

Comparing FAIR Assessment Tools and their Alignment with FAIR Implementation Profiles using Digital Humanities Datasets

Andre Valdestilhas^{1,*†}, Menzo Windhouwer^{2,†}, Ronald Siebes^{1,†} and Shuai Wang^{1,3,†}

¹Department of Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

²KNAW - Humanities Cluster, Oudezijds Achterburgwal 185, 1012 DK Amsterdam, The Netherlands

³University Library, Maastricht University, Grote Looiersstraat 17, 6211 JH Maastricht, The Netherlands

Abstract

FAIR principles serve as guidelines for implementing data and metadata to improve Findability, Accessibility, Interoperability, and Reusability. In recent years, numerous tools have been developed to assess how well datasets adhere to each FAIR principle. However, due to their diverse designs, these tools interpret the FAIR principles differently and provide varying assessment results, which can be confusing. Many communities publish datasets that follow similar data management practices, and some of these common practices have recently been compiled into community standards known as FAIR Implementation Profiles (FIPs). This paper compares the metrics of FAIR assessment tools with FIP. We illustrate these differences by analyzing the assessment results of two datasets in the Digital Humanities domain and further explore how these results compare with their corresponding FIPs.

Keywords

FAIR assessment, Digital Humanities, FAIR Implementation Profile

1. Introduction

The adoption of Semantic Web technologies and FAIR (Findable, Accessible, Interoperable, and Reusable) principles[24] by the social sciences and humanities (SSH) represents a transformative approach to data management and research practices. This paradigm shift improves the interoperability and reuse of data, facilitating collaborative efforts across disciplines, and promoting innovative research outputs [24]. Communities interpret FAIR principles in different ways that align with their specific needs, preferences, and available resources, leading to different practices for data management. To assess the progress towards FAIR data and understand implementation choices, various FAIR Assessment Tools (FATs) have been designed to automate the evaluation of published datasets. Thus, a community's implementation choices can be reflected collectively in the assessment results (i.e. the FAIRness) of the datasets they publish. Alternatively, in recent years, research communities have published their implementation choices in the format of FAIR Implementation Profiles (FIPs) [14]. For a community,¹ its FIP is a formalized, machine-actionable document of a collection of the different choices (a.k.a. community standards) regarding each FAIR principle. These FIPs can be used to compare implementation choices between different communities. However, as far as the authors know, little research has been done to compare FAIR assessment results with the FIPs. In this paper, we present a pilot study in the field of Digital Humanities (DH). We compare the assessment results of the representative datasets of two DH communities with their corresponding FIPs.

SemDH'25: Second International Workshop of Semantic Digital Humanities, June 1–5, 2025, Portoroz, Slovenia

*Corresponding author.

[†]These authors contributed equally.

✉ a.valdestilhas@vu.nl (A. Valdestilhas); menzo.windhouwer@di.huc.knaw.nl (M. Windhouwer); r.m.siebes@vu.nl (R. Siebes); shuai.wang@maastrichtuniversity.nl (S. Wang)

🌐 <https://github.com/firmao> (A. Valdestilhas); <https://shuai.ai> (S. Wang)

🆔 0000-0002-0079-2533 (A. Valdestilhas); 0000-0002-2204-4018 (M. Windhouwer); 0000-0001-8772-7904 (R. Siebes); 0000-0002-1261-9930 (S. Wang)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹As a pilot project, we limit our study to only research communities that published one FIP. In reality, a community may work on data of various kinds, which corresponds to multiple FIPs.

Due to the interdisciplinary nature of the research field, DH researchers often encounter various types of data. The datasets used and published in one project can differ significantly from those in another, leading to complex approaches to data management. Some datasets are available through modern data infrastructures, while others can only be accessed on outdated websites, contributing to the diversity of implementation choices. It has been observed that certain implementation choices may depend on the selected data infrastructure, particularly in terms of metadata. To complicate matters further, the interpretation of FAIR principles can vary among different assessment tools, resulting in inconsistent scores [26]. How do the assessment results of DH datasets differ between FATs, and what does the difference imply? How does the difference between datasets impact findability, interoperability, and other aspects? How does a FIP capture the diversity of choices between datasets published by researchers in a community? Addressing these questions would not only guide better implementation decisions in data management but also ease the comparison between communities.

