



# FOSE: a framework for open science evaluation

Alexander Walther<sup>1\*</sup> and Jasper J. F. van den Bosch<sup>2</sup>

<sup>1</sup> Cognition and Brain Sciences Unit, Medical Research Council, Cambridge, UK

<sup>2</sup> Institute of Medical Psychology, Goethe-University Frankfurt, Frankfurt am Main, Germany

## Edited by:

Nikolaus Kriegeskorte, Medical Research Council Cognition and Brain Sciences Unit, UK

## Reviewed by:

Satrajit S. Ghosh, Massachusetts Institute of Technology, USA  
Jason Priem, University of North Carolina at Chapel Hill, USA

## \*Correspondence:

Alexander Walther, Cognition and Brain Sciences Unit, Medical Research Council, 15 Chaucer Road, Cambridge CB2 7EF, UK.  
e-mail: alexander.walther@mrc-cbu.cam.ac.uk

Pre-publication peer review of scientific literature in its present state suffers from a lack of evaluation validity and transparency to the community. Inspired by social networks, we propose a framework for the open exchange of post-publication evaluation to complement the current system. We first formulate a number of necessary conditions that should be met by any design dedicated to perform open scientific evaluation. To introduce our framework, we provide a basic data standard and communication protocol. We argue for the superiority of a provider-independent framework, over a few isolated implementations, which allows the collection and analysis of open evaluation content across a wide range of diverse providers like scientific journals, research institutions, social networks, publishers websites, and more. Furthermore, we describe how its technical implementation can be achieved by using existing web standards and technology. Finally, we illustrate this with a set of examples and discuss further potential.

**Keywords:** open evaluation, peer review, social networking, standard

## INTRODUCTION

The success of scientific ideas critically depends on their successful publication. An unpublished idea, innovative, and promising as it might be, remains just that; only after publication it becomes a legitimate part of the scientific consciousness. A central gate-keeper function between the multiplicity of ideas and their manifestation as scientific publications is assigned to formal reviews governed by scientific journals. The current publishing system hinges on voluntary pre-publication peer review, with reviewers selected by the editorial staff. Peer review is undeniably a vital means of research evaluation for it is based on mutual exchange of expertise. Its role in the current system, however, has been the subject of concern with regard to accuracy, fairness, efficiency, and the ability to assess the long-term impact of a publication for the scientific community (Casati et al., 2010). For instance, studies suggest that peer review does not significantly improve manuscript quality (Goodman et al., 1994; Godlee et al., 1998) and that it is susceptible to biases to affiliation (Peters and Ceci, 1982) and gender (Wenneras and Wold, 1997). These concerns seem to be partly caused by the fact that the reviewer selection only includes a small sample from all peers potentially available. Aggravating this situation, no common agreements exist to provide reviewers with uniform guidelines, let alone binding rules, and no established standards by which those rules can be designed—peer review is in fact largely conducted at the discretion of the reviewers themselves. Given that reviewers vary considerably with respect to assessment and strictness, manuscript evaluation in the present model is highly dependent on reviewer selection. The lack of validity is further compounded by review and reviewer confidentiality, rendering them elusive to follow-up inspection.

Having been published, a scientific paper is exposed to interested scholars and hence goes through an ongoing process of *open evaluation*. When compared to journal-guided procedures, post-publication peer review is more suitable for evaluating

research impact, as scientists constantly need to consider which work they choose to accept, refute or expand upon. Over time, publications are thereby empirically detached from affixed quality labels like journal impact (but high-impact publications remain predominantly requested when it comes to promotions and grant applications, as journal publishing has traditionally been the main means of disseminating scientific knowledge). Although part of every individual scientist's everyday work, this communal effort has so far failed to develop into a cohesive framework within which research evaluation can be managed systematically and efficiently. A first step towards challenging this state was made feasible through the technological advancements of the web 2.0, constituting a change toward more openness between both researchers themselves, and researchers and public. This has manifested in the establishment of online open access formats and data repositories, and the growing recognition of scientific blogs and social networks for massive-scale scholarly exchange (e.g., <http://thirdreviewer.com>, <http://peerevaluation.org>). Such examples demonstrate the potential of exploiting the broad communication resources and simple usability of web-based technology by translating it into scientific practice. Smith (1999, 2003) reviews the current state of net-based publishing, concluding that all the activities of traditional journal publishing could be carried out collaboratively by existing web services. In a similar vein, overlay journals utilize the web to compile distributed information about one particular topic (Enger, 2005; Harnad, 2006). These studies indicate the high potential of distributed networking for a framework of open evaluation.

