FAIR Data Assessment Using LLMs: The Fair-Way

Anmol Sharma
RWTH Aachen University
Aachen, Germany
anmol.sharma@rwth-aachen.de

Sulayman K. Sowe
Fraunhofer FIT
Aachen, Germany
sulayman.k.sowe@fit.fraunhofer.de

Soo-Yon Kim
RWTH Aachen University
Aachen, Germany
kim@dbis.rwth-aachen.de

Sayed Hoseini
Hochschule Niederrhein
Krefeld, Germany
Sayed.Hoseini@hs-niederrhein.de

Fidan Limani
ZBW - Leibniz Information Centre for Economics
Kiel, Germany
f.limani@zbw.eu

Zeyd Boukhers
Fraunhofer FIT
Sankt Augustin, Germany
zeyd.boukhers@fit.fraunhofer.de

Christoph Lange
Fraunhofer FIT
Sankt Augustin, Germany
christoph.lange-bever@fit.fraunhofer.de

Stefan Decker
RWTH Aachen University
Aachen, Germany
decker@dbis.rwth-aachen.de

Abstract

As part of modern research practices, the FAIR data principles have become essential for data discoverability, usability, and sharing. Existing implementations for automatically assessing FAIR adherence (FAIRness) often suffer from limited usability, inconsistent accuracy, and difficult-to-interpret results, as they require explicit rules to cover for specific FAIR assessment frameworks, which are not easy to generalize. This paper introduces Fair-Way, an open source tool that leverages Large Language Models (LLMs) to automate FAIRness assessment. Fair-Way applies a divide-and-conquer approach to decompose the assessment process into fine-grained tasks, as well as to split the metadata into manageable chunks. Evaluation demonstrates that Fair-Way achieves performance comparable to existing tools, while outperforming them in several key metrics. Moreover, Fair-Way generalizes across FAIR assessment indicators without requiring explicitly programmed logic and supports both structured and unstructured metadata in diverse formats. Finally, it enables user-defined, domain-specific tests, which are typically not supported by other systems. Overall, Fair-Way represents a scalable and flexible solution to accelerate FAIR data practices across research domains.

CCS Concepts

• **Information systems → Information systems applications**; **Digital libraries and archives**;

Keywords

FAIR Assessment, Large Language Models, FAIRification, Research Data Management

ACM ISBN 979-8-4007-2040-6/2025/11
https://doi.org/10.1145/3746252.3760811

ACM Reference Format:
Anmol Sharma, Sulayman K. Sowe, Soo-Yon Kim, Sayed Hoseini, Fidan Limani, Zeyd Boukhers, Christoph Lange, and Stefan Decker. 2025. FAIR Data Assessment Using LLMs: The Fair-Way. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3746252.3760811

1 Introduction

Given the exponential growth in research data volume, the Findable, Accessible, Interoperable, and Reusable principles (FAIR) [1] are crucial for enhancing their discoverability, reusability, and the machine-actionability [2]. Consequently, many initiatives across domains adopt these guiding principles, such as demonstrated in the works on FAIR compliance in agriculture [3], FAIR in health [4], FAIR for AI research [5], or FAIR for LLM training data [6]. In each of these cases, the principles are tailored both in scope (supporting only certain FAIR aspects which are deemed to be of higher importance than others) and in context (i.e., to a specific case or domain). Yet, evaluating the extent to which research data aligns with FAIR principles generally remains a challenge [7, 8, 9].

Manual assessment of FAIRness is time-consuming, subjective, and often inconsistent across domains [10]. Furthermore, existing automated assessment tools are relatively limited in terms of the metadata types they handle, and information extraction rules they provide [8], leading to inconsistent results for sources with different characteristics from varying domains. Studies on FAIR evaluation tools [7, 9] underscore the need for a more comprehensive solution, capable of handling different types of sources, including structured and unstructured metadata, a multitude of metadata formats, and excellent information extraction abilities. Furthermore, enabling the metadata assessment prior to its publication results in a FAIRer dataset, a feature missing from automated FAIR assessment tools.

