

Leonardo Guerreiro Azevedo¹, Gabriel Banaggia*, Julio Tesolin^{*,2,} Renato Cerqueira¹ ¹IBM Research – Brazil ²PPGI/IME lga@br.ibm.com, gbannagia@gmail.com, jcctesolin@ime.eb.br, rcerq@br.ibm.com

IBM Research

* For authors marked with *: Work done while at IBM Research

An Appraisal of Automated Tools for FAIRness Evaluation





Contents

I. Context

II. Contribution

III. Tools appraisal

IV. Conclusion





There are several mechanisms to support the design of FAIR data

- -Guidelines
- -Questionnaires
- –Semi-automated tools
- –Automated tools



Mechanisms' goals

-Characterize digital objects related to the FAIR principles

And/or

– Evaluate digital object's FAIRness level

Context



Manual and semi-automated tools

- -Strengths
 - Essential for
 - Overall understanding and
 - Appreciation of the research life cycle

-Weaknesses

- Time consuming
- Requires experience and technical skills
- Carries difficulties when inspections is needed
- Does not scale for several digital objects







FAIRassist.org

esource 🗸	Execution Type	Key Features	Organisation	Target Objects	Readi
ar Data Rating Tool	Manual - questionnaire	Based on rating systems and maturity models	CSIRO OzNome	Datasets	

An Improved Questionnaire for FAIR Characterization

AutoFAIR	Semi-automated	A portal for automating FAIR a	ssessments for bioinforma	atics resources	Department of Computer Information Systems, Faculty of ICT, University of Malta	Bioinformatics resources	Publi
RIN Metadata Curation Dashboard	Automated	ZB MED-Publikationsportal Lebenswissenschaften	Fachrepositor Freier Zugang zu Wissensresso	Fium Lebenswissenschaften urcen aus den Lebenswissenschaften	ENMANAGEMENT DIGIT	Datasets	Doc
		Bestand 》 图 An Improved Questionnaire for FAIR Charact An Improved Questionnaire for FAIR Chara Paper - An Improved Questionnaire for FA 》 Slides - An Improved Questionnaire for F 》 Supplement: FAIR Characterization of Mat 》 Supplement: FAIR Characterization of Pub 》 Supplement: FAIR Characterization Questi 》 Supplement: Comparisons of the FAIR ques 》	erization An Improved Guerreiro Azevedo, Leon Guerreiro Azevedo, Leon (() () () () () () () () ()	Questionnaire for FAIR Characteriza ardo ● Tesolin, Julio Banaggia, Gabriel ● Cerqueira, Renato ● Azevedo et al. (DaMaLOS 2023, Date of all (DaMaLOS 2023, Date of all (2023 ■ Kongressschrift 2023 ■ Kongressschrift 2023-06-16 http://obid.org/resources/99372482780206441#1 [12] Workshop on Metadata and Research (objects) Management for Linked Open Science (3.: 2023 : Online) [12] https://creativecommons.org/licenses/by/4.0/ [12] ● Postprint Verlagsversion begutachtet Englisch	tion 2023)		



Context





Manual and semi-automated tools

- -Strengths
 - Essential for
 - Overall understanding and
 - Appreciation of the research life cycle

-Weaknesses

- Time consuming
- Requires experience and technical skills
- Carries difficulties when inspections is needed
- Does not scale for several digital objects





Automated tools

-Strengths

- Performs evaluation without human intervention
- Scale when evaluating several digital objects
- More objective
- Allow comparison of distinct digital objects
- -Weaknesses
 - Requires precise definition of metrics and evaluation tests
 - May be difficult to fit if community standards are not defined
 - May result on using domain-agnostics concepts
 - May not fit community needs

Contribution

Goal

Analyze automated tools for FAIRness assessment

Steps

- -Search for existing tools in the literature
 - Discover the tools
 - Elicit requiments
- Examine tools regarding elicited requirements

Literature review

Abbreviated systematic literature review

Research questions

- RQ1: What are the existing automated tools for FAIRness evaluation?
- RQ2. Which requirements do these tools meet?

Search string ("Tool" OR "Automated") AND ("Assessment" OR "Evaluation") AND ("FAIR- ness" OR "FAIRification") AND ("FAIR Principles" OR "FAIR Data")

Search on Scopus, IEEE and ACM digital libraries

-32 works found

– Exclution and inclusion criteria endup with 4

- Krans et al. (2022)
- Peters-Von Gehlen et al. (2022)
- Slamkov et al. (2022)
- Sun et al. (2022)
- Gaps on exiting works
 - Abstract characterization and comparison of tools
 - Do not propose or use requirements

Literature review

RQ1: What are the existing automated tools for FAIRness evaluation?

