2024 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) | 979-8-3315-0494-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/WI-IAT62293.2024.00106

FAIR Metrics for motivating ethics in peer review

Adam Craig
Dept. of Physics
Hong Kong Baptist University
Kowloon Tong, Hong Kong
agcraig@hkbu.edu.hk
0000-0002-7646-4384

Carl Taswell
Jacobs School of Engineering
University of California, Sand Diego
La Jolla, California, USA
ctaswell@eng.ucsd.edu
0000-0002-9386-4574

***Abstract*—Every year, peer reviewers perform countless hours of uncompensated, anonymous labor in order to maintain the integrity of the scholarly literature. However, the high volume of research output in need of review and the scarcity of experts' time make it difficult to maintain the quality of peer review.**

We previously introduced Fair Attribution to Indexed Reports (FAIR) Metrics that quantify how well a scholarly work cites and discusses prior literature, how many novel concepts it introduces, and how free it is of plagiarism and misattributions. Unlike lexical plagiarism detection, FAIR Metrics analysis relies on identifying statements with equivalent meanings. Using the FAIR Metrics module of the PDP-DREAM Ontology, we recorded the analyses in searchable, machine-readable FAIR Metrics semantic records. This approach has the potential to strengthen the integrity of scholarly publishing by providing a more transparent and systematic way to trace the origins of ideas.

Furthermore, FAIR Metrics analysis can provide a basis for integrated multimedia idea plagiarism detection. Figures and tables often serve as visual abstracts that convey the most important points of a work, making their inclusion necessary for a complete analysis of a paper. Instead of having separate metrics of similarity for comparing prose text, tables, figures, and other forms of media, FAIR Metrics analysis involves extracting the claims that each part of the work is communicating.

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 the additional classes and properties needed to record FAIR Metrics analysis of a review, and demonstrate usage with three example reviews.

***Index Terms*—Bibliometrics, Scientific publishing, Ethical aspects, Semantic Web**

I. Prior work evaluating peer reviews

Many works have studied the quality and effectiveness of peer review, but most rely on subjective ratings from the editor or author, which vary between raters and are difficult to interpret. For example, [4] 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) and that the correlation between the author's rating and the mean of the editors' ratings was even lower, 0.28 (95% CI 0.14-0.41). Even the correlations between first and second ratings of the same peer review by the same editor were only 0.66 to 0.88 [4]. As early as 2002, [2] performed two systematic reviews of similar studies and found that the design of questionnaires varied widely, showing disagreement among researchers as to what criteria peer review should satisfy. They suggested that some of the key functions were "(1) selecting submissions for publication (by a particular journal) and rejecting those with irrelevant, trivial, weak, misleading, or potentially harmful content, and (2) improving the clarity, transparency, accuracy, and utility of the selected submissions" [2]. However, assessing how well it performs those functions remains an elusive goal. In 2016, [3] conducted a systematic review of randomized controlled trials of the effects of different interventions on the quality and timeliness of peer review. Due to the small number of trials found in the literature, the heterogeneity of study design, and the variability of results, the authors could not offer any recommendations regarding such commonly touted approaches as reporting guideline checklists, addition of a statistical peer reviewer, training of reviewers, anonymization of authors, or open review [3].

To the best of our knowledge, only two authors have attempted to develop measures of peer review based on measurable qualities of the editorial process or resulting reviews. One, [5], describes two versions of a metric, both of which depend primarily on the number of reviewers and the number of editors overseeing the peer review process [5]. They differ in that one only depends on the counts of reviewers, editors in chief, and assistant editors involved, while the other weights the score according to some measure of the expertise of the reviewers and editors, such as their respective Hirsch indices [5]. As such, the unweighted metric reflects the number of people involved in the peer review, while the weighted metric reflects how influential those involved in the review process are collectively. This does not tell us anything about the quality of any individual review. Even a world-renowned expert on a topic may be too busy staying world-renowned to carefully read and thoroughly comment on another researcher's work. The other, [6], did analyze the content of each individual review directly but only looked at the amount of text and the tone, classifying each comment as either positive or negative and constructive or nonconstructive. Our approach instead investigates whether the reviewer's comments agree with observable facts about the content of the work under review (target work), the standards of the conference or journal to which the authors submitted it (venue), or the state of knowledge in the relevant field of study (domain knowledge). We have not found any other instance of a set of metrics of the factual accuracy of a peer review.

