LD, RDFa) and its registration in major research data registries like DataCite, through practical tests. FAIR Evaluator, in contrast, focuses solely on resource discoverability via a machine-driven search using Microsoft Bing, offering a simpler and narrower evaluation.

Both F-UJI and FAIR Evaluator aim to assess discoverability, but their methodologies differ significantly: F-UJI incorporates metadata standards and registry indexing, while FAIR Evaluator limits its assessment to search engine accessibility. FAIR-Checker's omission of F4 restricts its scope in evaluating findability. The tools' varying priorities and techniques emphasize the importance of selecting an appropriate assessment tool based on evaluation goals.

4.5.5 Comparative Analysis of A1 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI uniquely evaluates the A1 sub-principle, employing a detailed approach with three metrics to assess metadata and data retrievability through standardized protocols. Its evaluation involves practical tests to verify access levels and conditions, metadata accessibility via resolvable landing page links, and data accessibility through standardized protocols like HTTP. FAIR Evaluator and FAIR-Checker, in contrast, do not include any metrics or tests for A1, focusing on other aspects of FAIRness. This significant difference highlights a gap in their assessment frameworks. F-UJI's comprehensive focus on retrievability distinguishes it as the only tool addressing A1, offering a robust evaluation method that ensures both metadata and data meet accessibility standards.

4.5.6 Comparative Analysis of A1.1 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

FAIR Evaluator and FAIR-Checker assess the A1.1 sub-principle using different approaches, while F-UJI does not measure it. FAIR Evaluator separates data and metadata into distinct tests, resolving GUIDs via protocols like HTTP, with detailed compliance and explicit support for identifiers such as DOIs and URLs. FAIR-Checker, in contrast, consolidates the assessment into a single metric, providing broader scoring without specifying supported identifiers.

The key difference lies in granularity: FAIR Evaluator offers a detailed and explicit assessment, while FAIR-Checker focuses on simplicity and generality. Both tools emphasize accessibility via open protocols, aligning with FAIR principles, but F-UJI does not address A1.1, leaving a gap in its evaluation framework.

4.5.7 Comparative Analysis of A1.2 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

FAIR Evaluator and FAIR-Checker assess A1.2 using distinct methodologies, while F-UJI does not provide any measurement for this sub-principle. FAIR Evaluator evaluates both data and metadata, focusing on authentication and authorization in resolution protocols and the presence of access-related metadata properties. It conducts separate tests for data and metadata GUIDs, supporting standard identifiers like DOIs and Handles, and examines properties such as Dublin Core's accessRights.

FAIR-Checker, in contrast, focuses solely on metadata by identifying specific access policy properties like odrl:hasPolicy, dct:rights, and dct:accessRights. It provides a more granular scoring system based on the relevance of detected properties but does not assess authentication and authorization functionality in resolution protocols.

Both tools ensure that access is governed by proper policies and protocols and emphasize the importance of metadata in compliance assessments. However, FAIR Evaluator adopts a broader approach, while FAIR-Checker provides a more detailed analysis of metadata. F-UJI's omission of A1.2 leaves a gap in its assessment framework.

4.5.8 Comparative Analysis of A2 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

FAIR Evaluator, F-UJI, and FAIR-Checker differ in their approaches to assessing the A2 sub-principle. FAIR Evaluator evaluates metadata persistence through a dedicated metric, testing for a persistence policy and assigning a score based on its presence. In contrast, both F-UJI and FAIR-Checker exclude A2 from their assessments, with F-UJI citing the challenge of testing repository-level policies at the dataset level.

Despite these differences, all tools acknowledge the difficulty of evaluating A2 due to its reliance on repository-level policies and the limitations of automated methods. While FAIR Evaluator provides a basic mechanism for A2 measurement, F-UJI and FAIR-Checker omit it entirely, highlighting the need for standardized approaches to ensure consistent evaluation of metadata persistence.

4.5.9 Comparative Analysis of I1 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI, FAIR Evaluator, and FAIR-Checker employ distinct methodologies to assess compliance with the I1 sub-principle, though they share a focus on structured and machine-readable formats. F-UJI evaluates metadata exclusively, testing for parsable, structured formats like RDF or JSON-LD embedded in landing pages or accessible via protocols. FAIR Evaluator applies a broader scope, assessing both metadata and data through granular weak and strong maturity indicators, with strong criteria requiring ontology-grounded representations. FAIR-Checker simplifies the evaluation by combining metadata and data into a single metric, assessing machine-readable formats and RDF triples with weak and strong compliance levels.

Similarities include the emphasis on structured, machine-readable formats and the use of binary scoring for individual tests, ensuring straightforward interpretation. Differences lie in their scope and complexity: F-UJI focuses narrowly on metadata with strict format requirements, FAIR Evaluator provides detailed, flexible criteria for both metadata and data, and FAIR-Checker offers a simpler evaluation combining metadata and data into one metric. These variations reflect differing approaches, suggesting the possibility of deploying a comprehensive FAIRness assessment using a combination of approaches.

4.5.10 Comparative Analysis of I2 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

The I2 sub-principle is assessed by F-UJI, FAIR Evaluator, and FAIR-Checker using distinct methodologies. F-UJI focuses on detecting namespaces of semantic resources in metadata, providing a straightforward binary assessment (0 or 1) of vocabulary usage without evaluating term resolution or quality. FAIR Evaluator emphasizes resolution, using maturity indicators to test whether metadata terms resolve and, in stronger cases, whether they resolve to FAIR-compliant data. FAIR-Checker examines ontology classes and properties, offering a more granular assessment by verifying registry recognition and completeness, with a broader scoring range (0, 1, or 2).

Despite their differences, all tools share the objective of ensuring semantic interoperability by verifying compliance with FAIR principles, relying on registries or resolvable terms for validation. F-UJI's approach is simpler, focusing on namespaces, FAIR Evaluator highlights term resolution, and FAIR-Checker evaluates ontology completeness and registry recognition. Together, these tools offer complementary perspectives.

4.5.11 Comparative Analysis of I3 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI, FAIR Evaluator, and FAIR-Checker assess I3 compliance using distinct yet complementary techniques. F-UJI focuses on explicit and machine-readable links in metadata, with practical tests verifying the presence of explicitly mentioned related resources and ensuring they are actionable. FAIR Evaluator emphasizes Linked Data compatibility by examining outward references to third-party resources in Linked Data-compatible metadata. FAIR-Checker evaluates the diversity of URL authorities within RDF metadata, requiring at least three distinct authorities to highlight link diversity.

Despite these differences, all tools aim to assess metadata interconnectedness, a crucial aspect of FAIRness. F-UJI prioritizes explicitness and machine readability, FAIR Evaluator focuses on semantic web compatibility, and FAIR-Checker highlights link diversity. Together, these approaches provide a comprehensive evaluation of I3 compliance, addressing different facets of metadata interconnection.

4.5.12 Comparative Analysis of R1 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI is the only tool that assesses the R1 sub-principle, employing a detailed evaluation framework. It uses practical tests to verify metadata richness by checking for minimal information, verifiable data descriptors, and alignment between metadata and data attributes, with scores ranging from 0 to 4. FAIR Evaluator and FAIR-Checker, in contrast, do not include metrics or tests for R1, focusing their methodologies on aspects such as metadata accessibility and identification, leaving metadata richness unaddressed. Despite their shared reliance on metadata for FAIRness evaluation, F-UJI uniquely provides a comprehensive assessment framework for R1, aligning closely with its requirements and offering insights into descriptive richness that the other tools do not address.

4.5.13 Comparative Analysis of R1.1 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI, FAIR Evaluator, and FAIR-Checker assess R1.1 using distinct methodologies. F-UJI employs a single metric to check for license information in metadata and verify its validity, such as alignment with SPDX or community-specific standards. Its scoring is granular, allowing values of 0, 1, or 2, providing a detailed evaluation. FAIR Evaluator uses two "Maturity Indicators," with a strong test for explicit license pointers validated against linked data predicates and a weak test using regular expressions to detect licenses in key-value or linked data formats. Scores are binary (0 or 1). FAIR-Checker, on the other hand, detects license properties from a predefined list (e.g., schema:license, dct:license) and scores based on the number of properties identified, without validating their content.

All three tools focus on identifying license information in metadata and use recognized vocabularies to guide their assessments. However, they differ in approach: F-UJI prioritizes license validity, FAIR Evaluator distinguishes between explicit and implicit licenses through strong and weak tests, and FAIR-Checker adopts a simpler property-detection method. Despite these differences, the shared goal is to ensure license information is present and accessible in metadata.

4.5.14 Comparative Analysis of R1.2 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

The tools differ significantly in their approaches to R1.2. F-UJI offers a detailed and structured assessment with clearly defined tests and scoring, making it the most comprehensive tool for evaluating provenance. FAIR-Checker provides only a general acknowledgment of provenance without detailed criteria or scoring, while FAIR Evaluator omits R1.2 entirely, leaving a gap in its coverage. These differences underscore the varying depth and focus of the tools in assessing provenance effectively.

4.5.15 Comparative Analysis of R1.3 Assessment in F-UJI, FAIR Evaluator, and FAIR-Checker

F-UJI and FAIR-Checker employ distinct approaches to assess R1.3, while FAIR Evaluator does not measure this sub-principle. F-UJI utilizes a detailed methodology, detecting community-specific metadata standards and validating compliance with catalogs such as FAIRsharing and the RDA Metadata Standards Catalog. In contrast, FAIR-Checker focuses on ontology usage, conducting ”weak” evaluations to check if at least one ontology class or property is recognized in major registries like OLS, BioPortal, or LOV, and ”strong” evaluations to ensure all are recognized.

The primary difference between the tools lies in their scope and granularity. F-UJI provides a comprehensive, multi-step evaluation covering both metadata standards and file formats, making it suitable for detailed assessments but requiring extensive metadata. FAIR-Checker offers a simpler, ontology-centered approach that is less resource-intensive. Their scoring systems also differ, with F-UJI assigning binary scores for each test and FAIR-Checker using a broader range of 0, 1, or 2 based on weak and strong evaluations. A similarity between the two tools is their reliance on external registries to validate compliance, though their focus and registries differ.

In summary, F-UJI delivers a more structured and detailed evaluation, while FAIR-Checker provides a simpler assessment focused on ontologies. The absence of R1.3 assessment in FAIR Evaluator highlights a gap in its methodology, underscoring the varied focus and capabilities of these tools.

4.6 Concluding Remarks

In this chapter, a comparative analysis of three automated FAIRness assessment tools—F-UJI, FAIR Evaluator, and FAIR-Checker—was presented. Their inputs, outputs, and metrics were examined, and key findings were inferred. This analysis reveals that while all tools address critical aspects of FAIRness, none provides complete coverage of all sub-principles. For example, A2 is assessed solely by FAIR Evaluator, as F-UJI removed it due to its perceived impracticality for automatic evaluation based solely on data and metadata. Similarly, R1.3 is measured only by F-UJI and FAIR-Checker, with FAIR Evaluator omitting this sub-principle. These differences underscore the distinct focus and limitations of each tool.

F-UJI offers the most detailed evaluations, particularly excelling in its multi-step assessments of metadata standards and file formats. FAIR Evaluator demonstrates robust coverage of several sub-principles, such as A1.1 and I1, with its inclusion of strong and weak evaluations. FAIR-Checker provides a simpler, ontology-based assessment, particularly for sub-principles like R1.3. A notable similarity between F-UJI and FAIR-Checker is their reliance on external registries to validate metadata compliance, although their focus and selected registries differ. FAIR Evaluator, while addressing a broad range of sub-principles, omits certain aspects, such as R1.3, where its methodology could benefit from expansion.

In conclusion, these findings highlight that the selection of a FAIRness assessment tool may depend on the depth and focus required for the analysis. F-UJI is well-suited for detailed evaluations of metadata standards and file formats, while FAIR-Checker offers a simpler, ontology-based perspective. FAIR Evaluator, despite certain limitations, provides valuable insights across several principles, particularly through its structured approach to metadata and data accessibility. Nevertheless, the complementary nature of these tools indicates that a combined or enhanced methodology is advisable to achieve comprehensive and consistent coverage of the FAIR principles. In Chapter 6, further analysis will be conducted to explore this prospect in greater detail.

Chapter 5

Data Collection

5.1 Chapter Overview

In order to evaluate automatic FAIRness assessment, appropriate open-access knowledge graph datasets must first be identified. Various sources provide knowledge graphs, such as KG-Hub[1](Caufield et al. 2023), which is a collection of biological and biomedical knowledge graphs, the OpenAIRE Graph[2] (Manghi et al. 2022), a collection of research products linked together, and the European Union Open Data Portal[3] (European Union 2024), a central access point to European open data. However, these sources are focused on specific fields of knowledge and are not representative of a broad range of domains.

To include a diverse set of knowledge graphs covering both general and domain-specific knowledge, the Linked Open Data Cloud[4] (Jentzsch, Cyganiak, and Bizer 2011; Schmachtenberg, Bizer, and Paulheim 2014) was selected as the dataset source. This chapter outlines the Linked Open Data Cloud, including an explanation of its suitability, and is structured as follows: Section 5.2 discusses the rationale for selecting the LOD Cloud as the data source for this research. With the LOD Cloud established as the foundation, Section 5.3 explores its metadata underpinnings. Next, the

[1]https://kghub.org, date accessed: 18th November 2024.
[2]https://graph.openaire.eu/, date accessed: 18th November 2024.
[3]https://data.europa.eu/en, date accessed: 18th November 2024.
[4]https://lod-cloud.net, date accessed: 25th June 2024.

process for selecting appropriate metadata links for FAIRness assessment is detailed in Section 5.4. Finally, Section 5.6 presents the FAIRness assessment process for the LOD Cloud.

5.2 Overview of the LOD Cloud

The collection of publicly available knowledge graphs on the web is commonly known as the Web of Linked Data (Bizer, Heath, and Berners-Lee 2023) or the Linked Open Data (LOD) Cloud (Caraballo et al. 2016). It serves as a global information space rich with structured data (Akhtar et al. 2020). Since 2007, the Web of Linked Data has expanded from a dozen datasets to a data space containing over a thousand datasets. The central idea of Linked Data is that data publishers facilitate data discovery and integration by adhering to best practices (Heath and Bizer 2011) in three key areas (Schmachtenberg, Bizer, and Paulheim 2014):

1. Linking: By creating RDF links, publishers connect datasets into a global data graph, enabling applications to navigate and discover new data through these links.

2. Vocabulary Usage: Publishers should use terms from widely-used vocabularies to improve data interpretation. When using proprietary vocabularies, terms should link back to RDF schema or OWL definitions, with references to common vocabularies for consistency.

3. Metadata Provision: Datasets should include metadata for self-description, especially provenance metadata to indicate data origin and quality. Licensing and dataset-level metadata, such as a VoID file, should also be provided, particularly when additional access methods (e.g., SPARQL endpoints, data dumps) are available.

Over the years, several studies have evaluated the LOD Cloud's adherence to best practices. These studies, such as Debattista, Attard, et al. 2019; Debattista, Lange, et al. 2018; Nogales, Angel Sicilia-Urban, and García-Barriocanal 2017; Debattista, Auer, and Lange 2016; Vandenbussche et al. 2017; Polleres et al. 2018 and Assaf, Troncy, and Senart 2015, have used the LOD Cloud graph or the LOD network[5] (Jentzsch, Cyganiak, and Bizer 2011; Schmachtenberg, Bizer, and Paulheim 2014) as a snapshot of the Web of Linked Data. The LOD Cloud graph, often referred to simply as the LOD Cloud in various research papers[6] (Debattista, Attard, et al. 2019; Debattista, Lange, et al. 2018; Nogales, Angel Sicilia-Urban, and García-Barriocanal 2017; Debattista, Auer, and Lange 2016), was first designed in October 2007 and has since been updated multiple times (Geiger and Von Lucke 2012; Zhang 2013; Debattista, Attard, et al. 2019). It is considered to be the largest collection of interlinked datasets on the web (Assaf, Troncy, and Senart 2015; Debattista, Attard, et al. 2019; Hitzler 2021). Structurally, the LOD Cloud is an undirected graph G=(S,E), where S represents a set of knowledge graphs in the LOD Cloud, and E includes an edge (t,u) if there is at least one linkset from t to u, or from u to t (Caraballo et al. 2016).

The LOD Cloud was proposed by the Semantic Web community as a foundational reference to track the sources of datasets published and linked on the Web (Haller et al. 2020). Over the years, this prominent representation of LOD has not only served as an emblem for the Web of Linked Data, but also as a starting point for numerous studies and applications (Debattista, Attard, et al. 2019).

The LOD Cloud offers datasets across a wide range of domains, from life sciences to media and social networking (Assaf, Troncy, and Senart 2015; Hitzler 2021). It is widely regarded as a major contribution to promoting Linked Data and Semantic Web technologies, not only in academia but also, to some extent, in industry (Debattista, Attard, et al. 2019). Although the LOD Cloud is often seen as a comprehensive resource for knowledge graphs on the Web of Data (Debattista, Auer, and Lange 2016; Zaveri et al. 2016), it does not encompass all knowledge graphs on the Web. Rather, it provides a representative sampling that offers insight into the

[5] https://lod-cloud.net, date accessed: 25th June 2024.
[6] Similarly, the LOD Cloud Graph is hereafter referred to as the LOD Cloud.

breadth and depth of the Web of Linked Data (Hitzler 2021; Miller and Pretorius 2022).

The "Linked Open Data Cloud" search query on Google Scholar[7] returned 5290 results on 18th November, 2024, indicating the widely utilized nature of this web resource. Moreover, although various sources of interlinked data of varying quality are available across the web, there is no central resource listing all available Linked Data access points on the Web (Haller et al. 2020). This research has therefore selected the LOD Cloud for the following reasons:

- It covers a broad range of knowledge domains (Assaf, Troncy, and Senart 2015; Hitzler 2021);
- It provides principles for publishing interlinked knowledge graphs on the Web as FAIR datasets (Haller et al. 2020).
- It is widely used in both academic research and also in industry (Debattista, Attard, et al. 2019; Debattista, Lange, et al. 2018; Nogales, Angel Sicilia-Urban, and García-Barriocanal 2017; Debattista, Auer, and Lange 2016);

The LOD Cloud is updated monthly, typically with an increasing number of datasets. Knowledge graphs that adhere to the open data definition—allowing them to "be freely used, modified, and shared by anyone for any purpose" (Open Knowledge Foundation 2015)—and follow the principles of Linked Data (Berners-Lee 2006b) are eligible for inclusion in the LOD Cloud (Debattista, Attard, et al. 2019), provided they meet specific criteria (McCrae et al. 2024). These criteria include the use of resolvable http:// or https:// URIs, which must point to RDF data in one of the widely-used formats such as RDFa, RDF/XML, Turtle, or N-Triples, with or without content negotiation. The dataset must contain at least 1,000 triples, excluding smaller datasets like FOAF files, and it must connect via at least 50 RDF links to another dataset already included in the LOD Cloud, either by referencing external URIs or being referenced itself. Finally, the dataset must be fully accessible

[7] https://scholar.google.com

through RDF crawling, an RDF dump, or a SPARQL endpoint. In this research, the LOD Cloud is accessed on 25 June, 2024 is used.

5.3 Metadata Foundations of the LOD Cloud

As mentioned in the previous section, since its creation, the LOD Cloud has been a point of reference to the Linked Data community (Debattista, Lange, et al. 2018). It serves as a catalog for data consumers to discover datasets for reuse and/or linking (Debattista, Attard, et al. 2019; Nogales, Angel Sicilia-Urban, and García-Barriocanal 2017). For this to be possible, datasets on the LOD Cloud include metadata that helps users understand how the data can be accessed and used (Debattista, Attard, et al. 2019).

The LOD Cloud was originally created using source metadata from the now-discontinued Datahub catalog[8] (Open Knowledge Foundation 2007; Haller et al. 2020; Schmachtenberg, Bizer, and Paulheim 2014). An LOD catalog provides metadata for the LOD Cloud datasets, with Datahub and the Mannheim Linked Open Data Catalog[9] (University of Mannheim, Data and Web Science Group 2014b) being two of the most popular examples (Caraballo et al. 2016).