Recent developments in natural language processing and the emergence of Large Language Models (LLMs) have opened new avenues for simplifying data interactions [11]. LLMs have demonstrated remarkable capabilities in understanding and generating

Type	Tool	Open Source	Metric
Self-Assessment	ARDC	–	–
Self-Assessment	EUDAT	–	–
Semi-Automated	FAIR-Shake	Yes	FAIRness Maturity Indicators
Automated	FAIR Evaluator	Partially	FAIRness Maturity Indicators
Automated	F-UJI	Yes	FAIRsFAIR
Automated	FAIR-Checker	Yes	RDA FAIR Maturity Model

Table 1: Comparison of existing FAIR assessment tools.

human-like text, making them suitable for various applications [12, 13, 14], including data management [15]. LLMs show outstanding capabilities to execute complex instructions, act as agents, generate actionable recommendations, and handle diverse information formats to gain insights into data and metadata.

In this paper, we introduce an LLM-based approach for automatic FAIR assessment, as well as the corresponding tool, *Fair-Way*, capable of handling, processing, assessing, and creating different types of metadata. Fair-Way supports the assessment of published and unpublished datasets and reduces the reliance on explicitly programmed implementation for FAIR assessment, giving it a wide application range of domains and accelerating the FAIRification process. Additionally, Fair-Way enables simple domain-specific user-provided tests for vocabularies and domain-specific standards, which is difficult to assess with existing automated tools.

In summary, we make the following contributions:

- We present a novel approach for the adoption of LLMs for automatic FAIR assessment.
- We develop Fair-Way, a tool based on this approach, that supports multiple metadata formats for both published and unpublished data, including domain-specific assessments through user-provided test prompts. Fair-Way's code, prompts, and evaluation datasets are available on GitHub and a demo showcasing the application working can be accessed here.
- We conduct an evaluation of open and closed-source LLMs for FAIR assessment tasks.

The rest of this paper is structured as follows: Section 2 reviews related work, section 3 describes our methodology and the implementation of the tool, section 4 presents an analysis of the system's capabilities and limitations, and section 5 summarizes key contributions and future directions.

2 Related Work

FAIR assessment utilizes frameworks that define indicators and associated tests to measure a resource's adherence to FAIR principles. Prominent examples include the Research Data Alliance's FAIR Data Maturity Model (DMM) [16], the FAIRsFAIR Data Object Assessment Metrics (DOAM) [17], and the FAIRness Maturity Indicators (MIs) [18]. These frameworks provide community-agreed metrics that FAIR assessment tools evaluate.

Building on these community metrics, various tools have been developed for FAIR assessments, ranging from manual questionnaires to fully automated systems [10], as summarized in Table 1. Manual questionnaire-based tools, such as those by ARDC [19] and EUDAT [20], rely on self-assessment and promote understanding, but are time-consuming, require prior FAIR knowledge, and lack quantitative rigor. Semi-automated tools address some of the limitations mentioned above, by allowing some level of automated evaluation; FAIRshake [21, 22], for example, uses MIs and allows users to define custom assessment "rubrics". For fully automated assessment, we have tools like F-UJI [23, 24], based on the DOAM metrics, which checks for domain-agnostic indicators. However, extending it to new terminologies, metadata formats, or domain-specific checks necessitates programming and modifications to its source code. Another automated tool, FAIR-Checker [25, 26], employs the DMM metrics, using Semantic Web technologies like knowledge graphs for metadata analysis, but designed specifically for bio-informatics artifacts. While existing automated tools mark progress, they face challenges in generalizing across diverse metadata standards, languages, and formats. They require explicit programmed logic for each indicator test and resource type [24, 26]. This inherent inflexibility, including customization calls for broader applicability and the need for a more generalizable solution.

3 Methodology

To enable reliable and automated FAIR data assessment using LLMs, our methodology combines a FAIR assessment framework with an LLM-based processing pipeline. For this, it was crucial that the FAIR assessment framework includes practical, automated tests for every metric. Thus, we chose FAIRsFAIR DOAM [17] as, besides supporting automated tests, they are domain-agnostic, community driven, well documented and used by assessment tools like F-UJI.

To automate evaluation of our chosen metrics, we adopted a design science approach to develop and refine an LLM-driven pipeline. Initial experiments employed open source LLMs, like Microsoft Phi-4 (14b), Meta Llama3.1 (8b), and Qwen2.5-Coder (14b) [27, 28, 29]. The models were prompted with simple instructions plus full dataset metadata to perform a test evaluation on. We also specified a required JSON response format for each test to enable complete automation. Such an example instruction is *"Check if metadata includes descriptive elements, like creator, title, publisher, or publication date."*, for which the model was asked to execute the given instructions and respond strictly in the provided JSON schema.