The variance of interpretation of FAIR principles captured by different FATs and the differences mentioned above between communities, we would like to study the following research questions (RQ). **RQ1:** How do the FATs and FIP differ in the interpretation of the FAIR principles? **RQ2:** How do the FAIR assessment results of DH datasets differ between FATs? **RQ3:** How do the FAIR assessment results of representative datasets from various DH communities align with their respective FIPs? In particular, we will compare the assessment results of two selected representative datasets within DH communities. In addition, we will examine the aspects included in the FIPs and investigate how the assessment results correspond to these profiles. The evaluation results reveal disparities in FAIR assessment outcomes since different DH communities have developed unique approaches.

The remainder of this paper is as follows: in Section 2, we introduce the FATs and the FIP concept in more detail. Section 3 describes the methodology and implementation decisions: the selected DH communities, their FIPs, and the representative datasets. Section 4 provides a concise comparison of the FATs' assessment metrics with the FIP and studies how much the assessment results align with the FIPs. Finally, Section 5 discusses options to mitigate the problems identified.²

2. Preliminaries and Related Work

The past years have witnessed the development of many FAIR assessment tools³ using metrics tailored to different needs with different interfaces [2]. A recent study compared 20 FATs and their 1180 relevant metrics and highlighted the different characteristics of the tools and the trends over time [4]. Seven automatic FATs are web services that can perform evaluations on data sets: AutoFAIR [3], FAIR Checker [8], FAIR Enough [7], FAIR Evaluator [25], F-UJI [11], FAIR EVA [1], and FAIROs [10]. FAIROs is intended for a specific format of digital objects, namely RO-Crate [4]. AutoFAIR is further tailored towards bioinformatics and FAIR Checker for life sciences. Some of these FATs could be adapted to specific disciplines [4], but none are specialized in social science and humanities. Sun et al. compared FAIR Evaluator, FAIR Checker, and F-UJI, focusing on the characteristics of the evaluation tools, the FAIRness evaluation metrics, as well as the testing results using some public datasets [18]. The EOSC FAIR Metrics subgroup reported that these FATs have comparability issues, leading to inconsistency [26]. They reported by example that the same dataset could end up with completely different assessment results [26]. A closer examination shows that FATs studied used different numbers of tests and the distribution of tests differs, with F-UJI having most of its tests in the "Reusability" domain while the FAIR Evaluator has most of its tests on Findability and Interoperability [26]. A recent study reported that some items in the assessment metrics of 20 FATs do not always have a one-to-one matching with the FAIR principle (i.e. could be one-to-many), have misalignment, or are not about FAIR at all [4]. Moreover, the assessment result could be different even if the metrics align [12].

Alternatively, FAIRness can be assessed by manual questionnaires such as the FAIR Data Self-

²The supplementary material is available on Zenodo with DOI: 10.5281/zenodo.15261773.

³By 18 April 2025, 30 tools and services are listed and compared on FAIRassist: <https://fairassist.org/>.

Assessment Tool⁴ and FAIR Implementation Profiles (FIPs) [14]. This paper focuses on FIPs: a FIP serves as a comprehensive framework that captures the strategies for the management of digital assets of FAIR Implementation Community (FIC), a self-identified organization (often with more than one person) with a common interest that aspires to the creation of FAIR data and services [14]. The creation of FIP starts with a FIP template consisting of 21 questions that address various facets of the FAIR principles, focused explicitly on datasets and their associated metadata. The answer to each question is restricted to FAIR Enabling Resources (FERs): a list of tools, services, licenses, infrastructures, and other services and resources that help researchers, data stewards, and institutions make data FAIR. These answers are considered community standards. These questions prompt thoughtful responses that illustrate how resources are allocated and utilized to facilitate adherence to the FAIR principles. FIP not only captures the commitment of the FIC to implementing FAIR practices but also serves as a valuable tool for evaluating and enhancing the effectiveness of these efforts over time. In recent years, FIP gained popularity in many domains, including Social Science and Humanities (SSH) [22]. These FIPs can serve as references for decisions in data management, especially for the R1.3 principle: (Meta)data meet domain-relevant community standards. The applications of FIPs for data management include providing suggestions for Data Management Plans [16] and decisions for the upcycling of legacy data [21]. To facilitate the creation of FIPs, the FIP Wizard⁵, a user-friendly web interface, is used to guide users through the questions to document their FAIR implementation strategies. This process fosters introspection within communities, aiding in the recognition of their strengths and aspects that require enhancement regarding FAIR practices. Furthermore, FIPs can raise awareness within a particular field and act as references for other communities to create policies that align with FAIR principles. The resulting FIPs are published as nanopublications [14]. Published FIPs, FERs, and other related resources can be found on FIR Connect’s search engine.⁶ Next, we introduce the FIPs of two communities.