The principle of exchanging evaluation content through a data format and protocol has been put forward by Rodriguez et al. (2006). However, important elements of a framework, such as topic ("subject domain") attribution, review evaluation, reviewer selection, and other aspects central to the evaluation process, are in that system based on recursively data-mining the references of a paper. Even more so, single review elements are not evaluated on

their own worth, instead they are weighted by the reviewers life-time “influence”. Riggs and Wilensky (2001) come one step closer in that their rating of reviewers is based on the agreement with other reviews, yet they also do not differ between single reviews by the same reviewer. We deem these aspects a part of the evaluation, and think they should therefore be done by peers, case-by-case.

In the following article, we suggest utilizing the advantages of web-based communication in order to implement a framework of post-publication peer review. First, we outline its requirements, standard and protocol, serving to unify services dedicated to the evaluation of published research, and pinpoint its potential to help overcome shortcomings of the current reviewing system. We particularly emphasize the importance of provider-independence, with potentially infinite implementations. Finally, we illustrate our approach with a minimal working example and discuss it further.

## REQUIREMENTS FOR A NETWORK DEDICATED TO OPEN EVALUATION

Based on the weaknesses of pre-publication assessment, we formulate six criteria, which any net-based design aiming to attain large-scale open evaluation should be bound to fulfill in order to usefully complement the current state:

### ACCESSIBILITY FOR ALL AND EACH REGISTERED USER IS ENTITLED TO REVIEW

Open evaluation content should be open to everyone with an interest in science. Just as with pre-publication review, it builds on the expertise of peers; however, peer review does not need to be limited by external reviewer selection. Quite to the contrary, a network of open evaluation should recognize every user as a potential reviewer in order to most effectively serve to amass criticism. This should include scholars from topic-related and -unrelated fields as well as the educated public.

### EACH PUBLICATION CAN BE SUBJECT TO AN INFINITE NUMBER OF REVIEWS

A single review only represents a single opinion. At best, it was carried out objectively, identifies all flaws in a manuscript and contains helpful suggestions for improvement; even in this ideal case, standards between reviewers vary. At worst, a reviewer conducts reviewing according to career interests. The continuum between these two extremes is vast and impossible to ignore. However, objectivity can arguably be enhanced by incorporating many opinions. Hence, the number of reviews pertaining to a given publication needs to be unrestrained. If a large number of reviews conform to a particular opinion, it is likely that their assessment deserves notice. Even more importantly, if the dissimilarity between reviews is high, this indicates the need for further feedback from competent peers. Separate reviews can finally be consolidated into one complete assessment whose outcome reliably approximates the actual value of a publication.

### EACH REVIEW NEEDS TO BE DISCLOSED

The value of a scientific study depends on its recognition by other scientists. Careful feedback from the community is indispensably valuable for both the executing scientist and the recipient, as they

help to put results in perspective and can motivate adjustments or new research. In a network of open science evaluation, a review should be understood as just another type of publication directed to a topic-interested audience, including the author. Therefore, its disclosure is necessary in order to gauge its reception among other users, which will determine the overall quality. Since that way it is more likely to be scrutinized, it also serves as an incentive for thorough reviewing.

### REVIEWS NEED TO BE ASSESSABLE

Each review has to be considered potentially subjective, incomplete or faulty; consequently, it needs to be the subject of evaluation, just as with scientific publications. Here, a review of a review is termed *meta review*. Meta reviews can be either quantitative or qualitative (see “Standard”). Thereby, existing reviews can be rated and sorted by their overall reception. It further reduces repetitiveness and prevents trolling (i.e., posting off-topic comments). As per definition, any given meta review can again be target of another meta review, and so on.

Note that our design advocates information *gathering* rather than *re-computation*: after publication, an article has usually already received some level of attention; evaluation is carried out by individuals and journal clubs, lab meetings and other events devoted to research. Existing assessments thus often only need to be collected, and can then be analyzed and shared.

### REVIEWER EXPERTISE NEEDS TO BE DIFFERENTIATED THROUGH COMMUNITY JUDGMENT

For each user, expertise is bound to specific entities, such as a scientific method or theory, and among those, pronounced to varying degrees. User accounts thus need to feature a discernible expertise profile. An expertise profile should reflect scientific topics that were addressed by the user in submitted reviews and own publications. In turn, the attribution of these topics should be performed by peers.

