A large-scale empirical analysis of FAIR compliance in biomedical KGs using KGHeartBeat and FAIR-Checker*

Sana Latif[1,*]

[1]University of Salerno - Via Giovanni Paolo II, 132 - 84084 Fisciano (SA)

Abstract

The increasing availability of automated tools for assessing data quality and FAIR compliance has enabled large-scale evaluations of publicly available datasets. However, a major challenge remains: many of these tools operationalize FAIR principles differently, resulting in assessments that are not directly comparable. Consequently, the same dataset may receive different evaluations depending on the tool used, potentially leading to inconsistent interpretations of its FAIRness and quality. This work investigates this potential misalignment through a quantitative comparative analysis of FAIR compliance assessment tools. In particular, it compares two state-of-the-art systems, namely KGHeartBeat and FAIR-Checker, to evaluate the extent to which their results align when applied to the same datasets. As an empirical case study, it focuses on biomedical knowledge graphs from the Biomedical Linked Open Data (BLOD) collection. The analysis considers a subset of 1,209 datasets from a larger corpus of 1, 314 datasets, selecting those that can be successfully evaluated by both tools. We then perform statistical correlation analyses using Spearman measures to examine the degree of agreement between the metrics produced by the two systems. The results reveal moderate overall FAIR compliance across biomedical datasets, with stronger performance in Findability and weaker performance in Reusability. More importantly, the analysis highlights varying levels of correlation between metrics, revealing both areas of alignment and significant discrepancies in how different tools operationalize specific FAIR principles. This study provides a large-scale empirical comparison of automated FAIR assessment tools and offers insights into the strengths and limitations of current evaluation methodologies. Our findings highlight the need for greater methodological alignment to ensure that FAIR assessments are reliable, interpretable, and comparable across tools.

Keywords

FAIR principles, Knowledge graphs, Biomedical Linked Data, Empirical investigation, Correlation analysis

1. Introduction

The FAIR Principles are a set of guidelines designed to improve the infrastructure of digital research data, making it Findable, Accessible, Interoperable, and Reusable for both humans and machines [1]. The FAIR principles have emerged as a transformative framework to address issues related to data management, integration, and reuse by providing guidelines for effective data stewardship across scientific disciplines. In this context, Knowledge Graphs (KGs) have become a cornerstone of FAIR data implementation, serving as the underlying structure for FAIR Digital Objects, supporting structured metadata, semantic interoperability, and automated data discovery.

In biomedical research, the implementation of FAIR principles is especially critical due to the sensitive nature of health data, the diversity of data types, and the central role of data quality in supporting both scientific research and clinical decision-making. To contextualize the types of biomedical datasets considered in this work, we adopt the categorization proposed in the European Health Data Space (EHDS)[1] regulatory framework. This classification organizes health-related datasets into seven categories: *Clinical & Patient Data*, *Omics & Molecular Data*, *Medical Imaging & Signals*, *Public Health & Surveillance*, *Biobank & Research Data*, *Behavioral & Social Data*, and *Terminologies & Metadata*. These categories reflect the diversity and complexity of biomedical data infrastructures.

The growing availability of automated tools for assessing data quality and FAIR principles compliance

QKG@ESWC 2026: Workshop on Quality of Knowledge Graphs at ESWC 2026, May 11, Dubrovnik, Croatia
*Corresponding author.
slatif@unisa.it (S. Latif)
0000-0002-6213-3576 (S. Latif)

[1]EHDS:https://www.european-health-data-space.com

has enabled large-scale evaluations of publicly available datasets. Among them, KGHeartBeat [2] provides a comprehensive framework that maps FAIR principles to established Linked Data quality dimensions, enabling automated and large-scale evaluation of KGs. At the same time, tools such as FAIR-Checker [3] offer alternative operationalizations of FAIR indicators, implementing different evaluation strategies and metrics. While these tools aim to support FAIR assessments, a fundamental challenge remains: different tools often operationalize FAIR principles and quality indicators in distinct ways. As a result, the FAIRness of a dataset may vary depending on the assessment tool used, making results difficult to compare and potentially leading to inconsistent interpretations of data quality and reusability. Addressing this lack of alignment is essential for ensuring that automated FAIR evaluations are reliable, transparent, and comparable across platforms and domains.

This work investigates whether different FAIR assessment tools produce consistent evaluations when applied to the same datasets through a large-scale empirical analysis of the Biomedical Linked Open Data (BLOD) cloud. Specifically, the study focuses on a subset of $1,209$ datasets from a larger corpus of $1,314$ datasets, selecting those that can be successfully evaluated by both tools. Hence, this study provides both (i) a quantitative overview of FAIR compliance across biomedical KGs and (ii) an empirical investigation of the consistency of automated FAIR evaluation tools. In particular, this work addresses the following research questions (RQs):

RQ1 To what extent do different automated FAIR evaluation tools agree when assessing the same (biomedical) KGs?

RQ2 What is the current level of FAIR compliance across biomedical datasets in the BLOD cloud?

2. Related Work

The FAIR Data Principles, formalized by Wilkinson et al. [4] after the 2014 Lorentz workshop, address the persistent challenge of siloed or poorly managed research data. FAIR comprises four pillars: Findability (globally unique identifiers and searchable metadata), Accessibility (standardized retrieval protocols and persistent metadata), Interoperability (machine-readable formats with linked vocabularies), and Reusability (rich provenance, licensing, and discipline-specific metadata). Mons et al. [5] emphasized that FAIR provides flexible, domain-specific guidelines and extends beyond open data by allowing controlled access where necessary. The FAIR Metrics Group[2] and major initiatives such as the European Open Science Cloud[3], the NIH Data Commons[4], the ELIXIR life sciences infrastructure[5], and the BD2K Initiative[6] demonstrate the global momentum toward improved research data management practices.