The CLARIN (Common Language Resources INfrastructure) is one of the oldest European ERICs (European Research Infrastructure Consortia) and serves the community of researchers developing and interested in Language Resources and Tools (LRT)⁷. Nowadays, its network spans 243 institutions in 24 countries in Europe and South Africa and their associated communities of researchers. Over the years, this infrastructure has developed a set of requirements for datasets and their metadata. These requirements are vetted by the CLARIN Technical Centre Committee in the “Checklist for CLARIN B-Centres”⁸. To become a CLARIN B centre, a candidate centre does a self-assessment, which gets reviewed by the CLARIN Assessment Committee resulting, in general, in successfully being granted the B centre status. This procedure is repeated every 3 years. Although these requirements predate the FAIR principles, they largely overlap. So, upon creation of a FIP for CLARIN, these requirements have become the basis, resulting in the answers to the FIP questionnaire as shown in Table 1⁹. It also includes some resources under development, which are in *italics*, indicating future use. It does show that long-established communities, like the LRT one as now represented by CLARIN, often have a well-established and still actively developed set of technologies in place to meet the FAIR principles. CLARIN develops and maintains the ‘CLARIN Virtual Language Observatory’ [20], providing access to the joint metadata domain of the European CLARIN infrastructure on LRT. The FIP was created in collaboration with the FAIR Expertise Hub using the FIP template 4.3.4.

The ODISSEI Portal Community is the subcommunity of ODISSEI¹⁰, the Dutch Research Infrastructure for Social Sciences and Economics, which is a collaborative consortium of 45 member organisations, including social sciences faculties in the Netherlands, Central Statistics Office (CBS), public research agencies, and research institutes. This community advances the implementation of FAIR principles by

⁴<https://ardc.edu.au/resource/fair-data-self-assessment-tool/>

⁵<https://fip-wizard.ds-wizard.org/>

⁶<https://fairconnect.pro/search-fair-nanopublications/>

⁷<https://www.clarin.eu/>

⁸<http://hdl.handle.net/11372/DOC-78>

⁹Answering a FIP question by items taken from a registry is a feature not yet implemented in the FIP Wizard. Thus, the FIP in the supplementary material does not have it.

¹⁰<https://odissei-data.nl/>

developing and maintaining the ODISSEI Portal [6].¹¹ Different from CLARIN, this community does not produce datasets, but collects and provides metadata through their portal. Thus, their FIP mostly focuses on metadata. The FIP was created in a similar fashion as CLARIN with the same template.

3. Methodology and Implementation Decisions

Our methodology consists of three steps: 1) align assessment metrics of FATs with FIPs, 2) compare the assessment results of representative datasets, and 3) reflect the assessment results with FIPs.

Step 1. To better understand how the FATs align with FIPs, we compare the metrics of the two FATs and align them with each other and with the FIP template (RQ1). This will clarify why the assessment results are different and how they can be explained. As explained in Section 2, there are many automated FATs. Given the comparability issues as addressed by [26], in this paper, we use FAIR Enough and FAIR Checker as they both use the FAIR Maturity Indicators (MI) [13] as their assessment criteria. They are selected for their state-of-the-art development, usability, and robust performance. We leave the comparison of the assessment results with the remaining FATs for future work. FAIR Enough was built on the FAIR Test Python library.¹² and evaluates how much online resources follow the FAIR principles [7] with integrated insights from previous assessment initiatives, such as F-UJI [11], FOOPS! [9] and FAIR Evaluator [25]. Assessments can be performed using the web service¹³ as well as the FAIR Extension[15], its Google Chrome extension. The FAIR Checker¹⁴ [8] is another assessment tool that includes a collection of SPARQL queries for evaluating FAIR principles and a SHACL constraints generator for enhancing metadata completeness. It could be used as a web service as well as through its RESTful API [8]. Moreover, it further specializes in digital objects in life sciences. During alignment, where such compatibility issues occur, we choose the closest possible one-to-one match while preventing misalignment.