Adding a dataset to the LOD Cloud was initially done by uploading dataset metadata directly to datahub.io with the LOD Cloud tag. More recently, this process involves filling out a form with the required fields. The datahub.io method generated DCAT metadata (World Wide Web Consortium 2014a), while the form submission maps the data into a combination of VoID and DCAT metadata (Debattista, Attard, et al. 2019). The mandatory fields in the metadata are as follows:

- Title – The dataset's name in text form.

- Description – A brief text summary of the dataset, potentially including details on its usage and a human-readable license.

[8]https://old.datahub.io, date accessed: 26th June 2024.
[9]http://linkeddatacatalog.dws.informatik.uni-mannheim.de, date accessed: 26th June 2024.

- Creator – A resource identifying the dataset's creator or publisher, who can be contacted with any questions about the dataset.

- Website – A web page that provides a detailed, human-readable description of the dataset.

- Full Download – A resource describing the full data dump of the dataset, including its media type (e.g., "application/rdf+xml" for an RDF/XML data dump). Using Linked Data resources for media types, which offer semantic descriptions for RDF serializations, is recommended but optional if a SPARQL endpoint is available (Peroni 2016).

- SPARQL Endpoint – Similar to the full download, this is a resource that provides access details for the SPARQL endpoint, potentially including various SPARQL protocols. This is optional if a full download is available.

- Domain – A text description of the dataset's domain, such as financial or geospatial.

- License – A machine-readable resource outlining the legal terms for reusing the dataset. The use of proper machine-readable licenses, such as those in Rodriguez-Doncel, Villata, and Gomez-Perez 2014 or Creative Commons semantic URIs, is required (Debattista, Attard, et al. 2019).

- 606 include namespaces from which 161 are valid.

Additional fields such as DOI, example resources, data catalog, number of triples, and links can be included (Debattista, Attard, et al. 2019). Metadata descriptions of datasets can be easily retrieved from the catalog's Linked Data interface (Debattista, Lange, et al. 2018).

5.4 LOD Cloud KG Selection and De-duplication

As of June 25, 2024, there were 1312 knowledge graphs in the LOD Cloud. Each knowledge graph has an HTML page in the LOD Cloud with information in up to seven sections, as follows:

1. Title: KG's title.
2. About this dataset: Description of the KG's context and content plus its license and Keywords.
3. Contact Details: Contact point/s and KG's website.
4. Download Links.
5. Data Facts: A table including information about the KG's size (number of triples), namespace, number of links to other KGs.
6. Data quality estimation by Luzzu.
7. Download metadata as JSON, RDF/XML, Turtle, N-Triples.

Having reviewed all available KGs in the LOD Cloud, a search for possible duplication was undertaken. This involved identifying (1) KGs with identical titles, and (2) KGs with similar titles (using a case-, space-, and special character-insensitive search) and duplicate namespaces or websites. This duplicate identification process produced the following KGs.

Table 5.1: KGs which were examined as potential duplications

#	**Title**
885	Chinese WordNet (as part of Open Multilingual WordNet).
936	Chinese WordNet (as part of Open Multilingual WordNet).
1004	Naturopathy knowledge Graph (Ontology and Dataset) - RDF distribution of the naturopathy dataset.

Continued on next page

Table 5.1: KGs which were examined as potential duplications (Continued)

#	**Title**
1239	Naturopathy Knowledge Graph (Ontology and Dataset) - RDF distribution of the naturopathy dataset.
440	Greek Children Art Museum dataset.
1059	Greek Children Art Museum dataset.
207	Lexvo.
422	Lexvo.org.
259	Data about business entities from the ARES system - business registry of the Czech Republic.
52	Cell line ontology.
124	Cell line ontology.
307	Cell Line Ontology.
19	Gene Regulation Ontology.
39	Gene Regulation Ontology.
756	School of Electronics and Computer Science, University of Southampton.
915	School of Electronics and Computer Science, University of Southampton.
892	UniProt.
1157	UniProt.
1265	Data about Czech business entities from the ARES system - Trade Licensing Register.
370	VIVO.
895	VIVO.

Four KGs were found to be duplicates and were consequently resolved in advance of the analysis phase, rendering a total of 1308 KGs. The remaining potential duplicates were determined to be different datasets. Specific details of this potential duplication analysis are as follows:

1. KGs 885 and 936 come from different projects and are different.

2. KGs 1004 and 1239 have the same DOI. This shows that they are descriptions

of the same KG. Among these two KGs, 1004 was as it presented greater details in respect of the description and associated information.

3. KGs 440 and 1059 have the same DOI and are describing the same KG. 1059 provided more detailed information and was therefore retained.

4. KGs 207 and 422 have the same website and the same contact point, but their sizes are different and they do not have DOIs. They appear to be two different components of one resource that are represented in a single website. So, they were treated as two different KGs.

5. KGs 52, 124, and 307 are different KGs despite their lookalike titles. They have different webpages and different descriptions.

6. KGs 19 and 39 have the same size, same data facts, similar description, and the same contact point. However their webpages are different and one is not valid. Despite searching through the Google search engine, no other valid webpage or information could be found. Therefore, it looks like they are the same resource, mentioned two times in the LOD Cloud. 19 was retained as it contains the valid webpage.

7. KGs 756 and 915 have different websites, namespaces, contact points, and sizes and are treated as two different KGs.

8. 892 and 1157, while having different contact points and data facts (size and properties), are two versions of the same KG. Only one Uniport could be retrieved through Google search engine. This KG is updated every few weeks and none of the Uniport metadata, mentioned in the LOD Cloud, are up-to-date. Accordingly, KG 1157 was retained for evaluation.

9. KGs 370 and 895 have different websites, contact points, number of triples; one has the namespace included but the other does not, one has an SPARQL endpoint but the other does not. They are treated as two different KGs.

Some of the retrieved KGs identified only a subset of metadata elements, and for the 1308 distinct KGs, the following summary statistics outline the general position in respect of metadata availability

- 148 include no description.
- 578 do not include any contact points.
- 1016 include websites from which 569 links are valid.
- 827 include download links from which 388 are valid.
- 513 include SPARQL Endpoints, from which 106 are valid.
- 213 mentioned size 0 for the corresponding KG. This means that the information is not available.
- 1258 include links to other KGs.
- 569 include a section, titled “Data Quality Estimation by Luzzu”. However, this section is empty in all the 569 html pages.
- All 1308 html pages include a JSON file which contains the same information recorded in the html pages (machine-readable metadata). None of the pages includes metadata in RDF/XML, Turtle, or N-Triples formats, despite them being mentioned in the html pages.

It is interesting that there exists inconsistency in the metadata provision in the LOD Cloud KGs. This suggests that although principles for KG submission are provided, these principles are not always faithfully adopted. This observation potentially undermines the quality of some of the LOD Cloud KGs, especially in the context of missing metadata related to provenance where KGs are to be adopted for the purpose of knowledge elaboration or model training. The impact of this shortcoming is examined in greater detail in the evaluation chapter.

5.5 Overview of Automated FAIRness Assessment

While it is possible to manually assess the FAIRness of individual KGs, from a practical perspective, the analysis of many KGs warrants the use of automation so that scalability considerations can be addressed. The automation of any assessment process also helps to raise the consistency of measurement.

FAIRness assessment using automatic tools is a metadata-dependent process. To assess a KG's FAIRness with FAIR-Checker (Gaignard et al. 2023), F-UJI (Devaraju and Huber 2020), and FAIR Evaluator (Mark D Wilkinson, Dumontier, Sansone, et al. 2019)—three tools suitable for evaluating KG FAIRness—a link to its RDF metadata is required. For this purpose, to the best of our knowledge, five possible resources are available as follows.

1. The original **Datahub catalog** which includes the RDF metadata for the majority of the LOD Cloud's KGs. The Datahub catalog is a robust, open-source data management platform developed by Open Knowledge International, built on the CKAN (Open Knowledge International 2024) system. Datahub enables users to search for datasets, register and manage dataset groups, and receive updates on datasets of interest.

   CKAN, supports managing and publishing open data websites by organizing datasets with metadata and associated resources. CKAN uses the `ckanext-dcat` extension[10], allowing data publishers to expose and consume metadata as RDF using DCAT[11]. To create a dataset in CKAN, the following information is required: (CKAN Project 2024).

   (a) Title – A unique, brief title.

   (b) Description – A summary of the dataset.

   (c) Tags – Keywords to improve discoverability.

   (d) License – Usage rights of the data.

[10]https://github.com/ckan/ckanext-dcat, date accessed: 18th November 2024.

[11]Data Catalog Vocabulary (DCAT) is an RDF vocabulary for interoperability between web data catalogs - https://www.w3.org/TR/vocab-dcat-3/, date accessed: 18th November 2024.

(e) Organization – The owner of the dataset.

(f) Resources – A file or link containing the data.

Additional recommended fields include:

- Name, Description, and Format for each resource.
- Visibility – Public or Private.
- Author/Author e-mail – Contact for data creator.
- Maintainer/Maintainer e-mail – Secondary contact.
- Custom fields – Additional information as needed.

2. **The Mannheim Linked Data Catalog** provides an overview of Linked Datasets available on the Web as effective in August 2014, combining metadata from a web crawl (April 2014) and community-contributed information from the Datahub catalog on CKAN. The catalog includes 1,091 datasets, with metadata indicating aspects like licensing, provenance, and vocabulary use. The catalog served as the foundation for the August 2014 LOD Cloud. This catalog is static; updates should be made in the Datahub catalog on CKAN, following the W3C guidelines (University of Mannheim, Data and Web Science Group 2014a).

3. FAIRsharing[12] (Sansone et al. 2019) is a curated, informative resource on data and metadata standards linked to databases and data policies. It guides users in confidently discovering, selecting, and utilizing these resources; supports producers in enhancing the visibility, adoption, and citation of their resources; and provides reliable content accessible to both humans and tools, facilitating data management tasks.

4. The KG's Website or webpage.

5. Link to the LOD Cloud html page related to the KG.

[12]https://fairsharing.org/, date accessed: 18th November 2024.

For KGs that receive updates, the metadata on Datahub, the Mannheim catalog, and the LOD Cloud HTML pages is likely to be outdated. Therefore, websites likely to receive updates, are identified using FAIRsharing, a platform designed to provide FAIR metadata and DOIs for digital resources. The FAIRsharing platform was searched for the LOD Cloud's KGs. Metadata for 115 LOD Cloud KGs was found on FAIRsharing, all of which were KGs with websites, making them likely to receive updates. Accordingly, For the KGs which receive updates, their FAIRsharing DOIs were used for FAIRness assessment.

At the next level, if a KG has a website (indicating it may receive updates) but is not registered on FAIRsharing, or if it is not likely to be updated, its Datahub link is examined. If no Datahub link is available, the Mannheim link is examined. If the Mannheim link is also unavailable, the KG's webpage/site listed on the LOD Cloud is used. Finally, if none of the above are available, the KG's corresponding LOD Cloud HTML page link is used as a last resort. This process, depicted in Figure 5.1, is an innovation of this research and formalizes the process of identifying the most up to date metadata corresponding to each LOD Cloud KG.

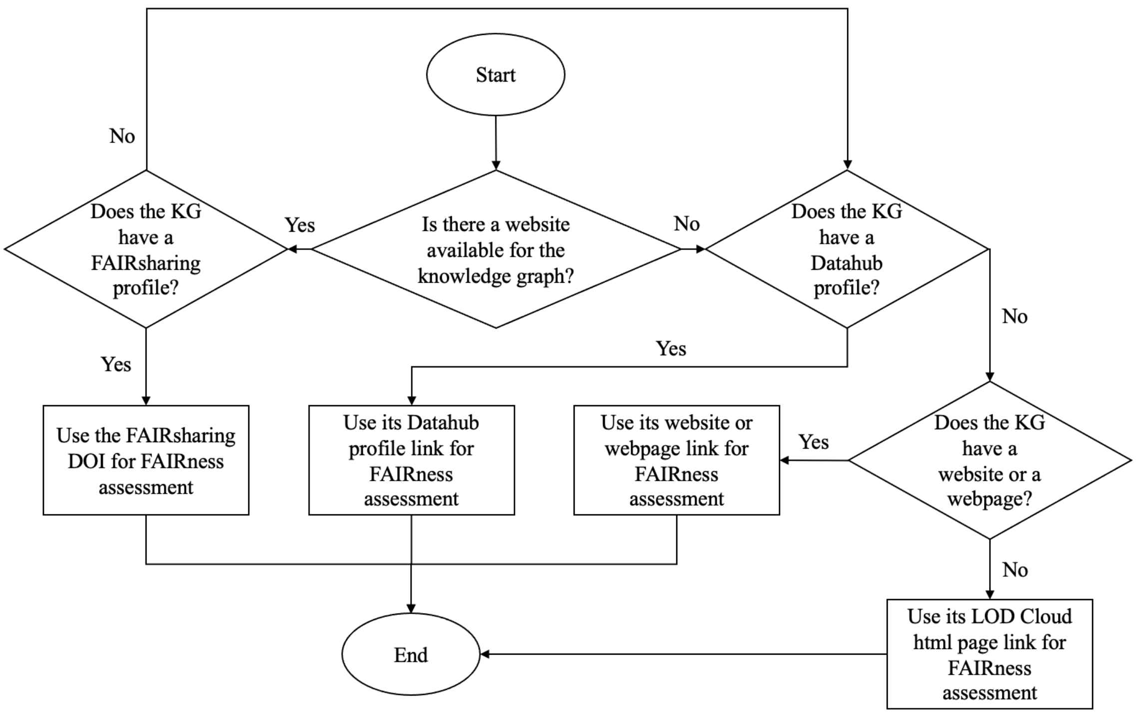


Figure 5.1: Decision tree outlining the link selection process for the LOD Cloud FAIRness assessment.

Through this selection process, 115 FAIRsharing links, 1116 Datahub links, 62 KG website links, and 15 LOD Cloud HTML page links for FAIRness assessment of the LOD Cloud are collected.

5.6 Automated FAIRness Assessment in this Study

To assess the FAIRness of the KGs in the LOD Cloud using F-UJI, the tool was installed on a macOS machine following the instructions on the F-UJI GitHub page[13], and the assessment was conducted using a Python script for sequential evaluation. F-UJI provides a JSON output for each assessment, which then was parsed in Python to extract the FAIRness assessment results.

To assess the FAIRness of the LOD Cloud using FAIR-Checker and FAIR Evaluator, separate Python scripts were used to send GET requests to the corresponding tools' APIs to help with sequential assessments, with random delays between requests to avoid overwhelming the FAIR-Checker server. Similar to F-UJI, these tools also provide JSON outputs for each assessment, which then were parsed using separate regular expressions in Python. Listings 5.1 and 5.2 show the regular expressions used for parsing FAIR-Checker results and FAIR Evaluator results, respectively.

Listing 5.1: Regular expression used for parsing FAIR-Checker results.

```
metrics = ["F1A", "F1B", "F2A", "F2B", "A1.1", "A1.2", "I1",
    "I2", "I3", "R1.1", "R1.2", "R1.3"]
pattern = rf"'http://www.w3.org/ns/dqv#isMeasurementOf':
    \[\{{'@id': 'https://fair-checker.france-bioinformatique.
    fr/data/{metric}'\}}\], 'http://www.w3.org/ns/dqv#value':
    \[\{{'@value': (\d)\}}\]"
```

Listing 5.2: Regular expression used for parsing FAIR Evaluator results.

[13] https://github.com/pangaea-data-publisher/fuji, date accessed: 18th November 2024.

```
pattern = r'\"(http://fairdata\.services:3333/FAIR_Evaluator/
   metrics/\d+)\"\:\[\{\"@id\":\"http://tests:8080//tests
   /[^\"]+\",\"@type\":\[\"http://fairmetrics\.org/resources/
   metric_evaluation_result\"\],\"http://semanticscience\.org
   /resource/SIO_000300\":\[\{\"@type\":\"http://www\.w3\.org
   /2001/XMLSchema#int\",\"@value\":\"(\d+)\"\}'
```

When using FAIR-Checker, the LOD Cloud HTML links were used for 9 KGs—including 450, 574, 587, 761, 959, 1044, 1262, 1279, and 1309—which initially had their website or webpage links used for FAIRness assessment. This adjustment was due to errors received from the FAIR-Checker server. Similarly, while using FAIR Evaluator, further selection refinement was required arising from server-side shortcomings, with the following using the corresponding LOD Cloud HTML link (in place of the originally identified Datahub links):

- Websites or webpages links for 29 KGs, including 88, 195, 204, 206, 250, 278, 324, 460, 546, 549, 595, 612, 616, 653, 678, 820, 856, 956, 1014, 1082, 1083, 1087, 1108, 1130, 1143, 1212, 1252, 1259, and 1291.
- The LOD Cloud HTML links for seven KGs, including 266, 490, 701, 796, 1047, 1115, 1127, and 1273.

Table 5.2 presents an overview of the utilized metadata sources for each FAIRness assessment tool.

Table 5.2: Metadata sources for assessing the FAIRness of the LOD Cloud across different tools.

Metadata source	**F-UJI**	**FAIR-Checker**	**FAIR Evaluator**
Datahub	1116	1116	1080
FAIRsharing	115	115	115
KG's website/page	62	53	90
lod-cloud.net	15	24	23

5.7 Concluding Remarks

This chapter has outlined the data collection process, indicating how automated assessment relies on metadata, along with the details surrounding data selection based on the capabilities of the three freely available public-domain automated FAIRness assessment tools: F-UJI, FAIR-Checker and FAIR Evaluator. In the next chapter, this systematically collected data is used to provide a statistical view of the consistencies and inconsistencies of FAIRness assessment results in these three tools and provide a consistent FAIRness measure.

Chapter 6

Evaluation

6.1 Chapter Overview

As demonstrated in the literature (Chapter 2) and the comparative analysis presented in Chapter 4, the absence of a standardized FAIRness assessment approach, and the consequent proliferation of diverse evaluation methods, have resulted in inconsistencies and confusion in selecting assessment methods and analyzing their results.

To address these gaps, this chapter presents a rigorous statistical analysis of the FAIR scores produced by three selected automated FAIRness assessment tools that may be used to assess the FAIRness of knowledge graphs (KGs): F-UJI, FAIR Evaluator, and FAIR-Checker. Unlike prior works that have suggested these inconsistencies qualitatively, this analysis employs a novel statistical approach to quantify the variations. This analysis is conducted in parallel with the methodology outlined in Chapter 3, using the FAIRness Assessment Data (FAD) derived from assessing the FAIRness of KGs within the Linked Open Data (LOD) Cloud (see Chapter 5).

This quantitative approach represents a significant advancement over previous research by introducing objectivity and reproducibility into the analysis of FAIRness tool outputs. By applying statistical techniques such as distribution analysis, correlation measures, and dimensionality reduction, this work not only identifies inconsistencies across tools but also characterizes their extent and structure in a measurable way. This contributes to the field by offering a methodologically grounded framework for evaluating the reliability and complementarity of FAIRness assessment tools—an area that has, until now, lacked empirical rigor. Accordingly, this analysis forms a key contribution of the thesis, laying the foundation for more standardized and data-driven evaluations in future FAIR-related research.

In the subsequent step, the range of overall FAIR scores is analyzed to identify the FAIRness measures that contribute most significantly to this variation. This analysis facilitates a clearer interpretation of the FAIRness results obtained from different tools and provides a better understanding of the combination of measures that could inform the development of a combined method. A smaller range between overall FAIR scores obtained from the tools indicates higher consistency among the tools, thereby enhancing the reliability of the FAIRness assessment. It is proposed that higher overall FAIR scores along with a smaller range imply that the knowledge graph (KG) is FAIRer, making it more dependable for reuse in training potential AI models.

The structure of this chapter is outlined as follows. In Section 6.2, a comparative analysis of the overall FAIR results is presented. The processes involved in feature engineering are discussed in Section 6.3 and Section 6.4. In Section 6.5, the modeling of the range of overall FAIRness scores is presented, and the results are analyzed to identify the most effective set of metrics for predicting this range. Finally, concluding remarks are provided in Section 6.6.