Subsequent efforts translated FAIR principles into measurable metrics and automated evaluation methods [6, 7]. Our study builds on the Linked Data quality assessment framework proposed by Zaveri et al. [8], which underpins the KGHeartBeat monitoring platform [2]. This framework evaluates datasets across multiple quality dimensions, including accessibility (availability, interlinking, licensing, performance, and security), contextual aspects (data volume, completeness, and understandability), dynamicity (currency and timeliness), intrinsic quality (accuracy, consistency, and conciseness), representational characteristics (interoperability, interpretability, and versatility), and trust (source reliability, reputation, and contributor verifiability). Compared to existing monitoring systems such as SPARQLES [9], YummyData [10], and Luzzu [11], KGHeartBeat offers broader coverage across quality dimensions and supports automated monitoring across heterogeneous Linked Data datasets without requiring significant workflow redesign. Similarly, FAIR-Checker [3] is an automated system that assesses the FAIRness of digital resources by analyzing dataset metadata and accessibility properties. FAIR-Checker evaluates indicators aligned with the FAIR principles, including the presence of persistent identifiers, structured metadata, standardized access protocols, machine-readable formats, and licensing information. The

[2]FAIR Metrics Group: http://fairmetrics.org
[3]EOSC: https://research-and-innovation.ec.europa.eu/
[4]NISO: https://www.niso.org/niso-io/2017/11/nih-data-commons
[5]Elixir: https://elixir-europe.org/
[6]BD2K:https://commonfund.nih.gov/bd2k

tool primarily relies on metadata-driven analysis, inspecting whether datasets expose FAIR-enabling elements such as globally unique identifiers, shared vocabularies, and provenance information through web-accessible metadata sources.

While substantial effort has been devoted to evaluating individual Linked Data quality metrics, relatively little attention has been paid to understanding their interrelationships. Early observations by authors of [12, 8] suggested that several quality dimensions may be inherently interdependent rather than independent. More recent studies by Pellegrino et al. [13, 14] investigated FAIR compliance and explored correlations between quality dimensions and FAIR principles in Linked Data datasets using automated assessment tools. However, most FAIR evaluations still assess indicators independently, focusing on measurable aspects such as endpoint availability or licensing information, while more complex factors such as metadata exposure, registry inclusion, and provenance remain less explored. Consequently, systematic empirical analyses of relationships between FAIR principles remain scarce, particularly in domain-specific Linked Data ecosystems such as biomedical KGs.

An analysis of biomedical studies shows that clinical and patient data studies [15, 16, 17] generally achieve higher FAIR compliance, while omics datasets e.g., genomes and molecular datasets [18, 19, 20] show variable coverage. Cross-domain studies [21] confirm the broader applicability of FAIR frameworks. Several works provide assessment tools [7, 18, 10, 22] leveraging Luzzu [11] or similar platforms [23]. Overall, despite progress, no study offers a large-scale, quantitative FAIR evaluation across biomedical KGs using multiple automated tools—a gap addressed by our work.

3. Methodology

This section describes the methodology used to evaluate the FAIR compliance of biomedical datasets in the BLOD cloud and to analyze relationships between FAIR indicators obtained from different automated assessment tools. The evaluation process consists of three main steps. First, we construct a subset of the BLOD cloud containing biomedical Linked Open Data datasets that can be consistently evaluated by both assessment frameworks. In addition, we selected KGHeartBeat and FAIR-Checker as the primary FAIR assessment tools due to their suitability for evaluating Linked Open Data datasets at scale. KGHeartBeat provides comprehensive monitoring of Linked Data resources, including endpoint availability, metadata accessibility, and operational dataset characteristics. FAIR-Checker complements this approach by focusing on metadata-level FAIR indicators, such as persistent identifiers, structured metadata, machine-readable formats, and licensing information. Together, the two tools provide complementary perspectives on FAIR compliance by combining dataset-level operational checks with metadata-driven evaluation. Other FAIR evaluation tools, such as F-UJI [24], are primarily designed to assess FAIRness for individual digital objects or repository records through landing-page analysis and metadata inspection. As our study focuses on the large-scale evaluation of biomedical KGs and Linked Open Data endpoints, KGHeartBeat and FAIR-Checker provide more suitable metrics and automation capabilities for analyzing FAIR characteristics across hundreds of datasets. Second, we establish a mapping between the FAIR indicators implemented by KGHeartBeat and those evaluated by FAIR-Checker to ensure conceptual comparability between the two tools. Third, we perform statistical analysis of the resulting FAIR scores to investigate potential relationships between FAIR principles and their associated metrics.

3.1. Subset of BLOD Cloud for Correlation Analysis between KGHeartBeat and FAIR-Checker FAIR Principles Results

The complete BLOD cloud contained $1,314$ biomedical-linked datasets, which are available on the BLOD Cloud GitHub page[7]. All datasets were evaluated for FAIRness using two automated assessment tools: KGHeartBeat and FAIR-Checker. The evaluation results are available online at [8]

[7]BLOD Dataset https://github.com/sanalatif0806/BLOD
[8]BLOD FAIRness Results https://shorturl.at/h9f5s

KGHeartBeat successfully performed the FAIR assessment for the entire set of $1,314$ datasets, providing a comprehensive overview of dataset availability, endpoint status, and several FAIR indicators. However, when the same dataset list was evaluated with FAIR-Checker, the tool successfully analyzed $1,209$ datasets. The remaining 105 datasets could not be evaluated because they were not discoverable by the FAIR-Checker tool. As it does not perform general web crawling; instead, it evaluates datasets accessible through supported entry points, such as persistent identifiers (e.g., DOIs), FAIR data repositories, or machine-readable metadata exposed via standard protocols (e.g., RDF or SPARQL endpoints). In our case, these datasets were either not registered in repositories indexed by FAIR-Checker, lacked resolvable identifiers, or did not expose machine-readable metadata in a format compatible with the tool. As a result, FAIR-Checker was unable to retrieve the necessary metadata to compute FAIR metrics for these datasets.