Step 2. We choose the representative datasets of the two communities and obtain their assessment results for comparison to answer RQ2. Although it is easy to obtain assessment results of an individual dataset and aggregate the scores corresponding to each subprinciple [12], there is no aggregation scheme, as far as the authors are aware, that compiles the results of all the datasets and other digital assets a community publishes into one aggregated assessment result. Such an aggregation is not as simple as taking the average of the assessment results of datasets. There are many factors to be taken into account: the evolution of data management strategies, different versions of datasets, the change of data infrastructures, duplicates, exceptional datasets, different licenses, etc. As a pilot study, we focus on communities' representative datasets, which are to be assessed by selected FATs. Apart from the regular communities that produce and publish data, it was observed that some communities have an impact on datasets and their management by providing data infrastructures and services. These communities are often not the curators of datasets themselves. Thus, we choose one community of each kind. Our paper takes a similar approach as [18] but with a focus on DH datasets. Next, we introduce the two chosen DH communities and their representative datasets, respectively. We then obtain the assessment results of representative datasets using the two selected tools. To answer RQ2, we rely on the answer to RQ1 and examine three aspects: 1) coverage of the FAIR principles, 2) the difference in assessment results, and 3) the format and reuse of the results.

For CLARIN, we choose the **Awetí dataset**¹⁵ in The Language Archive at the Max Planck Institute for Psycholinguistics [5]. The MPI-PL offers resources for long-term archiving language resources and tool development. They have been the birthplace of the CLARIN infrastructure and they are the primary archive of the DOBES project funded by the Volkswagen Foundation for multimodal resources on endangered languages collected by trained field linguists. The Language Archive has been a CLARIN

¹¹<https://portal.odissei.nl/>

¹²<https://github.com/MaastrichtU-IDS/fair-test>

¹³<https://fair-enough.semanticscience.org/>

¹⁴<https://fair-checker.france-bioinformatique.fr/>

¹⁵<https://hdl.handle.net/1839/74209f7d-c6ff-4129-8afd-64c78f4d300e>

Principle	FIP Question	CLARIN FERs	ODISSEI FERs
F1-MD	What globally unique, persistent, resolvable identifiers do you use for metadata records?	DOI, Handle	DOI, Handle
F1-D	What globally unique, persistent, resolvable identifiers do you use for datasets?	Handle, DOI	none
F2	Which metadata schemas do you use for findability?	CMDI (https://www.clarin.eu/cmdi), Dublin Core	DDI, DCAT, <i>Croissant</i>
F3	What is the schema that links the persistent identifiers of your data to the metadata description?	CMDI resource proxy	DOI
F4-MD	Which service do you use to publish your metadata records?	CLARIN VLO (https://vlo.clarin.eu/)	Zenodo and ODISSEI Portal
F4-D	Which service do you use to publish your datasets?	depends	none
A1.1-MD	Which standardized communication protocol do you use for metadata records?	HTTPS, OAI-PMH	HTTPS, OAI-PMH, SPARQL API
A1.1-D	Which standardized communication protocol do you use for datasets?	HTTPS	<i>Data Access Broker</i>
A1.2-MD	Which authentication & authorisation service do you use for metadata records?	none	<i>SURFConeXT, SURF SRAM</i>
A1.2-D	Which authentication & authorisation service do you use for datasets?	SAML, Open ID Connect	SURF SRAM, CBS Microdata Authentication
A2	What metadata preservation policy do you use?	Core Trust Seal (CTS) & CLARIN B centre status	DataCite DOI Policy
I1-MD	What knowledge representation language (allowing machine interoperation) do you use for metadata records?	CMDI & its concept links, SKOS-based vocabularies	JSON-LD, RDFS, SKOS, JSON
I1-D	What knowledge representation language (allowing machine interoperation) do you use for datasets?	depends	not applicable
I2-MD	Which structured vocabularies do you use to annotate your metadata records?	CMDI vocabularies, SKOS-based vocabularies	ELST (Social Science Thesaurus), CESSDA Topic Classification; CBS taxonomy, Schema.org, Dublin Core, DDI Controlled vocabularies
I2-D	Which structured vocabularies do you use to encode your datasets?	vocabularies registered in the <i>CLARIAH FAIR Vocabulary Registry</i> (https://registry.vocabs.dev.clariah.nl/)	not applicable
I3-MD	What semantic model do you use for your metadata records?	CMDI concept link	DCAT, SKOS, DataCite Metadata Scheme, DDI Codebook
I3-D	What semantic model do you use for your datasets?	a model registered in the <i>Standard Information System</i> (https://standards.clarin.eu/sis/)	not applicable
R1.1-MD	Which usage license do you use for your metadata records?	Creative Commons (CC0)	Creative Commons (CC0)
R1.1-D	Which usage license do you use for your datasets?	depends	Not applicable
R1.2-MD	What metadata schema do you use for describing the provenance of your metadata records?	none	none
R1.2-D	What metadata schema do you use for describing the provenance of your datasets?	depends	not applicable