### MANDATORY USER AUTHENTICATION

To ascertain the level of participation, account should be taken of the user’s authentication. Reading and submission of reviews should be enabled upon registration, where the user authenticates themselves with an unidentifiable credential such as a valid email address. Additionally, users may find it worthwhile to indicate their academic status with a validated academic email address, branding the account as “validated scientist” which may initially increase their perceived trustworthiness. Note that in general, authentication does not imply general disclosure of identity. In fact, authentication is necessary to unambiguously attribute the content to its real author, regardless of the level of anonymity. Nonetheless, it is conceivable that a user might prefer to disclose his identity (McNutt et al., 1990; Justice et al., 1998; Godlee, 2002; Bachmann, 2011), as it may add further credibility and acknowledgment to their effort. In order to illustrate and stimulate this initiative we formulate the following tentative initial features.

## EXCHANGING OPEN EVALUATION THROUGH AN IMPLEMENTATION-INDEPENDENT FRAMEWORK

The above described requirements initially invite the idea of an implementation as one platform. However, when comparing to

the current situation of both pre- and post-publication evaluation and distribution, such a platform seems economically unviable. Multiple institutions and companies compete for a role in these processes, so that their united participation is unlikely. In addition, dependence on one such system might be incongruous with the scientific principle of independent research. Therefore, we suggest that the above described structure should be implemented through a framework of: (1) a standard for the structure of the evaluation data and (2) a protocol for their communication, tentatively called *Framework for Open Science Evaluation* or FOSE (**Figure 1**). Both protocol and standard should be in the public domain. Ideally, such a framework should in time be agreed upon by all relevant parties. These include essentially the same organizations that may implement FOSE, such as academic institutions, publishers, funding agencies, scientist interest groups, etc. Supporting resources, such as open source software libraries, that implement the framework with an *Application Programming Interface (API)* and descriptive documentation could further promote the usage. This approach then supports the development of concrete platforms that make use of this framework, enabling them to share and integrate the evaluation content.

### RESPONSIBILITY

To develop and maintain this framework, an organization should be in place. This organization could be modeled after the successful W3C (<http://www.w3.org/>) organization which is responsible for the arguably daunting task of agreeing upon standards and protocols for the World Wide Web. Representatives of groups and organizations with an interest should be invited to participate, as their use and compliance is important to the success of this approach. Such members could include publishers, major research institutions, funding agencies, and the scientific community in general.

### STANDARD

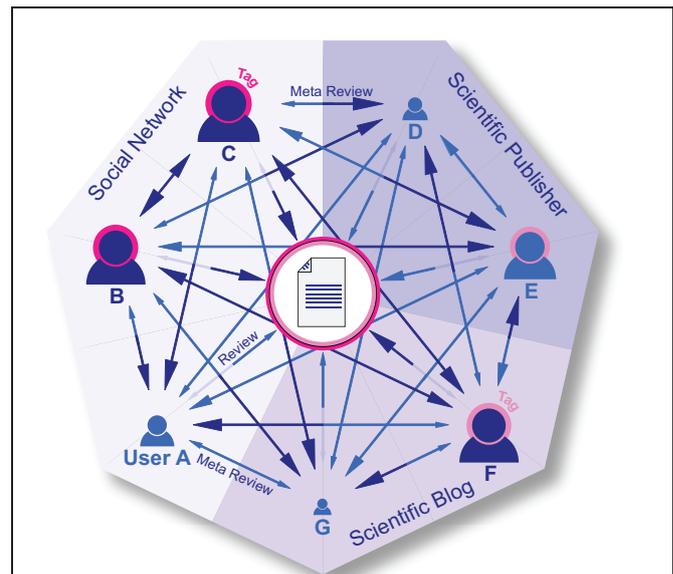
The framework structures the review process with *evaluation elements*. These predefined contributions can themselves contain standardized attributes.

### Qualitative review

In the context of the proposed design, we define a qualitative review as any textual feedback of undefined format (length, level of detail). The content need not necessarily be of the appraising kind—questions with the goal of clarification and journal-club-style summaries are just some of the alternative content types that are at place here. Potential attributes may include a *creation timestamp* indicating the date and time it was submitted.

### Quantitative review/rating

Content should also be evaluable with ordinal ratings. This way it can be quantitatively compared and sorted. Potential attributes include *contextualizing phrase*, which is a “category of feedback” indicating what aspect of the content the rating applies to, or *scale type*, denoting the range and order of the used scale as compared to a general reference scale.