- -Tools referenced in the works
 - Krans et al. (2022)
 - Peters-Von Gehlen et al. (2022)
 - Slamkov et al. (2022)
 - Sun et al. (2022)

Search for existing tools in the literature

Tool	Automated?
F-UJI	Yes
FAIR Evaluator	Yes
FAIR Enough*	Yes
FAIR-Checker	Yes
ARDC's FAIR Data Self Assessment Tool	No
Checklist for Evaluation of Dataset Fitness for Use	No
CSIRO's 5°Oz Data tool	No
DANS's SATIFYD	No
Data Stewardship Wizard	No
EUDAT's Checklist	No
FAIRdat	No
FAIRenough	No
FAIRshake	No
GARDIAN	No
RDA's Simple Grid	No
Semi-automated workflow for FAIR maturity indicators	No

8 Itor.

Literature review

RQ2. Which requirements do these tools meet?

- -Requirements
 - Guide the appraisal and development of tools
 - Crucial for making objective FAIRness evaluations and improving digital objects
- -Requirements elicited from
 - The works (Krans et al., Peters-Von Gehlen et al., Slamkov et al., and Sun et al.)
 - Tools documentation (F-UJI, FAIR Evaluator, FAIR Enough, FAIR Checker)

Elicited requirements (23 requirements)

Req	Requirement: The tool should
R1	be fully automated.
R2	give a FAIRness score /grade.
R10	be customizable according to the type of digital object and community.
R12	provide a visual representation (e.g., a badge) of the FAIR assessment results.
R14	rely on FAIR-enabling services.
R15	offer guidance on how it is used (e.g., providing user manual, help, and publications).
R18	disclose its rating system (e.g., evidences and rationale).
R19	be informative, i.e., teach the user about FAIR.
R20	give recommendations on how to improve the FAIRness of the evaluated resource.
R23	support versioning of FAIRness assessment.

Appraisal of the tools

Evaluation by reading tools' documentation –Web pages

- –GitHub pages
- -Papers

Examine tools regarding elicited requirements

Req	Keyword	F-UJI	FAIR Evaluator	FAIR	FAIR Checker
R1	Automated				
R2	Score				
R10	Customizable			\bigcirc	
R12	Badge		X	×	
R14	FAIR- enabling services				
R15	Guidance			X	
R18	Rating system				
R19	Teach				
R20	Recommenda tions	X	X	X	
R23	Versioning	X	X	X	X

Requirement is totally supported
 Requirement is not supported
 Requirement is partially supported





R12. The tool should provide a visual representation (e.g., a badge) of the FAIR assessment results.



Without a badge, the user does have the whole assessment in a visual representation.

Pain question: What is the best representation that present the results' overview for all evaluation levels (principles, metrics, tests)?





Recommendations

R20: The tool should give recommendations on how to improve the FAIRness of the evaluated resource

X	F-UJI	× FAIR Enough	FAIR Evaluator		FAIR Checker
FSF-R1.1-01M - N FAIR level: Score: Output: Metric tests:	Metadata includes license information under which data can be reused. 0 of 3 0 of 2 1 1 Test: Test name: Score: Maturity: Result: FsF-R1.1-01M-1 Licence information is given in an appropriate metadata element 0 Image: Comparison of the co	F1 - FAIR Metrics Gen2 - Data Identifier Persistence https://w3id.org/FAIR_Tests/tests/gen2_data_identifier_persistence - Version: Hvst-1.4.4:Tst-0.2.2 - Metric to test if the unique identifier of the data resource is likely to be persistent. Known schema are registered in FAIRSharing (https://fairsharing.org/standards/rq=&selected_facets=type_exactidentifier%20schema). For URLs that don't follow a schema in FAIRSharing we test known URL persistence schemas (purl, ocic, fdip, purlz, w3id, ark). Test result URL: http://tests:8080/tests/gen2_data_identifier_persistence#https://doi.pangaea.de/10.1594/PANGAEA.893286/result-2024-05-23T22:50:28+00:00 WARN: HTTP error 400 Bad Request encountered when trying to resolve https://doi.pangaea.de/10.1594/PANGAEA.893286?format=metadata_iso19139 WARN: Unable to resolve https://doi.pangaea.de/10.1594/PANGAEA.893286?format=metadata_iso19139 using HTTP Accept header {"Accept"=>"text/turtle, application/ld+json, application/rdf+xml, text/xtml+xml, application/curturla_application/curuurla_applicapplication/curturla_application/curturla_applica	GEN2 DATA IDENTIFIER PERSISTENCE Gen2 data IDENTIFIER PERSISTENCE Status: Failure Principle tested: Description: Metric test created on: by (updated on). Test accuted on: reb 10, 2020 Test results INFO: Found a DDI. INFO: Found a DDI. INFO: Attempting to resolve https://doi.org/10.25504/FAIRsharing.j0t0pe using HTTP Headers {"Accept"=>"text/turtle, application/rdf+xa], appli	F18 Persistent IDs	Image: Section on Identifiers or in BDMMs:
Debug messages:	Level: Message: WARNING License information unavailable in metadata INFO Will consider all SPDX licenses as community specific licenses for FsF-R1.1-01M WARNING Skipping SPDX and community license verification since license information unavailable in metadata	<pre>text/rdf+n3, text/rdf+turtle, application/n-triples"}. WARN: HTTP error 400 Bad Request encountered when trying to resolve https://doi.pangaea.de/10.1594/PANGAEA.893286?format=metadata_dif WARN: Unable to resolve https://doi.pangaea.de/10.1594/PANGAEA.893286?format=metadata_dif using HTTP Accept header {"Accept"=>"text/turtle, application/ld+json, application/rdf+xml, text/xhtml+xml, application/n3, application/rdf+n3, application/n-triples"}. WARN: HTTP error 400 Bad Request encountered when trying to resolve</pre>	<pre>INFO: Found turtle text/turtle type of content when resolving https://doi.org/10.25504/FAIRsharing.j0t0pe using HTTP Accept header {"Accept"=>"text/turtle, application/ld+json, application/rdf+xml, text/xhtml+xml, application/n3, application/ndf+n3, application/turtle, application/x-turtle, text/n3, text/turtle, text/rdf+n3, text/rdf+turtle, application/n-riples"}. INFO: parsing as linked data. INFO: The response message body component appears to contain RDF::Turtle::Format. INFO: Attempting to resolve https://doi.org/10.25504/FAIRsharing.j0t0pe using HTTP Headers {"Accept"=>"*/*"}. INFO: Found html text/html type of content when resolving https://doi.org/10.25504/FAIRsharing.j0t0pe using HTTP Accept header {"Accept"=>"*/*"}.</pre>		
					To ensure that the used identification scheme is persistent, you should build your resource ID with a namespace that can be found in identifiers and the incince oriented registry). Examples of persistent identifiers for a United entry is: <u>https://dentifiers.org/united/288483</u> with unityot as namespace of for a PubMed publication. <u>https://dentifiers.org/united/288483</u> with unityot as namespace of for a PubMed publication. <u>https://dentifiers.org/united/288483</u> with unityot as namespace of for a licentifier can be either the URL itself or encoded in the metadata as a dct:identifier or schemaidentifier property. Learn more about persistent identifiers in the identifiers.org documentation or in the <u>FAIRE</u> <u>CookBook section on identifiers</u> or in <u>BDMkit</u> .

