

A Large-Scale Empirical Analysis of

FAIR Compliance in Biomedical KGs

Using KGHeartBeat and FAIR-Checker

Sana Latif · University of Salerno



Motivation & Problem

The Core Challenge

Automated FAIR assessment tools operationalize the same principles differently. The same dataset can receive conflicting evaluations depending on the tool used — leading to unreliable, non-comparable FAIRness scores.

Why FAIR Matters

- Guides digital research data infrastructure
- Critical in biomedical research (sensitive data, diverse types)
- Knowledge Graphs (KGs) are cornerstones of FAIR implementation

Why Tool Comparison Matters

- Tools differ in scope: metadata-level vs. operational checks
- Discrepant scores confuse stakeholders and researchers
- No large-scale cross-tool study exists for biomedical KGs
- Alignment is needed for reliable, transparent evaluation

Research Questions & Dataset

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?

Dataset — BLOD Cloud

1,314

Total BLOD
datasets

1,209

Evaluated by
both tools

105

Not discovered
by FAIR-Checker

Methodology

01

Dataset Construction

Select BLOD subset evaluable by both KGHeartBeat & FAIR-Checker (1,209 datasets)

02

Metric Mapping

Establish conceptual correspondences between FAIR indicators across both tools (Table 1)

03

Statistical Analysis

Shapiro-Wilk normality test → Spearman rank correlation (non-parametric, monotonic)

KGHeartBeat: endpoint availability + metadata + operational access **FAIR-Checker:** metadata-level FAIR indicators (PID, structured metadata, licensing)

FAIR Metric Mapping — Shared Principles Only

FAIR Dim.	FAIR-Checker Metric	KGHeartBeat Metric
Findability	F1A — Unique ID present; F1B — ID resolves	F1-M — DOI in search engine (persistent)
Findability	F2A — Structured metadata (JSON-LD/RDFa); F2B — Shared vocab used	F2a-M — SPARQL/VoID/DCAT available; F2b-M — Attributes covered
Accessibility	A1.1 — RDF metadata accessible via HTTP	A1.1-M — Metadata sources from F2a-M reachable
Accessibility	A1.2 — Access rights / conditions declared	A1.2 — Security discoverable via SPARQL
Interoperability	I1 — Resource parseable as RDF/JSON-LD/OWL	I1-M — VoID/DCAT description present
Interoperability	I2 — Ontology found in BioPortal/OLS/LOV	I2 — #FAIR vocabularies / #total vocabularies
Interoperability	I3 — schema:sameAs or outgoing link found	I3-D — Contains link to another dataset
Reusability	R1.1 — License property found in metadata	R1.1 — License explicitly reported
Reusability	R1.2 — Provenance property found in metadata	R1.2 — Publisher info explicitly reported
Reusability	R1.3 — Bioschemas SHACL profile satisfied	R1.3-M — VoID/DCAT description present

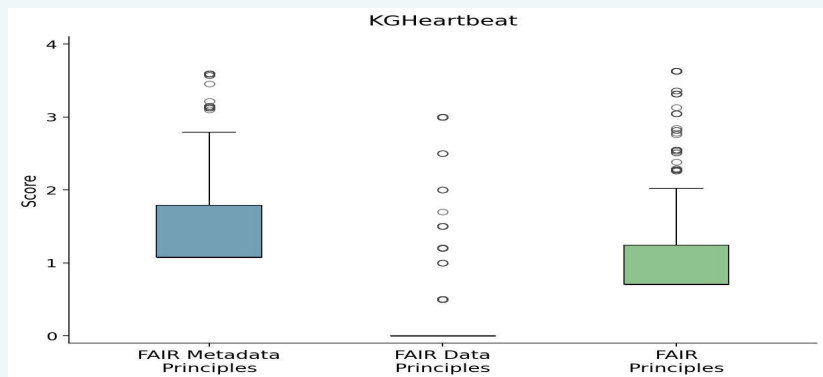
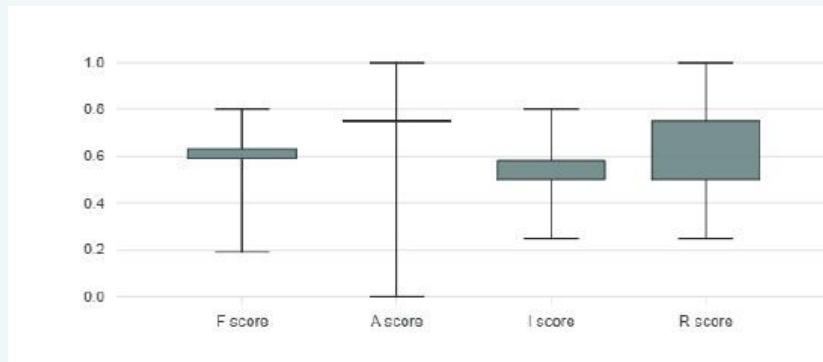
KGHeartBeat-only metrics excluded from comparison: F1-D, F3-M, F4-M, A1.1-D, A2-M, I1-D, R1.3-D