**FIGURE 1 | Open evaluation can be organized through a uniform framework of open scientific social networking.** The figure explains the concept of an implementation-independent framework in the context of open evaluation. Importantly, reviewers can fulfill their roles via independent providers that all conform to the standard of the framework; an implementation as a single social network website is rendered unnecessary herewith. Three sectors of the pie chart correspond to three platforms implementing FOSE: a social network (either specifically academic or for other purposes), a scientific publisher, and a scientific blog. User icons labelled A to G represent participating reviewers. Arrows indicate their interactions with an original publication placed at the center (that is, submission of a *review*) and each other's evaluative content (submission of a *meta review*). Arrows coming from the center (semi-transparent) symbolize responses from the publication's author, who can participate with evaluation content. Light and dark pink halos represent thematic tags that have previously been assigned to both the publication and some users, namely reviewer B and C (dark pink) as well as E and F (light pink). Note that for simplicity, both tags have an equally high load. Icon size encodes authority (that is, a continuous variable indicating how proficiently a user has reviewed in the past, as seen by the community): larger size indicates higher authority, here exemplarily for a) tags shared with the publication in question (shown for users B, C, E and F), or b) other, publication-irrelevant tags (users A, D, G; not shown); it is important to stress that with respect to a given publication, reviews by users with either non-matching or no tags receive equal weights, irrelevant of their level of authority for other tags. The authority level for tags controls the weight and visibility of a user's review; review impact is here reflected by the size of the arrow heads, with a larger arrow head carrying more weight. Users B, C, and F already reached a critical authority threshold and were awarded an expert badge for their particular tag (see “Determining expertise and content classification” in Discussion); their nodes and arrows are therefore colored in dark blue. Since the publication has the same tags, their reviews have higher impact than those by other users. Note that E has indeed a tag but still lacks the necessary amount of positive ratings to appear as an expert. A, D, and G do not share any tags with the publication: their arrow heads are thus equally sized, indicating that they are not considered proficient in that scientific field.

### Tag

A tag element attributes a certain topic to a target, that is, to a publication or to other evaluation content. Tags could either be retrieved from publications or proposed by a reviewer. A key

attribute is *tag load*, a continuous variable reflecting the number of reviewers that agree that the target covers a given topic. A tag can also be the target of a review element.

One critical attribute present in all elements is the *user identity* that is the unique and abstract reference to the user who has authored the content. Another important attribute is the *target*, referring to the content that is evaluated with a given element. In case of a regular review it would refer to a publication, whereas meta reviews target existing evaluation elements.

## PROTOCOL

In order to link elements of the evaluation such as those described above, the protocol must be able to unambiguously refer to publications. More importantly, the same should apply to connecting the elements among each other. This referring can be done through identifiers that are globally unique. Moreover, such elements pertaining to a limited set, for example all those applying to a certain publication and its evaluation elements, should be available for discovery. This requires a fixed address structure. Additional rules should apply to the referencing to the author of these elements. For anonymous elements, the reference should allow getting enough information to gauge authority (see “Determining Expertise and Content Classification” in the Discussion). For named elements, the reference should ideally be human-readable. This ensures ease of carrying over or referencing to their own identity by users.

Formula or rules could dictate how to gauge meta-level measures such as authority and impact, preferable on a content-dependent basis, as categorized by tags. As an example, the average user rating could be weighted amongst other indicators of quality (e.g., number of citations, the reviewee’s ratio of highly ranked publications, and other estimates) and appear in global score rankings. If the network is supervised by the community alone, review rating also counteracts malpractice. The protocol should further allow some customization by its implementing platforms, for example through additional extra-standard elements or attributes thereof.

## TECHNICAL IMPLEMENTATION

Earlier work by Rodriguez et al. (2006) proposed to integrate the evaluation content with the existing OAI-PMH framework for exchange of publication metadata. However, as Smith (1999, 2003) suggests, these are separate parts of the scientific process and, therefore, need not be served by the same system; indeed, this may be deleterious to their independent development. As with the current publisher-organized system for reviewing, OAI-PMH has centralized elements, and they propose to use one authoritative provider (Rodriguez et al., 2006, line 155, “The [...] pre-prints.”) Contrarily, in FOSE, there is no inherent difference between a provider and a consumer. However, for harvesting publications related to evaluative content, OAI-PMH would be a prime candidate. These requirements can, however, be partially implemented by the use of existing standards or technologies.

## IDENTIFIERS

A standard scheme for providing evaluation elements with an address could be based on the representational state transfer

### Example 1 | Localizing

A specific review hosted by provider Smith and Jones:

```
smith-and-jones.com/fose/g3H2Ah4j
```

All reviews of the paper with doi 888.444 hosted by provider Eval.net:

```
eval.net/fose/reviews/by-doi/888.444/
```

(REST) resource identifier (Fielding, 2000). That is, the address to the evaluation element (the resource) could be a unified resource locator (URL) formed from the provider’s address and several pre-defined hierarchical elements. Discovery would use the same scheme (see **Example 1**). Publications could be identified by means of the widely used Digital Object Identifier (DOI, e.g., Rosenblatt, 1997), and implementations could likely benefit from using a service such as OAI-PMH for the discovery of publications to evaluate.