6.2 Consistency Analysis of Overall Scores from F-UJI, FAIR-Checker, and FAIR Evaluator

In this section, the overall FAIR scores of the KGs in the LOD Cloud, obtained using F-UJI, FAIR-Checker, and FAIR Evaluator, are analyzed to provide insights into the level of consensus and consistency among their respective FAIRness assessment approaches. This helps addressing the first research question[1]. The analysis reveals a 49% correlation between the overall results of FAIR-Checker and FAIR Evaluator, while their correlations with F-UJI were determined to be 11% and 16%, respectively. The observed correlations between the overall FAIR scores from F-UJI, FAIR-Checker, and FAIR Evaluator can be attributed to differences in their methodologies, scoring mechanisms, and operational focus (see Chapter 4 for further details). This variation highlights how differences in methodology and implementation influence the consistency of FAIRness assessments across tools.

To analyze the data for anomalies, Mahalanobis D^2 distance (Mahalanobis 1936) is used, which is a measure of the multivariate distance between a point and a multivariate mean, accounting for correlations and varying scales of the data. Unlike Euclidean distance, it considers the covariance of the dataset, making it ideal for identifying outliers and multivariate anomalies. The formula is:

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$$

where x is the data point, μ is the mean vector, and Σ is the covariance matrix. Figure 6.1 presents a scatter plot of Mahalanobis distances, highlighting outliers identified based on the defined threshold. Points above the red dashed line represent outliers in the dataset.

Anomalies in the overall FAIR scores within the dataset were examined using scatter plots, which depict the scores obtained from F-UJI (FUJI-FAIR), FAIR Evaluator (FE-FAIR), and FAIR-Checker (FC-FAIR) plotted against one another, as shown in Figure 6.2. In this figure, anomalies—represented by points deviating from the overall trend—are highlighted with red oval shapes. These anomalies demonstrate an increase in one overall score, while the other remains unchanged.

As an example, Table 6.1 presents a sample of three data points exhibiting anomalous overall FAIR score behaviors. In this sample, the overall FAIR scores

[1] How consistent are the results produced by automated knowledge graph FAIRness assessment tools?

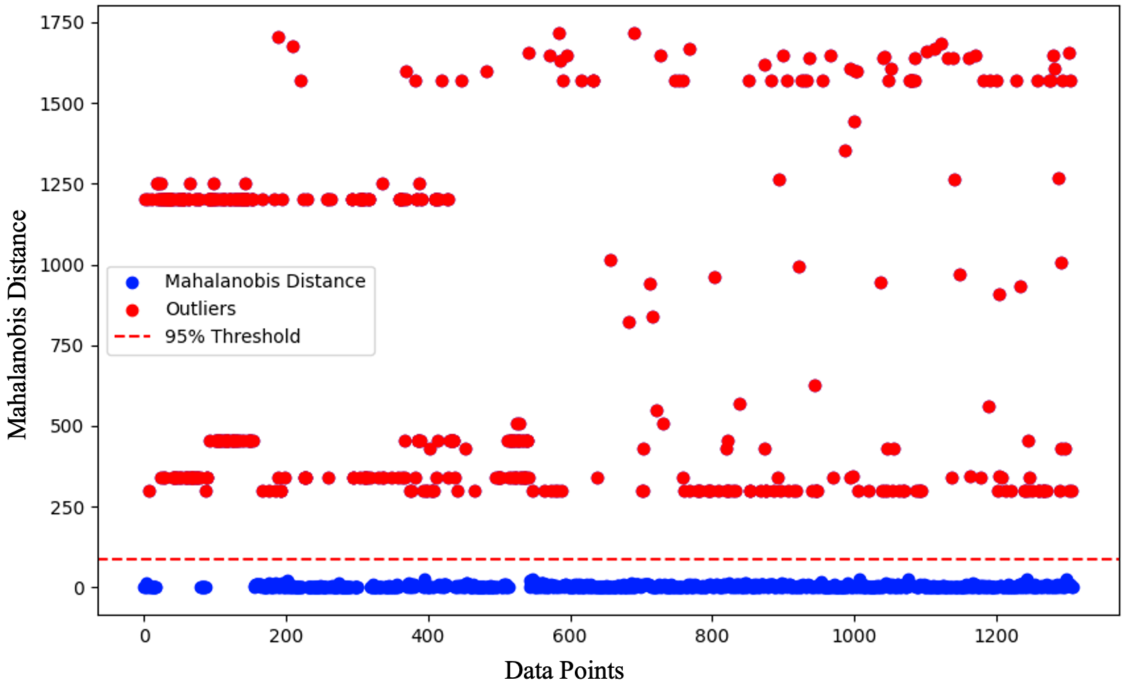


Figure 6.1: Scatter plot of Mahalanobis distances highlighting outliers based on the defined threshold.

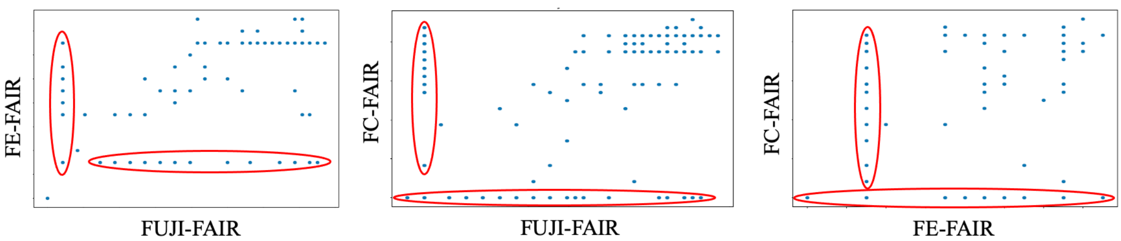


Figure 6.2: Scatter plots illustrating the overall FAIR scores obtained from F-UJI (FUJI-FAIR), FAIR Evaluator (FE-FAIR), and FAIR-Checker (FC-FAIR).

provided by FAIR-Checker vary across the data points, whereas the scores from F-UJI and FAIR Evaluator remain unchanged. For brevity, the table includes only the features responsible for the variations in FAIR-Checker's overall FAIR scores[2].

Table 6.1: Three data points with anomalous overall FAIR score behaviors.

#	A1.2	I3	R1.1	FC-FAIR	FE-FAIR	FUJI-FAIR
82	2	2	2	**83.33**	65	4.17
1005	0	0	0	**58.33**	65	4.17
1300	2	0	2	**75**	65	4.17

The results indicate that for A1.2, FAIR-Checker identified more than one access policy property—such as *odrl:hasPolicy*, *dct:rights*, *dct:accessRights*, *dct:license*, or

[2]Appendix A presents tables listing all the features associated with these three data points.

schema:license—in the metadata of KG 82 and KG 1300. However, it was unable to find any of these properties in the metadata of KG 1005. However, different I3 results across the three KGs suggest that FAIR-Checker identified at least three different URL authorities in the URIs of the RDF metadata for KG 82, but fewer than three in the metadata for KG 1005 and KG 1300.

In addition, varying R1.1 results across the three KGs is particularly intriguing, as all three tools assessed the same link from the same source. This suggests that FAIR-Checker was able to locate at least one of the license properties—such as *schema:license*, *dct:license*, *doap:license*, *dbpedia-owl:license*, *cc:license*, *xhv:license*, *sto:license*, or *nie:license*—in the metadata of KG 82 and KG 1300. However, it could not find any of these properties in the metadata of KG 1005.

It is important to note that this represents only a small sample of anomalous behavior in the dataset, where the overall FAIR-Checker score varies between points, while the scores from the other two tools remain unchanged. Similarly, there are other instances where the overall scores of one of the other tools change, while the scores from the remaining two remain constant.

Based on the correlation and anomaly analyses conducted, it can be concluded that the overall results of the three automatic FAIRness assessment tools lack consistency. This indicates that, at present, no reliable method exists for measuring the FAIRness of knowledge graphs, nor has an established approach been identified to determine the most effective method for such evaluations.

To address this gap, inline with addressing the second research question[3], this research utilizes the range of overall FAIR scores as an indicator of inconsistency among the results. A smaller range is proposed to signify lower inconsistency and greater reliability in FAIRness assessment outcomes. Accordingly, the objective is to determine whether it is possible to predict the range of the overall FAIR scores generated by the three FAIRness assessment tools suitable for KGs, namely F-UJI, FAIR Evaluator, and FAIR-Checker. This approach aims to identify the set of

[3]How can confidence in automated FAIRness assessment for knowledge graphs be increased through specific techniques?

FAIRness measures responsible for the variability in the overall FAIR scores, which define the range and can be utilized for more robust and comprehensive FAIRness assessments. A smaller range, combined with higher FAIR scores, would indicate greater reliability in the reuse and interoperability of the KGs and simplifies decision making in the use or not using of KGs for the model training purposes.

To achieve this, the overall scores were first standardized, and their range was included as a target feature in the dataset. Subsequently, the individual overall score features were removed to mitigate multicollinearity. Figure 6.3 presents the histogram of the range of the overall FAIRness assessment results.

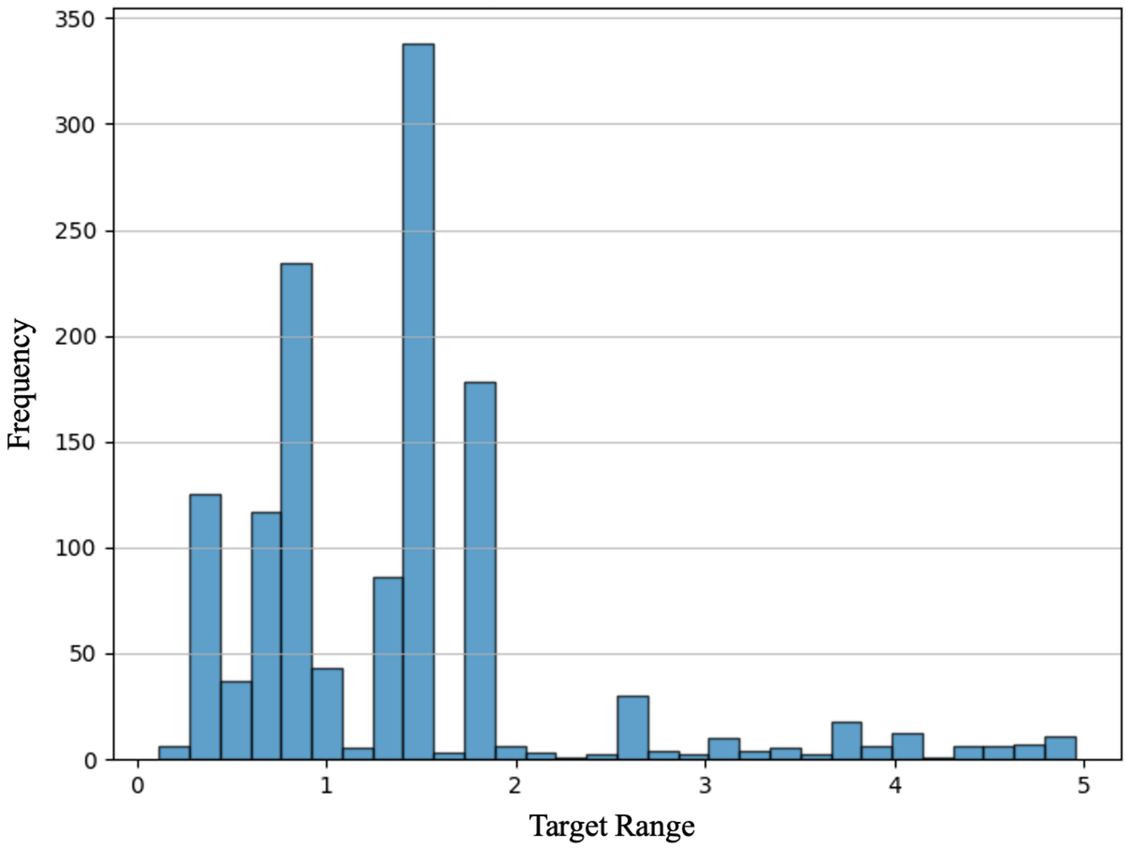


Figure 6.3: Histogram of the target range feature.

The histogram of the target variable reveals a right-skewed distribution, indicating that the mean of the distribution exceeds the median. This skewness suggests that higher values of the target variable are less frequent but remain significant. A notable peak is observed around values between 1 and 2, highlighting that a substantial portion of the target variable falls within this range. In the next sections, the feature engineering process is presented to facilitate data pre-processing for predictive analysis.

6.3 FAD Feature Engineering: Correlation Analysis of numerical features

The correlation analysis on FAD's numerical features, excluding non-numerical features (8 features), features with zero variance[4] (6 features), F-UJI aggregate features[5] (16 features), and the 'target range', revealed many highly correlated features within FAD. Below are groups of mutually highly correlated features (correlation $\geq$ 0.8) extracted from the correlation matrix.

- Group1: *FE-A1.1: FAIR Metrics Gen2 - Uses open free protocol for metadata retrieval*, *FE-A1.2: FAIR Metrics Gen2 - Metadata authentication and authorization*, *FE-F1: FAIR Metrics Gen2 - Unique Identifier*, *FUJI-FsF-F1-01D-1*;
- Group2: *FC-F1B*, *FE-F1: FAIR Metrics Gen2 - Identifier Persistence*, *FUJI-FsF-F1-02D-1*, *FUJI-FsF-F1-02D-2*, *FUJI-FsF-F4-01M-2*, *FUJI-FsF-I3-01M-2*;
- Group3: *FUJI-FsF-F2-01M-2*, *FUJI-FsF-F2-01M-3*, *FUJI-FsF-R1.1-01M-1*;
- Group4: *FC-A1.1*, *FC-F1A*, *FC-F2A*, *FC-F2B*, *FC-I1*, *FC-I2*, *FC-R1.3*;
- Group5: *FC-A1.2*, *FC-R1.1*;
- Group6: *FUJI-FsF-A1-02M-1*, *FUJI-FsF-F2-01M-1*, *FUJI-FsF-I1-01M-2*, *FUJI-FsF-I3-01M-1*, *FUJI-FsF-R1-01MD-1*, *FUJI-FsF-R1.2-01M-1*, *FUJI-FsF-R1.3-01M-3*;

[4] 'FUJI-FsF-F1-01D-2,' 'FUJI-FsF-A1-01M-3,' 'FUJI-FsF-R1-01MD-4,' 'FUJI-FsF-R1.3-01M-2,' 'FE-F4: FAIR Metrics Gen2 - Searchable in major search engine,' and 'FE-A2: FAIR Metrics Gen2 - Metadata Persistence'

[5] F-UJI provides aggregate results for FAIR principles and sub-principles. For purposes of comparability, these features were removed from FAD.

- Group7: *FUJI-FsF-A1-02M-1*, *FUJI-FsF-F2-01M-1*, *FUJI-FsF-F4-01M-1*, *FUJI-FsF-I1-01M-1*, *FUJI-FsF-I3-01M-1*, *FUJI-FsF-R1-01MD-1*, *FUJI-FsF-R1.2-01M-1*, *FUJI-FsF-R1.3-01M-3*;

- Group8: *FUJI-FsF-A1-03D-1*, *FUJI-FsF-F3-01M-1*, *FUJI-FsF-F3-01M-2*, *FUJI-FsF-F4-01M-1*, *FUJI-FsF-I1-01M-1*, *FUJI-FsF-I2-01M-2*, *FUJI-FsF-R1.2-01M-1*;

- Group9: *FE-A1.1: FAIR Metrics Gen2 - Uses open free protocol for data retrieval*, *FE-A1.2: FAIR Metrics Gen2 - Data authentication and authorization*, *FE-F3: FAIR Metrics Gen2 - Data Identifier Explicitly In Metadata*, *FE-I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)*, *FE-I3: FAIR Metrics Gen2 - Metadata contains qualified outward references*;

- Group10: *FE-F2: FAIR Metrics Gen2 - Grounded Metadata*, *FE-F2: FAIR Metrics Gen2 - Structured Metadata*, *FE-I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong)*, *FE-I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (weak)*, *FE-I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)*, *FE-I3: FAIR Metrics Gen2 - Metadata contains qualified outward references.*

To assess multicollinearity among the independent features, their Variance Inflation Factor (VIF) values were calculated. The results indicate that, as anticipated due to the high correlation between variables, only 12 out of 59 features exhibit VIF values below 10, while 22 features have VIF values equal to infinity[6]. This confirms the presence of significant multicollinearity among the majority of the features.

The presence of significant multicollinearity among the majority of features in the dataset has the potential to induce bias in the standard error of the parameter for any regression model and diminish the interpretability of predictors. To mitigate this issue, Principal Component Analysis (PCA), a dimensionality reduction

[6]See Appendix B for a table of VIF values.

technique, is applied in the next section to improve model performance and enhance interpretability.

6.4 FAD Feature Engineering: Principal Component Analysis (PCA)

To apply PCA to FAD, the data was first standardized (rescaled) to have a mean of 0 and a standard deviation of 1, using the formula:

$$z = \frac{x - \mu}{\sigma}$$

where μ and σ denote the mean and standard deviation of each feature, respectively. Standardization ensures that all features are on the same scale, thereby minimizing potential scale-related biases in the PCA results. For this analysis, the target variable was excluded from the dataset. The cumulative explained variance across the Principal Components (PCs) is illustrated in Figure 6.4[7].

Based on the cumulative explained variance scores, as illustrated in Figure 6.4, the first 15 PCs account for 95% of the total variance. Beyond this point, the cumulative explained variance shows minimal incremental improvement with the addition of further PCs. Consequently, the first 15 PCs will be used for predictive analysis.

[7] Appendix C lists the explained variance ratio and cumulative explained variance values for the FAD Principal Components.

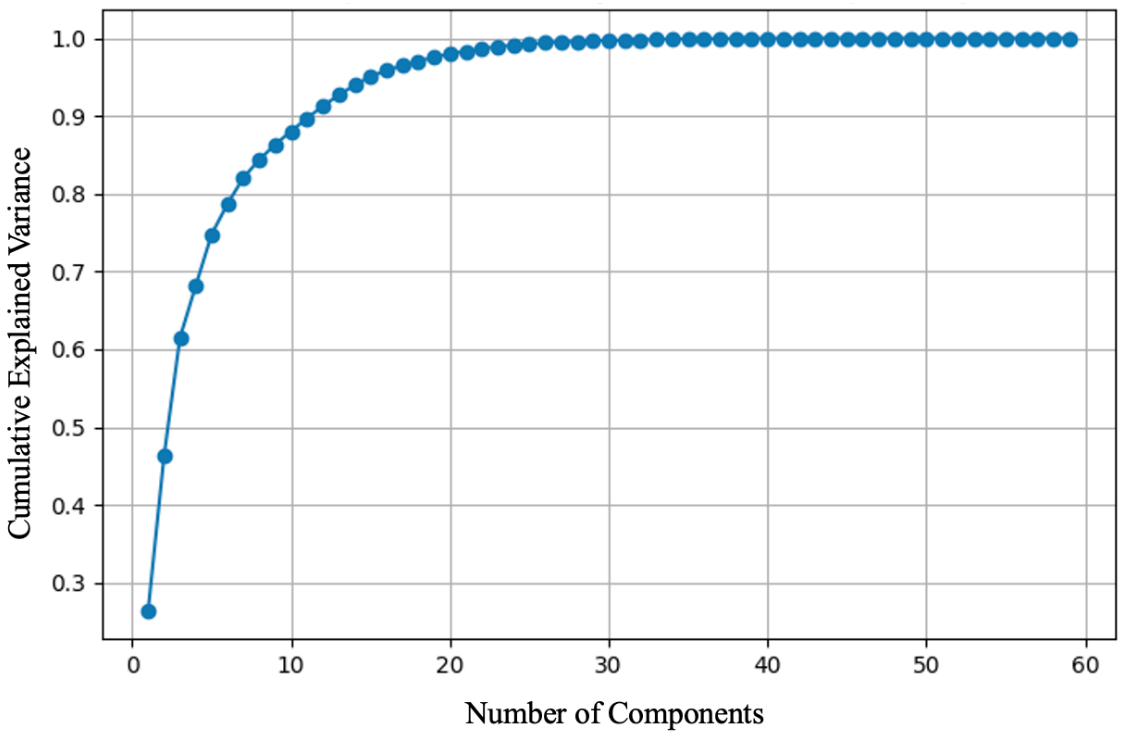


Figure 6.4: Cumulative explained variance by number of principal components.