For consistency in the comparative analysis and statistical evaluation, only the datasets that were successfully evaluated by both tools were considered in the subsequent analysis. Therefore, a subset of 1,209 datasets was used for further statistical processing, including the computation of aggregated FAIR scores and correlation analysis. The normality of the score distributions was assessed using the Shapiro-Wilk test applied to the scores produced by both tools. The results $W = 0.7644$, $p < 0.001$ for *FAIR-Checker*, and $W = 0.6192$, $p < 0.001$ for *KGHeartBeat* indicate statistically significant deviations from normality. Given the non-normal distribution of the data, the Spearman rank correlation coefficient was employed, as it does not assume normality and is suitable for measuring monotonic relationships between continuous variables. This enables the assessment of the strength and direction of association between FAIR scores produced by the two tools, as well as their level of agreement across the full range of FAIRness values.

3.2. Mapping between FAIR principles and KGHeartBeat

This section describes the mapping between the FAIR-Checker and KGHeartBeat evaluation frameworks. For each FAIR principle assessed by both tools, we identify the corresponding metrics and explain how each tool operationalizes the FAIR score. The mapping details are derived from the FAIR Cookbook[9] and the official KGHeartBeat documentation[10]. Table 1 maps the correspondence between the metrics implemented by FAIR-Checker and those used in KGHeartBeat. Since FAIR-Checker primarily evaluates metadata-level properties of datasets, the mapping focuses on the metadata-related metrics of KGHeartBeat. For example, the FAIR-Checker indicator A1.1 evaluates whether a dataset can be accessed through an open resolution protocol HTTP. In contrast, KGHeartBeat performs a more detailed analysis of this dimension by separating it into multiple indicators. Specifically, A1.1-D evaluates the availability and responsiveness of the dataset endpoint, while A1.1-M verifies whether the metadata sources identified in the F2a-M metric are actually accessible. Overall, the comparison highlights that FAIR-Checker focuses primarily on verifying the availability and structure of metadata describing linked data resources. KGHeartBeat, however, provides a more comprehensive assessment by evaluating both metadata availability and the operational accessibility of the dataset itself, including endpoint functionality and data access mechanisms as described in Table 1. To make the comparison fair, the mapping only focused on the metadata-related principles covered by both tools.

The complete details of all FAIR indicators measured by both tools are described below:

FAIR-Checker (F1A) maps to KGHeartBeat (F1B, F1-M) FAIR-Checker evaluates whether datasets use globally unique and persistent identifiers (e.g., DOIs, PURLs, URIs). KGHeartBeat's F1-M metric covers the same requirement by checking whether the dataset provides unique, stable, and resolvable identifiers. FAIR-Checker splits this into two metrics (uniqueness vs. persistence), while KGHeartBeat merges them into a single combined measure.

[9] FAIR Cookbook: https://faircookbook.elixir-europe.org/content/home.html
[10] KGHeartBeat: https://gabrielet0.github.io/CHe-CLOUD/fair_mapping.html

Table 1
Mapping between FAIR-Checker and KGHeartBeat Metrics with Scoring Functions

FAIR-Checker	FAIR-Checker Scoring	KGHeartBeat	KGHeartBeat Scoring
F1A	$\begin{cases} 1, & \texttt{schema:identifier}, \\ & \texttt{dct:identifier}, \\ & \text{or identifiers.org URI found} \\ 0, & \text{otherwise} \end{cases}$	F1-M	$\begin{cases} 1, & \text{dataset registered in a search} \\ & \text{engine with persistent DOI} \\ 0, & \text{otherwise} \end{cases}$
F1B	$\begin{cases} 1, & \text{identifier resolves via HTTP/HTTPS} \\ & \text{using identifiers.org registry} \\ 0, & \text{otherwise} \end{cases}$		
F2A	$\begin{cases} 1, & \text{JSON-LD, RDFa, or HTML} \\ & \text{microdata parseable from} \\ & \text{resource URL} \\ 0, & \text{otherwise} \end{cases}$	F2a-M	$\begin{cases} 1, & \text{SPARQL endpoint, searchable} \\ & \text{engine, or VoID/DCAT available} \\ 0, & \text{otherwise} \end{cases}$
F2B	$\begin{cases} 1, & \text{schema.org, Dublin Core, or} \\ & \text{Bioschemas property found in RDF} \\ 0, & \text{otherwise} \end{cases}$		
A1.1	$\begin{cases} 1, & \text{RDF metadata accessible via} \\ & \text{HTTP/HTTPS without} \\ & \text{authentication} \\ 0, & \text{otherwise} \end{cases}$	A1.1-M	$\begin{cases} 1, & \text{sources from F2a-M accessible} \\ & \text{and contain metadata} \\ 0, & \text{otherwise} \end{cases}$
A1.2	$\begin{cases} 1, & \texttt{dct:accessRights or} \\ & \texttt{schema:conditionsOfAccess} \\ & \text{found} \\ 0, & \text{otherwise} \end{cases}$	A1.2	$\begin{cases} 1, & \text{security requirements discoverable} \\ & \text{via SPARQL} \\ 0, & \text{otherwise} \end{cases}$
I1	$\begin{cases} 1, & \text{RDF, JSON-LD, or OWL format} \\ & \text{parseable from resource} \\ 0, & \text{otherwise} \end{cases}$	I1-M	$\begin{cases} 1, & \text{published according to} \\ & \text{VoID/DCAT specs} \\ 0, & \text{otherwise} \end{cases}$
I2	$\begin{cases} 1, & \text{ontology term found in BioPortal,} \\ & \text{OLS, or LOV registry} \\ 0, & \text{otherwise} \end{cases}$	I2	$\frac{\#\text{FAIR vocabularies}}{\#\text{total vocabularies}}$
I3	$\begin{cases} 1, & \texttt{schema:sameAs} \\ & \text{or qualified outgoing} \\ & \text{link property found} \\ 0, & \text{otherwise} \end{cases}$	I3-D	$\begin{cases} 1, & \text{contains link to another dataset} \\ 0, & \text{otherwise} \end{cases}$
R1.1	$\begin{cases} 1, & \texttt{schema:license}, \\ & \texttt{dct:license}, \\ & \texttt{doap:license}, \\ & \text{or } \texttt{cc:license} \text{ found} \\ 0, & \text{otherwise} \end{cases}$	R1.1	$\begin{cases} 1, & \text{license explicitly reported} \\ 0, & \text{otherwise} \end{cases}$
R1.2	$\begin{cases} 1, & \texttt{prov:wasGeneratedBy}, \\ & \texttt{pav:createdBy}, \\ & \texttt{dct:creator}, \\ & \text{or } \texttt{dct:contributor} \text{ found} \\ 0, & \text{otherwise} \end{cases}$	R1.2	$\begin{cases} 1, & \text{publisher info explicitly reported} \\ 0, & \text{otherwise} \end{cases}$
R1.3	$\begin{cases} 1, & \text{mandatory Bioschemas} \\ & \text{profile properties} \\ & \text{satisfied via SHACL} \\ 0, & \text{otherwise} \end{cases}$	R1.3-M	$\begin{cases} 1, & \text{published according to} \\ & \text{VoID/DCAT specs} \\ 0, & \text{otherwise} \end{cases}$