979-8-3315-0494-6/24/$31.00 ©2024 IEEE
DOI 10.1109/WI-IAT62293.2024.00106

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 fictitious example	1	0	1	0	0	1	1	1	-1	$\frac{1}{3}$
Review 1, "The Multimedia FAIR Metrics Grand Challenge"	2	2	0	0	0	0	0	N/A	N/A	0
Review 2, "The Multimedia FAIR Metrics Grand Challenge"	1	5	0	0	0	2	$-\frac{2}{3}$	0	-1	$-\frac{3}{4}$
Review 1, [1]	2	0	0	0	0	0	1	N/A	N/A	1
Review 2, [1]	7	0	0	0	0	0	1	N/A	N/A	1

II. FAIR METRICS OF CITATIONAL FAIRNESS

We previously applied this paradigm of searching for equivalent claims in two works to the problem of evaluating a work for plagiarism of a prior work. See [7] for the initial pilot evaluations and [8] for the extended version. While semantic methods of plagiarism detection exist, as reviewed in [9] and [10], the use of FAIR Metrics to assess an accusation of plagiarism involves the creation of a systematic, machine-readable report of precisely which claims in the target text were equivalent to which in the prior work and which claims had or lacked appropriate attribution [11]. This report consists of Resource Description Framework (RDF) triples with semantics provided by the FAIR Metrics module of the PDP-DREAM Ontology, an overarching ontology that covers all concepts and relationships relevant to the PORTAL-DOORS Project and its guiding design principles [11].

From its beginning, the PORTAL-DOORS Project has sought to aid integration of diverse media types by empowering users to create distributed repositories of both semantic and lexical data and metadata grouped by problem domain rather than by data type [12]. While researchers have developed various methods for detecting plagiarism of media other than text, mainly images, as in [13] and [14], only one approach of which we are aware attempts to extract information from multiple media types [15]. While it shares the basic idea of extraction into a common format, it relies on conversion of images, audio, and other media into plain text, after which it relies on lexical comparison [15]. Our workflow specifically requires the extraction and comparison not only of data but of meaning, something difficult for a machine but of which a human reviewer is readily capable. The human-curated semantic descriptions integrated into PORTAL-DOORS records concerning multiple media types have the potential to serve as a standard for future automated approaches to multimedia knowledge extraction and comparison.

In the present work, we extend this methodology to the problem of assessing the quality of a peer review, specifically whether it shows an understanding of the content of the work under review, the requirements of the publication venue, and the relevant problem domain. In this workflow, the meta-reviewer, whether a human or software agent, creates a machine-readable report of which specific claims of the review match or contradict which statements found in the target work, venue editorial guidelines, or prior literature. As with our previous use case of plagiarism detection, the creation of a semantic description of the findings of the analysis allows for greater transparency, identification, and correction of mistakes in both the review and the analysis of the review and quantification of the results via metrics [11]. The resulting set of human-curated records will serve as a resource for future design and testing of automated approaches to extraction of claims from a wide variety of media types into a common format and comparison of claims for equivalence of meaning. This represents a more systematic approach compared to the few existing collections of annotated peer reviews, such as the one described in [16], which indicates the function of the comment, such as praise or criticism, the section of the target work to which it refers, and the aspect on which it comments, such as novelty or theoretical soundness. This paper expands on the poster presented at IEEE eScience 2024 [17] with a full account of FAIR Metrics calculations for the example reviews and further information about the extensions to the FAIR Metrics module of the PDP-DREAM Ontology.

III. TECHNICAL CHALLENGES

Every step of the FAIR Metrics analysis process represents a distinct sub-problem. While the final two steps, the tallying of counts and calculation of ratio FAIR Metrics, are trivial, every step leading up to them poses substantial challenges. In the ideal automated system, the check for equivalence of claims would use an inference engine to perform a provably correct assessment of equivalence between two semantic representations of each pair. However, the utility of such inferences relies on the thoroughness and correctness of the rules encoded in the formal ontologies used for the semantic markup. See [18] for a review of these concepts. As such, the engineering of the inference engine itself and of the formal ontology, along with the curation of and management of the semantic records constitute four integral sub-problems. Additionally, extraction of semantic representations of knowledge conveyed represents a distinct sub-problem for each medium considered [19], [20]. Even closely related sub-types may require different algorithms [21]. For example, extraction of statements from figures stored as vector graphics may require different methods versus extraction from rasterized images.