KGHeartBeat-Only Metrics — Excluded from Cross-Tool Analysis

These 7 KGHeartBeat metrics have no FAIR-Checker counterpart and are excluded from the comparative analysis.

KGHeartBeat ID	FAIR Principle	Input	What It Measures	Why No FC Counterpart
F1-D	F1	Data	URI dereferenceability	FC does not evaluate data-level URIs
F3-M	F3	Metadata	Metadata attaches DOI(s) to data	FC merges F1/F3 into a single identifier check
F4-M	F4	Metadata	Dataset registered in any search engine	FC does not check registry presence separately
A1.1-D	A1.1	Data	Operational SPARQL endpoint or data dump	FC only checks metadata-level HTTP access
A2-M	A2	Metadata	Metadata persists even if data is removed	FC has no A2 indicator
I1-D	I1	Data	Standard & open representation format	FC evaluates format only at metadata level
R1.3-D	R1.3	Data	Data organized per SPARQL/dump/OWL/RDFS	FC evaluates community standards via metadata only

Results — KGHeartBeat Analysis



Key Observations

Findability (0.60)

Compact IQR — consistent adoption of persistent identifiers

Accessibility (0.55)

Wide variation (0.20–1.00), reflecting diverse auth mechanisms

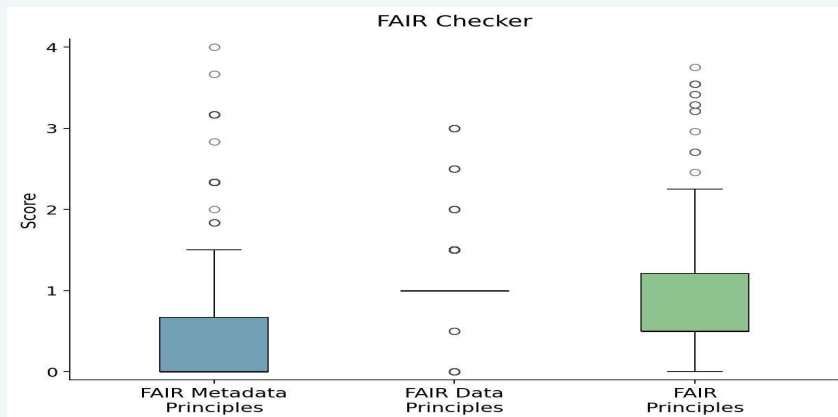
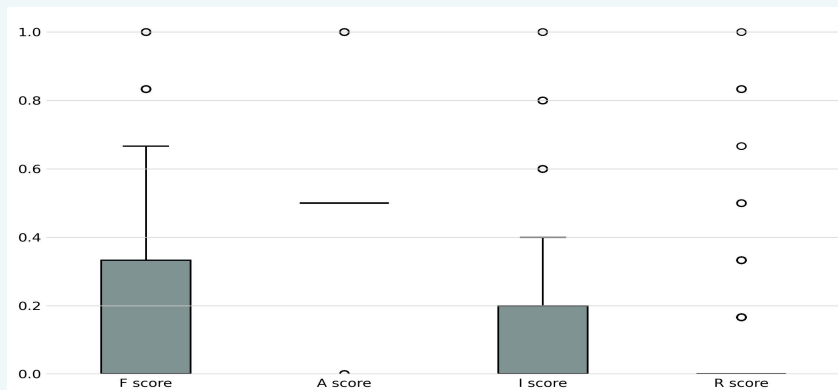
Interoperability (0.50)

Most compact distribution — partial, uniform semantic adoption

Reusability (0.65)

Highest median but high spread — inconsistent licensing/provenance

Results — FAIR-Checker Analysis



Key Observations

Findability (0.33)

Wide variability (Q1≈0, Q3≈0.67) — uneven discoverability

Accessibility (0.50)

Collapsed distribution — datasets either fully pass or fully fail

Interoperability (0.20)

Low median — partial, inconsistent semantic standard adoption

Reusability (0.00!)