6.5 Prediction Analysis of the Range Between Overall FAIR Scores from F-UJI, FAIR Evaluator, and FAIR-Checker

In this section, the PCA-transformed data, comprising the first 15 principal components, is utilized to predict the range of overall FAIR scores obtained from F-UJI, FAIR-Checker, and FAIR Evaluator. The analysis is performed to identify the top features contributing to the variation between the overall FAIR scores. This approach facilitates the development of a more comprehensive and reliable FAIRness assessment method by potentially integrating a combination of FAIR measures from different tools, thereby addressing the confusion and challenges associated with selecting a single assessment approach.

The prediction process begins with a multivariate linear regression. For this analysis, the Ordinary Least Squares (OLS) regression model[8] was employed. The

[8]A standard method that minimizes the sum of squared residuals.

regression was conducted using the Least Squares method to ensure an effective fit to the observed data (Legendre 1805).

The model explains 68.0% of the variance in the dependent variable R^2, with an Adjusted R^2 of 67.5%, confirming the model's complexity is justified. The F-statistic ($P < 0.001$, $f_{15} = 145.9$) indicates the model is statistically significant, with predictors collectively showing a strong relationship with the target variable. However, on the other hand, the residuals significantly deviate from normality, as indicated by the Omnibus statistic (471.339, $P < 0.001$), skewness (1.669, right-skewed), kurtosis (15.272, heavy tails), and the Jarque-Bera statistic (7049.392, $P < 0.001$). The Durbin-Watson statistic (1.987) suggests little to no autocorrelation, while the condition number (10.9) indicates no severe multicollinearity, aligning with the use of principal components as features.

To address the issue of non-normal residuals, the range of overall FAIR scores was modeled using Random Forest and Multilayer Perceptron (MLP) regressor, which are nonparametric models. For this purpose, the data was divided into 80% training and 20% validation. Hyperparameters were tuned using the Grid Search method. Table 6.2 presents the optimized hyperparameters.

Table 6.2: Optimized hyperparameters for the Random Forest and MLP models after tuning.

Prediction model	**Tuned hyperparameters**
MLP	Activation: relu alpha: 0.1 Hidden layer size: 30 Learning rate: Constant Max iterations: 500 Solver: lbfgs Best MSE: 0.00236
Random forest	Max depth: 10 Min samples leaf: 2 Min samples split: 5 N estimators: 50 Best MSE: 0.0368677

The training process was conducted using an MLP regressor and a Random

Forest model, utilizing the tuned hyperparameters presented in Table 6.2. Table 6.3 summarizes the mean results derived from ten independent runs of the models.

Table 6.3: Results of the MLP regressor and the random forest models for predicting the range between overall FAIR scores.

Model	CV results[9]		Training results					Vld[10] results	
	Mean MSE	Mean R^2	SSE	MSE	R^2	AIC	BIC	MSE	R^2
MLP	0.002	0.997	0.585	0.001	0.999	-6953.57	-4422.72	0.002	0.997
RF	0.047	0.935	11.744	0.011	0.985	-4620.62	-4372.99	0.053	0.935

Based on Table 6.3, the MLP model demonstrates better generalization (lower validation MSE and higher validation R^2) and performs well in cross-validation. Additionally, the AIC and BIC values indicate that the MLP model has a better balance between goodness-of-fit and complexity compared to the Random Forest model, as evidenced by its significantly lower values for both metrics. Accordingly, the MLP model is used to identify the most significant FAIR measures influencing the range between the overall FAIR scores. Table 6.4 presents the importance scores for each of the first 15 PCs used in the MLP regression modeling process, calculated using three approaches: Permutation (Breiman 2001), Shapley (Lundberg and S.-I. Lee 2017), and Ablation (Montavon, Samek, and Müller 2018) techniques. The importance scores for each technique represent the mean values derived from running the model ten times. The table also includes the mean importance score, calculated from the normalized scores, which ranks the PCs based on their contribution to the model's predictions. Normalization was applied to bring the scores from different techniques to a common scale, preventing differences in magnitude from disproportionately affecting the results. To achieve this, Min-Max normalization was employed, scaling the values from each technique to a range of [0, 1]. This approach ensures that the smallest value is mapped to 0, the largest to 1, and the relative distances between values are preserved, as demonstrated by the following formula.

$$x_{\text{normalized}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Table 6.4: Importance scores of the first 15 principal components (PCs) in the MLP regression model.

Rank	Permutation		SHAP values		Ablation		**Mean imp**[11]	
	PC	Imp[12]	PC	MSV[13]	PC	MA[14]	**PC**	**MNimp**[15]
1	PC2	1.0233	PC2	0.3642	PC2	0.0164	PC2	1
2	PC6	0.4293	PC1	0.2146	PC6	0.0110	PC6	0.5330
3	PC1	0.3771	PC6	0.1871	PC5	0.0045	PC1	0.3783
4	PC4	0.1461	PC4	0.1603	PC4	0.0033	PC4	0.2566
5	PC5	0.0532	PC5	0.0859	PC1	0.0031	PC5	0.1824
6	PC3	0.0187	PC3	0.0458	PC7	0.0009	PC7	0.0534
7	PC9	0.0124	PC7	0.0426	PC10	0.0007	PC3	0.0489
8	PC7	0.0083	PC9	0.0342	PC8	0.00065	PC8	0.0379
9	PC8	0.0066	PC8	0.0308	PC11	0.00064	PC9	0.0361
10	PC11	0.0039	PC11	0.0239	PC14	0.0004	PC11	0.0305
11	PC13	0.0008	PC13	0.0062	PC3	0.00033	PC10	0.0120
12	PC10	0.0005	PC10	0.0043	PC9	0.00032	PC13	0.0064
13	PC14	0.0004	PC12	0.0040	PC13	0.00031	PC14	0.0050
14	PC12	0.0002	PC14	0.0033	PC15	0.00028	PC12	0.0025
15	PC15	0.0001	PC15	0.0014	PC12	0.00023	PC15	0.0009

As shown in Table 6.4, a significant drop in the mean normalized importance scores is observed after PC5. Consequently, the top five features—PC2, PC6, PC1, PC4, and PC5—are identified as the most influential for the MLP model's predictive power and will be used to determine the key FAIR measures.

To identify the most important features within the top five most influential PCs, their corresponding PCA coefficients were analyzed. For each PC, a threshold was established based on the 80th percentile of the absolute values of the feature loadings, capturing the top 20% of features with the strongest positive or negative influences. Figures 6.5 to 6.9 illustrate the identified key contributing features for the top five PCs[16].

PC2 is identified as the most influential component in the MLP regression process, with its 13 most important features split between FAIR-Checker (seven) and FAIR Evaluator (six). The three most significant features—*'R1.3-Community standards'*, *'I2-Use shared ontologies'*, and *'F2B-Shared vocabularies for metadata'*—assess the use of recognized ontology classes or properties. Other notable metrics include *'I1-Machine-readable'*, *'F2A-Structured metadata'*, *'F1A-Unique IDs'*, and *'A1.1-*

[16]Appendix D lists the key contributing features for these five principal component.

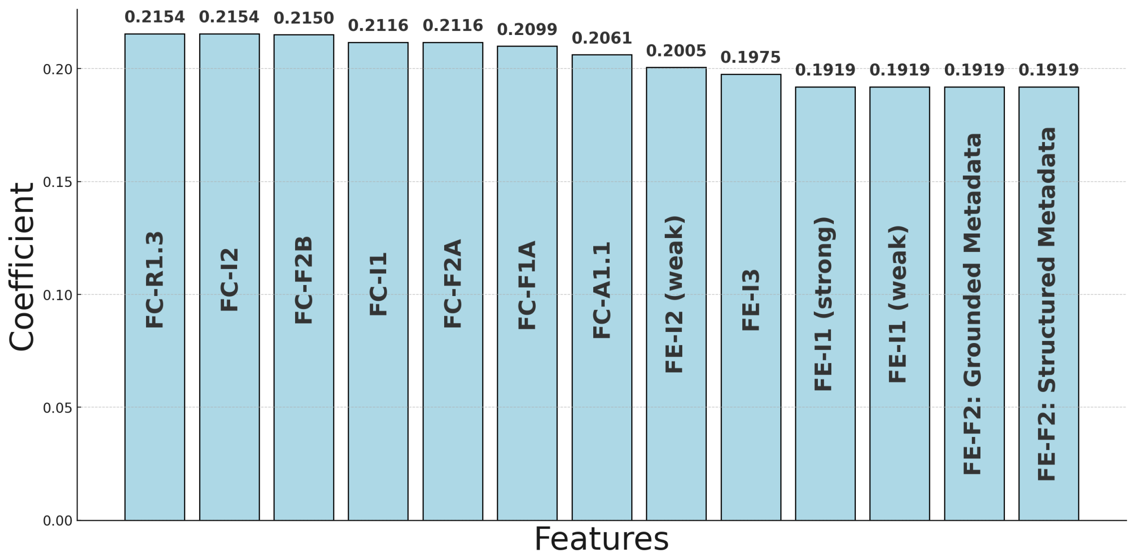


Figure 6.5: Key contributing features in Principal Component 2.

Open resolution protocol', which evaluate metadata properties or RDF triples.

The six key metrics from FAIR Evaluator include *'I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)'*, focusing on term resolution, and *'I3: FAIR Metrics Gen2 - Metadata contains qualified outward references'*, which assesses metadata links to third-party resources. Metrics like *'I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong)'* and *'I1 (weak)'* are also significant, with conceptual overlap but differing implementations compared to FAIR-Checker.

Additionally, PC2 includes F2-related metrics from both tools, such as *'Grounded Metadata'* and *'Structured Metadata'*, which align conceptually but differ in implementation. These six FAIR Evaluator metrics also appear as significant features in PC4, underscoring their broader influence in the analysis.

PC6, ranked fourth, includes features from all three tools—F-UJI, FAIR Evaluator, and FAIR-Checker—spanning all FAIR principles. Findability is supported by features like *'FsF-F2-01M-3'*, ensuring core descriptive metadata, and *'FAIR Metrics Gen2 - Metadata Identifier Explicitly In Metadata'*, which verifies metadata identifiers. Accessibility is addressed by *'FsF-A1-01M-1'* and *'FsF-A1-01M-2'*, ensuring access restrictions are clearly described and machine-readable, while *'A1.2 - Authorization procedure or access rights'* enhances accessibility reliability. Interop-

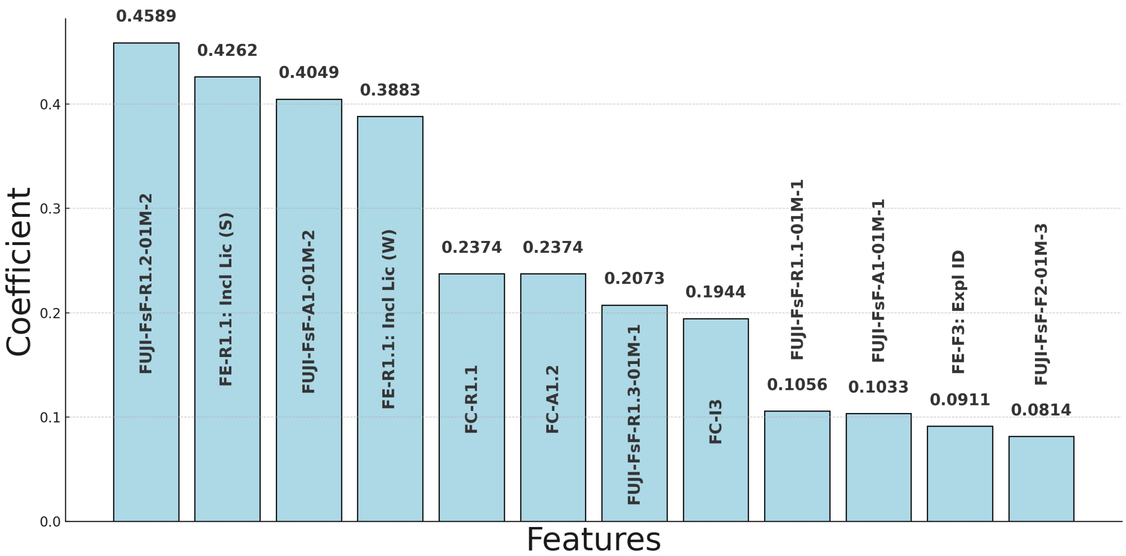


Figure 6.6: Key contributing features in Principal Component 6.

erability is reflected in *'I3 - External Links'*, which ensures metadata includes links to external authorities. Reusability is emphasized through licensing and provenance metrics, including *'R1.1: FAIR Metrics Gen2 - Metadata Includes License'* and *'FsF-R1.2-01M-2'*, which use ontologies like PROV-O. Overlapping features such as *'A1.2'* and *'FsF-R1.1-01M-1'* highlight their importance across components.

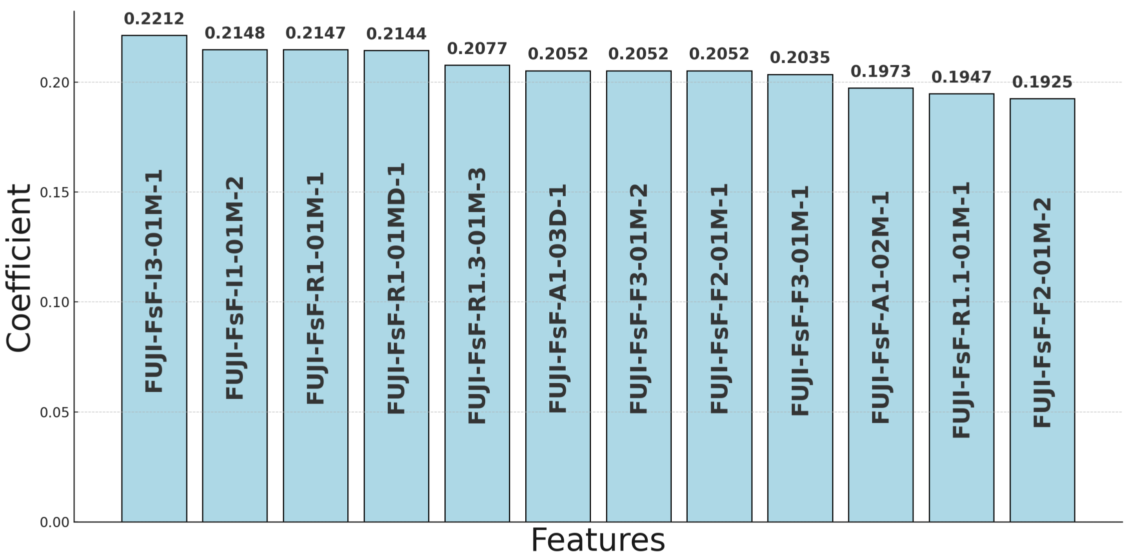


Figure 6.7: Key contributing features in Principal Component 1.

PC1, the second most influential principal component for predicting overall FAIR scores, is composed entirely of features derived from F-UJI FAIR tests across all four FAIR principles: Findability, Accessibility, Interoperability, and Reusability. These features, encompassing metadata quality, web accessibility, and data interoperability, provide a strong foundation for identifying inconsistencies in FAIRness assessments. PC1's emphasis on metadata availability, web protocols, structured data, and reusability standards underscores its critical role in evaluating and predicting resource FAIRness.

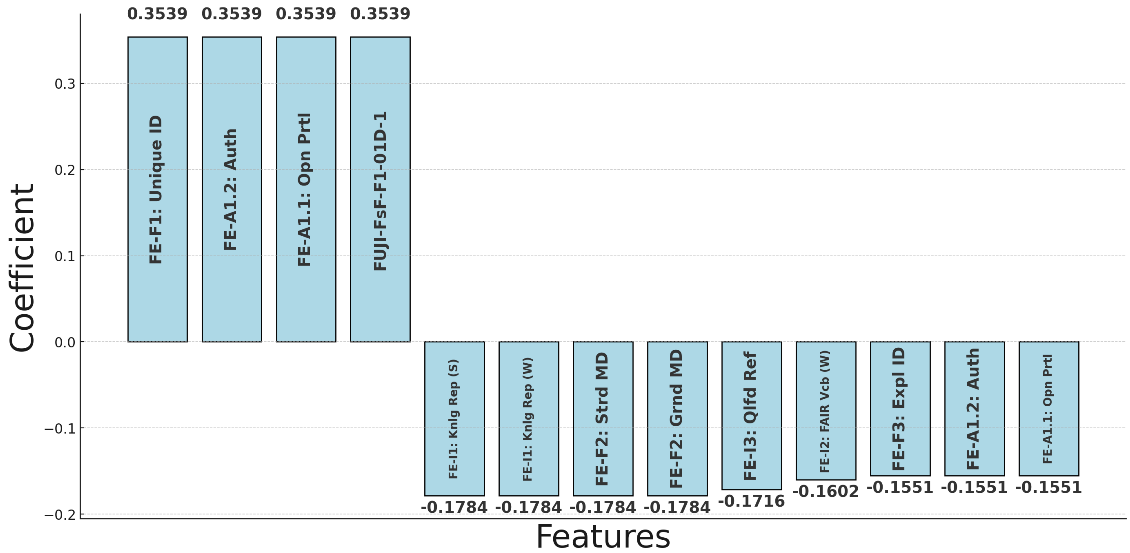


Figure 6.8: Key contributing features in Principal Component 4.

PC4 is primarily influenced by features from FAIR Evaluator tests, focusing on metadata quality, identifier uniqueness, and data accessibility. Key features include *'FAIR Metrics Gen2 - Unique Identifier'* and *'FsF-F1-01D-1'*, which assess identifier uniqueness, and *'FAIR Metrics Gen2 - Data Identifier Explicitly in Metadata'*, promoting resource discovery. Accessibility is ensured through features like *'FAIR Metrics Gen2 - Uses open free protocol for metadata retrieval'* and *'FAIR Metrics Gen2 - Metadata authentication and authorization'*, emphasizing secure and open protocols. Interoperability is addressed by features such as *'Metadata uses FAIR vocabularies'* and *'FAIR Metrics Gen2 - Metadata Knowledge Representation Language'*, which foster structured metadata and system integration. Overlapping features across PCs, such as *'FsF-F1-01D-1'* and *'Metadata Knowledge Representation Language'*, highlight their importance, making PC4 critical for assessing metadata structure, identification, and accessibility in FAIRness evaluation.

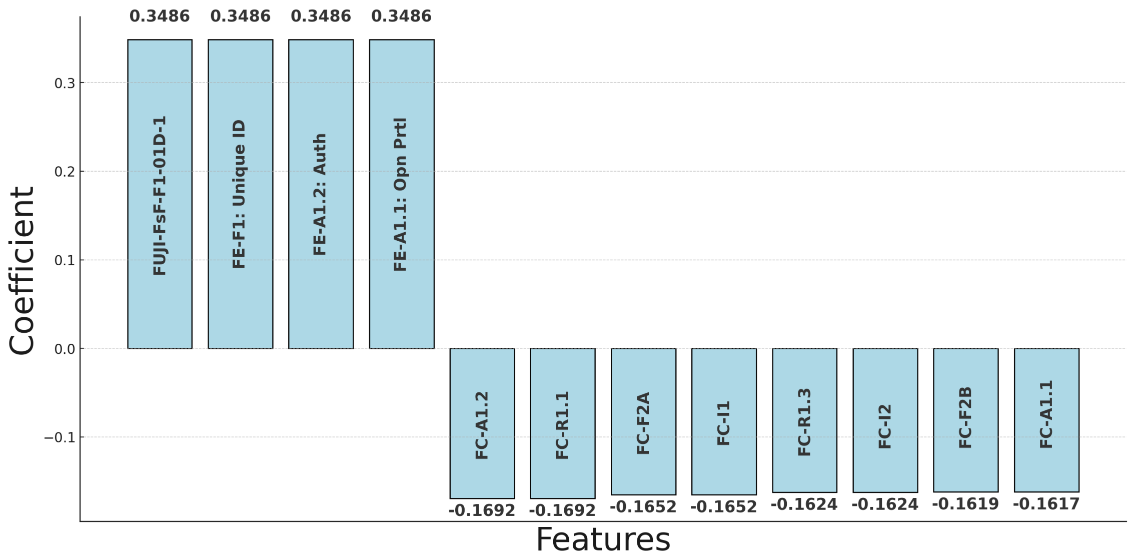


Figure 6.9: Key contributing features in Principal Component 5.