TABLE II
SUMMARY OF THE FAIR METRIC ANALYSIS OF THE SIMPLE EXAMPLE REVIEW

Statement from Review	About	Attribution	Statement from Source	Match	Count	Question
This work is out of scope.	N/A	N/A	N/A	N/A	None	None
It proposes a decision-tree-based expert system for retrieving drugs and drug targets relevant to a patient's symptoms.	Target	"A novel expert system for matching small disease symptoms to small molecule-target pairs"	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.	Yes	A_T	None
The scope of this conference is biomedical applications of AI.	Venue	AI4Biomed 2025 Call for Papers	Relevant submissions should employ some form of artificial intelligence and demonstrate one or more potential use cases for it in biology, medicine, or health.	Yes	A_V	None
Human-curated decision trees are not AI.	Domain	"On Defining Artificial Intelligence" [22]	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.	No	M_D	By which definition is the system described in the submission not AI?

IV. BENEFITS OF FAIR METRICS ANALYSIS

The stakeholders who stand to benefit from widespread use of FAIR Metrics analysis include the entire scholarly community. More systematic and transparent peer review, especially with automation accelerating parts of the process, will help distinguish researchers who uphold community standards from those who violate them. By making reviews more grounded in textual evidence, assessments will have more grounding in fact and less room for vague, biased, or politically motivated criticisms. This 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 [23]. 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 avid reviewers [24]. Additionally, this approach could lead to new innovations in the scholarly writing process itself, such as a claims-first approach in which researchers can perform a semantic search of existing works. Such an approach will help to automate many tasks in peer review, further alleviating the burden it imposes.

V. FAIR METRICS FOR PEER REVIEWS

Our FAIR Metrics of review quality serve as sanity checks to ascertain that the reviewer has grounded their recommendations regarding submission in the content of the work under review, the editorial policies of journal or conference, and the current state of knowledge in the field of study. This process requires peer review of the peer review by a meta-reviewer. Initially, this meta-reviewer will be a human reader, but we hope to automate some or all of the process in future iterations.

The first step is extraction of the key statements supporting the reviewer's recommendation. Consider the following fictional example review of a paper, "A novel expert system for matching small 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-curated decision trees are not AI [22]." The first sentence is the overall conclusion, and the three subsequent sentences are the statements supporting it.

We next classify each statement as an assertion about the work under review (the target work), about the policies of the journal or conference (the venue), or about relevant prior work (domain knowledge). In this example, 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.

We then classify each statement as correctly attributed or misattributed. To determine proper attribution of a statement about the target work, we search the target work itself for a statement equivalent to the reviewer's or a collection of statements of which the reviewer's is a reasonable summary. In this example, suppose that the target work does include the following passage: "We here introduce a novel expert system curated by a team of biochemists and pharmacologists that infers from a survey of patient symptoms a list of potentially relevant small molecule drugs and receptors on which they act via a decision tree." Since the reviewer's first supporting assertion matches this statement, it is correctly attributed.

To determine whether the reviewer has correctly attributed a statement about the venue, we need to search its editorial policies and the call for submissions. As with statements relating to the target, we need to locate an equivalent statement or a collection of statements that are collectively equivalent. Suppose that the venue to which the authors submitted is a conference and that the call for papers includes the following

TABLE III
SUMMARY OF THE FAIR METRIC ANALYSIS OF REVIEW 1 OF "THE MULTIMEDIA FAIR METRICS GRAND CHALLENGE"