FAIR-Checker (F2A) maps to KGHeartBeat (F2B, F2a-M) FAIR-Checker checks for the presence of structured metadata (F2A) and whether metadata uses shared vocabularies (F2B). KGHeartBeat's F2a-M metric consolidates these checks by verifying that metadata is available through standard, machine-readable sources such as VoID, DCAT, or other linked data standards. Both tools assess the quality and semantic structure of metadata, though with slightly different emphases.

FAIR-Checker (A1.1) maps to KGHeartBeat (A1.1-M) FAIR-Checker validates the presence of an open, standardized resolution protocol (e.g., HTTP, HTTPS). KGHeartBeat's A1.1-M metric focuses on whether the primary sources hosting the dataset are operational and return valid metadata. Both metrics ensure resources can be accessed programmatically.

FAIR-Checker (A1.2) maps to KGHeartBeat (A1.2) FAIR-Checker verifies the presence of authentication or authorization mechanisms when needed. KGHeartBeat evaluates similar properties, including HTTPS support and authentication requirements. Both assess openness and accessibility conditions.

FAIR-Checker (I1) maps to KGHeartBeat (I1-M) FAIR-Checker inspects whether datasets are available in machine-readable formats. KGHeartBeat's I1-M evaluates whether datasets expose VoID or DCAT descriptions, which are standard machine-readable formats for linked data. Both tools therefore, capture machine interpretability.

FAIR-Checker (I2) maps to KGHeartBeat (I2) Both FAIR-Checker and KGHeartBeat directly evaluate the usage of shared ontologies or FAIR vocabularies. These metrics are essentially equivalent across the two tools.

FAIR-Checker (I3) maps to KGHeartBeat (I3-D) FAIR-Checker checks for the presence of outbound links to other datasets. KGHeartBeat's I3-D metric quantifies the degree of interlinking, i.e., the number and variety of links to external datasets. Both metrics assess connectivity and integration within the Linked Open Data cloud.

FAIR-Checker (R1.1) maps to KGHeartBeat (R1.1) FAIR-Checker verifies the presence of a license in the metadata. KGHeartBeat measures whether a retrievable license is available. Both directly address the FAIR principle requiring accessible usage rights.

FAIR-Checker (R1.2) maps to KGHeartBeat (R1.2) FAIR-Checker checks that provenance information is provided. KGHeartBeat evaluates whether publisher information is available, which is part of the dataset provenance. The two metrics align closely on metadata accountability.

FAIR-Checker (R1.3) maps to KGHeartBeat (R1.3-M) FAIR-Checker evaluates whether the dataset adheres to community standards. KGHeartBeat links this concept to the presence of VoID/DCAT descriptions, which are widely adopted standards. Thus, both measure conformity to community-driven metadata best practices.

4. Results

4.1. BLOD FAIR Score Analysis by KGHeartBeat

By measuring the FAIR scores on the entire BLOD Cloud and collecting results for all the covered FAIR sub-principles, the analysis of KGHeartBeat in Figure 1 shows that the Findability score has a compact distribution with median ≈ 0.60 and a narrow interquartile range(IQR), indicating consistent performance across datasets. Accessibility exhibits wider variation with a median around 0.55 and whiskers extending from approximately 0.20 to 1.00, reflecting differences in authentication mechanisms and access protocols. Interoperability presents the most compact distribution with a median near 0.50 and short whiskers (approximately 0.25-0.40), suggesting relatively uniform but partial adoption of semantic standards. Reusability shows the highest median (≈ 0.65) but a spread comparable to Accessibility, with whiskers reaching 1.00, highlighting variability in licensing and provenance metadata.

As shown in Figure 2, for FAIR Metadata Principles, KGHeartBeat reports a Q1 of approximately 1.10, a median of 1.50, and a Q3 of 1.80, with the upper whisker reaching approximately 2.80 and numerous

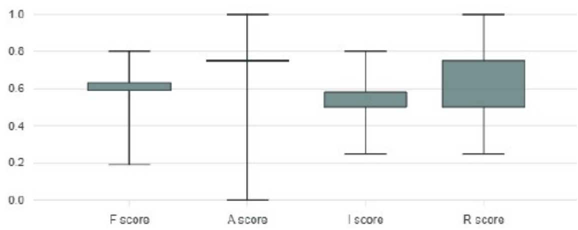

Figure 1: KGHeartBeat FAIR Score Distribution. Each boxplot represents the distribution of normalized scores (range [0,1]) across datasets for FAIR principles: F (Findability), A (Accessibility), I (Interoperability), and R (Reusability).