In one approach, framework elements themselves can be created with a locally (at the host provider) unique identifier, such as a simple integer number key, or automatically generated random character sequence (e.g., UUID5, <http://tools.ietf.org/html/rfc4122>). They can then be referred to externally (at another provider or in another element) through the standardized address scheme. This has some drawbacks, most importantly, such a scheme is likely subject to “link rot,” if the implementation ceases to exist. The alternative, however, of centrally-registered links, such as the DOI, comes with a dependence on the registration agency, and in the case of a commercial agency, such as crossref, a financial cost (this would weigh in heavily when used for each review element), and does not support collections or discovery. These drawbacks seem to go against the idea of distributed responsibility, which is central to FOSE.

## DATA FORMAT

The documents can be encoded with XML (<http://www.w3.org/XML/>; see **Example 2**) and their format, as defined in the section

### Example 2 | Encoding

XML document excerpt using a FOSE namespace.

Qualitative review of publication with DOI 888.444:

```
<fose:qualitative-review id='smith-and-jones.com/fose/g3H2Ah4j' created='03/05/2011' target-doi='888.444' author='smith-and-jones.com/afarnsworth'>
```

Interesting manuscript. Please use gender-neutral pronouns.

```
</fose:review>
```

Quantitative meta review of the above review:

```
<fose:quantitative-review created='14/05/2011' target='smith-and-jones.com/fose/g3H2Ah4j' author='eval.net/users/pbishop' scale=num>
```

```
8
```

```
</fose:review>
```

“Standard”, can be published and validated through the use of XML Schema (<http://www.w3.org/XML/Schema>).

Where (a) there is the need for the evaluation content to be machine readable (i.e., not just transferred, but “understood”), and (b) the structure of that content is complex enough not to be expressible in vanilla XML, it might be of advantage to publish that content in RDF. A strict XML scheme will allow RDF conversion from the XML. Further developments of standardization of data formats describing scientific knowledge (for instance along the line of nano-publications, Mons and Velterop, 2009) will naturally increase machine readability.

### A MINIMAL WORKING EXAMPLE

For the sake of simplicity, we illustrate the basic concept with an example of minimal complexity, involving one author (W. Bell), two scientist reviewers (O. Dunham and P. Bishop), and three independent services implementing FOSE (*The Journal of Foomatics*, the Medical Research Institution (NIX) employee site, *Eval.net*; **Figure 2**).

#### THE PUBLISHING AUTHOR

W. Bell’s research article “Effects of Cortexiphan on Inter-dimensional Travel” has successfully passed the pre-publication review and is now published in *The Journal of Foomatics*. A week later, Bell finds a notification of a new review in his inbox. This service is offered by *The Journal of Foomatics*, whose implementation of FOSE allows them to track reviews of their publications. In response to the review, he uses the *Journal of Foomatics* website