Near-zero for most datasets — systematic licensing/provenance gap

How We Measured Agreement — Correlation Methodology

1

Test for normality (Shapiro-Wilk)

Before choosing a correlation method, we tested whether scores from each tool follow a normal distribution. The Shapiro-Wilk test gave $W = 0.76$, $p < 0.001$ for FAIR-Checker and $W = 0.62$, $p < 0.001$ for KGHeartBeat. Both distributions are significantly non-normal — ruling out Pearson correlation, which assumes normality.

2

Choose Spearman rank correlation (non-parametric)

Spearman converts each dataset's score into a rank (1st, 2nd, 3rd... highest scorer) within each metric, then measures whether the two tools agree on the ranking. It does not require normal data and captures any monotonic relationship — if tool A ranks dataset X above dataset Y, does tool B do the same? Values range from -1 (perfect disagreement) to $+1$ (perfect agreement), with 0 meaning no relationship.

3

Two levels of analysis — metric-level and dimension-level

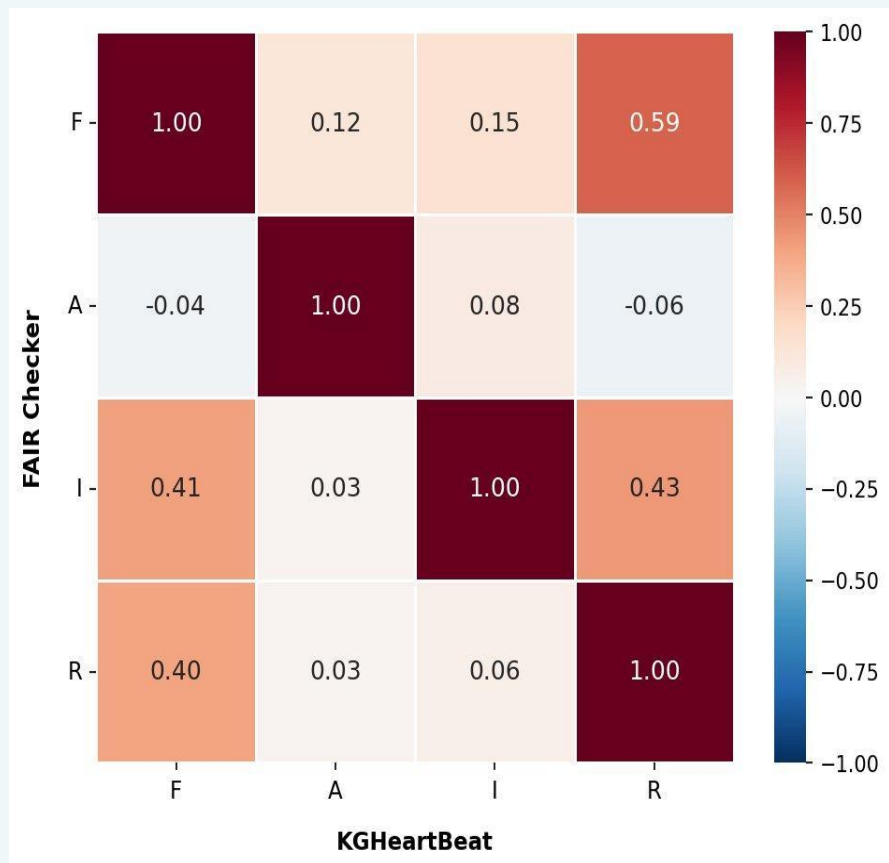
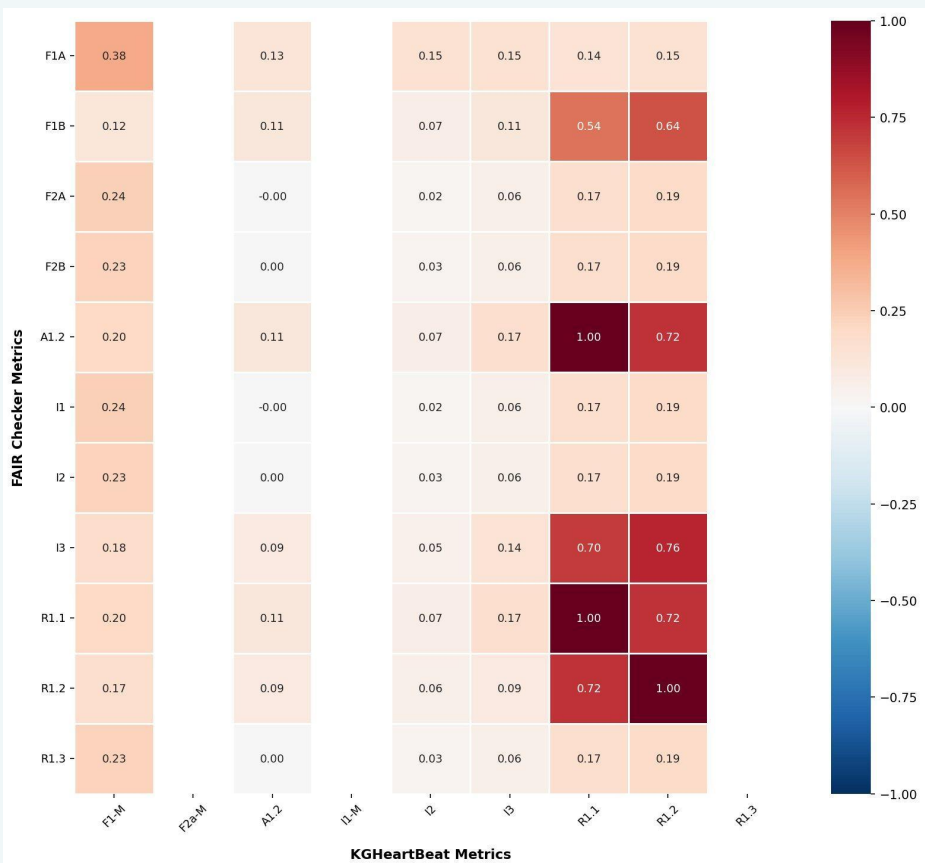
Metric-level (Fig. 5 in paper)