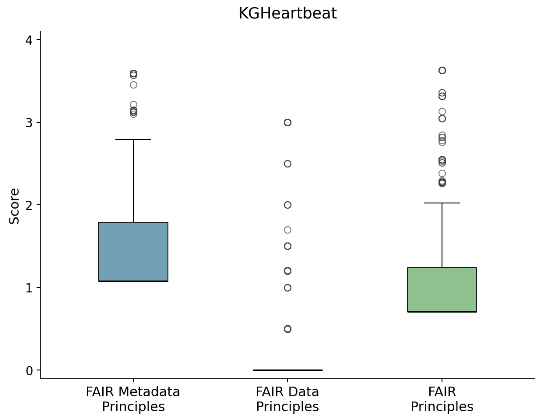

Figure 2: KGHeartBeat metrics FAIR Score. Each boxplot shows the distribution of aggregated scores across FAIR Metadata Principles, FAIR Data Principles, and All FAIR Principles.

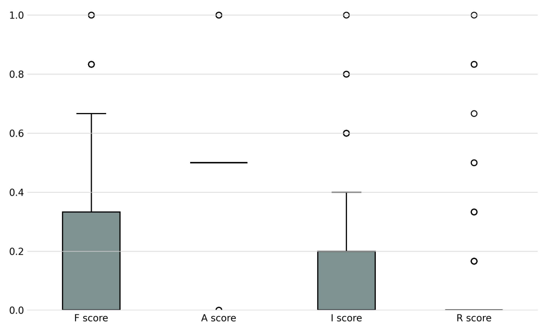

Figure 3: FAIR-Checker FAIR Score Distribution. Each boxplot represents the distribution of normalized scores (range [0,1]) across datasets for FAIR principles: F (Findability), A (Accessibility), I (Interoperability), and R (Reusability).

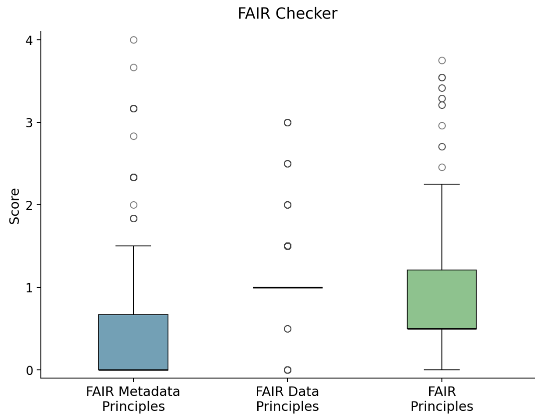

Figure 4: FAIR-Checker metrics FAIR Score. Each boxplot shows the distribution of aggregated scores across FAIR Metadata Principles, FAIR Data Principles, and All FAIR Principles.

outliers clustered between 3.10 and 3.60. The median value above 1.0 indicates moderate compliance with metadata-specific FAIR requirements. For FAIR Data Principles, the interquartile range collapses to a single median line at 0, with scattered outliers ranging from 0.50 to 3.00, indicating that the majority of datasets achieve no baseline data-level FAIR compliance and only a small subset reaches higher scores.

For overall FAIR Principles, KGHeartBeat reports a Q1 of approximately 0.75, a median of 1.05, and a Q3 of 1.25, with the upper whisker at approximately 2.00 and outliers extending to 3.60, suggesting moderate aggregate FAIR compliance with a long tail of well-performing datasets.

Overall, the results suggest that while many biomedical linked datasets partially comply with the FAIR principles, inconsistencies remain across several dimensions. Enhancements in standardized metadata, persistent identifiers, semantic vocabularies, and licensing documentation could significantly improve the FAIRness and usability of biomedical linked data resources.

4.2. BLOD FAIR Score Analysis by FAIR-Checker

By limiting to the BLOD Cloud subset of $1,209$ datasets as described in Section 3.1 and evaluating all FAIR principles as reported in Table 1, the analysis of FAIR-Checker results in Figure 3 reveals distinct patterns across the four FAIR dimensions. The Findability (F) score shows a median of approximately 0.33 with Q1 $\approx$ 0.00 and Q3 $\approx$ 0.67, indicating wide variability in dataset discoverability with outliers reaching 1.00. Accessibility (A) displays a collapsed distribution with no visible IQR and a median line at 0.50, with only two outliers at 0.00 and 1.00, suggesting that nearly all datasets either fully satisfy or completely fail the single accessibility indicator. Interoperability (I) shows a low median of approximately 0.20, with Q1 $\approx$ 0.00 and Q3 $\approx$ 0.20, and outliers at 0.60, 0.80, and 1.00, reflecting partial and inconsistent adoption of semantic standards. Reusability (R) exhibits a collapsed distribution with a median line at 0.00 and no visible IQR, with outliers scattered between 0.17 and 1.00, indicating that the vast majority of datasets provide no reusability-relevant metadata.

Figure 4 presents the aggregated scores across FAIR Metadata Principles, FAIR Data Principles, and all FAIR Principles. Overall, the FAIR Principles scores are concentrated in the low range, with a median of approximately 1.00, Q1 $\approx$ 0.50, and Q3 $\approx$ 1.25, with the upper whisker reaching approximately 2.25 and outliers extending to 3.75, indicating that most datasets achieve only minimal aggregate FAIR compliance. For FAIR Metadata Principles, FAIR-Checker shows a highly skewed distribution with Q1 $\approx$ 0.00, median $\approx$ 0.00, and Q3 $\approx$ 0.67, with the upper whisker reaching approximately 1.50 and outliers extending to 4.00. This indicates that the majority of datasets score at or near zero on metadata principles, with a small number achieving substantially higher scores.