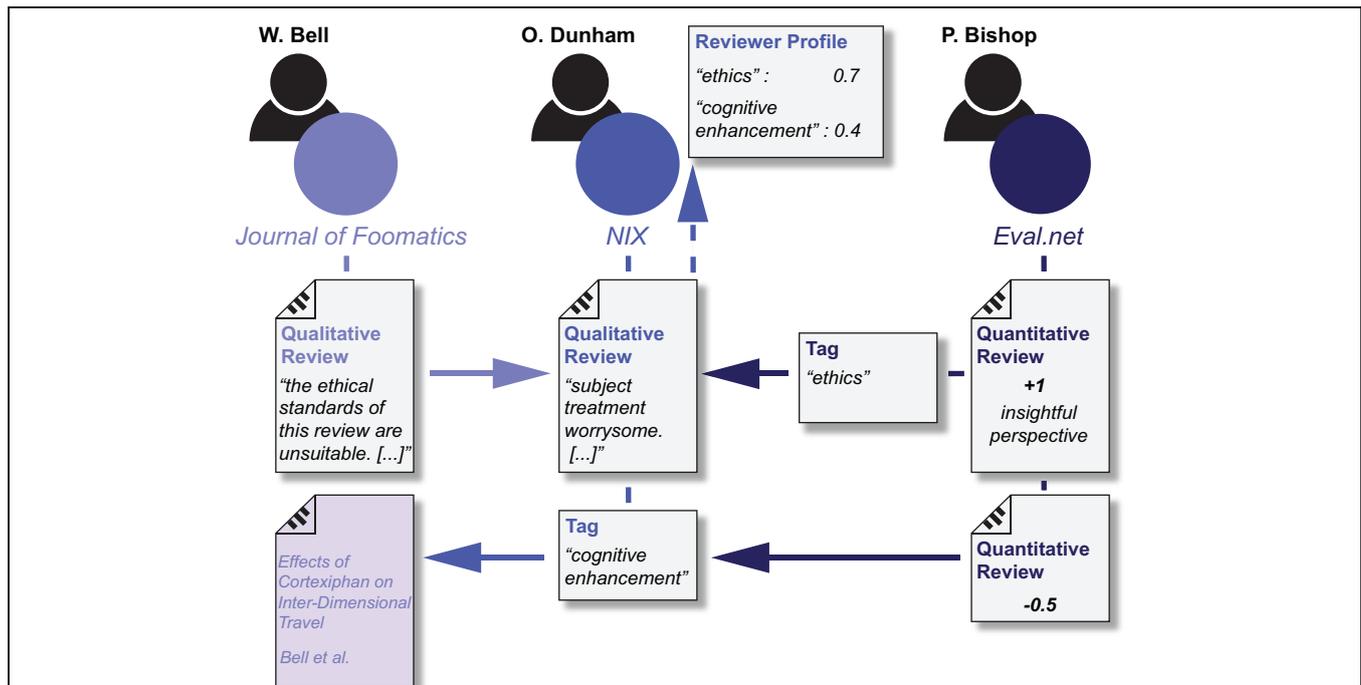
to comment on the referenced, unreasonably harsh, ethical standards. He further notices that another reviewer has supported the review’s assessment by giving it a positive rating.

#### THE REVIEWER

O. Dunham, a scientist working for the NIX, comes across Bell’s article during her literature research and, after reading, decides to publish some critical remarks about its ethical standards. She logs in to the NIX employee website and selects the tab “Review” where she finds textual and numerical rating elements; these features accord to the FOSE standard. Dunham submits a brief textual review on the ethical aspects of the subject’s involvement in the study. Moreover, she gives Bell’s publication the tag “cognitive enhancement” which starts out with an initial value of +1. Some days later, Dunham logs in to the NIX website and finds that a reviewer recently updated her review on Bell’s article and assigned a tag “ethics”. Thereby, the authority rating of her reviewer profile (see “Assessing Scientific Impact” in Discussion) in “ethics” increases, and less so for “cognitive enhancement”, as this tag was downrated by Bishop. Moreover, this renders all of her reviews in those fields more visible to the community.

#### THE META REVIEWER

A PhD student named P. Bishop logs in to a community website called *Eval.net*, which has been designed by a non-profit organization with the aim of facilitating the scientific discourse. Again, *Eval.net* subscribes to FOSE and has downloaded



**FIGURE 2 | A minimal working example of open evaluation in FOSE.**

Depiction of example interactions of evaluation contributors concerning one publication, with an author (W. Bell), a reviewer (O. Dunham), and a meta

reviewer (P. Bishop). All are registered users of independent providers.

Arrows indicate targets of the evaluation content pieces they contribute. See “A Minimal Working Example” for complete narrative.

Dunham's review from *NIX*. Being interested in related topics, Bishop comes across Bell's article and Dunham's review. After having read both, he thinks that Dunham's review is valuable and rates it +1, *insightful perspective*. He, however, disagrees with the tag "*cognitive enhancement*" and gives it a quantitative review -0.5. He further assigns Dunham's review the tag "*ethics*."

## DISCUSSION

### ONE FRAMEWORK FOR EXCHANGE OF RESEARCH EVALUATION

Open evaluation, by its very nature, is a diverse approach: it aims at sourcing article evaluations in large quantities in order to approximate the value of a publication. In the web 2.0, it appears counter-intuitive to bind this process to a single community website. In fact, users should be free to choose from a range of independent providers with services tailored to specific interests and needs (users differ in interests and thus frequent different websites). For instance, a university's personnel platform (See *NIX website* in *A minimal working example*) may offer single sign-on for their employees, or a publisher has immediate access to publications. We believe that provider-variety significantly increases the overall participation of scientists and non-scientists in peer review. However, the main holdback for current platforms might be their closed, egocentric approach, which due to commercial motivations will not be readily accepted by other, influential contenders, thereby scattering the content. Instead of competing with other platforms, a new approach should promote interoperability. As a consequence, they must subscribe to one established norm in order to integrate evaluation content between them. The attractiveness of the framework for potential implementers should then be access to other, existing content at other providers. This integration ensures that all evaluation content is accessible everywhere, enhancing its traceability and comparability (across borders of papers, publishers, providers).