PC5 encompasses a diverse set of features from F-UJI, FAIR Evaluator, and FAIR-Checker, reflecting the foundational elements of robust FAIRness evaluation aligned with all four FAIR principles. Findability is supported by metrics like *'FsF-F1-01D-1'* and *'FAIR Metrics Gen2 - Unique Identifier'*, ensuring resolvable and unique identifiers, while *'F2A - Structured Metadata'* enhances discoverability through structured metadata.

Accessibility features, such as *'A1.1 - Uses open free protocol for metadata retrieval'* and *'A1.2 - Metadata authentication and authorization'*, ensure reliable access via standard protocols with clear access rights. Interoperability is addressed through *'I1 - Machine-readable format'* and *'I2 - Use shared ontologies'*, ensuring machine-readable data linked to shared ontologies like OLS and BioPortal. Reusability is supported by *'R1.1 - Metadata includes license'* and *'R1.3 - Community standards'*, promoting legally and ethically reusable data aligned with community standards. Overlapping features, such as *'A1.2'* and *'FsF-F1-01D-1'*, underscore their broader significance in FAIRness assessment.

After analyzing the most important features in each of the highly influential PCs, a breakdown of the contribution of features from three tools—F-UJI, FAIR-Checker (FC), and FAIR Evaluator (FE)—to five principal components (PC1, PC2, PC4, PC5, and PC6) is provided in Table 6.5. F-UJI has the highest influence in PC1 (12 features) and a notable presence in PC6 (6 features), contributing a total of 20 features across all components. FAIR-Checker is most prominent in PC5 (8 features) and PC2 (7 features), with a total of 18 features. FAIR Evaluator dominates PC4 with 12 features and has balanced contributions across PC2, PC5, and PC6, contributing a total of 24 features, the highest among the tools.

Table 6.5: Tool Influence Across Principal Components.

Tools	**PC1**	**PC2**	**PC4**	**PC5**	**PC6**	**Total**
F-UJI	12	0	1	1	6	20
FC[17]	0	7	0	8	3	18
FE[18]	0	6	12	3	3	24

Table 6.6 summarizes the distribution of features across FAIR principles within the most influential Principal Components (PC1, PC2, PC4, PC5, PC6).

Table 6.6: FAIR principles influence across principal components.

FAIR	**PC1**	**PC2**	**PC4**	**PC5**	**PC6**	**Total**
F1	0	1	2	2	0	5
F2	1	4	2	2	1	10
F3	2	0	1	0	1	4
F4	0	0	0	0	0	0
A1	2	0	0	0	2	4
A1.1	0	1	2	2	0	5
A1.2	0	0	2	2	1	5
A2	0	0	0	0	0	0
I1	2	3	2	1	0	8
I2	0	2	1	1	0	4
I3	1	1	1	0	1	4
R1	1	0	0	0	0	1
R1.1	1	0	0	1	4	6
R1.2	1	0	0	0	1	2

R1.3	1	1	0	1	1	4

Based on Table 6.6, key observations include:

- Findability (F1, F2, F3): F2 is the most represented findability metric, appearing 10 times across PCs, particularly in PC2. F1 and F3 have moderate representation (5 and 4 occurrences, respectively), while F4 is absent in all PCs.

- Accessibility (A1, A1.1, A1.2, A2): Accessibility metrics are well-distributed, with A1.1 and A1.2 each appearing 5 times, primarily in PC4 and PC5. A1 is represented 4 times, while A2 is absent across all PCs.

- Interoperability (I1, I2, I3): I1 dominates interoperability metrics with 8 occurrences, heavily represented in PC2 and PC1. I2 and I3 each appear 4 times, reflecting moderate influence across PCs.

- Reusability (R1, R1.1, R1.2, R1.3): R1.1 is the most frequent reusability metric, appearing 6 times, with strong representation in PC6. R1.2 and R1.3 have limited occurrences, at 2 and 4, respectively, while R1 appears only once.

Overall, PC2 contributes the most diverse representation across FAIR principles, particularly for F2 and I1, while PC6 emphasizes reusability through R1.1. Metrics related to findability and interoperability dominate the PCs, highlighting their centrality in assessing FAIRness.

Among features in FAD that were used for prediction analysis, 17 were absent among the most important features, as follows.

- FAIR-Checker: *'F1B - Persistent IDs'*[19], *'R1.2 - Metadata includes provenance'*[20]

[19]Strong evaluation: Checking if there is either *schema:identifier* or *dct:identifier* property in metadata. Weak evaluation: Checking that at least one namespace from identifiers.org is in metadata

[20]Checking that at least one of the following provenance properties is found in metadata: *prov:wasGeneratedBy*, *prov:wasDerivedFrom*, *prov:wasAttributedTo*, *prov:used*,

- FAIR Evaluator: *'F1: FAIR Metrics Gen2 - Identifier Persistence'*, *'F1: FAIR Metrics Gen2 - Data Identifier Persistence'*, *'F4: FAIR Metrics Gen2 - Searchable in major search engines'*, *'I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (strong)'*

- *'F-UJI: FsF-F1-02D-1'*[21], *'FsF-F1-02D-2'*[22], *'FsF-F2-01M-2'*[23], *'FsF-F4-01M-1'*[24], *'FsF-F4-01M-2'*[25], *'FsF-I2-01M-2'*[26], *'FsF-I3-01M-2'*[27], *'FsF-R1-01MD-2'*[28], *'FsF-R1-01MD-3'*[29], *'FsF-R1.1-01M-2'*[30], *'FsF-R1.3-02D-1'* [31]

While these features represent core aspects of FAIRness, their absence among the most important features suggests they may be less influential in predicting the range of overall FAIR scores (reliability of FAIRness assessment) within the dataset.

Following a thorough examination of the most important features contributing to the five PCs with the greatest influence on predicting the range of overall FAIR scores, it is evident that these PCs collectively incorporate measures addressing all the FAIR principles, derived from a combination of tools. These features play a pivotal role in enhancing the predictive model's ability to assess the reliability of FAIRness assessments. This underscores the critical importance of utilizing these features in tandem across tools to achieve a comprehensive and robust FAIRness assessment, thereby enabling more reliable decision-making regarding the use of KG

prov:wasInformedBy, *prov:wasAssociatedWith*, *prov:startedAtTime*, *prov:endedAtTime*, *dct:hasVersion*, *dct:isVersionOf*, *dct:creator*, *dct:contributor*, *dct:publisher*, *pav:hasVersion*, *pav:version*, *pav:hasCurrentVersion*, *pav:createdBy*, *pav:authoredBy*, *pav:retrievedFrom*, *pav:importedFrom*, *pav:createdWith*, *pav:retrievedBy*, *pav:importedBy*, *pav:curatedBy*, *pav:createdAt*, *pav:previousVersion*, *schema:creator*, *schema:author*, *schema:publisher*, *schema:provider*, *schema:funder*.

[21] Identifier follows a defined persistent identifier syntax.

[22] The identifier is resolvable to a valid URI.

[23] Core data citation metadata is available (creator, title, publisher, publication_date, object_identifier, object_type) are specified through appropriate metadata fields.

[24] Metadata is given in a way major search engines can ingest it for their catalogs (JSON-LD, Dublin Core, RDFa).

[25] Metadata is registered in major research data registries (DataCite).

[26] Namespaces of known semantic resources (listed in LOD registry) can be identified in metadata.

[27] Related resources are indicated by machine readable links or identifiers.

[28] Verifiable data descriptors (file info, measured variables or observation types) are specified in metadata.

[29] Data content matches file type and size specified in metadata.

[30] Recognized license is valid (community specific or registered at SPDX).

[31] The format of a data file given in the metadata is listed in the long term file formats, open file formats or scientific file formats controlled list.

in modeling processes.

Consequently, it is proposed that these principal components be utilized not only as indicators of the reliability of FAIRness assessments but also as measures of FAIRness itself. To facilitate an analysis of the results, understanding how fluctuations in each core PC affect the range is essential. Table 6.7 presents the correlation between the core PCs and the range of overall FAIR scores derived from F-UJI, FAIR Evaluator, and FAIR-Checker.

Table 6.7: Correlation between core principal components and the overall FAIR scores and their range.

PC	**Overall FAIR scores range**	**FUJI-FAIR**	**FC-FAIR**	**FE-FAIR**
PC1	**-0.5142**	**0.9400**	0.4274	0.3433
PC2	-0.3142	-0.3361	**0.8758**	0.4559
PC4	**0.4287**	0.0061	-0.0775	**-0.5154**
PC5	-0.1592	0.0133	0.0462	**0.2817**
PC6	-0.1764	-0.0156	-0.1834	**0.5683**

As previously mentioned, a smaller range indicates greater consistency and reliability in FAIRness assessment results. As shown in Table 6.7, PC1 and PC2 exhibit a strong negative correlation with the range, meaning that as these components increase, the range decreases, and vice versa. In contrast, PC4 shows a strong positive correlation, indicating that an increase in PC4 corresponds to an increase in the range, and a decrease leads to a smaller range. PC5 and PC6 also have a negative correlation with the range, though this correlation is weaker compared to PC1 and PC2. These relationships provide insights into analyzing the results derived from these principal components. A KG with higher values for PC1, PC2, PC5, and PC6, combined with a lower value for PC4, can be considered reliably FAIR and suitable for use in model training.

A score that serves as an indicator for both overall FAIRness and reliability is of interest in this analysis. Based on the correlations presented in Table 6.7, the is proposed as an indicator of overall FAIRness and consistency, referred to hereafter as the FAIRness Consistency Indicator (FCI). PC1, PC2, PC6, and PC5 are included to reflect their positive correlations with the overall FAIR scores and negative correlations with the range of these scores. In contrast, PC4 has strong positive correlation with the range, weak positive correlation with FUJI-FAIR, and negative correlations with FE-FAIR and FC-FAIR.

Accordingly, the negative regression coefficient was used as the weight for each of the PCs. This adjustment ensures that consistency across the results is effectively captured in the indicator. Table 6.8 presents the correlation between the core PCs and the target range, along with their corresponding regression coefficients and derived weights.

Table 6.8: Correlation between core PCs and the target range plus their core PCs' regression coefficients.

PC	**Overall FAIR scores range**	**Regression coefficient**	**Weight**
PC1	-0.5142	-0.27	0.27
PC2	-0.3142	-0.19	0.19
PC4	0.4287	0.38	-0.38
PC5	-0.1592	-0.20	0.20
PC6	-0.1764	-0.27	0.27

The following is the FAIRness Consistency Indicator (FCI), which serves as a measure of both overall FAIRness and assessment reliability, and represents one of the contributions of this research.

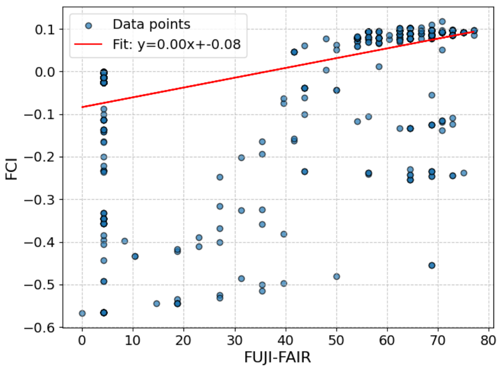
$$\mathrm{FCI} = \frac{\begin{array}{c}(0.27 \cdot \mathrm{PC1}) + (0.19 \cdot \mathrm{PC2}) + (0.2 \cdot \mathrm{PC5}) \\ +(0.27 \cdot \mathrm{PC6}) + (-0.38 \cdot \mathrm{PC4})\end{array}}{5}$$

Table 6.9 presents correlation between the proposed indicator and the overall FAIR scores and their range. As shown in the table, FCI demonstrates a strong positive correlation with all overall FAIR scores and a significant negative correlation with the range. These results suggest that the indicator effectively captures both overall FAIRness and consistency.

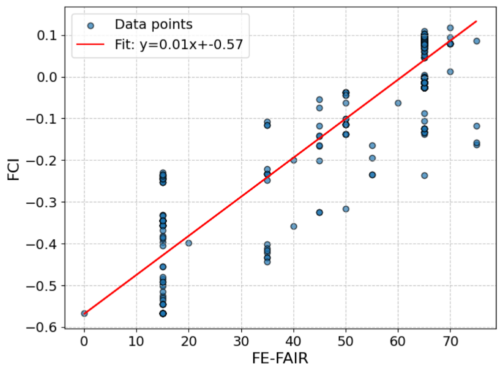
Figures 6.10 and 6.11 present scatter plots of the FCI against the overall FAIR scores (from F-UJI, FAIR Evaluator-FE, and FAIR-Checker-FC) and the range of overall FAIR scores, respectively.

Table 6.9: Correlation between the proposed indicator and the overall FAIR scores and their range.

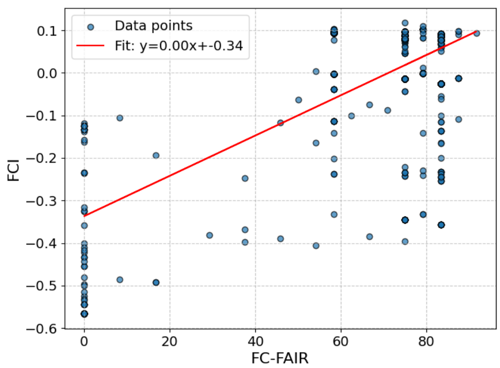
	Correlation with range[32]	**Correlation with FUJI-FAIR**	**Correlation with FC-FAIR**	**Correlation with FE-FAIR**
FCI	-0.7768	0.4854	0.6479	0.8835



(a) FCI Vs. FUJI-FAIR.



(b) FCI Vs. FE-FAIR.



(c) FCI Vs. FC-FAIR.

Figure 6.10: Scatter plots showing the novel FAIRness Consistency Indicator (FCI) indicator versus the overall FAIR scores from F-UJI, FAIR Evaluator (FE), and FAIR-Checker (FC).

Figure 6.11: Scatter plot of the novel FAIRness Consistency Indicator (FCI) indicator versus the range of overall FAIR scores.

The regression lines in Figure 6.10 exhibit positive slopes, indicating that higher FCI values correspond to increases in F-UJI, FE, and FC overall FAIR scores. Conversely, the regression line in Figure 6.11 displays a negative slope, consistent with the correlation analysis, showing that an increase in FCI corresponds to a decrease in the range.

The red circle in the figure shows the points were the FAIRness is high and the range is low and therefore the corresponding KGs could be reused for modeling purposes.

6.6 Concluding remarks

In this chapter, a comparative analysis of the overall FAIR scores obtained from three automatic FAIRness assessment tools—F-UJI, FAIR Evaluator, and FAIR-Checker—was conducted (See Section 6.2). This analysis highlighted inconsistencies in automatic FAIRness assessments, which were attributed to the diverse methods employed for measuring FAIRness. Therefore, for the first time the inconsistencies between the FAIRness assessment methods are shown, objectively.

To address these inconsistencies, a novel approach was introduced by reframing the issue as a prediction task (See Sections 6.3, 6.4, and 6.5). Specifically, the range of overall FAIRness scores was predicted as an indicator of the consistency and reliability of FAIRness assessments. A smaller range of the scores for a given Knowledge Graph (KG) signifies greater consistency and reduced ambiguity in its FAIRness evaluation. This, in turn, suggests a higher degree of FAIRness and improved suitability for safe reuse.

Furthermore, an optimal combination of features for this prediction was identified, leading to the proposal of an indicator, called FAIRness Consistency Indicator (FCI) indicator, that serves as a comprehensive measure of both the overall FAIR score and its consistency.

Chapter 7

Discussion and Conclusion

7.1 Chapter overview

This chapter presents the conclusion of this thesis, beginning with a summary of the research conducted to address the research questions in Section 7.2. The challenges encountered during the research are discussed in Section 7.3, followed by an outline of the research contributions in Section 7.4. Lastly, Section 7.5 addresses the limitations, proposes future research directions, and provides concluding remarks.

7.2 Summary of the Conducted Research

In this section, a summary of the work conducted to address the research questions is presented. Knowledge graphs formalize and classify knowledge, facilitating knowledge extraction, retrieval, and analysis. Consequently, their use in AI systems is increasingly prevalent. This growing use of knowledge graphs underscores the importance of ensuring their FAIRness to promote transparency in AI systems. In this research, as outlined in Section 2.13, the following gaps related to FAIRness assessment of KGs were identified through conducting a literature review (See Chapter 2 for more details).

1. There are inconsistencies in interpreting and implementing FAIR principles, metrics, and FAIRness assessment. However, different initiatives have tried to

harmonize the FAIRness assessment across various tools, inconsistencies still persist.

2. No systematic research has focused on a broad analysis of automated FAIRness measurement across a large variety of knowledge graphs, and therefore the true scale of measurement volatility is unknown. This lack of quantification raises considerable challenges for the community as it grapples with automated FAIRness measurement at scale.

These identified gaps have formed the following research questions in this thesis, which then were addressed in line with the objective quantitative methodology that was chosen for this research (See Chapter 3).

1. How consistent are the results produced by automated knowledge graph FAIRness assessment tools?

2. How can confidence in automated FAIRness assessment for knowledge graphs be increased through specific techniques?

Although concerns regarding consistency between different FAIRness assessment methods are noted in earlier peer-reviewed research, none of the existing research has provided a clear and statistically grounded view of the consistencies and inconsistencies among the automated FAIRness measurement tools. To address the identified gaps and consequently the first research question, the following steps were taken.

1. Automatic FAIRness assessment tool identification: A total of three automatic FAIRness assessment tools, which can be applied to assess the FAIRness of knowledge graphs, i.e., F-UJI, FAIR Evaluator, and FAIR-Checker, were identified through literature review (See Section 2.11).

2. Conducting comparative analysis of the measures employed by these three FAIRness assessment tools to elucidate how they operate and the measures they use (See Chapter 4).

3. FAIRness Assessment Data (FAD) collection (See Chapter 5)

   (a) Source identification: LOD Cloud was identified as the most comprehensive set of open access knowledge graphs, adopted across a range of research studies.

   (b) FAIRness assessment: The FAIRness of knowledge graphs within the LOD Cloud (a total of 1,308 knowledge graphs) was evaluated using the identified tools, forming a dataset, which then was utilized to analyze the consistencies and inconsistencies in automatic FAIRness assessment outcomes.

4. Consistency analysis (See Section6.2): The consistency of overall FAIR scores within FAD was analyzed using correlation analysis and Mahalanobis D^2 distance techniques, providing a clear objective view of the alignments and misalignments in state of the art automated FAIRness assessment.

To address the second research question, as a novel approach, the FAIRness assessment was converted into a prediction problem. Considering the range between the overall FAIR scores as the indicator for consistency between the tools, the goal was to find the set of FAIRness measures that most significantly contribute to predicting the range. This enabled identifying a combination of FAIRness measures from different tools that could be used for a robust FAIRness assessment. The following steps were taken:

1. FAD feature engineering (See Sections 6.3 and 6.4): After identifying multicollinearity in the data, Principal Component Analysis (PCA) was conducted to mitigate the implications.

2. Range prediction (See Section 6.5): The prediction analysis was performed using linear regression, random forest, and Multi-Layer Perceptron (MLP) regression techniques, resulting in:

   (a) Identification of features which most significantly contribute to the range.

(b) Proposing FAIRness Consistency Indicator (FCI), an indicator for both FAIRness and robustness/consistency of automated FAIRness measurements using a combination of the most important features in the prediction process.