Statement from Review	About	Attribution	Statement from Source	Match	Count	Question
Although the proposal underscores the importance of including statements extracted from various media types in future iterations,	Target	"The Multimedia FAIR Metrics Grand Challenge"	FAIR Metrics analysis, by providing a common framework for analysis of different media, such as both text and images, will improve the effectiveness of combining different plagiarism detection tools.	Yes	A_T	None
the initial focus on traditional scholarly article elements such as text, images, and tables may result in insufficient coverage of data diversity.	Target	"The Multimedia FAIR Metrics Grand Challenge"	Future iterations of this grand challenge will focus on extracting claims from different types of media and placing them in a shared semantic format that allows contrast and comparison.	No	M_T	What would the grand challenge need to cover on the first iteration to be worthy of having a first iteration?
This could limit the tool's effectiveness and applicability in handling multimedia and multi-format content.	N/A	N/A	N/A (It is trivially true that limiting the scope of media types covered in the first iteration limits the scope of media types to which the solutions may be applicable.)	N/A	N/A	N/A
While the proposal mentions future iterations focusing on different types of media,	Target	"The Multimedia FAIR Metrics Grand Challenge"	Future iterations of this grand challenge will focus on extracting claims from different types of media and placing them in a shared semantic format that allows contrast and comparison.	Yes	A_T	None
there's a concern about the sustainability of these efforts and their long-term impact on the scholarly community.	Target	"The Multimedia FAIR Metrics Grand Challenge"	Brain Health Alliance has sufficient funds to assure that our data repositories will remain publicly accessible for future decades of work on this important problem during the current era of information wars.	No	M_T	What evidence would show that the effort is sustainable for the 3 years required in the call for proposals?

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." In this case, the reviewer's second supporting assertion is a reasonable summary of this statement and thus correctly attributed.

Reviewers should include references for statements they make about existing domain knowledge. The meta-reviewer can then search the cited source for an equivalent statement or set of statements. If the reviewer does not provide a source or provides one that does not support the assertion, then we classify it as misattributed. In this example, the source for their third supporting statement, [22], discusses the challenge of arriving at a single definition of AI with an emphasis on the varied perspectives on "intelligence." It does not explicitly state that human-curated knowledge-based systems cannot be a form of AI and even contains a passage that contradicts any such attempt to narrow the definition to exclude them: "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." This justifies classifying the statement as misattributed.

Additionally, the meta-reviewer may include a question for the reviewer to indicate how they could make their reasoning clearer. In this example, a relevant question would be, "By which definition is the system described in the submission not AI?" For a summary of these analyses, see Table II.

To provide an example of usage in a real-world peer review scenario, we have performed analyses of the peer reviews we received in response to our ACM Multimedia Grand Challenge proposal and on the open reviews of [1] published alongside the original work in *Nature Communications*. For summaries of these analyses, see Tables III, IV, V, and VI. We summarize the total scores for all reviews in I. One notable feature common to all of these is the absence of any explicit references to the guidelines for submissions or to any published sources of domain knowledge.

The ACM Multimedia reviewers both express their doubts in vague, emotional terms such as "there's concern about the sustainability of these efforts" or "it's probably too overwhelming for a two-person team to host such a challenge" without any substantiating evidence while ignoring key sections of the text. The first ignores the statement about Brain Health Alliance's commitment of funds to maintaining the online resources needed for the Grand Challenge, while the second ignores the sections explaining the objectives and referencing prior work providing examples of FAIR Metric analyses. The use of FAIR Metrics highlights these issues by matching comments in the review to relevant sections of the original work.

By contrast, the *Nature Communications* reviewers do invoke specific facts, including both information from the work itself and outside domain knowledge, in order to explain what additional questions the authors should answer. For example, the first comment of the second review explains that "Ultrasound can permeabilize cells, and pressures and durations similar to those used here disrupted the blood-retinal barrier in one 2020 study using focused ultrasound." before asking the authors whether they have tested for such disruption. While

TABLE IV
SUMMARY OF THE FAIR METRIC ANALYSIS OF REVIEW 2 OF "THE MULTIMEDIA FAIR METRICS GRAND CHALLENGE"