Table 1: CLARIN & ODISSEI FIPs (*italic* FERs are planned future use)

B center for many years and provides its (meta)data compliant with the CLARIN requirements, so the FAIR assessment of the dataset could be a good representative of the CLARIN community.

For the ODISSEI Portal community, we select the dataset “Banen en lonen op basis van de Polisadministratie” (“Jobs and wages based on the Policy Administration” in English)¹⁶ contains data on jobs and wages in Dutch companies, derived from the Polisadministratie (the “Police administration” in English), which records all income relationships subject to wage tax. For simplicity, we refer to this dataset as **the ‘Banen’ dataset**. The dataset is managed by Statistics Netherlands (CBS) and accessible through the ODISSEI portal, requiring approval for use. This dataset is vital for the Semantic Web because it provides structured, detailed data about employment and wages in the Netherlands, which can be linked and integrated with other datasets to enhance understanding of labor markets, economic trends, and social factors. By using standardized formats and ontologies, the data can be made machine-readable and interoperable, facilitating automated analysis and richer insights across disciplines. Its availability as open data also supports transparent research and evidence-based policy-making. Integrating such datasets into the Semantic Web ensures that they are accessible, discoverable, and useful for a wide range of users and applications. Since the Banen dataset is not publicly accessible, some entries of assessment are expected to fail.

Step 3. Finally, our RQ3 requires us to use the assessment results and compare them with the FIPs regarding each aligned assessment metric. We discuss how FERs are used and how they are reflected in the assessment results. We compare the FIPs and discuss how to improve FAIRness scores.

4. Evaluation and Discussion

4.1. Comparing FATs’ assessment metrics with FIP questions

We answer RQ1 by comparing the evaluation metrics with the questions of FIP. They address FAIR at different levels and emphasize different aspects: the FAIR assessment has much more technical details on “behavior” of a digital asset. FATs detect if certain standards or specified resources are being used and return Success/Failure. However, the FIP is broader with open questions on the use of resources, i.e. community choices. For example, FAIR Checker would give the points when the condition is met: “Metadata includes provenance” by verifying that at least one provenance property from PROV, DCTerms, or PAV ontologies is found in metadata. In contrast, FIP asks “What metadata schema do you use for describing the provenance of your datasets?” For this reason, aggregating the assessment results requires the FATs to include the detected resources in their output. The assessment could be of different specificity. For F4, FIP asks which service is used to publish metadata/data, unlike FAIR Enough, which tests if the data can be found in Bing. For I1, FAIR-Checker verifies that at least one RDF triple can be found in metadata. In contrast, the corresponding FIP question is much more specific: “What knowledge representation language (allowing machine interoperation) do you use for metadata records?” For FAIR Enough, the metrics are different with a split of strong and weak for metadata and data, respectively. In addition, it was also noticed that the assessment of interoperability can be muddled. For example, assessment on the use of ontology is in I3 for FIP but is in I2 for FAIR Checker, and is not well-addressed by FAIR Enough. As shown above, there are compatibility issues between FAIR assessment metrics and FIP. This adds complexity to RQ2 and RQ3 to be shown below.

4.2. Comparing the assessment results of representative datasets

Next, we answer RQ2 (how the assessment results differ between selected FATs) by studying how the results differ per aspect of FAIR. We then study the differences in the assessment results and discuss the reasons. Finally, we evaluate the reuse of these results.