For each pair of conceptually mapped metrics (e.g. FC's R1.1 vs. KGH's R1.1), we computed the Spearman coefficient across all 1,209 datasets. This produces a matrix showing how well each individual metric from one tool correlates with its counterpart — and with other metrics — in the other tool.

Dimension-level (Fig. 6 in paper)

We also aggregated scores per FAIR dimension (F, A, I, R) for each tool, then computed pairwise Spearman correlations between all four dimensions across both tools. This reveals structural dependencies — e.g. whether datasets good at Interoperability tend also to score high on Reusability.

Spearman Correlation Analysis



Discussion & Conclusions

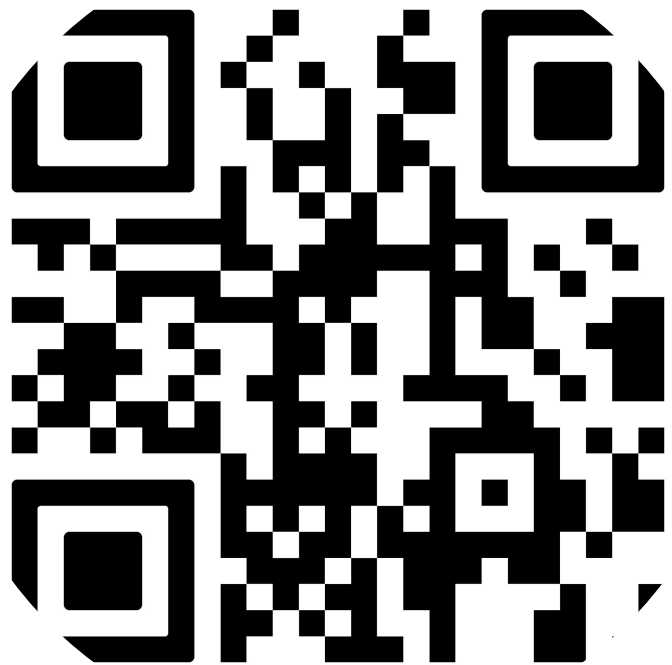
RQ1 — Tool Agreement

Partial agreement: strong correlations for licensing, provenance, access protocols. But significant discrepancies in scores (e.g., Reusability median 0.65 vs. 0.00). FAIR-Checker is metadata-focused; KGHeartBeat also covers operational aspects.

RQ2 — FAIR Compliance Level

Moderate but incomplete compliance. Findability performs best. Reusability most variable. Accessibility and Reusability show systemic weakness in FAIR-Checker. Core gaps: standardized access protocols, machine-readable licenses, provenance docs, shared vocabularies.

Future Work: Incorporate F-UJI tool · Longitudinal FAIR monitoring · Broader dataset coverage



Thank you