FAIR Metrics for Motivating Excellence in Peer Review

Adam Craig
Hong Kong Baptist University
Kowloon Tong, Hong Kong
orcid: 0000-0002-7646-4384

Carl Taswell
University of California San Diego
La Jolla, California, USA
orcid: 0000-0002-9386-4574

***Abstract*—Past attempts to measure the quality of peer review have relied on either subjective ratings or tangentially related factors such as the sheer number or length of reviews. Previously, we introduced the Fair Attribution to Indexed Reports (FAIR) Metrics to quantify adherence to good citation practices via systematic semantic comparison of statements in the target document to those found in cited and uncited prior reports. In the present work, we define new FAIR Metrics for assessing the quality of peer review, extend the FAIR Metrics module of the PDP-DREAM Ontology with additional classes and properties needed to record FAIR Metrics analysis of a peer review, and demonstrate use with a simple example.**

***Index Terms*—Bibliometrics, scholarly research, scientific publishing, ethics and integrity, peer review, semantic web.**

I. Introduction

Many works have studied the quality and effectiveness of peer review, but most rely on unreliable subjective ratings from the editor or author. For example, [1] found that the correlation in ratings of the same peer review among three editors was only 0.62 (95% confidence interval (CI) 0.50-0.71). To the best of our knowledge, only two authors have developed measures of peer review based on objective fact. One, [2], describes a metric based on the numbers of reviewers and editors involved, either unweighted or weighted by their Hirsch indices. The other, [3], analyzed the amount of text and the tone of each review, classifying comments as either positive or negative, constructive or nonconstructive.

We instead propose a new set of FAIR Metrics that measure how accurately the reviewer's comments reflect the content of the work under review, the policies of the publication venue, and the state of knowledge in the field of study. We previously introduced four FAIR Metrics focused on identifying plagiarism and misrepresentation of prior work [4]. The unifying feature of both sets of FAIR Metrics is a claim-by-claim comparison of documents with the results recorded in a Resource Description Framework (RDF) document referencing the FAIR Metrics module of the PDP-DREAM Ontology [5].

II. Methods

Evaluation of the FAIR Metrics of review quality currently requires peer review of the peer review by a human meta-reviewer. The evaluator must perform the following steps: 1) Extract the key statements supporting the reviewer's recommendation. 2) Classify each as being about the work under review (the target work), the policies of the journal or conference (the venue), or relevant prior work (domain knowledge). 3) Classify each statement as correctly attributed or misattributed by searching for one or more statements that either corroborate or refute the reviewer's claim. (a) For a statement about the target work, search the work under review. (b) For a statement about the venue, search its editorial policies and the call for submissions. (c) For a statement about domain knowledge, search the source the reviewer cites. If the reviewer does not cite a source, consider it misattributed. 4) Additionally, the meta-reviewer may include a question for the reviewer to indicate how they could make their reasoning clearer. 5) Tabulate the number of statements of each type: correctly attributed statements about the target work (A_T), misattributed statements about the target work (M_T), correctly attributed statements about the venue (A_V), misattributed statements about the venue (M_V), correctly attributed statements about domain knowledge (A_D), and misattributed statements about domain knowledge (M_D). 6) Use these counts to calculate four ratio FAIR Metrics of peer review quality: target ratio $f_T = (A_T - M_T)/(A_T + M_T)$, venue ratio $f_V = (A_V - M_V)/(A_V + M_V)$, domain ratio $f_D = (A_D - M_D)/(A_D + M_D)$, and justification ratio $f_J = (A_T + A_V + A_D - M_T - M_V - M_D)/(A_T + A_V + A_D + M_T + M_V + M_D)$. 7) Record the analysis in a resource description framework (RDF) document using classes and properties from the FAIR Metrics module of the PDP-DREAM Ontology [6]. The RDF documents also permit use of other ontologies. For example, authors may use the Dublin Core-compatible Bibliographic Ontology to add more detailed bibliographic information about prior work [7].

III. Results

To illustrate the use of the FAIR Metrics for peer reviw, consider the following fictional example review of a paper, "A novel expert system for matching disease symptoms to small molecule-target pairs", submitted to a conference, Artificial Intelligence for Biology and Medicine 2025 (AI4Biomed 2025): "This work is out of scope. It proposes a decision-tree-based expert system for retrieving drugs and drug targets relevant to a patient's symptoms. The scope of this conference is biomedical applications of artificial intelligence (AI). Human-

TABLE I
FAIR METRICS SCORES OF EXAMPLE REVIEWS

Example	A_T	M_T	A_V	M_V	A_D	M_D	f_T	f_V	f_D	f_J
Simple synthetic example	1	0	1	0	0	1	1	1	-1	$1/3$
Review 1 of "The Multimedia FAIR Metrics Grand Challenge"	2	2	0	0	0	0	0	0	0	0
Review 2 of "The Multimedia FAIR Metrics Grand Challenge"	1	5	0	0	0	2	$-2/3$	0	-1	$-3/4$
Review 1 of [8]	2	0	0	0	0	0	1	0	0	1
Review 2 of [8]	7	0	0	0	0	0	1	0	0	1

curated decision trees are not AI [9]." The first sentence is the overall conclusion, and the three subsequent sentences are the statements supporting it. The first of the three supporting statements is about the target work; the second is about the venue, and the third is about domain knowledge.