However, this approach resulted in many inconsistent and non-reproducible outputs, as part of model hallucinations and failure to adhere to instructions, especially with smaller models that often returned lengthy explanations instead of the requested output. This can be attributed to (1) a lack of specialized fine-tuning of open source models for metadata evaluation, (2) LLM context window limitations, as performance can significantly degrade with increased token count, (3) tokenization used in models, which typically treats a word plus its space as one token, impacting how a full prompt is tokenized, and (4) structured metadata (e.g., JSON) disproportionately expanding token counts. For instance, using the prompt above on metadata retrieved from this Zenodo record into a JSON format resulted in **1822** tokens using GPT-4o tokenizer[1] and it further inflates with few-shot examples on the given task. For our solution, to mitigate these challenges, we adopted a *Divide-and-Conquer* strategy, constraining the context window of LLMs via specific prompts and processing metadata in manageable chunks rather than complete sets.

[1] https://platform.openai.com/tokenizer

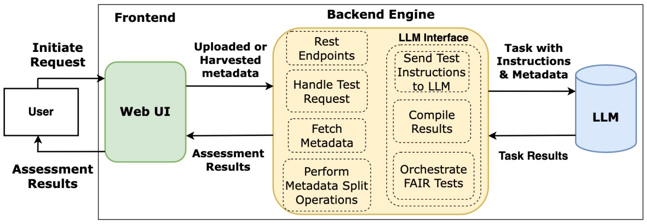

Figure 1: System Architecture

The *system's operational flow* begins with user input – either a resource URL (from supported repositories) or a local metadata file, and optional user defined domain-specific tests. Fair-Way then extracts metadata from online resources (via embedded content and repository APIs like Zenodo [30]) or by parsing uploaded files. The complete metadata is potentially chunked if determined to be large based on pre-defined metadata limits for each format, and subsequently assessed for each FAIR metric, where the LLM processes each chunk using tailored prompts and few-shot examples. Finally, successful chunks for a task are combined together to form a single assessment result. The key steps of the workflow include:

- **Metadata pre-processing** reduces tokens by pruning redundant whitespaces and validating structured metadata for syntax.
- **Chunking** divides large metadata files into smaller parts using the LangChain framework to limit tokens per LLM call, while small files are processed as whole. The LLM later synthesizes results from multiple chunks for a test into single final result.
- **Task decomposition** breaks test instructions into simpler, clear prompts for the LLM, each requiring a JSON response.
- **Result reuse** for certain metrics. Tests are performed on results obtained from other tests (e.g., using extracted file information to check for open data formats).

Fair-Way's *architecture* comprises of three core components (as shown in Figure 1): **Frontend** developed with VueJS for user interaction. **Backend** is a FastAPI-based RESTful service as the main engine encompassing multiple core components. **LLM Interface** is built within the backend, for managing requests to the chosen LLM service (Ollama [31] or the OpenAI) and compiling results. The inner workings of all the components are described in detail on our GitHub repository.

The *prompts* emulate the instructions of the chosen DOAM metric specification. Based on initial observations, we iteratively curated our prompts employing several prompt engineering techniques: **Specific Instructions** with clear and precise directives for each assessment test [32]. **Structured Output** enforces JSON model output for systematic parsing and complete automation of tests. **Chain-of-Thought** instructions are used by decomposing a test into sequential subtasks [33] to follow. **Two-Shot Prompting** examples [34] with varied inputs and expected outputs are provided for each test. Our experimentation found two-shot examples as a fine balance considering context window usage aiding non-fine-tuned LLMs in each FAIR test. **Special Prompts** are also used to combine results from multiple chunks and sources into a single result for a single test. These prompting techniques, combined with the pruning and splitting strategies for metadata, enable LLMs to perform automated FAIR assessments. For *domain-specific* testing, users are asked to select a vocabulary test or a domain standard test, specify their domain, and provide relevant vocabulary terms or a standard domain check (as prompt) for evaluation by LLM. As shown in Figure 2, a user can define a vocabulary check for the Ecology domain by providing a term to check and a brief description. These are evaluated independently in addition to the domain-agnostic DOAM metrics. The user-provided tests, composed of test types, test instructions and the corresponding domains, are then passed to the LLM as part of a special prompt to perform domain-specific testing. Collectively, these functionalities support researchers in curating metadata that adheres to FAIR principles, complies with domain-specific standards, and facilitates the publication of research artifacts to promote discoverability and seamless sharing.