Statement from Review	About	Attribution	Statement from Source	Match	Count	Question
The proposal's objective is not well defined	Target	"The Multimedia FAIR Metrics Grand Challenge"	We will allow the entry a total compute time of 8 hours to complete calculation of FAIR metrics on all 24 examples in the test pair of sets of plagiarizing and non-plagiarizing papers. For each example for which it produces a FAIR Metrics analysis report, we will award 1 point for correct formatting of the report and 1 point for correct classification of each case as plagiarism or non-plagiarism.	No	M_T	Are the requirements for a valid report in section 5 unclear or the concept of retraction for plagiarism?
and their objective is not proposed with valid examples.	Target	"The Multimedia FAIR Metrics Grand Challenge"	In our previous work demonstrating FAIR Metrics analysis of published scholarly articles, both retracted and non-retracted, we defined a workflow for focused analysis examining a target article for presence of ideas and information plagiarized from a specific comparison article [7].	No	M_T	Are the examples in [7] not valid, or should the authors recap them here?
Their goal seems to be too ambitious which is difficult to achieve during this challenge.	Target	"The Multimedia FAIR Metrics Grand Challenge"	Instead, we must establish expert consensus and will evaluate automated tools based on uniform formatting of the plagiarism analysis records and the ability to differentiate plagiarizing from non-plagiarizing documents.	No	M_T	If "the proposal's objective is not well defined," how is it too ambitious?
The authors present a workflow to help solve the problem,	Target	"The Multimedia FAIR Metrics Grand Challenge"	The goal of this grand challenge will be to automate this workflow. Here we describe the steps in more detail:	Yes	A_T	None
however, there is neither an explanation of how they design and why they use this workflow,	Target	"The Multimedia FAIR Metrics Grand Challenge"	FAIR Metrics semantic analyses require identifying statements with equivalent meaning. Recording the comparison of documents in searchable, machine-readable FAIR Metrics analysis records strengthens the integrity of scholarly publishing by providing a more transparent and systematic way to trace the origins of concepts, ideas, and creative contributions to the historical record of published literature.	No	M_T	Are you asking how the workflow serves this overall goal, or do you require more detail about how each step provides a necessary input to the next one?
nor is there evidence to support its effectiveness.	Target	"The Multimedia FAIR Metrics Grand Challenge"	In our previous work demonstrating FAIR Metrics analysis of published scholarly articles, both retracted and non-retracted, we defined a workflow for focused analysis examining a target article for presence of ideas and information plagiarized from a specific comparison article [7].	No	M_D	What evidence beyond that presented in [7] do you require?
By the way, the organizers' information is not clear and	Target	"The Multimedia FAIR Metrics Grand Challenge"	Adam Craig / agcraig@hkbu.edu.hk / Hong Kong Baptist University / Kowloon Tong, Hong Kong / Carl Taswell / ctaswell@health.ucsd.edu / Univ California San Diego / La Jolla, California, USA	No	M_T	What additional information would you recommend including?
it's probably too overwhelming for a two-person team to host such a challenge.	Domain	None provided	N/A	No	M_D	What studies have determined the optimal number of organizers for a grand challenge?
Since there have been similar challenges,	Domain	None provided	N/A	No	M_D	To what "similar challenges" are you referring?
I kindly recommend the authors submit to other Workshops.	N/A	N/A	N/A	N/A	N/A	Given the criticisms above, why would other workshops accept it?

the lack of a citation makes confirming this information more difficult, the statement is still concrete enough that one can search for supporting or contradicting evidence.

In the present analysis, we consider such additional facts to be supporting statements, not the key claims of the review. We also, discount expressions of sentiment, such as "The presented study by Lu and colleagues on noninvasive ultrasonic stimulation in the context of vision restoration is quite intriguing." and minor corrections, such as to grammar and spelling. Instead, we consider each specific comment to have a single core claim regarding what aspect of the article requires improvement. If the claim is that the authors should include details or analyses that are absent from the submitted draft, then, instead of identifying a specific quote from the original work to match or mismatch to the claim, finding a "match" involves verifying that the information is absent from the text. Since the authors' responses to the comments clarify what they have added to the final draft of the paper, we quote the responses where they acknowledge that the relevant information was missing from the original version.

Finally, while both *Nature Communications* reviews have final f_J scores of 1, Review 2 addressed more than three times as many substantive issues with the work as did Review 1. In a case like this where both reviews made legitimate critiques, the raw A_T counts help to quantify which reviewer assessed the article from more angles.