For FAIR Data Principles, the distribution collapses to a single median line at 1.00 with no visible IQR, and outliers scattered between 0.00 and 3.00, indicating low but non-zero compliance for most datasets with high individual variability.

The limited variability in these dimensions suggests that these deficiencies are widespread rather than dataset-specific. To improve FAIR compliance, priority should be given to implementing standardized access protocols, providing clear usage licenses, establishing comprehensive provenance documentation, and adopting shared semantic vocabularies. Addressing these issues would significantly enhance the discoverability, accessibility, and reusability of biomedical linked data resources within the research community.

4.3. KGHeartBeat and FAIR-Checker FAIR principles correlation analysis

We performed the correlation analysis at both the FAIR dimension level and the individual metric level to examine relationships between FAIR indicators. To quantify the strength of relationships between FAIR indicators, we employ the Spearman rank correlation coefficient. It produces values ranging from -1 to 1, where values close to 1 indicate strong positive monotonic relationship, values close to -1 indicate strong negative monotonic relationship, and values near 0 indicate weak or no monotonic relationship. This method is used in this research because the FAIR score distributions were found to be non-normal based on the Shapiro-Wilk test, making Spearman correlation the appropriate non-parametric alternative. The correlation analysis was performed on the subset of datasets successfully evaluated by both tools, and pairwise Spearman coefficients were computed for individual FAIR metrics as well as for aggregated FAIR dimensions.

This analysis enables us to identify dependencies and potential interactions between FAIR indicators and to evaluate the level of agreement between the two automated assessment frameworks. The metric level correlation is presented in Figure 5, which shows strong correlations among Reusability metrics R1.1 and R1.2 and Accessibility metric A1.2, as well as between I3 and Reusability indicators. These relationships indicate that datasets providing machine-readable licensing, provenance information, and external links tend to exhibit stronger reuse potential. Conversely, metadata-related metrics F2A and F2B show weak correlations with most other indicators, highlighting uneven adoption of structured metadata practices across datasets. It is important to note that the correlation matrix in Figure 5 compares metrics across two different tools (KGHeartBeat and FAIR-Checker). Therefore, the diagonal does not necessarily equal 1, as each cell represents the correlation between conceptually related but differently operationalized metrics from the two systems, rather than the same variable.

Figure 6 presents correlations computed within the same aggregation framework, resulting in a symmetric matrix where each axis represents the FAIR dimensions (F, A, I, R). The correlations are computed using aggregated FAIR dimension scores derived from both tools. It presents a strong positive correlation among Findability, Interoperability, and Reusability. In particular, Interoperability

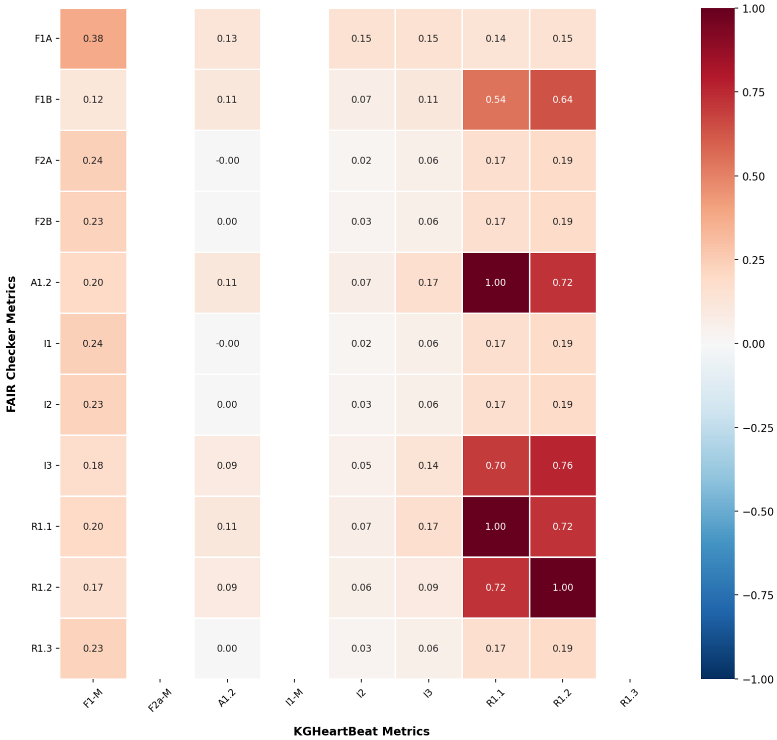

Figure 5: Spearman correlation between FAIR metrics across KGHeartBeat and FAIR-Checker.

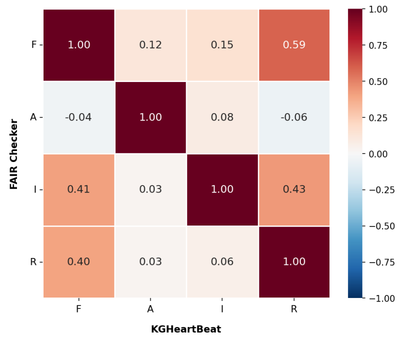

Figure 6: Spearman correlation between aggregated F

and Reusability show a very high correlation of 0.99, while Findability also correlates strongly with Interoperability (0.90) and Reusability (0.90). These results indicate that datasets performing well in one of these dimensions tend to perform well in the others, likely due to shared dependencies on structured metadata, linking, and semantic standards. In contrast, Accessibility (A) shows weak correlations with other dimensions (0.11–0.13), suggesting that access mechanisms operate largely independently from other FAIR properties.