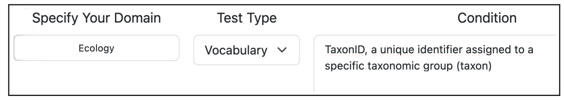

Figure 2: Domain Specific Testing

LLM	Temp.	Mean Structural Accuracy	Mean Sequence Match	Exact Match Accuracy	Mean BERT Score
GPT-4o	0.3	1.0	**0.557**	**0.905**	**0.937**
Mistral-Small	0.3	1.0	0.523	0.84	0.869
Phi-4	0.7	1.0	0.39	0.824	0.89
llama3.3	0.5	0.989	0.52	0.603	0.888
Qwen2.5-Coder	0.3	1.0	0.457	0.778	0.909

Table 2: Fair-Way LLM FAIR Assessment Comparison

4 Evaluation & Discussion

Our evaluation is twofold: (i) identifying the most suitable LLM for FAIR assessment by comparing several LLMs, and (ii) benchmarking Fair-Way against existing automated tools with the same metrics. To select Fair-Way's optimal LLM, we evaluated GPT-4o and a few open-weights models, prioritizing performance, deployability, and structured output capabilities. We created a benchmark consisting of 15 datasets across five domains with manually curated ground truths from harvested metadata for each dataset example for each of utilized DOAM metrics. The curated ground truths are available in our (GitHub) repository. Performance was assessed using four metrics, based on a 0 (lowest) to 1 (highest) scale: **Mean Structural Accuracy** for adherence to the required JSON output schema. **Mean Sequence Match** for accuracy of extracted entity lists (1.0=exact set match, 0.5=partial overlap, 0.0=empty set intersection). **Exact Match Accuracy** for verbatim correctness of specific entities embedded in metadata. **Mean BERT Score** for semantic similarity of long texts (e.g., comments, summaries, etc.)

Table 2 summarizes the LLM comparison on our benchmark, showcasing best metrics per model with the optimal temperature. We standardized hyperparameters *Top-p* (0.9) and *context window* (max. 5800 tokens), varying the temperature parameter (range: 0.3–0.9) to identify optimal settings for each model. Final scores were averaged over two independent runs per dataset for each model and temperature combination to mitigate variability. OpenAI's GPT-4o achieved the highest scores across all metrics, emerging as the optimal model for FAIR assessment. Lower temperatures yielded more precise outputs for most models, excluding Phi-4 [27]; Mistral-small [35] was the top-performing open source model.

FAIRsFAIR DOAM Indicators	Earth Sciences Dataset		Finance Dataset		Climate Science Dataset	
	F-UJI	Fair-Way	F-UJI	Fair-Way	F-UJI	Fair-Way
F1-01D: Data assigned a globally unique identifier	✓	✓	✓	✓	✓	✓
F1-02D: Data assigned a persistent identifier	✓	✓	✓	✓	✓	✓
F2-01M: Metadata includes descriptive core elements (creator, title, summary etc.)	✓	✓	✓	✓	✓	✓
F3-01M: Metadata includes data identifier to access contents	–	✗	✓	✓	✓	✓
A1-01M: Metadata contains access level and access conditions	–	✓	–	✓	✓	✓
I1-01M: Metadata represented using formal knowledge representation language	✓	✓	✓	✓	✓	✓
I2-01M: Metadata uses semantic resources	–	✗	–	✗	✗	✗
I3-01M: Metadata has links between data and related entities	✓	✓	✓	✓	✓	✓
R1-01M: Metadata specifies the content of the data (files, variables)	✓	✓	✓	✓	✓	✓
R1.1-01M: Metadata includes reusability license information	✓	✓	✓	✓	✓	✓
R1.2-01M: Metadata includes provenance information about data creation or generation	✓	✓	✓	✓	✓	✓
R1.3-02D: Data file formats open and scientific	–	✓	–	✓	–	✓
Correct/ Total Metrics	**8/12**	**10/12**	**9/12**	**11/12**	**9/12**	**11/12**

Table 3: FAIR Assessment Result Comparison (F-UJI vs Fair-Way)
✓ = Correct, – = "Failed to Complete" and ✗ = Incorrect

The Table 3 summarizes the metric completion comparison between Fair-Way (using GPT-4o model) and F-UJI [23] as standalone tools. The comparison used three randomly selected datasets from different domains: an (Earth Sciences) [36], a (Finance) [37], and a (Climate Sciences) [38]. We implemented 12 out of 17 FAIRsFAIR DOAM metrics (v0.5) with respective scores for each metric, focusing on those amenable to LLM-based information extraction and reasoning, excluding metrics like FsF-A1-02M (protocol accessibility) better suited for explicitly programmed logic. Suffixes D and M denote data and metadata metrics, respectively.