It was noticed that they both skipped some subprinciples. There are only 12 metrics by FAIR Checker in comparison with a more sophisticated assessment of 22 metrics by FAIR Enough. They overlap on

¹⁶<https://doi.org/10.57934/0b01e410807c72a7>

10 metrics over FAIR principles including F1, F2, A1, I1, I2, I3, and R1.1. Both FATs have more metrics about metadata than data. Furthermore, more metadata-data pairs are in the metrics of FAIR Enough.

Looking at the overall result, both FATs give higher marks to the **Banen** dataset: FAIR Checker assigns 70.83% to Banen in comparison with Awetí's 16.67%. The score by FAIR Enough is 3/22 for Awetí and 10/22 for Banen. Regarding their overlap, for Awetí, they agree on 7 out of 10 overlapping metrics. Conversely, they agree on merely half of the criteria for Banen, even by matching FAIR Checker's '1/2' with FAIR Enough's 'Success'. The difference lies mostly in the identifier, the communication protocol, the authentication and authorization, external links and outward references, and the licenses. This addresses the difference in implementation design between FATs, which could be further explored. Notice that the FATs have a high preference for metadata represented via Linked Data. Only F-UJI, which is run as part of the FAIR Enough run, recognizes the use of CMDI by CLARIN and is able to, although still very limited, interact with it. It was also noticed that FAIR Enough failed to detect the use of Handle thus the assessment result for F1 has some errors.

The assessment results could be downloaded from FAIR Checker in the format of CSV while FAIR Enough uses JSON. Their results are not machine-actionable not to mention interoperable. The explanation of the assessment result of each metric is in text rather than a structured form. None of their assessment results can be directly used to enrich their corresponding FIP without human interpretation, which reduces the reuse.

4.3. Comparing the assessment results with FIPs

Finally, for RQ3, we compare the assessment results with their corresponding FIPs. We further elaborate on our observations, analyze the differences, and address means to improve the FAIRness scores.

Findability. Both CLARIN and ODISSEI use globally unique, persistent identifiers like DOIs and Handles for metadata. The two datasets use DOI and Handle, and are aligned with the community standards captured by FIP. CLARIN relies on CMDI and Dublin Core for metadata schemas, while ODISSEI uses DDI, DCAT, and Croissant. The use of multiple schemas in ODISSEI may enhance compatibility with different platforms. Here, the Awetí dataset lost points for both FATs, in contrast, the Banen dataset gained points. Metadata indexing varies: CLARIN's metadata is indexed in the Virtual Language Observatory (VLO), while ODISSEI's metadata is found in Zenodo and its own portal. Both are aligned with community standards: Awetí can be found in VLO and Banen can be found in the ODISSEI Portal. However, this aspect was assessed by neither FAT.

Accessibility. Both use HTTPS for accessing metadata, ensuring security. CLARIN and ODISSEI support OAI-PMH for metadata exchange, but ODISSEI also provides APIs (Dataverse and SPARQL), increasing accessibility. Authentication for metadata is stricter in ODISSEI, using SURF systems, whereas CLARIN has no authentication for metadata records. Dataset authentication varies: CLARIN supports SAML and OIDC, while ODISSEI relies on SURF-SRAM and CBS Microdata Authentication, showing differences in security approaches. The FATs diverge in their authentication assessment outcomes for both datasets, underscoring the differences in implementation.

Interoperability. CLARIN and ODISSEI are different in knowledge representation languages, semantic models, and metadata schema. Despite that FATs' the assessment results agree with each other, the used FERs are not highlighted in the assessment result for comparison, which could be a barrier for the assessment of metadata interoperability.

Reusability Both infrastructures adopt Creative Commons CC0 licenses for metadata, promoting open reuse. Dataset licensing is less uniform, depending on specific contexts in CLARIN, while ODISSEI does not specify dataset licenses, potentially limiting clarity on reuse conditions. Provenance tracking differs: CLARIN lacks provenance metadata schemas, whereas ODISSEI is planning to use a customized JSON schema, offering more explicit tracking of metadata origins. Also in ODISSEI the versions of the various software components used to ingest and enrich the metadata is provided and made available via persistent identifiers.