5. Discussion

Regarding RQ1, namely *To what extent do different automated FAIR evaluation tools agree when assessing biomedical knowledge graphs?*, the mapping between FAIR-Checker and KGHeartBeat metrics presented in Table 1 shows that both tools evaluate similar FAIR principles but operationalize them differently. FAIR-Checker primarily focuses on evaluating FAIR compliance at the metadata level, examining the presence and structure of metadata elements such as persistent identifiers, metadata vocabularies, licensing information, and provenance descriptions. In contrast, KGHeartBeat evaluates both metadata properties and the operational characteristics of datasets, including endpoint responsiveness, accessibility of primary metadata sources, and the degree of interlinking with other datasets.

These methodological differences explain the discrepancies observed in the FAIR score distributions presented in Figures 1 for KGHeartBeat, and in Figures 3 for FAIR-Checker. As shown in the KGHeartBeat results Figures 1, most datasets achieve moderate FAIR scores with relatively consistent Findability performance and greater variability in Accessibility, Interoperability, and Reusability. In contrast, the FAIR-Checker analysis in Figures 3 produces substantially lower scores across all dimensions, with particularly low medians for Accessibility and Reusability. This difference reflects the fact that FAIR-Checker evaluates whether metadata descriptions comply with FAIR standards, whereas KGHeartBeat additionally evaluates operational aspects of dataset accessibility and integration.

While aggregated FAIR scores provide a convenient quantitative summary, they inevitably simplify a multidimensional concept into a reduced representation. Moreover, FAIR assessment tools operationalize the principles according to specific design choices, such as required vocabularies, metadata formats, or access protocols. Consequently, FAIR scores may partially reflect compliance with the assumptions embedded in each tool rather than an absolute measure of FAIRness.

Furthermore, collapsing multiple indicators into a single FAIR score may reduce interpretability, as different combinations of strengths and weaknesses across dimensions can lead to similar aggregate values. Therefore, FAIR scores should be interpreted with caution and complemented with fine-grained analysis at the metric level.

In this work, we do not treat FAIR scores as absolute indicators of quality, but rather as comparative signals to analyze alignment and discrepancies between different assessment methodologies.

Despite these differences in evaluation scope, the correlation analysis demonstrates that the two tools capture partially consistent patterns in FAIR compliance. The Spearman correlation analysis at the metric level, shown in Figure 5, reveals strong relationships among certain indicators, particularly between the Reusability metrics R1.1 and R1.2 and the Accessibility metric A1.2. These correlations suggest that datasets with machine-readable licensing, provenance information, and stable access mechanisms tend to exhibit stronger reuse potential and perform consistently across both tools. In contrast, metadata-related indicators such as F2A and F2B display weak correlations with most other metrics, indicating uneven adoption of structured metadata practices and differences in how metadata quality is operationalized by the two frameworks.

At the FAIR principle level, the correlation results presented in Figure 6 reveal strong positive relationships among Findability, Interoperability, and Reusability. In particular, Interoperability and Reusability show a very high correlation (0.99), while Findability also correlates strongly with Interoperability (0.90) and Reusability (0.90). These findings suggest that improvements in structured metadata, semantic vocabularies, and dataset linking can simultaneously enhance multiple FAIR dimensions. In contrast, Accessibility shows weak correlations with the other dimensions (0.11–0.13), indicating that access mechanisms are often implemented independently from metadata quality and interoperability practices. The extremely low Accessibility scores observed in FAIR-Checker are likely related to the fact that many BLOD datasets expose SPARQL endpoints without explicit metadata describing access protocols or authentication requirements.

Regarding RQ2, namely *What is the current level of FAIR compliance across biomedical datasets in the BLOD cloud?*, the overall FAIR score distributions indicate that biomedical datasets within the BLOD cloud generally exhibit moderate but incomplete FAIR compliance. As illustrated in the KGHeartBeat analysis in Figures 1, Findability performs relatively well due to the widespread adoption of persistent identifiers and metadata descriptions. However, Reusability shows the greatest variability, reflecting inconsistent licensing information, provenance metadata, and documentation practices. The FAIR-Checker analysis further highlights systemic weaknesses, particularly in the Accessibility and Reusability dimensions, where scores are consistently close to zero in Figures 3. These findings reinforce that FAIR scores should not be interpreted as definitive measures of dataset quality, but rather as tool-dependent indicators that highlight different aspects of FAIR compliance.

These findings suggest that while many biomedical linked datasets satisfy some basic FAIR requirements, significant gaps remain in the practical implementation of FAIR principles. In particular, improvements in standardized access protocols, machine-readable licensing, provenance documentation, and the adoption of shared semantic vocabularies would substantially improve the FAIRness and usability of biomedical knowledge graphs.

6. Conclusion, Future Work and Limitations

This study presents a framework for evaluating FAIR compliance in biomedical KGs, applied to 1,314 datasets from the BLOD cloud using KGHeartBeat and FAIR-Checker. The results show moderate overall FAIR compliance, with Findability achieving the highest median score of approximately 0.60 by KGHeartBeat due to widespread use of persistent identifiers, while Reusability shows the greatest variability, with scores ranging from near zero (FAIR-Checker) to moderate levels (KGHeartBeat), reflecting inconsistencies in licensing, provenance, and community standards adoption. The comparative analysis reveals partial agreement between the two tools, with strong Spearman correlations for indicators related to licensing, provenance, and access protocols, but differences in metric definitions and coverage (e.g., F3 and A2) limit direct comparability. Overall, the results indicate that FAIR-Checker and KGHeartBeat provide complementary perspectives on FAIR assessment. FAIR-Checker highlights deficiencies in metadata quality and documentation, while KGHeartBeat captures the operational aspects of dataset accessibility and integration. Together, these tools provide a more comprehensive evaluation of FAIR maturity across biomedical linked datasets in the BLOD cloud. While the findings indicate that biomedical datasets in the BLOD cloud partially comply with FAIR principles, significant gaps remain in metadata standardization, licensing documentation, provenance information, and semantic interoperability. A key limitation of this study is the use of only two evaluation tools and the differences in how they operationalize FAIR indicators. Future work will extend this analysis by incorporating additional assessment tools such as F-UJI [24] and by conducting longitudinal studies to monitor changes in FAIR compliance across biomedical datasets over time.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] R. Williams, T. P. V. Staa, A. M. Gallagher, T. Hammad, H. G. Leufkens, F. D. Vries, Cancer recording in patients with and without type 2 diabetes in the clinical practice research datalink primary care data and linked hospital admission data: A cohort study, BMJ Open 8 (2018). doi:10.1136/bmjopen-2017-020827.