Without giving recommendations (e.g., recipes or standard schemas), one misses the opportunity to increase the FAIRness of data

× F-UJI, FAIR Evaluator, FAIR Enough: present a log of the execution without explicit recommendations

V FAIR Checker: a set of recommendations for FAIRness improvements with links to training resources, such as FAIR-Cookbook

? Main question: How to present FAIRness improvements recommendations to be followed by the non-technical users?





Customization

R10: The tool should be be customizable according to the type of digital object and community



Without the ability to customize the tool, evaluation is limited to agnostic parameters, i.e., does not handle community-specific needs.

I? Main question: How to create FAIRness assessment tools that is easily adaptable by non software development users?



13

Appraisal results

Tools analysis

– Similar responses for 15 requirements

- R1 to R8, R10, R13, R16, R18, R19, R22, and R23
- Different responses for 8 requirements
 - R9, R11, R12, R14, R15, R17, R20, R21

Fulfillment

- -74%: F-UJI
- -70%: FAIR Checker
- -63%: FAIR Evaluator and FAIR Enough

Tools main strenghts

- Employ good software development practices
- Use state-of-the art technologies in
 - Software Engineering
 - Semantic Web

Tools main weaknesses

– Reporting features should be improved

-Storage of results and versioning are not implemented



Conclusion

No tool meets all requirements and stands out as state-of-the-art

Choosing the best tool is challenging

There is room to solve the gaps by

• Evolving existing tools

Or

- Developing a new tool
- The proposed requirements as a base for appraising automated tools for FAIRness assessment

To make a choice of tool or implementation

–Start by

- Using requirements, like the ones we proposed
- Identifying the most critial needs
- Reading the details of our appraisal

-Then

- Understand the difficulties to customize an existing tool
- Test the tools in practice
- Make a decision
 - To use or improve a tool or develop your own

15

Conclusion

Limitation

- -Search scope limited to academic works
 - Scopus, IEEE, and ACM
- -Rationale
 - They include papers from relevant journals and conferences in Computer Science

Future work

- Broaden the literature review
 - Include other digital libraries
 - Include gray literature (e.g., FAIRassist.org)
- Detail the requirements
- Develop benchmarks considering our proposal of requirements
- -Appraise the automated tools in practice



Thank you

Leonardo Guerreiro Azevedo

<u>lga@br.ibm.com</u> <u>https://ibm.biz/leonardo</u>

Gabriel Bannagia gbannagia@ibm.com

Julio Tesolin jcctesolin@ime.eb.br

Renato Cerqueira rcerq@br.ibm.com

ibm.com

IBM Research / May, 2024 / © 2024 IBM Corporation

IBM Research – Brazil Ventura Tower, Centro Rio de Janeiro