For *Findability*, both tools performed similarly on identifier detection (FsF-F1-01D/02D) and core metadata vocabulary checks (FsF-F2-01M). In *Accessibility*, Fair-Way identified public access conditions (FsF-A1-01M) for all datasets, while F-UJI failed for two Zenodo datasets, succeeding only in the Climate Sciences example. For *Interoperability*, both tools identified embedded structured metadata (FsF-I1-01M) however, both systems failed for the metric (FsF-I2-01M). F-UJI failed to check semantic resource namespaces and Fair-Way produced false positives. Notably, Fair-Way demonstrated superior performance in depth of information extracted for related entities (FsF-I3-01M), identifying citations, ORCID and RoR IDs, while F-UJI detected only version even though both succeeded in the test. In the *Reusability* assessment, both tools correctly identified licenses (FsF-R1.1-01M) and provenance (FsF-R1.2-01M) information. However, F-UJI provided incomplete information regarding dataset file formats (FsF-R1.3-02D) across all three datasets, a metric Fair-Way successfully completed.

Overall, Fair-Way completed more common metrics successfully than F-UJI: 10/12 vs. 8/12 for Earth Sciences; 11/12 vs. 9/12 for Finance; and 11/12 vs. 9/12 for Climate Sciences. This demonstrates Fair-Way's performance is at least on par with, and often superior to, F-UJI, particularly on metrics like FsF-A1-01M and FsF-R1.3-02D while providing more accurate and detailed responses by extracting richer information from metadata.

5 Conclusion

This paper introduced Fair-Way, an open source tool leveraging LLMs to automate and improve FAIR assessment by providing detailed assessment results. Our evaluation showed that Fair-Way achieves performance comparable to, and in several aspects superior to, existing tools, particularly in its ability to handle diverse metadata formats and generalize across assessment indicators.

Despite its promising capabilities, Fair-Way has limitations. The reliance on non-fine-tuned LLMs introduces a risk of hallucination on certain metrics, potentially affecting assessment precision. Additionally, the current iteration of Fair-Way employs purely LLM-based assessment tests, although certain common metrics are easily verifiable with explicitly implemented logic. To further improve accuracy and evaluation times the system would utilize a more hybrid approach utilizing MCP and tool calling to verify results of certain metrics. Furthermore, the computational demands of LLMs can lead to slower evaluation times, especially if powerful GPU hardware is unavailable.

Future work will address these limitations and expand Fair-Way's utility by offering actionable suggestions to improve metadata FAIRness post-assessment. To mitigate hallucination and improve task-specific accuracy, fine-tuning LLMs for FAIR assessment is also a significant research direction. Building on this foundation, additional FAIR metrics – challenging for traditional logic but amenable to LLMs will also be explored. These advancements aim to further establish Fair-Way as a scalable, flexible, and comprehensive solution for promoting FAIR data practices across research disciplines.

Acknowledgments

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), under the project NFDI4DataScience, grant No. 460234259

Generative AI Disclosure

During the preparation of this work, the authors used OpenAI's generative AI ChatGPT, DeepL, and Grammarly to improve the writing, make suggestions, and re-phrase. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Mark Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3 (Mar. 2016). DOI: 10.1038/sdata.2016.18.

[2] Kristin M Tolle, D Stewart W Tansley, and Anthony JG Hey. "The fourth paradigm: data-intensive scientific discovery [point of view]". In: *Proceedings of the IEEE* 99.8 (2011), pp. 1334–1337.

[3] Luiza Petrosyan et al. "FAIR degree assessment in agriculture datasets using the F-UJI tool". In: *Ecological Informatics* 76 (2023), p. 102126. ISSN: 1574-9541. DOI: https://doi.org/10.1016/j.ecoinf.2023.102126. URL: https://www.sciencedirect.com/science/article/pii/S1574954123001553.