Once we have classified all supporting statements, we tabulate six counts:

- A_T the number of correctly attributed statements about the target work
- M_T the number of misattributed statements about the target work
- A_V the number of correctly attributed statements about the venue
- M_V the number of misattributed statements about the venue
- A_D the number of correctly attributed statements about domain knowledge
- M_D the number of misattributed statements about domain knowledge

We use these counts to calculate four ratio FAIR Metrics of peer review quality:

$$f_T = \frac{A_T - M_T}{A_T + M_T} \text{ target ratio}$$
$$f_V = \frac{A_V - M_V}{A_V + M_V} \text{ venue ratio}$$
$$f_D = \frac{A_D - M_D}{A_D + M_D} \text{ domain ratio}$$
$$f_J = \frac{A_T + A_V + A_D - M_T - M_V - M_D}{A_T + A_V + A_D + M_T + M_V + M_D} \text{ justification ratio}$$

For counts and ratio scores of the examples, see Table I.

Finally, we record the analysis in a resource description framework (RDF) document using the FAIR Metrics module of the PDP-DREAM Ontology. We previously introduced the PDP-DREAM Ontology as a comprehensive ontology of concepts relating to the PORTAL-DOORS Project and its guiding design principles [11] and have added to it a module for elements used to record FAIR Metrics analyses [7].

For an RDF document of the simple example FAIR Metrics analysis embedded in a PDP Nexus record, see http://npds.portaldoors.net/nexus/fidentinus/Submission1Review1. We are managing this record in the Fidentinus repository, which we have reserved for records of resources known or suspected to contain plagiarism or other misrepresentations. To view this diristry in the curation web app, see https://portaldoors.net/NPDS/NexusService/AnonResreps/Diristry/Fidentinus.

While much work remains before we can automate the creation, curation, and discovery of FAIR Metrics analysis records, the Nexus-PORTAL-DOORS-Scribe Cyberinfrastructure provides an enterprise-grade platform for managing these and other rich metadata records and now has a tool for import and export of resource information to and from multiple semantic web and citation manager formats [25]. The ability to manage both semantic and lexical metadata and populate repositories with bibliographic information imported from other platforms constitute a foundation for future development of a user-friendly interface for human evaluators and AI-powered tools that will generate a first draft of the analysis.

VI. Compatibility with Dublin Core and BIBO

The Nexus, PORTAL, and DOORS specifications provide a flexible message-level protocol for sharing records describing a wide variety of resources [26]. As such, they lack specialized fields for bibliographic information but support embedding additional text in formats such as BibTeX [27]. Similarly, the PDP-DREAM Ontology and its sub-modules have only classes and properties needed for their respective functions. Specifically, the FAIR Metrics module facilitates the description of the claims a document contains, attributions of claims to prior works, and whether claims match ones found in prior works, cited or otherwise. As another example, the Provenance submodule features classes and properties for tracking versions of a work and assigning contributor roles [28].

To include more detailed bibliographic information in semantic records, authors can employ other ontologies alongside the PDP-DREAM ontology. For example, the Bibliographic Ontology (Bibo) is a Dublin Core-compatible ontology that supports rich bibliographic markup and has equivalents for all permitted fields in a BibTeX record [30]. The British National Bibliography Linked Data platform and the Deutsche Nationalbibliothek have both used it to publish bibliographic records as semantic resource descriptions [31]. The Document class in the PDP-DREAM Ontology is semantically equivalent to the Document class in Bibo, allowing assignment of properties from both ontologies to a document. For example, when recording the analysis of the reviews of [1], we include the properties http://purl.org/dc/terms/isPartOf *Nature Communications*, http://purl.org/dc/terms/publisher 'Springer Nature', http://purl.org/ontology/bibo/doi https://doi.org/10.1038/s41467-024-48683-6, and http://purl.org/dc/terms/issued '2024-05-27' (from https://www.dublincore.org/specifications/bibo/bibo/bibo.rdf.xml, 2024-05-27).