7.3 Challenges

This section presents the challenges encountered during the course of this research. A key challenge in conducting generalizable and robust research on knowledge graphs is the absence of a comprehensive and up-to-date repository. Although several repositories exist, there is no official governing entity to maintain a unified and up-to-date resource with valid metadata that can be readily utilized for research purposes. Given the increasing role of knowledge graphs in training machine learning models, this can be considered a highly undesirable situation and an area which required considerable effort in this work.

In this research, based on an extensive literature review, the LOD Cloud was identified as the most comprehensive source of open-access knowledge graphs. However, it does not encompass all knowledge graphs available on the internet, and furthermore, it does not contain an exclusive set of knowledge graphs (which themselves can be stored in multiple repositories, possibly presenting with multiple conflicting versions). Additionally, challenges were encountered, including invalid metadata descriptions and dead or outdated datasets listed on the LOD Cloud, as noted by Debattista, Attard, et al. 2019. The issue of outdated metadata has been further highlighted in previous studies, including those by Haller et al. 2020 and Akhtar et al. 2020.

The LOD Cloud diagram was originally generated from the source metadata descriptions available in the Datahub repository[1] (Open Knowledge Foundation 2007). Consequently, some metadata entries may not be updated to reflect current resources and access points. Additionally, sources providing a Linked Data access point may

[1] https://old.datahub.io, date accessed: 05/12/2024.

not be listed in the Datahub repository, and therefore might be excluded from the LOD Cloud diagram (Debattista, Attard, et al. 2019). On the other hand, metadata for new datasets imported into Datahub is typically added manually as textual descriptions, making it susceptible to errors such as inconsistency and duplication (Debattista, Lange, et al. 2018), a challenge encountered in this research.

The absence of a unified, valid, and continuously maintained metadata repository, combined with the proliferation of diverse metadata resources for knowledge graphs in the LOD Cloud, was a significant challenge to achieve a valid FAIRness assessment of these knowledge graphs. In addition, FAIRness assessment tools do not adequately account for the timeliness of the metadata being evaluated. As a result, higher FAIRness assessment scores may be obtained from outdated metadata for a knowledge graph compared to up-to-date metadata that is less comprehensive. For instance, assessing 'DBpedia[2]' using outdated metadata that provides more complete, albeit invalid, provenance information may yield better results than assessing the same knowledge graph with up-to-date metadata containing less comprehensive provenance information. While all practically possible efforts were undertaken in this work to reduce the effect of this challenge, this research highlights that although some progress has been made in automated knowledge graph FAIRness evaluation, the broader domain still suffers from the absence of an immutable and centralized configuration management construct that governs knowledge graph versioning. Until the community and associated stakeholders define a solution to this challenge, knowledge graph based research will continue to prove difficult to robustly evaluate.

FAIR concepts are not entirely distinct but rather they are overlapping, leading to confusion in the implementation of FAIRness assessment measures. For example, FAIR-Checker employs the same strategy to evaluate F2, I2, and R1.3, reflecting the lack of clear, standardized definitions and implementations for FAIR concepts. Furthermore, in this research, six measures were found to consistently have zero values across all knowledge graphs in the LOD Cloud. This finding highlights the

[2] A knowledge graph extracted from Wikipedia articles. https://www.dbpedia.org, date accessed: 05/12/2024.

inadequate provision and maintenance of metadata by knowledge graph publishers. While AI and data governance and accountability have gained increased attention in recent years, adherence to these principles still requires significant improvement.

Finally, an important factor contributing to continuous FAIRness assessment and compliance is the ease of use of the FAIRness assessment tools. While F-UJI is well-documented, easy to install, and straightforward to reuse, the evidence of this research indicates that the other two automated tools are less flexible and more challenging to install and operate.

7.4 Contributions

Contributions in this study are listed as follows.

- This research provides the first model of the inconsistency of automatic KG FAIRness measurement. For the first time, this research frames the problem of consistency between automatic KG FAIRness assessment methods as a prediction problem, utilizing the range between the overall FAIRness scores as an indicator of consistency across the tools (See Section 6.5).

- This research provides a statistically grounded view of inconsistencies between three open access automatic knowledge graph FAIRness assessment tools, as opposed to descriptive analysis based research undertaken in earlier related studies (See Chapter 6).

- This research provides the first structured process for identifying the best available metadata sources for automated FAIRness assessment of KGs, while also highlighting two key challenges, i.e., the lack of up-to-date metadata and the absence of governance authority and oversight.

- This research assesses the FAIRness of knowledge graphs in the LOD Cloud for the first time as an open access dataset for research and analysis (See Chapter 5).

- This thesis proposes a novel measure, called FAIRness Consistency Indicator (FCI), that is an indicator for both FAIRness and the consistency/robustness of FAIRness assessment. This measure combines the FAIRness factors that most significantly contribute to the range between the overall FAIRness scores obtained from the three tools. A higher value of the proposed measure indicates greater FAIRness and improved robustness. The net result of this contribution is that for the first time, it is possible to identify knowledge graphs that are statistically more likely to be present with higher consistent FAIRness. This has positive implications for various models trained using knowledge graphs, supporting the development of trustworthy AI systems reliant on FAIR training data. This contributes to FAIRness evaluation through using statistical methods.
- The first thorough and systematic comparative analysis of the three automatic FAIRness assessment tools which are suitable for FAIRness assessment, analyzing the nuances between different sets of measures (See Chapter 4).
- Finally, this research research contributes to FAIR education and awareness.

7.5 Research Limitations and Future Directions

This research is limited to the automatic FAIRness assessment tools that are open access and applicable to knowledge graph FAIRness assessment. Also, the focus of this research is on the consistent and robust FAIRness assessment for knowledge graphs and the results might not be generalizable to other data types. Nevertheless, the methodology employed could be adapted for use with other data types. Additionally, no threshold for FAIRness or consistency was established in this study.

Accordingly, future analyses could focus on proposing thresholds or sets of thresholds to define high/low or acceptable/unacceptable FAIRness levels, tailored to specific applications and domains. Additionally, since the novel method proposed in this work integrates measures from multiple tools, further research could explore the development of a unified tool that combines these features and their implementations.

In this study, FAIR principles were analyzed as a critical factor promoting reliability in datasets and trust in the models build upon them. It is hoped that this thesis provides a solid and scientifically reproducible foundation for further research and exploration in the area of FAIRness principles. Given the ever-increasing volume of data being produced, and its importance in terms of describing knowledge that is key for human endeavor, appropriate steps must be taken to increase confidence in the FAIRness of datasets so that our collective efforts to improve are built from solid foundations. This research has been motivated by this foundational consideration, and it aspires to robustly and clearly improve our capability to measure FAIRness in knowledge graphs.

Appendix A

A Sample of Anomalous Behavior in FAIRness Assessment Data (FAD)

Tables A.1 to A.9 show a complete view of three data points with anomalous overall FAIR score behaviors, mentioned in Section 6.2.

Table A.1: Three data points with anomalous overall FAIR score behaviors - F-UJI findability results.

#	F1-1D-1	F1-1D-2	F1-2D-1	F1-2D-2	F1	F2-1M-1	F2-1M-2	F2-1M-3	F2	F3-1M-1	F3-1M-2	F3	F4-1M-1	F4-1M-2	F4	F
82	1	0	0	0	50	0	0	0	0	0	0	0	0	0	0	14.29
1005	1	0	0	0	50	0	0	0	0	0	0	0	0	0	0	14.29
1300	1	0	0	0	50	0	0	0	0	0	0	0	0	0	0	14.29

Table A.2: Three data points with anomalous overall FAIR score behaviors - FAIR Evaluator findability results.

#	F1-U[a]	F1-P[b]	F1-DP[c]	F2-S[d]	F2-G[e]	F3-D[f]	F3-M[g]	F4[h]
82	1	0	0	1	1	1	0	0
1005	1	0	0	1	1	1	0	0
1300	1	0	0	1	1	1	0	0

Table A.3: Three data points with anomalous overall FAIR score behaviors - FAIR-Checker findability results.

#	F1A	F1B	F2A	F2B
82	2	0	1	2
1005	2	0	1	2
1300	2	0	1	2

[a] Unique Identifier
[b] Identifier Persistence
[c] Data Identifier Persistence
[d] Structured Metadata
[e] Grounded Metadata
[f] Data Identifier is Explicitly in Metadata
[g] Metadata Identifier is Explicitly in Metadata
[h] Searchable in Major Search Engine

Table A.4: Three data points with anomalous overall FAIR score behaviors - F-UJI accessibility and interoperability results.

#	A1-1M-1	A1-3D-1	A1-1M-3	A1-2M-1	A1	A	I1-1M-1	I1-1M-2	I1	I2-1M-2	I2	I3-1M-1	I3-1M-2	I3	I
82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A.5: Three data points with anomalous overall FAIR score behaviors - FAIR Evaluator accessibility and interoperability results.

#	A1.1-D[a]	A1.1-M[b]	A1.2-D[c]	A1.2-M[d]	A2[e]	I1[f]-Weak	I1-Strong	I2[g]-Weak	I2-Strong	I3[h]
82	1	1	1	1	0	1	1	1	1	1
1005	1	1	1	1	0	1	1	1	1	1
1300	1	1	1	1	0	1	1	1	1	1

Table A.6: Three data points with anomalous overall FAIR score behaviors - FAIR-Checker accessibility and interoperability results.

#	A1.1	A1.2	I1	I2	I3
82	2	**2**	1	2	**2**
1005	2	**0**	1	2	**0**
1300	2	**2**	1	2	**0**

[a] Uses Open Free Protocol for Data Retrieval
[b] Uses Open Free Protocol for Metadata Retrieval
[c] Data Authentication and Authorization
[d] Metadata Authentication and Authorization
[e] Metadata Persistence
[f] Metadata Knowledge Representation Language
[g] Metadata Uses FAIR Vocabularies
[h] Metadata Contains Qualified Outward References

Table A.7: Three data points with anomalous overall FAIR score behaviors - F-UJI reusability results and the overall FAIR score.

#	R1-1MD-1	R1-1MD-2	R1-1MD-3	R1-1MD-4	R1	R1.1-1M-1	R1.1-1M-2	R1.1	R1.2-1M-1	R1.2-1M-2	R1.2	R1.3-1M-1	R1.3-1M-2	R1.3-1M-3	R1.3-2D-1	R1.3	R	FAIR
82	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.17
1005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.17
1300	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4.17

Table A.8: Three data points with anomalous overall FAIR score behaviors - FAIR Evaluator reusability results and the overall FAIR score.

#	R1.1[a]-Strong	R1.1-Weak	FAIR
82	0	0	65
1005	0	0	65
1300	0	0	65

Table A.9: Three data points with anomalous overall FAIR score behaviors - FAIR-Checker reusability results and the overall FAIR score.

#	R1.1	R1.2	R1.3	FAIR
82	**2**	2	2	83.33
1005	**0**	2	2	58.33
1300	**2**	2	2	75

[a] Metadata Includes License

Appendix B

Variance Inflation Factor (VIF) Analysis

Table B.1 lists the VIF values calculated for FAD features, mentioned in Section 6.3.

Table B.1: VIF values of numerical features in FAD.

#	**Feature**	**VIF**
1	FUJI-FsF-F1-01D-1	inf
2	FUJI-FsF-F1-02D-1	inf
3	FUJI-FsF-F1-02D-2	inf
4	FUJI-FsF-F2-01M-1	inf
5	FUJI-FsF-F2-01M-2	37.094440
6	FUJI-FsF-F2-01M-3	37.953441
7	FUJI-FsF-F3-01M-1	inf
8	FUJI-FsF-F3-01M-2	inf
9	FUJI-FsF-F4-01M-1	199.449917
10	FUJI-FsF-F4-01M-2	152.344475
11	FUJI-FsF-A1-01M-1	89.209754
12	FUJI-FsF-A1-03D-1	inf
13	FUJI-FsF-A1-01M-2	3.272561
14	FUJI-FsF-A1-02M-1	inf
15	FUJI-FsF-I1-01M-1	213.161364

Continued on next page

Table B.1: VIF values of numerical features in FAD. (Continued)

#	**Feature**	**VIF**
16	FUJI-FsF-I1-01M-2	117.985288
17	FUJI-FsF-I2-01M-2	100.947612
18	FUJI-FsF-I3-01M-1	197.098400
19	FUJI-FsF-I3-01M-2	33.067159
20	FUJI-FsF-R1-01MD-1	211.903380
21	FUJI-FsF-R1-01MD-2	1.484375
22	FUJI-FsF-R1-01MD-3	1.058182
23	FUJI-FsF-R1.1-01M-1	132.402574
24	FUJI-FsF-R1.1-01M-2	38.155289
25	FUJI-FsF-R1.2-01M-1	223.391595
26	FUJI-FsF-R1.2-01M-2	3.389110
27	FUJI-FsF-R1.3-01M-1	2.513935
28	FUJI-FsF-R1.3-01M-3	416.188409
29	FUJI-FsF-R1.3-02D-1	5.089485
30	FC-F1A	365.995118
31	FC-F1B	6.464984
32	FC-F2A	inf
33	FC-F2B	3590.714448
34	FC-A1.1	425.452392
35	FC-A1.2	inf
36	FC-I1	inf
37	FC-I2	8685.697545
38	FC-I3	2.432197
39	FC-R1.1	inf
40	FC-R1.2	443.343802
41	FC-R1.3	6011.938526
42	FE-F1: FAIR Metrics Gen2 - Unique Identifier	inf
43	FE-F1: FAIR Metrics Gen2 - Identifier Persistence	185.758157

Continued on next page

Table B.1: VIF values of numerical features in FAD. (Continued)

#	**Feature**	**VIF**
44	FE-F1: FAIR Metrics Gen2 - Data Identifier Persistence	1.076201
45	FE-F2: FAIR Metrics Gen2 - Structured Metadata	inf
46	FE-F2: FAIR Metrics Gen2 - Grounded Metadata	inf
47	FE-F3: FAIR Metrics Gen2 - Data Identifier Explicitly in Metadata	inf
48	FE-F3: FAIR Metrics Gen2 - Metadata Identifier Explicitly in Metadata	1.981906
49	FE-A1.1: FAIR Metrics Gen2 - Uses open free protocol for data retrieval	inf
50	FE-A1.1: FAIR Metrics Gen2 - Uses open free protocol for metadata retrieval	inf
51	FE-A1.2: FAIR Metrics Gen2 - Data authentication and authorization	inf
52	FE-A1.2: FAIR Metrics Gen2 - Metadata authentication and authorization	inf
53	FE-I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (weak)	inf
54	FE-I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong)	inf
55	FE-I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)	539.926328
56	FE-I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (strong)	189.422056
57	FE-I3: FAIR Metrics Gen2 - Metadata contains qualified outward references	543.461412
58	FE-R1.1: FAIR Metrics Gen2 - Metadata Includes License (strong)	7.049212
59	FE-R1.1: FAIR Metrics Gen2 - Metadata Includes License (weak)	8.321246

Appendix C

Explained variance ratio and cumulative explained variance for FAD Principal Components

Table C.1 lists the explained variance ratio and cumulative explained variance for FAD Principal Component, mentioned in Section 6.4.

Table C.1: Explained variance ratio and cumulative explained variance for each Principal Component.

PC	**Explained variance ratio**	**Cumulative explained variance**
1	2.6507e-01	0.2651
2	1.9800e-01	0.4631
3	1.5175e-01	0.6148
4	6.7249e-02	0.6821
5	6.5838e-02	0.7479
6	3.9147e-02	0.7871
7	3.3629e-02	0.8207
8	2.2991e-02	0.8437

Continued on next page

Table C.1: Explained variance ratio and cumulative explained variance for each Principal Component. (Continued)

PC	Explained variance ratio	Cumulative explained variance
9	1.9580e-02	0.8633
10	1.7017e-02	0.8803
11	1.6559e-02	0.8968
12	1.5745e-02	0.9126
13	1.4027e-02	0.9266
14	1.2854e-02	0.9395
15	1.1077e-02	0.9505
16	7.8382e-03	0.9584
17	6.4105e-03	0.9648
18	5.4365e-03	0.9702
19	5.0895e-03	0.9753
20	4.3140e-03	0.9796
21	3.4739e-03	0.9831
22	3.1418e-03	0.9862
23	2.5539e-03	0.9888
24-43	1.8345e-03 to 5.8833e-05	0.9906 to 0.9999
44-59	4.9430e-05 to 0.0000e+00	1.0000

Appendix D

Key Contributing Features for the five most important Principal Component in the prediction process

Table D.1 lists the most significantly contributing features across key Principal Components, i.e., PC2, PC1, PC4, PC6, and PC5, analyzed in Section 6.5.

Table D.1: The most significantly contributing features across key Principal Components (PC2, PC1, PC4, PC6, PC5).

Rank	**Feature**	**PC2**	**PC1**	**PC4**	**PC6**	**PC5**
1	FC-R1.3	✓				✓
2	FC-I2	✓				✓
3	FC-F2B	✓				✓
4	FC-I1	✓				✓
5	FC-F2A	✓				✓
6	FC-F1A	✓				
7	FC-A1.1	✓				✓

(continued)

Rank	Feature	PC2	PC1	PC4	PC6	PC5
8	FE-I2: FAIR Metrics Gen2 - Metadata uses FAIR vocabularies (weak)	✓		✓		
9	FE-I3: FAIR Metrics Gen2 - Metadata contains qualified outward references	✓		✓		
10	FE-I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (strong)	✓		✓		
11	FE-I1: FAIR Metrics Gen2 - Metadata Knowledge Representation Language (weak)	✓		✓		
12	FE-F2: FAIR Metrics Gen2 - Grounded Metadata	✓		✓		
13	FE-F2: FAIR Metrics Gen2 - Structured Metadata	✓		✓		
14	FUJI-FsF-I3-01M-1		✓			
15	FUJI-FsF-I1-01M-2		✓			
16	FUJI-FsF-R1.2-01M-1		✓		✓	
17	FUJI-FsF-R1-01MD-1		✓			
18	FUJI-FsF-R1.3-01M-3		✓		✓	
19	FUJI-FsF-A1-03D-1		✓			
20	FUJI-FsF-F3-01M-2		✓			
21	FUJI-FsF-F3-01M-1		✓			

(continued)

Rank	Feature	PC2	PC1	PC4	PC6	PC5
22	FUJI-FsF-A1-02M-1		✓			
23	FUJI-FsF-F2-01M-1		✓			
24	FUJI-FsF-I1-01M-1		✓			
25	FUJI-FsF-R1.1-01M-1		✓		✓	
26	FE-F1: FAIR Metrics Gen2 - Unique Identifier			✓		✓
27	FE-A1.2: FAIR Metrics Gen2 - Metadata authentication and authorization			✓		✓
28	FE-A1.1: FAIR Metrics Gen2 - Uses open free protocol for metadata retrieval			✓		✓
29	FUJI-FsF-F1-01D-1			✓		✓
30	FE-F3: FAIR Metrics Gen2 - Data Identifier Explicitly In Metadata			✓		
31	FE-A1.2: FAIR Metrics Gen2 - Data authentication and authorization			✓		
32	FE-A1.1: FAIR Metrics Gen2 - Uses open free protocol for data retrieval			✓		
33	FE-R1.1: FAIR Metrics Gen2 - Metadata Includes License (strong)				✓	

(continued)

Rank	Feature	PC2	PC1	PC4	PC6	PC5
34	FE-R1.1: FAIR Metrics Gen2 - Metadata Includes License (weak)				✓	
35	FC-R1.1				✓	✓
36	FC-A1.2				✓	✓
37	FUJI-FsF-A1-01M-2				✓	
38	FUJI-FsF-R1.3-01M-1				✓	
39	FC-I3				✓	
40	FUJI-FsF-R1.1-01M-1				✓	
41	FUJI-FsF-A1-01M-1				✓	
42	FE-F3: FAIR Metrics Gen2 - Metadata Identifier Explicitly In Metadata				✓	
43	FUJI-FsF-F2-01M-3				✓	

Bibliography

Abdollahi, Behnoush and Olfa Nasraoui (2018). "Transparency in fair machine learning: the case of explainable recommender systems". In: *Human and machine learning: Visible, explainable, trustworthy and transparent*, pp. 21–35.