[4] Paul Guillot, Martin Bøgsted, and Charles Vesteghem. "FAIR sharing of health data: a systematic review of applicable solutions". In: *Health and Technology* 13 (Nov. 2023). DOI: 10.1007/s12553-023-00789-5.

[5] E. A. Huerta et al. "FAIR for AI: An interdisciplinary and international community building perspective". In: *Scientific Data* 10.1 (July 2023). ISSN: 2052-4463. DOI: 10.1038/s41597-023-02298-6. URL: http://dx.doi.org/10.1038/s41597-023-02298-6.

[6] Shaina Raza et al. "FAIR Enough: Develop and Assess a FAIR-Compliant Dataset for Large Language Model Training?" In: *Data Intelligence* 6.2 (May 2024), pp. 559–585. ISSN: 2641-435X. DOI: 10.1162/dint_a_00255. eprint: https://direct.mit.edu/dint/article-pdf/6/2/559/2458950/dint_a_00255.pdf. URL: https://doi.org/10.1162/dint%5C_a%5C_00255.

[7] N.A. Krans et al. "FAIR assessment tools: evaluating use and performance". In: *NanoImpact* 27 (2022), p. 100402. ISSN: 2452-0748. DOI: https://doi.org/10.1016/j.impact.2022.100402. URL: https://www.sciencedirect.com/science/article/pii/S2452074822000246.

[8] Chang Sun, Vincent Emonet, and Michel Dumontier. "A comprehensive comparison of automated FAIRness Evaluation Tools". English. In: *Semantic Web Applications and Tools for Health Care and Life Sciences*. Vol. 3127. CEUR Workshop Proceedings. Publisher Copyright: Copyright © 2022 for this paper by its authors.; 13th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, SWAT4HCLS 2022 ; Conference date: 10-01-2022 Through 14-01-2022. Rheinisch-Westfaelische Technische Hochschule Aachen * Lehrstuhl Informatik V, Jan. 2022, pp. 44–53. URL: https://www.swat4ls.org/workshops/leiden2022/.

[9] Leonardo Candela, Dario Mangione, and Gina Pavone. "The FAIR Assessment Conundrum: Reflections on Tools and Metrics". In: *Data Science Journal* (May 2024). DOI: 10.5334/dsj-2024-033.

[10] Alban Gaignard et al. "FAIR Assessment Tools: An evaluation of assessment tools of data sets according to the FAIR principles". In: *Journal of Biomedical Semantics* (July 2023). DOI: 10.11588/heidok.00033145.

[11] Tom Brown et al. "Language models are few-shot learners". In: *Advances in neural information processing systems* 33 (2020), pp. 1877–1901.

[12] Yupeng Chang et al. "A Survey on Evaluation of Large Language Models". In: *ACM Trans. Intell. Syst. Technol.* 15.3 (Mar. 2024). ISSN: 2157-6904.

[13] Josh Achiam et al. "Gpt-4 technical report". In: *arXiv preprint arXiv:2303.08774* (2023).

[14] G Bharathi Mohan et al. "An analysis of large language models: their impact and potential applications". In: *Knowledge and Information Systems* 66.9 (2024), pp. 5047–5070.

[15] Guoliang Li, Xuanhe Zhou, and Xinyang Zhao. "LLM for Data Management". In: *Proc. VLDB Endow.* 17.12 (Aug. 2024), pp. 4213–4216. ISSN: 2150-8097. DOI: 10.14778/3685800.3685838. URL: https://doi.org/10.14778/3685800.3685838.

[16] RDA FAIR Data Maturity Model Working Group. "FAIR Data Maturity Model". In: (2020). DOI: 10.15497/RDA00050.

[17] Anusuriya Devaraju et al. *FAIRsFAIR Data Object Assessment Metrics*. English. ", Oct. 2020. DOI: 10.5281/zenodo.4081213.

[18] Mark Wilkinson et al. "Evaluating FAIR Maturity Through a Scalable, Automated, Community-Governed Framework". In: (May 2019). DOI: 10.1101/649202.

[19] Australian Research Data Commons. *ARDC-Fair-Tool*. en. https://ardc.edu.au/resource/fair-data-self-assessment-tool/. Accessed: 2024-10-24.