[2] M. A. Pellegrino, A. Rula, G. Tuozzo, Kgheartbeat: a knowledge graph quality assessment tool, in: European Semantic Web Conference, Springer, 2024, pp. 276–280.

[3] A. Gaignard, T. Rosnet, F. De Lamotte, V. Lefort, M.-D. Devignes, Fair-checker: supporting digital resource findability and reuse with knowledge graphs and semantic web standards, Journal of Biomedical Semantics 14 (2023) 7.

[4] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[5] B. Mons, C. Neylon, J. Velterop, M. Dumontier, L. O. B. da Silva Santos, M. D. Wilkinson, Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud, Information Services and Use 37 (2017) 49–56.

[6] A. Devaraju, R. Huber, An automated solution for measuring the progress toward fair research data, Patterns 2 (2021).

[7] A. Gaignard, T. Rosnet, F. D. Lamotte, V. Lefort, M. D. Devignes, Fair-checker: supporting digital resource findability and reuse with knowledge graphs and semantic web standards, Journal of Biomedical Semantics 14 (2023). doi:10.1186/s13326-023-00289-5.

[8] A. Zaveri, G. Ertaylan, Linked data for life sciences, Algorithms 10 (2017) 126.

[9] P.-Y. Vandenbussche, J. Umbrich, L. Matteis, A. Hogan, C. Buil-Aranda, Sparqles: Monitoring public sparql endpoints, Semantic web 8 (2017) 1049–1065.

[10] Y. Yamamoto, A. Yamaguchi, A. Splendiani, Yummydata: Providing high-quality open life science data, Database 2018 (2018). doi:10.1093/database/bay022.

[11] J. Debattista, S. Auer, C. Lange, Luzzu—a methodology and framework for linked data quality assessment, Journal of Data and Information Quality (JDIQ) 8 (2016) 1–32.

[12] A. Hogan, J. Umbrich, A. Harth, R. Cyganiak, A. Polleres, S. Decker, An empirical survey of linked data conformance, Journal of Web Semantics 14 (2012) 14–44.

[13] M. A. Pellegrino, G. Tuozzo, How fair is FAIR? understanding lod cloud fairness through correlation patterns, in: Proceedings of the 34th ACM International Conference on Information and Knowledge Management, 2025, pp. 2315–2325.

[14] M. A. Pellegrino, A. Rula, G. Tuozzo, Are quality dimensions correlated? An empirical investigation over linked data, in: International Semantic Web Conference, Springer, 2025, pp. 23–41.

[15] N. Queralt-Rosinach, R. Kaliyaperumal, C. H. Bernabé, Q. Long, S. A. Joosten, H. J. van der Wijk, E. L. Flikkenschild, K. Burger, A. Jacobsen, B. Mons, M. Roos, Applying the fair principles to data in a hospital: challenges and opportunities in a pandemic, Journal of Biomedical Semantics 13 (2022). doi:10.1186/s13326-022-00263-7.

[16] V. Pezoulas, C. Androutsos, D. I. Fotiadis, Towards the development of a fair-compliant biomedical ontology for colorectal cancer, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 1234–1241. doi:10.1109/EMBC53108.2024.10782443.

[17] K. McGlinn, M. A. Rutherford, K. Gisslander, L. Hederman, M. A. Little, D. O'Sullivan, Fairvasc: A semantic web approach to rare disease registry integration, Computers in Biology and Medicine 145 (2022). doi:10.1016/j.compbiomed.2022.105313.

[18] A. Ammar, C. Evelo, E. Willighagen, Fair assessment of nanosafety data reusability with community standards, Scientific data 11 (2024) 503.

[19] A. Pereira, J. R. Almeida, R. P. Lopes, J. L. Oliveira, Querying semantic catalogues of biomedical databases, Journal of Biomedical Informatics 137 (2023). doi:10.1016/j.jbi.2022.104272.

[20] S. Soleymani, N. Gravel, L. C. Huang, W. Yeung, E. Bozorgi, N. G. Bendzunas, K. J. Kochut, N. Kannan, Dark kinase annotation, mining, and visualization using the protein kinase ontology, PeerJ 11 (2023). doi:10.7717/peerj.16087.

[21] G. Tuozzo, Navigating the LOD subclouds: Assessing linked open data quality by domain, in: Companion Proceedings of the ACM on Web Conference 2025, 2025, pp. 2141–2148.

[22] S. Zhang, N. Benis, R. Cornet, Assessing resolvability, parsability, and consistency of rdf resources: a use case in rare diseases, Journal of Biomedical Semantics 14 (2023). doi:10.1186/s13326-023-00299-3.

[23] C. Y. Chen, L. W. Stevenson, G. C. Stewart, D. L. Bhatt, M. Desai, J. D. Seeger, L. Williams, J. J. Jalbert, S. Setoguchi, Real world effectiveness of primary implantable cardioverter defibrillators implanted during hospital admissions for exacerbation of heart failure or other acute co-morbidities: Cohort study of older patients with heart failure, BMJ (Online) 351 (2015). doi:10.1136/bmj.h3529.

[24] R. Huber, A. Devaraju, F-uji: an automated tool for the assessment and improvement of the fairness of research data, in: EGU General Assembly Conference Abstracts, 2021, pp. EGU21–15922.