AI Standards Hub (2024a). *Algorithmic Transparency Standard*. `https://aistandardshub.org/guidance/algorithmic-transparency-standard/`. (Visited on 05/06/2024).

AI Standards Hub (2024b). *Output from workshop on ISO/IEC standards for AI transparency and explainability – ISO/IEC AWI 12792 and ISO/IEC AWI TS 6254*. `https://aistandardshub.org/forums/topic/output-from-workshop-on-iso-iec-standards-for-ai-transparency-and-explainability/`. (Visited on 05/06/2024).

Akhtar, Usman et al. (2020). "A cache-based method to improve query performance of linked Open Data cloud". In: *Computing* 102, pp. 1743–1763.

Alvarez-Romero, Celia et al. (2022). "FAIR4Health: findable, accessible, interoperable and reusable data to foster health research". In: *Open Research Europe* 2.

Amdouni, Emna, Syphax Bouazzouni, and Clement Jonquet (2022). "O'FAIRe makes you an offer: metadata-based automatic FAIRness assessment for ontologies and semantic resources". In: *International Journal of Metadata, Semantics and Ontologies* 16.1, pp. 16–46.

Amdouni, Emna and Clement Jonquet (2021). "FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies".

In: *Research Conference on Metadata and Semantics Research*. Springer, pp. 67–80.

Andersen, Jennie, Sylvie Cazalens, and Philippe Lamarre (2021). "On the way to measure KG transparency: Formalizing transparency-Requirements and first models". Doctoral dissertation. LIRIS UMR 5205, INSA Lyon.

Archer, Margaret et al. (2016). "What is critical realism?" In: *Perspectives: A Newsletter of the ASA Theory Section* Fall.2017.

Arnold, Matthew et al. (2019). "FactSheets: Increasing trust in AI services through supplier's declarations of conformity". In: *IBM Journal of Research and Development* 63.4/5, pp. 6–1.

Assaf, Ahmad, Raphal Troncy, and Aline Senart (2015). "What's up LOD cloud? Observing the state of linked open data cloud metadata". In: *The Semantic Web: ESWC 2015 Satellite Events: ESWC 2015 Satellite Events*. Springer, pp. 247–254.

Athens Research Center for Innovation in Information, Communication and Knowledge Technologies (July 2021). *EOSC Future*. DOI: 10.3030/101017536.

Austin, Claire et al. (Apr. 2019). *WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations*. DOI: 10.15497/rda00034.

Australian Research Data Commons (2020). *AIR Data Self Assessment Tool*. URL: https://ardc.edu.au/resource/fair-data-self-assessment-tool/ (visited on 06/19/2024).

Bahim, Christophe, Carlos Casorrán-Amilburu, et al. (2020). "The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments". In: *Data Science Journal* 19, p. 41. DOI: 10.5334/dsj-2020-041.

Bahim, Christophe, Makx Dekkers, and Brecht Wyns (Jan. 2020). *Results of an Analysis of Existing FAIR Assessment Tools*. DOI: 10.15497/rda00035.

Barclay, Iain et al. (2019). "Quantifying transparency of machine learning systems through analysis of contributions". In: *arXiv preprint arXiv:1907.03483*.

Basereh, Maryam, Annalina Caputo, and Rob Brennan (2023). "Automatic transparency evaluation for open knowledge extraction systems". In: *Journal of Biomedical Semantics* 14.1, p. 12.

Berners-Lee, Tim (2006a). *Linked Data*. URL: `https://www.w3.org/DesignIssues/LinkedData`.

Berners-Lee, Tim (2006b). *Linked Data – Design Issues*. URL: `https://www.w3.org/DesignIssues/LinkedData.html` (visited on 10/21/2024).

Berners-Lee, Tim, Roy Fielding, and Larry Masinter (Jan. 2005). *Uniform Resource Identifier (URI): Generic Syntax*. RFC 3986. Internet Engineering Task Force (IETF). URL: `https://datatracker.ietf.org/doc/html/rfc3986` (visited on 08/25/2024).

Bertino, Elisa, Ahish Kundu, and Zehra Sura (2019). "Data transparency with blockchain and AI ethicsgebru2021datasheets". In: *Journal of Data and Information Quality (JDIQ)* 11.4, pp. 1–8.

Bertino, Elisa, Shawn Merrill, et al. (2019). "Redefining data transparency: A multidimensional approach". In: *Computer* 52.1, pp. 16–26.

Bhaskar, Krish (1979). "A multiple objective approach to capital budgeting". In: *Accounting and Business Research* 10.37, pp. 25–46.

Bhaskar, Roy et al. (1998). "Critical realism". In: *Proceedings of the standing conference on realism and human sciences, Bristol, UK*. Vol. 4, pp. 1–.

Bioschemas (2024). *Bioschemas Profiles*. URL: `https://bioschemas.org/profiles/` (visited on 06/18/2024).

BioSharing (2024). *BioSharing*. URL: `https://fairsharing.org` (visited on 06/18/2024).

Bishop, Bradley Wade and Carolyn Hank (2018). "Measuring FAIR principles to inform fitness for use". In: *International Journal of Digital Curation* 13.1, pp. 35–46.

Bizer, Christian, Tom Heath, and Tim Berners-Lee (2023). "Linked Data - The Story So Far". In: *Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web*. 1st ed. New York, NY, USA: Association for Computing Machinery, pp. 115–143. ISBN: 9798400707940. URL: https://doi.org/10.1145/3591366.3591378.

Bonino-da-Silva-Santos, Luiz Olavo (2022). *FAIR Digital Object Framework Documentation, Working Draft*. URL: https://fairdigitalobjectframework.org (visited on 06/02/2024).

Bourne, Philip E et al. (2015). "The NIH big data to knowledge (BD2K) initiative". In: *Journal of the American Medical Informatics Association* 22.6, pp. 1114–1114.

Breiman, Leo (2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32. DOI: 10.1023/A:1010933404324.

Bryman, Alan (2016). *Social Research Methods*. Oxford University Press.

Cambridge University Press (2024). *Cambridge Dictionary*. URL: https://dictionary.cambridge.org/ (visited on 07/31/2024).

Caraballo, Alexander Arturo Mera et al. (2016). "Automatic creation and analysis of a linked data cloud diagram". In: *Web Information Systems Engineering–WISE 2016: 17th International Conference, Shanghai, China, November 8-10, 2016, Proceedings, Part I 17*. Springer, pp. 417–432.

Caufield, J Harry et al. (June 2023). "KG-Hub—building and exchanging biological knowledge graphs". In: *Bioinformatics* 39.7, btad418. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btad418.

Choy, Looi Theam (2014). "The Strengths and Weaknesses of Research Methodology: Comparison and Complimentary between Qualitative and Quantitative Approaches". In: *IOSR Journal of Humanities and Social Science* 19.4, pp. 99–104.

CKAN Project (2024). *CKAN User Guide*. URL: https://docs.ckan.org/en/latest/user-guide.html (visited on 10/21/2024).

Clarke, Daniel JB et al. (2019). "FAIRshake: toolkit to evaluate the FAIRness of research digital resources". In: *Cell systems* 9.5, pp. 417–421.

Coleti, Thiago Adriano et al. (2020). “TR-model. A metadata profile application for personal data transparency”. In: *IEEE Access* 8, pp. 75184–75209.

Creswell, John W. and J. David Creswell (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.

Creswell, John W. and J. David Creswell (2018). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches*. 5th. California: Sage Publications.

da-Cruz, Nuno Ferreira et al. (2016). “Measuring local government transparency”. In: *Public Management Review* 18.6, pp. 866–893.

Danermark, Berth (2002). “Interdisciplinary research and critical realism: The example of disability research”. In: *Alethia* 5.1, pp. 56–64.

Daneshjou, Roxana et al. (2021). “Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review”. In: *JAMA dermatology* 157.11, pp. 1362–1369.

DataCite (2022). *DataCite*. URL: `https://datacite.org` (visited on 06/18/2024).

Dataverse Project (2024). *Dataverse*. URL: `https://dataverse.org/` (visited on 06/18/2024).

Datta, Anupam, Shayak Sen, and Yair Zick (2016). “Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems”. In: *2016 IEEE symposium on security and privacy (SP)*. IEEE, pp. 598–617.

David, Romain et al. (Nov. 2018). *Operationalizing and evaluating the FAIRness concept for a good quality of data sharing in Research: the RDA-SHARC-IG (SHAring Rewards and Credit Interest Group*. Version 1.0. DOI: `10.5281/zenodo.1745374`.

Debattista, Jeremy, Judie Attard, et al. (2019). “Is the LOD cloud at risk of becoming a museum for datasets? Looking ahead towards a fully collaborative and sustainable LOD cloud”. In: *Companion Proceedings of The 2019 World Wide Web Conference*, pp. 850–858.

Debattista, Jeremy, Søren Auer, and Christoph Lange (2016). "Luzzu—a methodology and framework for linked data quality assessment". In: *Journal of Data and Information Quality (JDIQ)* 8.1, pp. 1–32.

Debattista, Jeremy, Christoph Lange, et al. (2018). "Evaluating the quality of the LOD cloud: An empirical investigation". In: *Semantic Web* 9.6, pp. 859–901.

Denzin, Norman K. and Yvonna S. Lincoln, eds. (2011). *The SAGE Handbook of Qualitative Research*. 3rd ed. Thousand Oaks, CA: SAGE Publications. ISBN: 9781412974172.

Devaraju, Anusuriya and Robert Huber (2020). *F-UJI - An Automated FAIR Data Assessment Tool*. DOI: 10.5281/zenodo.6361400.

Devaraju, Anusuriya, Robert Huber, et al. (July 2020). *FAIRsFAIR data object assessment metrics*. Version 0.3. DOI: 10.5281/zenodo.

Devaraju, Anusuriya, Mustapha Mokrane, et al. (2021). "From conceptualization to implementation: FAIR assessment of research data objects". In: *Data Science Journal* 20.1, pp. 1–14.

Diakopoulos, Nicholas (2016). "Accountability in algorithmic decision making". In: *Communications of the ACM* 59.2, pp. 56–62.

Dørpinghaus, Jens and Andreas Stefan (2019). "Knowledge extraction and applications utilizing context data in knowledge graphs". In: *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, pp. 265–272.

Dudovskiy, John (2020). *Positivism - Research Methodology*. URL: https://research-methodology.net/research-philosophy/positivism/ (visited on 12/01/2024).

Duque-Ramos, Astrid et al. (2011). "OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies". In: *Journal of research and practice in information technology* 43.2, pp. 159–176.

Dutch Data Archiving and Networked Services (2020). *SATIFYD Self-Assessment Tool*. URL: https://dans.knaw.nl/en/satifyd/ (visited on 06/19/2024).

Dutch Institute for Public Health and Environment (2024). *RIVM*. URL: https://www.rivm.nl/en (visited on 06/19/2024).

Dutch Organisation for Health Research and Development (2024). *Dutch Organisation for Health Research and Development*. URL: https://www.zonmw.nl/en (visited on 06/18/2024).

Eijssen, Lars et al. (2015). "The Dutch Techcentre for Life Sciences: enabling data-intensive life science research in the Netherlands". In: *F1000Research* 4.

EOSC (2024). *Knuth: Computers and Typesetting*. URL: https://eosc.eu/advisory-groups/fair-metrics-and-data-quality/ (visited on 06/17/2024).

EOSC Task Force FAIR Metrics and Data Quality (2020). *EOSC Task Force FAIR Metrics and Data Quality Charter*. Version 1.0. URL: https://www.eosc.eu/sites/default/files/2021-12/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf.

EOSC-Nordic (2024). *EOSC-Nordic*. URL: https://www.eosc-nordic.eu (visited on 06/18/2024).

EOSC-Pillar (2024). *EOSC-Pillar*. URL: https://www.eosc-pillar.eu (visited on 06/18/2024).

EOSC-synergy (2024). *EOSC-synergy*. URL: https://www.eosc-synergy.eu (visited on 06/18/2024).

European Commission (2019). *Ethics guidelines for trustworthy AI*. URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (visited on 06/17/2024).

European Commission (2024a). *Research and innovation*. URL: https://commission.europa.eu/research-and-innovation_en?pg=open-science-cloud (visited on 06/17/2024).

European Commission (2024b). *European Open Science Cloud (EOSC) - What the cloud is, how it was developed and being implemented*. URL: https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en (visited on 06/17/2024).

European Commission Directorate-General for Research and Innovation (July 2016). *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*. Version 3.0. URL: `https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf`.

European Parliament and Council of the European Union (May 4, 2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council.* of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). URL: `https://data.europa.eu/eli/reg/2016/679/oj` (visited on 04/13/2023).

European Parliament and Council (2024). "Artificial Intelligence Act". In: *OJ L* 2024.1689.

European Parliamentary Research Service (2019). *A governance framework for algorithmic accountability and transparency.*

European Union (2024). *European Union Open Data Portal.* `https://data.europa.eu/euodp/en/home`. (Visited on 10/21/2024).

European Union Agency For Network and Information Security (ENISA) (Dec. 2018). *Recommendations on Shaping Technology According to GDPR Provisions: Exploring the Notion of Data Protection by Default.* `https://www.aepd.es/documento/recomendations-shaping-technology-according-gdpr-provisions-2.pdf`. (Visited on 06/17/2024).

ExPaNDS (2024). *ExPaNDS*. URL: `https://expands.eu/` (visited on 06/18/2024).

FAIR-IMPACT (2024). *FAIR-IMPACT: Fostering FAIR Data Practices and Services*. URL: `https://fair-impact.eu/` (visited on 06/18/2024).

FAIRdom (2024). *FAIRdom*. URL: `https://fair-dom.org/` (visited on 06/18/2024).

FAIRplus (2019-2022). *The IMI FAIRplus. FAIR Cookbook.* URL: `https://fairplus.github.io/the-fair-cookbook` (visited on 06/17/2024).

FAIRsFAIR (2019-2022). *FAIRsFAIR - Fostering FAIR Data Practices in Europe.* URL: `https://www.fairsfair.eu/the-project` (visited on 06/17/2024).

FAIRsFAIR Project (2024). *FAIRsFAIR Data Object Assessment Metrics*. https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments. (Visited on 11/22/2024).

FAIRsharing Team Data Readiness Group (2024). *FAIRassist: Assisting with FAIR Data Practices*. URL: https://fairassist.org/ (visited on 06/18/2024).

Felzmann, Heike et al. (2020). "Towards transparency by design for artificial intelligence". In: *Science and engineering ethics* 26.6, pp. 3333–3361.

FORCE11 (2011). *About FORCE11*. URL: https://force11.org/info/about-force11/ (visited on 06/02/2024).

FORCE11 (2024). *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0*. URL: https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0/ (visited on 06/02/2024).

G7 (2023). *G7 2023 Leaders' Communique*. URL: https://www.g7uk.org/g7-leaders-communique-2023/ (visited on 08/25/2024).

Gaignard, Alban et al. (2023). "FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards". In: *Journal of Biomedical Semantics* 14.1, pp. 1–14.

Garijo, Daniel, Oscar Corcho, and María Poveda-Villalón (2021). "FOOPS!: An Ontology Pitfall Scanner for the FAIR principles." In: *ISWC (Posters/Demos/Industry)*.

Gasser, Urs and Virgilio AF Almeida (2017). "A layered model for AI governance". In: *IEEE Internet Computing* 21.6, pp. 58–62.

Gatti, Roberta et al. (2024). "Data transparency and GDP growth forecast errors". In: *Journal of International Money and Finance* 140, p. 102991.

Gebru, Timnit et al. (2021). "Datasheets for datasets". In: *Communications of the ACM* 64.12, pp. 86–92.

Geiger, Christian Philipp and Jörn Von Lucke (2012). "Open government and (linked)(open)(governm
In: *JeDEM-eJournal of eDemocracy and open Government* 4.2, pp. 265–278.

GeoData (2020). *GeoData dataset*. Hosted by PANGAEA. DOI: `10.1594/PANGAEA.908011`. URL: `https://doi.org/10.1594/PANGAEA.908011`.

Glesne, Corrine (2016). *Becoming qualitative researchers: An introduction*. One Lake Street, Upper Saddle River, New Jersey 07458: Pearson.

Goodman, Bryce and Seth Flaxman (2017). "European Union regulations on algorithmic decision-making and a "right to explanation"". In: *AI magazine* 38.3, pp. 50–57.

Gorski, Philip S. (2018). "After positivism: Critical realism and historical sociology". In: *Critical realism, history, and philosophy in the social sciences*. Emerald Publishing Limited, pp. 23–45.

Gruber, Thomas R (1993). "A translation approach to portable ontology specifications". In: *Knowledge acquisition* 5.2, pp. 199–220.

Guidotti, Riccardo et al. (2018). "A survey of methods for explaining black box models". In: *ACM computing surveys (CSUR)* 51.5, pp. 1–42.

Haibe-Kains, Benjamin et al. (2020). "Transparency and reproducibility in artificial intelligence". In: *Nature* 586.7829, E14–E16.

Haller, Armin et al. (2020). "What are links in linked open data? A characterization and evaluation of links between knowledge graphs on the web". In: *Journal of Data and Information Quality (JDIQ)* 12.2, pp. 1–34.

Hasnain, Ali and Dietrich Rebholz-Schuhmann (2018). "Assessing FAIR data principles against the 5-star open data principles". In: *The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers 15*. Springer, pp. 469–477.

Health Research Board (2024). *Health Research Board*. URL: `https://www.hrb.ie` (visited on 06/18/2024).

Heath, Tom and Christian Bizer (2011). *Linked data: Evolving the web into a global data space*. Vol. 1. Morgan & Claypool Publishers.

Hitzler, Pascal (2021). "A review of the semantic web field". In: *Communications of the ACM* 64.2, pp. 76–83.

Ho, CWL et al. (2019). “Governance of automated image analysis and artificial intelligence analytics in healthcareshin2020user”. In: *Clinical radiology* 74.5, pp. 329–337.

Hodson, Simon et al. (June 2018). *FAIR Data Action Plan. Interim recommendations and actions from the European Commission Expert Group on FAIR data.* Version Jun 7, 2018. DOI: 10.5281/zenodo.1285289.

Information Commissioner’s Office (July 2020). *Guidance on the AI auditing framework: Draft guidance for consultation.* URL: https://ico.org.uk/media/2617219/ico-guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf (visited on 06/13/2024).

International Organization for Standardization (2023). *ISO/IEC DTS 12791 (Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks).* https://www.iso.org/standard/84110.html. Status: Under development, Stage: Proof sent to secretariat or FDIS ballot initiated: 8 weeks [50.20], Edition: 1, Number of pages: 24, Technical Committee: ISO/IEC JTC 1/SC 42, ICS: 35.020.

International Organization for Standardization (2024). *Building a responsible AI: How to manage the AI ethics debate.* https://www.iso.org/artificial-intelligence/responsible-ai-ethics. (Visited on 05/06/2024).

International Organization for Standardization and International Electrotechnical Commission (2023a). *ISO/IEC 23894:2023 (Information technology — Artificial intelligence — Guidance on risk management).* https://www.iso.org/standard/77304.html. Status: Published, Publication date: 2023-02, Stage: International Standard published [60.60], Edition: 1, Number of pages: 26, Technical Committee: ISO/IEC JTC 1/SC 42, ICS: 35.020.

International Organization for Standardization and International Electrotechnical Commission (Dec. 2023b). *ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system.* International Standard published [60.60]. Edition: 1, Number of pages: 51, Technical Committee: ISO/IEC JTC

1/SC 42, ICS: 35.020, 03.100.70. URL: https://www.iso.org/standard/ISO42001.

Internet Engineering Task Force (2024). *Internet Engineering Task Force*. URL: https://www.ietf.org/ (visited on 06/18/2024).

Iturbide, Maialen et al. (2022). “Implementation of FAIR principles in the IPCC: the WGI AR6 Atlas repository”. In: *Scientific data* 9.1, p. 629.