5. Conclusion and Future Work

In this paper, we answered our first research question by critically examining the evaluation metrics of two popular FAIR assessment tools. We highlighted the differences and addressed the compatibility issues. The second research question was answered by a detailed comparison of FATS' results from three aspects. This difference reflects the FATS' evaluation metrics and implementation choices. Finally, for RQ3, we compare FATS' results against FIPs. Using FATS to assess the representative datasets shows that these domain-specific solutions pose problems for these generic tools. The FATS don't understand enough of some of the technologies chosen by the infrastructure, e.g., CMDI. However, to go beyond being FAIR within one's own community aiming for a better score with the more generic FATS will improve the maturity of FAIRness, e.g. by also making a core part of the metadata available as linked data in the landing page of a dataset.

European projects, e.g., FAIRCORE4EOSC [19] and OStrails [17], are developing solutions based on FAIR testing execution flows that address individual tests[23]. This approach allows a community to mix and match tests from the various tools to make FAIR assessment better align with their FIP. It also allows a community to implement some of these tests with some more specialized criteria to suit their solutions better. Both infrastructures exhibit domain-specific implementations of FAIR principles, highlighting how different research communities tailor their metadata, authentication, and interoperability approaches. These differences suggest that achieving FAIRness is not a one-size-fits-all approach. Enhancing dataset indexing, licensing policies, and provenance metadata could improve FAIR compliance while ensuring alignment with FATS. Neither fully meets all FAIR criteria, as dataset indexing is inconsistent, and dataset licensing and provenance tracking remain underdeveloped.

In conclusion, this paper presents a proof-of-concept work towards a more complete evaluation and analysis on a larger scale of communities' FAIR practices. Despite that the assessment results can be easily obtained for an individual dataset, it remains challenging for DH researchers to interpret and take advantage of these disparate assessment results. Our approach can be adapted and applied to other domains beyond SSH and with alternative FATS. Although the assessment results and their comparison with FIP can be different, the approach can be scalable when an aggregation method is implemented for assessment results (and that of various digital objects) is available.

6. Acknowledgments

This publication is part of the project Social Science and Humanities Open Cloud for the Netherlands (SSHOC-NL) with file number 184.036.020 of the research programme National Roadmap for Large-Scale Research Facilities which is (partly) financed by the Dutch Research Council (NWO). The authors would like to thank Liliana Melgar for the discussions and her assistance with proofreading.

References

- [1] F. Aguilar Gómez and I. Bernal. FAIR EVA: Bringing institutional multidisciplinary repositories into the FAIR picture. *Sci. Data*, 10(1):764, Nov. 2023.
- [2] M. Antonioletti, C. Wood, N. Chue Hong, E. Breitmoser, K. Moraw, and M. Verburg. Comparison of tools for automated fair software assessment, Aug. 2024.
- [3] J. Bonello, E. Cachia, and N. Alfino. Autofair-a portal for automating fair assessments for bioinformatics resources. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1865(1):194767, 2022.
- [4] L. Candela, D. Mangione, and G. Pavone. The fair assessment conundrum: Reflections on tools and metrics. *Data Science Journal*, 23(1), 2024.
- [5] S. Drude, D. Broeder, P. Trilsbeek, and P. Wittenburg. The language archive: A new hub for language resources. In *LREC 2012: 8th International Conference on Language Resources and Evaluation*, pages 3264–3267. European Language Resources Association (ELRA), 2012.
- [6] T. Emery, R. Braukmann, M. Wittenberg, J. van Ossenbruggen, R. Siebes, and L. van de Meer. The odyssey portal: Linking survey and administrative data. *Royal Netherlands Academy of Arts and Sciences (KNAW)*, 2020.