[20] Sarah Jones and Marjan Grootveld. *How FAIR are your data?* July 2021. DOI: 10.5281/zenodo.5111307. URL: https://doi.org/10.5281/zenodo.5111307.

[21] *FairShake*. en. https://fairshake.cloud/metric/14/. Accessed: 2024-10-17.

[22] Daniel J. B. Clarke et al. "FAIRshake: toolkit to evaluate the findability, accessibility, interoperability, and reusability of research digital resources". In: *bioRxiv* (2019). DOI: 10.1101/657676. eprint: https://www.biorxiv.org/content/early/2019/06/03/657676.full.pdf. URL: https://www.biorxiv.org/content/early/2019/06/03/657676.

[23] FAIRsFAIR Group. *F-UJI*. en. https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool. Accessed: 2025-03-10.

[24] Devaraju, Anusuriya and Mokrane, Mustapha and Cepinskas, Linas and Huber, Robert and Herterich, Patricia and Vries, Jerry and Akerman, Vesa and L'Hours, Hervé and Davidson, Joy and Diepenbroek, Michael. "From Conceptualization to Implementation: FAIR Assessment of Research Data Objects". In: *Data Science Journal* 20 (Feb. 2021). DOI: 10.5334/dsj-2021-004.

[25] Bioinformatique France. *FairChecker*. en. https://fair-checker.france-bioinformatique.fr/. Accessed: 2025-03-10.

[26] Assmann Cora ; Gerlach Roman ; Lang Kevin ; Neute Nadine ; Rex Jessica. "FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards". In: *Journal of Biomedical Semantics* 14 (Apr. 2023). DOI: 10.1186/s13326-023-00289-5.

[27] Marah Abdin et al. *Phi-4 Technical Report*. 2024. arXiv: 2412.08905 [cs.CL]. URL: https://arxiv.org/abs/2412.08905.

[28] Llama AI Team @ Meta. *The Llama 3 Herd of Models*. 2024. arXiv: 2407.21783 [cs.AI]. URL: https://arxiv.org/abs/2407.21783.

[29] Binyuan Hui et al. *Qwen2.5-Coder Technical Report*. 2024. arXiv: 2409.12186 [cs.CL]. URL: https://arxiv.org/abs/2409.12186.

[30] European Organization For Nuclear Research and OpenAIRE. *Zenodo.org*. en. Accessed: 2025-04-10. 2013. DOI: 10.25495/7GXK-RD71. URL: https://www.zenodo.org/.

[31] *Ollama*. https://github.com/ollama/ollama. Accessed: 2025-03-10.

[32] Sondos Mahmoud Bsharat, Aidar Myrzakhan, and Zhiqiang Shen. *Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4*. 2024. arXiv: 2312.16171 [cs.CL]. URL: https://arxiv.org/abs/2312.16171.

[33] Jason Wei et al. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. 2023. arXiv: 2201.11903 [cs.CL]. URL: https://arxiv.org/abs/2201.11903.

[34] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL]. URL: https://arxiv.org/abs/2005.14165.

[35] Albert Q. Jiang et al. *Mistral 7B*. 2023. arXiv: 2310.06825 [cs.CL]. URL: https://arxiv.org/abs/2310.06825.

[36] Hee-Sung Jung. *Ice/Water classification maps from study entitled "An Arctic sea ice concentration data record on a 6.25 km polar stereographic grid from 3 years of Landsat-8 imagery" [Data set]*. en. Accessed: 2025-04-10. 2025. DOI: 10.5281/zenodo.14791442. URL: https://zenodo.org/records/14791443.

[37] Tamiru Gabusho, Chala Gelana, and Ismael Hussein. *Effect of Financial Inclusion on Rural-Urban Households' Economic Welfare West Hararghe Zone, Evidence's From Some Selected Woredaa [Data set]*. en. Accessed: 2025-04-10. 2025. DOI: 10.5281/zenodo.14731831. URL: https://zenodo.org/records/14731832.

[38] Larocca Conte Gabriele; Aleksinski Adam; Liao Ashley. *Data from: Eocene shark teeth from peninsular Antarctica: Windows to habitat use and paleoceanography [Data set]*. en. Accessed: 2025-04-10. 2024. DOI: 10.5061/dryad.qz612jmq2. URL: https://datadryad.org/dataset/doi:10.5061/dryad.qz612jmq2.