TABLE V
SUMMARY OF THE FAIR METRIC ANALYSIS OF REVIEW 1 OF [1]

Statement from Review	About	Attribution	Statement from Source	Match	Count	Question
"However, in practice, it takes time to process the pattern generation algorithm from the acquired images and to perform spatial correction feedback, which makes the reviewer wonder if the positioning can follow normal eye movement."	Target	[1]	"To ensure accurate and effective stimulation on the retina, we implemented an auto-alignment technique in U-RP, relying on ultrasound 3D imaging and automated position detection."	Yes	A_T	None
"Consideration is needed regarding the appropriate frequency and spatial resolution that can be expected to be achieved when this device is applied to actual humans."	Target	[1]	"By investigating the performance, efficiency, and safety of U-RP, we have paved the path for U-RP from rodent animal research to human study. Future studies are desired to develop optimized and wearable stimulation devices and evaluate the quality of regenerated vision in humans."	Yes	A_T	None

TABLE VI
SUMMARY OF THE FAIR METRIC ANALYSIS OF REVIEW 2 OF [1]

Statement from Review	About	Attribution	Statement from Source	Match	Count	Question
"Have the authors performed an experiment to exclude the possibility of blood-retinal barrier disruption?"	Target	[1]	No such experiment reported. Authors' response: "After considering these two factors, we don't expect blood-retina barrier disruption induced by the ultrasonic retina prosthesis[...]"	Yes	A_T	None
"The immunolabeling for microglial cells experiment lacks a positive control."	Target	[1]	"Three healthy rats were used in the safety examination study. The unstimulated eye was used in the negative control group. One healthy rat was used in the positive control group."	Yes	A_T	None
"Figure 5B. it appears that this animal responded to light with anticipatory licks on day 7. However, this animal is also described as being "blind" in the figure legend."	Target	[1]	Not present in the published version but acknowledged in authors' response to comment: "Results in Figure 5b are from a blind rat and they only had ultrasound stimulation. We have corrected the figure legend of Fig. 5b by changing 'light stimulation' to 'ultrasound stimulation' in the revised manuscript"	Yes	A_T	None
"Figure 5C. Please clarify in the figure legend the day corresponding to the data that is shown?"	Target	[1]	Day is present in the published version but absence acknowledged in authors' response to comment: "We have added in the figure legend of Fig. 5c: 'c, Lick response heatmaps recorded in Day 8."	Yes	A_T	None
"Figure 5D. Earlier in the figure, different sessions across several days are shown. To which sessions or days do these data refer?"	Target	[1]	Published version of 5D makes this clear. Ambiguity acknowledged in authors' response to comment: "The x-axis label of Figure 5d has been corrected from 'session' to 'Day'."	Yes	A_T	None
"It is not always clear how many animals were studied for each condition for each set of studies."	Target	[1]	Published version still does not list number of animals separately for every result. Ambiguity acknowledged in authors' response to comment: "We have clarified this information in the Method – Animals of the revised manuscript: [...]"	Yes	A_T	None
"Supplemental Figure 13 shows changes in gene expression for several proteins. Could these changes be indicative of pathology or a response to a noxious stimulus? What is the significance of the changes in Itgb3, for example?"	Target	[1]	Due to addition of more figures, this is now Supplementary Figure 21. It shows results of a gene enrichment analysis. S13a lists 8 differentially expressed clusters associated with either intracellular signalling or mechanical stimulation. S13b and c depict Itgb3 as by far the most differentially expressed gene in cluster GO:0071260 "cellular response to mechanical stimulus". The text does not compare these results to responses to noxious stimuli or explain the role of Itgb3.	Yes	A_T	None

REFERENCES

[1] 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, May 2024.

[2] T. Jefferson, E. Wager, and F. Davidoff, "Measuring the quality of editorial peer review," *Jama*, vol. 287, no. 21, pp. 2786–2790, 2002.

[3] R. Bruce, A. Chauvin, L. Trinquart, P. Ravaud, and I. Boutron, "Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis," *BMC medicine*, vol. 14, pp. 1–16, 2016.

[4] 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.

[5] 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.

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

[7] A. Craig, A. Athreya, and C. Taswell, "Example evaluations of plagiarism cases using fair metrics and the pdp-dream ontology," *IEEE eScience 2023*, 2023.

[8] A. Craig, A. Athreya, and C. Taswell, "Managing lexical-semantic hybrid records of fair metrics analyses with the npds cyberinfrastructure," *Brainiacs Journal*, vol. 4, 2023.