The target work includes this passage: "We here introduce a novel expert system curated by a team of biochemists and pharmacologists that takes as input a survey of patient symptoms and uses a decision tree to retrieve a list of potentially relevant small molecule drugs and receptors on which they act." The reviewer's first supporting assertion matches this statement and thus is correctly attributed.

The call for papers for the conference includes the following text: "Relevant submissions should employ some form of artificial intelligence and demonstrate one or more potential use cases for it in biology, medicine, or health." The reviewer's second supporting assertion is a reasonable summary of this statement and thus correctly attributed.

The source for the third supporting statement, [9], discusses the challenge of arriving at a single definition of AI with an emphasis on the varied perspectives on "intelligence." It does not say that human-curated knowledge-based systems cannot be a form of AI and even contains a passage that contradicts this narrowing of the definition: "To the larger community of computer science and information technology, AI is usually identified by the techniques grown from it, which at different periods may include theorem proving, heuristic search, game playing, expert systems, neural networks, Bayesian networks, data mining, agents, and recently, deep learning." As such, the statement is misattributed. A relevant question to include would be, "By which definition is the system described in the submission not AI?" The counts are $A_T = 1$, $M_T = 0$, $A_V = 1$, $M_V = 0$, $A_D = 0$ and $M_D = 1$. The ratios are $f_T = 1/1$, $f_V = 1/1$, $f_D = -1/1$, and $f_J = 1/3$.

For the RDF record of this example analysis, see http://npds.portaldoors.net/nexus/fidentinus/Submission1Review1 in the Fidentinus repository, which we have reserved for records of resources known or suspected to contain plagiarism or other misrepresentations. It also contains two real-world examples of FAIR Metrics of reviews of a Grand Challenge proposal the authors of the present work submitted to ACM Multimedia 2024, and the Avicenna repository at www.portaldoors.net includes records of the two open reviews of [8].

IV. Discussion

By making reviews more grounded in textual evidence, the FAIR Metrics for peer review of peer review discourage the use of vague, biased, or politically motivated criticisms. This approach will make open peer review more viable and increase the value of the reviews as works in their own right, potentially even leading to the counting of peer reviews among a scholar's output as opposed to the anonymous *pro bono* labor it is today [10]. By motivating more researchers to participate in peer review, this approach will help to distribute the work more evenly, whereas it currently falls disproportionately on a small number of more motivated reviewers [11]. Additionally, this approach could lead to new innovations in the scholarly writing process itself, such as composing a knowledge graph representing the key claims of a paper first, then using software to generate natural language representations tailored to the reader. Such an approach will help to automate many tasks in peer review, further alleviating the burden it imposes.

References

[1] A. P. Landkroon, A. M. Euser, H. Veeken, W. Hart, and A. J. P. Overbeke, "Quality assessment of reviewers' reports using a simple instrument," *Obstetrics & Gynecology*, vol. 108, no. 4, pp. 979–985, 2006.

[2] A. Etkin, "A new method and metric to evaluate the peer review process of scholarly journals," *Publishing research quarterly*, vol. 30, pp. 23–38, 2014.

[3] A. Dobele, "Assessing the quality of feedback in the peer-review process," *Higher Education Research & Development*, vol. 34, no. 5, pp. 853–868, 2015.

[4] A. Craig and C. Taswell, "Formulation of FAIR Metrics for primary research articles," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 12 2018, pp. 1632–1635.

[5] A. Craig, A. Ambati, S. Dutta, P. Kowshik, S. Nori, S. K. Taswell, Q. Wu, and C. Taswell, "DREAM Principles and FAIR Metrics from the PORTAL-DOORS Project for the semantic web," in *2019 IEEE 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*. IEEE, 6 2019, pp. 1–8. [Online]. Available: https://portaldoors.org/pub/docs/ECAI2019DREAMFAIR0612.pdf

[6] A. Craig, A. Athreya, and C. Taswell, "Example evaluations of plagiarism cases using FAIR Metrics and the PDP-DREAM Ontology," *IEEE eScience 2023*, 2023.

[7] M. Dabrowski, M. Synak, and S. R. Kruk, "Bibliographic ontology," in *Semantic digital libraries*. Springer, 2009, pp. 103–122.

[8] G. Lu, C. Gong, Y. Sun, X. Qian, D. S. R. Nair, R. Li, Y. Zeng, J. Ji, J. Zhang, H. Kang, L. Jiang, J. Chen, C.-F. Chang, B. B. Thomas, M. S. Humayun, and Q. Zhou, "Noninvasive imaging-guided ultrasonic neurostimulation with arbitrary 2D patterns and its application for high-quality vision restoration," *Nature Communications*, vol. 15, no. 4481, May 2024.

[9] P. Wang, "On defining artificial intelligence," *Journal of Artificial General Intelligence*, vol. 10, no. 2, pp. 1–37, 2019.

[10] A. Craig, C. Lee, N. Bala, and C. Taswell, "Motivating and maintaining ethics, equity, effectiveness, efficiency, and expertise in peer review," *Brainiacs Journal of Brain Imaging And Computing Sciences*, vol. 3, no. 1, pp. 1–21, 2022. [Online]. Available: https://BrainiacsJournal.org/arc/pub/Craig2022MMEEPR

[11] M. Kovanis, R. Porcher, P. Ravaud, and L. Trinquart, "The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise," *PloS one*, vol. 11, no. 11, p. e0166387, 2016.