### MASSIVE AND IMMEDIATE FEEDBACK TO THE AUTHOR, THE SCIENTIFIC COMMUNITY, AND THE JOURNALS

Two shortcomings of the journal-based system have been widely mentioned: first, it obscures the discussion between authors and reviewers; second, there are few possibilities to comment on the result (Wicherts et al., 2012). Open evaluation can attenuate these weaknesses by administering feedback from the community to authors, reviewers, and publishers. We think that the proposed framework can help organize this process in a principled way. Feedback can be submitted online on institutional websites and platforms of third-party suppliers, and then collectively analyzed. *Ad-hoc* networking will enable users—authors and reviewers, scientists and the educated public—to engage in discussions among themselves. In combination, these features craft a highly transparent and dynamic alternative to established means of article evaluation, such as response letters and scientific meetings, and have several advantages over them: first, authors will receive unfiltered criticism by the scientific community in a quantitative and qualitative manner. The higher the participation, the more meaningfully does the evaluation in sum approximate the benefit for the community. To reflect that, numerical ratings could be related to the number of submitted reviews. However,

even few reviews are likely to contain valuable feedback, considered that they come from a vast pool of potential reviewers. Second, any feedback given is instantly visible to the scientific community and can thus be challenged and questioned. As a consequence, reviews can be commented and rated in turn. Third, the network generally encourages discussions about scientific publications. An additional advantage of post-publication review is feedback for journals. The ratings obtained in open evaluation can be compared to the editor's assessment used for the publication decision. If there is a large discrepancy, the journal could change its assessment policy or reviewer selection and instruction.

### DETERMINING EXPERTISE AND CONTENT CLASSIFICATION

FOSE sources reviews as globally as possible, setting few constraints on general participation. This raises the issue of trustworthiness, as reviewers clearly differ in their authority with respect to a particular scientific field. Drawing a bold line between affiliated and unaffiliated scholars (as by email authentication only) is insufficient to resolve this and would allay criticism from the latter. In fact, it is equally necessary to distinguish between experts and laymen within a given scientific field. Therefore, in a framework with no prior reviewer selection, expertise is only determinable *post-hoc*. Natural sources for this assessment are reviews and meta reviews. Someone should arguably be considered proficient in a specific theory, method, research area, etc., when he or she has garnered a critical mass of positive evaluation on publications, i.e., articles and reviews. By reaching a certain threshold, a user could be awarded an "expert" badge. This would provide a communally determined credential distinguishing proficient users from others. Consequently, their reviews should carry greater weight and be most visible; they could be branded *expert review* and analyzed both separately and jointly with non-expert reviews, with scores displayed in the user statistics. Complementary to that binary classification, expertise should also be recognized as a continuum. The level at which any user is authorized to contribute could generally refer to a variable named *authority*: again, one's authority should depend on the average quality of one's reviews submitted, and reviews by users with lower *authority* should be less visible (**Figure 1**). One critical ingredient in this formula is content classification. To restrict *authority* to a given field, contributions must be classified as covering a particular theme. In the proposed framework, peer-based tagging fills this role. Thereby, the community can discuss their opinion on the attribution of these tags through quantitative and qualitative reviews. The use of tags leaves space for advanced indexing methods, such as hierarchical relations between tags, and is a further step in the direction of more advanced semantic markup, such as nano-publications (Mons and Velterop, 2009). One's individual reviewer profile could be determined by the union of all tags targeting the user's contributions, weighted by participation and content quality. This accrued information could also be used by automatic filters in order to suggest other publications for review. In a more general sense, by jointly crowd-sourcing scientific classification *and* evaluation, tags can be utilized to meaningfully and reliably index scientific literature.