- [7] V. Emonet, R. Çelebi, J. Yang, and M. Dumontier. Towards an extensible fairness assessment of fair digital objects. *Research Ideas and Outcomes*, 8:e94988, 2022.
- [8] A. Gaignard, T. Rosnet, F. De Lamotte, V. Lefort, and M.-D. Devignes. Fair-checker: supporting digital resource findability and reuse with knowledge graphs and semantic web standards. *Journal of Biomedical Semantics*, 14(1):7, 2023.
- [9] D. Garijo, O. Corcho, and M. Poveda-Villalón. Foops!: An ontology pitfall scanner for the fair principles. In *ISWC (Posters/Demos/Industry)*, page 0, 2021.
- [10] E. González, A. Benítez, and D. Garijo. FairOS: Towards fair assessment in research objects. In G. Silvello, O. Corcho, P. Manghi, G. M. Di Nunzio, K. Golub, N. Ferro, and A. Poggi, editors, *Linking Theory and Practice of Digital Libraries*, pages 68–80, Cham, 2022. Springer International Publishing.
- [11] R. Huber and A. Devaraju. F-uji: an automated tool for the assessment and improvement of the fairness of research data. In *EGU General Assembly Conference Abstracts*, pages EGU21–15922, 2021.
- [12] P. Lamarre, J. Andersen, A. Gaignard, and S. Cazalens. A generic framework to better understand and compare fairness measures. In *Knowledge Engineering and Knowledge Management: 24th International Conference, EKAW 2024, Amsterdam, The Netherlands, November 26–28, 2024, Proceedings*, page 291–308, Berlin, Heidelberg, 2024. Springer-Verlag.
- [13] B. RDA FAIR Data Maturity Model Working Group et al. Fair data maturity model: specification and guidelines. *Research Data Alliance*, 2020.
- [14] E. Schultes, B. Magagna, K. M. Hettne, R. Pergl, M. Suchánek, and T. Kuhn. Reusable fair implementation profiles as accelerators of fair convergence. In G. Grossmann and S. Ram, editors, *Advances in Conceptual Modeling*, pages 138–147, Cham, 2020. Springer International Publishing.
- [15] P. H. Serrano and V. Emonet. The fair extension: A web browser extension to evaluate digital object fairness. *Research Ideas and Outcomes*, 8:e95006, 2022.
- [16] N. K. Singh, S. Wang, A. Maineri, T. Hofstra, M. Bruyneel, S. v. de Sandt, R. Siebes, J. van Ossenbruggen, and T. Kuhn. Aligning data management plans with community standards using fair implementation profiles. In M. Sfakakis, E. Garoufallou, M. Damigos, A. Salaba, and C. Papatheodorou, editors, *Metadata and Semantic Research*, pages 159–171, Cham, 2025. Springer Nature Switzerland.
- [17] M. Suchanek, J. Martinkova, M. Wilkinson, D. Garijo, T. Miksa, P. Manghi, R. Dyume, M. R. Mazon, K. Moilanen, M. Heinonen, et al. Ostrails d2. 1: Analysis of the ostrails interoperability framework tools compliance. ., 2024.
- [18] C. Sun, V. Emonet, and M. Dumontier. Comparison results of FAIR evaluation tools (version V0.1). <https://doi.org/10.5281/zenodo.5539823>, Sept. 2021.
- [19] T. Suominen. Faircore4eosc and the metadata schema and crosswalk registry. ., 2023.
- [20] D. van Uytvanck, H. Stehouwer, and L. Lampen. Semantic metadata mapping in practice: the virtual language observatory. In N. C. C. Chair, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, page 0, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).
- [21] S. Wang, A. Maineri, N. K. Singh, and T. Hofstra. Fair implementation profiles for the social sciences and their use cases. *Open Conference Proceedings*, 5, Mar. 2025.
- [22] S. Wang, A. Maineri, N. K. Singh, and T. Kuhn. Fair implementation profiles for social science. In E. Garoufallou and F. Sartori, editors, *Metadata and Semantic Research*, pages 284–290, Cham, 2024. Springer Nature Switzerland.
- [23] M. Wilkinson and D. Garijo. D1.2 fairness reference model for digital objects v1, Jan. 2025.
- [24] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3:160018, 2016.
- [25] M. D. Wilkinson, M. Dumontier, S.-A. Sansone, L. O. Bonino da Silva Santos, M. Prieto, D. Batista, P. McQuilton, T. Kuhn, P. Rocca-Serra, M. Crosas, et al. Evaluating fair maturity through a scalable, automated, community-governed framework. *Scientific data*, 6(1):174, 2019.
- [26] M. D. Wilkinson, S.-A. Sansone, G. Marjan, J. Nordling, R. Dennis, and D. Hecker. Fair assessment tools: Towards an “apples to apples” comparisons, Dec. 2022.