Jacobsen, Annika et al. (2020). “FAIR principles: interpretations and implementation considerations”. In: *Data intelligence* 2.1-2, pp. 10–29.

Jagodnik, Kathleen M et al. (2017). “Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop”. In: *Journal of biomedical informatics* 71, pp. 49–57.

Jentzsch, Anja, Richard Cyganiak, and Christian Bizer (2011). *State of the LOD Cloud*. URL: http://lod-cloud.net/state/ (visited on 06/21/2024).

Jia, Shengbin et al. (2019). “Triple trustworthiness measurement for knowledge graph”. In: *The World Wide Web Conference*, pp. 2865–2871.

Jobin, Anna, Marcello Ienca, and Effy Vayena (2019). “The global landscape of AI ethics guidelines”. In: *Nature Machine Intelligence* 1.9, pp. 389–399.

Kontokostas, Dimitris et al. (2014). “Test-Driven Evaluation of Linked Data Quality”. In: *Proceedings of the 23rd International Conference on World Wide Web*. Association for Computing Machinery, New York, NY, USA, pp. 747–758.

Krans, N. A. et al. (2022). “FAIR assessment tools: evaluating use and performance”. In: *NanoImpact* 27, p. 100402.

Lawson, Tony (1997). *Economics and Reality*. London and New York: Routledge. Chap. Subjectivism.

Leaders, G (2016). *G20 Leaders ‘Communique (Hangzhou Summit)*.

Lee, Min Kyung et al. (2015). “Working with machines: The impact of algorithmic and data-driven management on human workers”. In: *Proceedings of the 33rd* *annual ACM conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, United States, pp. 1603–1612.

Legendre, Adrien-Marie (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris: F. Didot.

Lepri, Bruno et al. (2018). "Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges". In: *Philosophy & Technology* 31.4, pp. 611–627.

Lia, Matteo et al. (2023). "euFAIR: A Digital Tool for Assessing the FAIR Principles". In: *International Conference on Conceptual Modeling*. Springer, pp. 49–58.

Liefgreen, Alice et al. (2023). "Beyond ideals: why the (medical) AI industry needs to motivate behavioural change in line with fairness and transparency values, and how it can do it". In: *AI & SOCIETY*, pp. 1–17.

Lundberg, Scott M. and Su-In Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, pp. 4765–4774. URL: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Mahalanobis, P. C. (1936). "On the generalized distance in statistics". In: *Proceedings of the National Institute of Sciences (India)* 2.1, pp. 49–55.

Manghi, Paolo et al. (Dec. 2022). *OpenAIRE Graph Dataset*. Version 5.0.0. Zenodo. DOI: 10.5281/zenodo.7488618.

Margolis, Ronald et al. (2014). "The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data". In: *Journal of the American Medical Informatics Association* 21.6, pp. 957–958.

McCrae, John P. et al. (2024). *Linked Open Data Cloud*. URL: http://lod-cloud.net (visited on 10/21/2024).

Menke, Joe et al. (2022). "Establishing institutional scores with the rigor and transparency index: large-scale analysis of scientific reporting quality". In: *Journal of Medical Internet Research* 24.6, e37324.

Mihindukulasooriya, Nandana (2020). "A framework for linked data quality based on data profiling and rdf shape induction". PhD thesis. ETSI_Informatica.

Miller, Grant and Laurette Pretorius (2022). "Institutional Repositories in the Linked Open Data Cloud". In: *Journal of the Digital Humanities Association of Southern Africa* 4.02.

Miranda Azevedo, Ricardo de and Michel Dumontier (2020). "Considerations for the conduction and interpretation of FAIRness evaluations". In: *Data Intelligence* 2.1-2, pp. 285–292.

Molavi-Vasse'i, Ramak and Jesse McCrosky (Mar. 2023). *AI Transparency in Practice*. Accessed: 2025-05-13. URL: https://foundation.mozilla.org/en/research/library/ai-transparency-in-practice/ai-transparency-in-practice/.

Mons, Barend, Cameron Neylon, et al. (2017). "Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud". In: *Information services & use* 37.1, pp. 49–56.

Mons, Barend, Erik Schultes, et al. (2020). *The FAIR principles: First generation implementation choices and challenges*.

Mons, Bernard (2017). *GO-FAIR initiative – Dutch Techcentre for Life Sciences*. URL: http://www.dtls.nl/go-fair/ (visited on 06/17/2024).

Montavon, Grégoire, Wojciech Samek, and Klaus-Robert Müller (2018). "Methods for Interpreting and Understanding Deep Neural Networks". In: *Digital Signal Processing* 73, pp. 1–15. DOI: 10.1016/j.dsp.2017.10.011.

Morley, Jessica et al. (Nov. 2019). *The Debate on the Ethics of AI in Health Care: A Reconstruction and Critical Review*. SSRN. Preprint. DOI: 10.2139/ssrn.3486518. URL: https://ssrn.com/abstract=3486518 (visited on 12/01/2024).

Muralidhar, Deepa et al. (2023). “Elements that Influence Transparency in Artificial Intelligent Systems-A Survey”. In: *IFIP Conference on Human-Computer Interaction*. Springer, pp. 349–358.

Nogales, Alberto, Miguel Angel Sicilia-Urban, and Elena García-Barriocanal (2017). “Measuring vocabulary use in the Linked Data Cloud”. In: *Online Information Review* 41.2, pp. 252–271.

O’Donoghue, Tom (2018). *Planning Your Qualitative Research Thesis and Project: An Introduction to Interpretivist Research in Education and the Social Sciences*. 2nd. eBook edition. London: Routledge. ISBN: 9781351165563. DOI: `10.4324/9781351165563`. URL: `https://doi.org/10.4324/9781351165563` (visited on 12/01/2024).

Oliveira, Natalia Queiroz de et al. (2021). “A practical approach of actions for FAIRification workflows”. In: *Research Conference on Metadata and Semantics Research*. Springer, pp. 94–105.

Open Knowledge Foundation (2007). *DataHub*. URL: `https://old.datahub.io` (visited on 10/21/2024).

Open Knowledge Foundation (2015). *The Open Definition*. URL: `http://opendefinition.org` (visited on 10/21/2024).

Open Knowledge International (2024). *CKAN: The open-source data portal software*. URL: `https://ckan.org` (visited on 10/21/2024).

OpenAIRE (2018). *OpenAIRE*. URL: `https://www.openaire.eu/about` (visited on 06/17/2024).

Panch, Trishan, Heather Mattie, and Leo Anthony Celi (2019). “The “inconvenient truth” about AI in healthcare”. In: *NPJ digital medicine* 2.1, pp. 1–3.

Papadopoulou, Elli et al. (Mar. 2024). *Report on FAIR Evaluation community survey*. DOI: `10.5281/zenodo.10797765`. (Visited on 06/17/2024).

Peroni, Silvio (2016). *Media type as Linked Open Data*. URL: `http://www.sparontologies.net/mediatype/` (visited on 10/21/2024).

Peters-von Gehlen, Karsten et al. (2022). “Recommendations for discipline-specific FAIRness evaluation derived from applying an ensemble of evaluation tools”. In: *Data Science Journal* 21, pp. 7–7.

Petrosyan, Luiza et al. (2023). “FAIR degree assessment in agriculture datasets using the F-UJI tool”. In: *Ecological Informatics* 76, p. 102126.

Polleres, Axel et al. (2018). “A More Decentralized Vision for Linked Data”. In: *Proceedings of the 2nd Workshop on Decentralizing the Semantic Web, co-located with the International Semantic Web Conference (ISWC), DeSemWebISWC*. Vol. 2165. Monterey, CA, USA: CEUR-WS.org, p. 8.

Poveda-Villalón, María et al. (2020). “Coming to Terms with FAIR Ontologies”. In: *International Conference on Knowledge Engineering and Knowledge Management*. Springer, pp. 255–270.

PresQT (2024). *Preservation Quality Tool Developers*. URL: `https://presqt.crc.nd.edu` (visited on 06/18/2024).

Psillos, Stathis (2007). “Causal explanation and manipulation”. In: *Rethinking explanation*. Dordrecht: Springer Netherlands, pp. 93–107.

Pushkarna, Mahima, Andrew Zaldivar, and Oddur Kjartansson (2022). “Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AIcoleti2020tr”. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, New York, NY, USA, pp. 1776–1826.

pySHACL Developers (2024). *pySHACL: Python SHACL Validator*. URL: `https://github.com/RDFLib/pySHACL` (visited on 06/18/2024).

Queralt-Rosinach, Núria et al. (2022). “Applying the FAIR principles to data in a hospital: challenges and opportunities in a pandemic”. In: *Journal of biomedical semantics* 13.1, p. 12.

Ray, Christopher (2017). *Logical Positivism*, pp. 243–251. DOI: `https://doi.org/10.1002/9781405164481.ch37`.

RDA FAIR Data Maturity Model Working Group (June 2020). *FAIR Data Maturity Model: Specification and Guidelines*. Research Data Alliance. RDA Recommendation. DOI: `10.15497/rda00050`. (Visited on 12/01/2024).

RDFlib Developers (2024). *RDFlib: A Python Library for Working with RDF*. URL: `https://github.com/RDFLib/rdflib` (visited on 06/18/2024).

Reddy, Sandeep et al. (2020). "A governance model for the application of AI in health care". In: *Journal of the American Medical Informatics Association* 27.3, pp. 491–497.

Reinanda, Ridho, Edgar Meij, Maarten de Rijke, et al. (2020). "Knowledge graphs: An information retrieval perspective". In: *Foundations and Trends® in Information Retrieval* 14.4, pp. 289–444.

Reitz, Kenneth (2024). *Requests: HTTP for Humans*. URL: `https://pypi.org/project/requests/` (visited on 06/18/2024).

Richards, Keith (2003). *Qualitative Inquiry in TESOL*. London: Palgrave Macmillan. ISBN: 9781403903894. DOI: `10.1057/9781403903894`. (Visited on 12/01/2024).

Ritchie, Jane et al. (2013). *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. 2nd ed. London: SAGE Publications. ISBN: 9781446209127.

Rodriguez-Doncel, Victor, Serena Villata, and Asuncion Gomez-Perez (2014). "A dataset of RDF licenses". In: *Legal Knowledge and Information Systems - JURIX 2014: The Twenty-Seventh Annual Conference, Jagiellonian University, Krakow, Poland, 10-12 December 2014*. Ed. by Rinke Hoekstra. Vol. 271. Frontiers in Artificial Intelligence and Applications. IOS Press, pp. 187–188. DOI: `10.3233/978-1-61499-468-8-187`.

Ryan, Gemma (2018). "Introduction to positivism, interpretivism and critical theory". In: *Nurse Researcher* 25.4, pp. 41–49.

Al-Saadi, Hashil (2014). "Demystifying Ontology and Epistemology in research methods". In: *Research Gate* 1.1, pp. 1–10.

Samuel, Sheeba, Frank Löffler, and Birgitta König-Ries (2020). "Machine learning pipelines: provenance, reproducibility and FAIR data principles". In: *International Provenance and Annotation Workshop*. Springer, pp. 226–230.

Sansone, Susanna-Assunta et al. (2019). "FAIRsharing as a community approach to standards, repositories and policies". In: *Nature biotechnology* 37.4, pp. 358–367.

Sayer, Andrew (1992). *Method in Social Science: A Realist Approach*. 2nd ed. London: Routledge. ISBN: 9780415076071.

Schmachtenberg, Max, Christian Bizer, and Heiko Paulheim (2014). "Adoption of the linked data best practices in different topical domains". In: *The Semantic Web–ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I 13*. Springer, pp. 245–260.

Science Europe (2024). *Science Europe*. URL: `https://scienceeurope.org` (visited on 06/18/2024).

Scrapinghub (2024). *Extruct: A Library for Extracting Embedded Metadata from HTML*. URL: `https://pypi.org/project/extruct/` (visited on 06/18/2024).

Selenium Project (2024). *SeleniumHQ Browser Automation*. URL: `https://www.selenium.dev/` (visited on 06/18/2024).

Shin, Donghee (2019). "Toward fair, accountable, and transparent algorithms: Case studies on algorithm initiatives in Korea and China". In: *Javnost-The Public* 26.3, pp. 274–290.

Sofi-Mahmudi, Ahmad and Eero Raittio (2022). "Transparency of COVID-19-Related Research in Dental Journals". In: *Frontiers in oral health* 3, p. 871033.

Soler Garrido, Josep et al. (2023). *Analysis of the preliminary AI standardisation work plan in support of the AI Act*. Tech. rep. JRC132833. Luxembourg: Publications Office of the European Union. DOI: `10.2760/5847`.

Suchánek, Marek and Robert Pergl (2018). "Data stewardship wizard for open science". In: *Data A Znalosti WIKT*, pp. 121–125.

Sun, Chang, Vincent Emonet, and Michel Dumontier (2022). "A comprehensive comparison of automated FAIRness Evaluation Tools". In: *13th International*

Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, pp. 44–53.

Sun, Tara Qian and Rony Medaglia (2019). "Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare". In: *Government Information Quarterly* 36.2, pp. 368–383.

Swiss National Science Foundation (2024). *Swiss National Science Foundation.* URL: `https://www.snf.ch/en/FKhU9kAtfXx7w9AI/page/home` (visited on 06/18/2024).

Szarkowska, Kinga et al. (2021). "Quality Assessment of Knowledge Graph Hierarchies using KG-BERT." In: *DL4KG@ ISWC*.

Tennis, Joseph T. (2008). "Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework". In: *Knowledge Organization* 35.2/3, pp. 102–112.

University of Mannheim, Data and Web Science Group (2014a). *Linked Data Catalog - About.* URL: `http://linkeddatacatalog.dws.informatik.uni-mannheim.de/about` (visited on 10/21/2024).

University of Mannheim, Data and Web Science Group (2014b). *Mannheim Linked Data Catalog.* URL: `http://linkeddatacatalog.dws.informatik.uni-mannheim.de` (visited on 10/21/2024).

Vandenbussche, Pierre-Yves et al. (2017). "SPARQLES: Monitoring public SPARQL endpoints". In: *Semantic Web* 8.6, pp. 1049–1065. DOI: `10.3233/SW-170254`. URL: `https://doi.org/10.3233/SW-170254`.

Viinikkala, Mika (2004). *Ontology in Information Systems.* Accessed: 2025-05-13. URL: `http://www.cs.tut.fi/~kk/webstuff/Ontology.pdf`.

Vogt, Lars et al. (2024). "FAIR 2.0: Extending the FAIR Guiding Principles to Address Semantic Interoperability". In: *arXiv preprint arXiv:2405.03345*.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi (2017). "Transparent, explainable, and accountable AI for robotics". In: *Science robotics* 2.6, eaan6080.

Wang, Lucy Lu et al. (2020). "Cord-19: The covid-19 open research dataset". In: *ArXiv*.

Wang, Richard Y and Diane M Strong (1996). "Beyond accuracy: What data quality means to data consumers". In: *Journal of management information systems* 12.4, pp. 5–33.

Wang, Weiyu and Keng Siau (2018). "Artificial Intelligence: A Study on Governance, Policies, and Regulations". In: *MWAIS 2018 Proceedings*. 40. URL: `https://aisel.aisnet.org/mwais2018/40`.

Welsh, Ciaran et al. (2021). "libOmexMeta: enabling semantic annotation of models to support FAIR principles". In: *Bioinformatics* 37.24, pp. 4898–4900.

Whyte, Angus et al. (Sept. 2021). *D3.2 FAIR Data Practice Analysis*. Version 1.0. DOI: `10.5281/zenodo.5362079`.

Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1, pp. 1–9.

Wilkinson, Mark D, Michel Dumontier, Susanna-Assunta Sansone, et al. (2019). "Evaluating FAIR maturity through a scalable, automated, community-governed framework". In: *Scientific data* 6.1, p. 174.

Wilkinson, Mark D, Susanna-Assunta Sansone, Marjan Grootveld, et al. (Jan. 2024). *Report on FAIR Signposting and its Uptake by the Community*. DOI: `10.5281/zenodo.10490289`.

Wilkinson, Mark D, Susanna-Assunta Sansone, Grootveld Marjan, et al. (Dec. 2022). *FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons*. DOI: `10.5281/zenodo.7463421`.

Wilkinson, Mark D, Susanna-Assunta Sansone, Erik Schultes, et al. (2018). "A design framework and exemplar metrics for FAIRness". In: *Scientific data* 5.1, pp. 1–4.

Wilkinson, Mark D. et al. (Dec. 2022). *Community-driven Governance of FAIRness Assessment: An Open Issue, an Open Discussion*. Version Final. DOI: `10.5281/zenodo.7390482`.

Winfield, Alan et al. (2022). *IEEE Standard for Transparency of Autonomous Systems*. DOI: 10.1109/IEEESTD.2022.9726144. URL: https://uwe-repository.worktribe.com/output/9233640.

Winfield, Alan F et al. (2019). "Machine ethics: The design and governance of ethical AI and autonomous systems [scanning the issue]". In: *Proceedings of the IEEE* 107.3, pp. 509–517.

Wirtz, Bernd W, Jan C Weyerer, and Carolin Geyer (2019). "Artificial intelligence and the public sector—applications and challenges". In: *International Journal of Public Administration* 42.7, pp. 596–615.

Wolf, Christine T (2020). "From knowledge graphs to knowledge practices: On the need for transparency and explainability in enterprise knowledge graph applications". In: *Proceedings of the KG-BIAS Workshop 2020 at AKBC 2020*.

World Data Center for Climate (2024). *WDCC*. URL: https://cera-www.dkrz.de/WDCC/ui/cerasearch/ (visited on 06/19/2024).

World Wide Web Consortium (2013). *SPARQL 1.1 Overview*. URL: https://www.w3.org/TR/sparql11-overview/ (visited on 06/18/2024).

World Wide Web Consortium (2014a). *Data Catalog Vocabulary (DCAT)*. Tech. rep. W3C Recommendation, 16 January 2014. World Wide Web Consortium (W3C). URL: https://www.w3.org/TR/vocab-dcat/.

World Wide Web Consortium (2014b). *RDF 1.1 Concepts and Abstract Syntax*. https://www.w3.org/TR/rdf11-concepts/. Available at: https://www.w3.org/TR/rdf11-concepts/.

World Wide Web Consortium (2017). *SHACL: Shapes Constraint Language*. URL: https://www.w3.org/TR/shacl/ (visited on 06/18/2024).

World Wide Web Consortium (2024). *World Wide Web Consortium*. URL: https://www.w3.org/ (visited on 06/18/2024).

Wu, Fei and Daniel S Weld (2010). "Open information extraction using wikipedia". In: *Proceedings of the 48th annual meeting of the association for computational* *linguistics*. Association for Computational Linguistics, United States, pp. 118–127.

Wynants, Laure et al. (2020). "Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal". In: *BMJ* 369. Accessed: 2025-05-13, p. m1328. DOI: 10.1136/bmj.m1328. URL: https://www.bmj.com/content/369/bmj.m1328.

Yu, Jonathan and Simon Cox (2017). "5-Star Data Rating Tool". In: *CSIRO Software Collection* 6. DOI: 10.4225/08/5a12348f8567b.

Zaveri, Amrapali et al. (2016). "Quality Assessment for Linked Data: A Survey". In: *Semantic Web* 7.1, pp. 63–93. DOI: 10.3233/SW-150175.

Zenodo (2024). *Zenodo*. URL: https://zenodo.org/ (visited on 06/18/2024).

Zerilli, John, Umang Bhatt, and Adrian Weller (2022). "How Transparency Modulates Trust in Artificial Intelligence". In: *Patterns* 3.4, pp. 1–10. DOI: 10.1016/j.patter.2022.100455.

Zhang, Lishan (2013). "How structured data (Linked Data) help in Big Data Analysis-Expand Patent Data with Linked Data Cloud". In: *Electrical Engineering and Computer Sciences University of California at Berkeley*.

Žukauskas, Pranas, Jolita Vveinhardt, and Regina Andriukaitienė (2018). "Philosophy and paradigm of scientific research". In: *Management culture and corporate social responsibility* 121.13, pp. 506–518.