[9] T. Foltỳnek, N. Meuschke, and B. Gipp, "Academic plagiarism detection: a systematic literature review," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, pp. 1–42, 2019.

[10] F. Khaled and M. S. H. Al-Tamimi, "Plagiarism detection methods and tools: An overview," *Iraqi Journal of Science*, pp. 2771–2783, 2021.

[11] 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.

[12] C. Taswell, "DOORS to the semantic web and grid with a PORTAL for biomedical computing," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, pp. 191–204, 3 2007. In the Special Section on Bio-Grid published online 3 Aug. 2007.

[13] S. Arrish, F. N. Afif, A. Maidorawa, and N. Salim, "Shape-based plagiarism detection for flowchart figures in texts," *arXiv preprint arXiv:1403.2871*, 2014.

[14] N. Meuschke, C. Gondek, D. Seebacher, C. Breitinger, D. Keim, and B. Gipp, "An adaptive image-based plagiarism detection approach," in *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, pp. 131–140, 2018.

[15] S. Butakov and V. Scherbinin, "The toolbox for local and global plagiarism detection," *Computers & Education*, vol. 52, no. 4, pp. 781–788, 2009.

[16] T. Ghosal, S. Kumar, P. K. Bharti, and A. Ekbal, "Peer review analyze: A novel benchmark resource for computational analysis of peer reviews," *Plos one*, vol. 17, no. 1, p. e0259238, 2022.

[17] A. Craig and C. Taswell, "Fair metrics for motivating excellence in peer review," in *2024 IEEE 20th International Conference on e-Science (e-Science)*, pp. 1–2, IEEE, 2024.

[18] T. Rattanasawad, K. R. Saikaew, M. Buranarach, and T. Supnithi, "A review and comparison of rule languages and rule-based inference engines for the semantic web," in *2013 International Computer Science and Engineering Conference (ICSEC)*, pp. 1–6, IEEE, 2013.

[19] J. L. Martinez-Rodriguez, A. Hogan, and I. Lopez-Arevalo, "Information extraction meets the semantic web: a survey," *Semantic Web*, vol. 11, no. 2, pp. 255–335, 2020.

[20] G. Paliouras, C. D. Spyropoulos, and G. Tsatsaronis, *Knowledge-driven multimedia information extraction and ontology evolution: bridging the semantic gap*, vol. 6050. Springer, 2011.

[21] G. Martens, *Extraction and representation of semantic information in digital media*. Ghent University, 2011.

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

[23] A. Craig, C. Lee, N. Bala, and C. Taswell, "Motivating and maintaining ethics, equity, effectiveness, efficiency, and expertise in peer review," *Brainiacs Journal*, vol. 3, no. 1, 2022.

[24] 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.

[25] S. K. Taswell, A. Anand, M. Montes-Soza, and C. Taswell, "Babblenewt: A simplified, consistent, and interoperable reference citation format for bibliographic metadata," in *2023 Guardians Workshop (Guardians)*, pp. 1–4, IEEE, 2023.

[26] A. Craig, S. H. Bae, T. Veeramacheneni, S. K. Taswell, and C. Taswell, "Web service APIs for Scribe registrars, Nexus diristries, PORTAL registries and DOORS directories in the NPD system," in *Proceedings 9th International SWAT4LS Conference*, 2016.

[27] S. K. Taswell, K. Uhegbu, S. Mashkoor, S. Dutta, and C. Taswell, "Storing bibliographic data in multiple formats with the NPDS cyber-infrastructure," *Proceedings of the Association for Information Science and Technology*, vol. 57, 10 2020.

[28] A. Craig and C. Taswell, "Managing bibliometrics with contributor roles and literature provenance for npds metadata records," *Brainiacs Journal*, 2023.

[29] D. Beckett and B. McBride, "RDF/XML syntax specification (revised)," *W3C recommendation*, vol. 10, no. 2.3, 2004.

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

[31] M. T. Biagetti, "A comparative analysis and evaluation of bibliographic ontologies," in *Challenges and Opportunities for Knowledge Organization in the Digital Age*, pp. 499–510, Ergon-Verlag, 2018.