## ASSESSING SCIENTIFIC IMPACT

An interoperable framework lends itself ideally to assessing the scientific impact of a publication. An article that is highly relevant to many scientists will attract more attention and will receive more and, on average, higher ratings than one being less so. Quantitative reviews in the form of numerical ratings can be summed up in statistics linked to user profiles. Statistics could feature different components, such as the average rating given to all publications, separate averaged ratings for reviews by scientists and the general public as well as their union, average meta review ratings, evaluation of replicability, etc. It will be one major assignment of the development of a universal standard to define meaningful scales for quantitative reviews. Contrary to pre-publication review, this will bring about an assessment model of scientific literature, in which evaluation is sought from *many* people and within a technically infinite period. Hence, it remains amendable over time: statistics of a publication can always be up- or downgraded by a new review. They are thereby more likely to reflect the actual scientific impact of a publication on the scientific endeavour. This approach makes a contribution toward counterbalancing the current focus of science on journal prestige, contrary to which it cooperatively approximates the value of a scientific publication based on actual relevance. It is desirable that these advantages are also accounted for in practical ramifications associated with publications. To that end, we believe that post-publication reviews can just as well serve as a valid reference in hiring processes and grant applications as publications in prestigious journals. As they provide an independent quality indicator for each publication, they should be referred to in an author's quantitative evaluation and used to put journal prestige in perspective (e.g., when an article in a low-impact journal receives a lot of attention or vice versa). Similarly, positively recognized reviewing can provide a reference on a user's scientific expertise even in the absence of own publications. The possibility to refer to one's user statistics as a meaningful reference in turn will incentivize participation in such a framework: influential research or expertise in a particular field is likely to be recognized by the community, even more so over time. One's reviewer profile could be added to one's CV in order to distinguish oneself. In the same vein, authors of scientific publications should be enabled to refer to their reception in order to add another plus to their resume. Moreover, as it reduces attention to journal impact,

it will adjust the allocation of scarce resources (i.e., positions, grants, etc.) on the basis of scientific soundness; hence an excellent article will be more likely to receive credit, regardless where it was published.

## OPENING THE SCIENTIFIC DEBATE

Scientific progress critically depends on the interaction among scientists. Traditionally, the dissemination of scientific content is achieved mainly by two means: scientific conferences, closed to the outside and limited to a usually pre-selected group of scientists; and publications, whose review process is only visible to a handful of people. Furthermore, only few scientific findings are chosen to be translated to the larger public (which at that point have been subjected to massive informational filtering and simplification). Hence, from society's perspective, the production of new knowledge can hardly be seen as participative or fully transparent. This status quo is inadequate, as scientific work arguably depends on society's endorsement, which requires mutual understanding, hence transparency. In that vein, the boundaries between science and the public domain have recently been blurred by the emergence of net-based communication (see "Introduction"). FOSE contributes to this development in that it utilizes social networking in order to help transcend the barrier between science and general public by open user-to-user propagation of scientific knowledge. Even more so, it is able to integrate criticism of scientific papers from *non-scientist* reviewers, yielding a more complete picture of research evaluation.

## TOWARD OPEN EVALUATION OF SCIENCE

The organization of research evaluation will always be competed for by a multitude of players; integrating their contributions into a cohesive framework promises the most efficient way to aggregate peer review, and ultimately to reliably reflect scientific impact. Accordingly, the FOSE way to open evaluation is *open*, that is by exchange between providers, and through *evaluation*, that is by having peers recursively evaluate content. These two principles rest on a standard for structuring this content and a protocol for its exchange.

## ACKNOWLEDGMENTS

Both authors would like to thank their host institutes for funding their work.

## REFERENCES

- Bachmann, T. (2011). Fair and open evaluation may call for temporarily hidden authorship, caution when counting the votes, and transparency of the full pre-publication procedure. *Front. Comput. Neurosci.* 5:61. doi: 10.3389/fncom.2011.00061
- Casati, F., Marchese, M., Mirylenka, K., and Ragone, A. (2010). Reviewing peer review: a quantitative analysis of peer review. Technical Report DISI-10-014, *Ingegneria e Scienza dell'Informazione*.
- Enger, M. (2005). *The Concept of 'Overlay' in Relation to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*. Master's thesis, Universitetet i Tromsø, Tromsø, Norway.
- Fielding, R. T. (2000). *Architectural Styles and the Design of Network-based Software Architectures*. Doctoral dissertation, University of California, Irvine.
- Godlee, F. (2002). Making reviewers visible. *Am. Med. Assoc.* 287, 2762–2765.
- Godlee, F., Gale, C. R., and Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports a randomized controlled trial. *Am. Med. Assoc.* 280, 237–240.
- Goodman, S. N., Berlin, J., Fletcher, S. W., and Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at annals of internal medicine. *Ann. Intern. Med.* 121, 11–21.
- Harnad, S. (2006). Research journals are already just quality controllers and certifiers: so what are "overlay journals"? Available online at: <http://goo.gl/rzunU>
- Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., and Rennie, D. (1998). The PEER investigators. Does masking author identity improve peer review quality: a randomized controlled trial. *Am. Med. Assoc.* 280, 240–242.
- McNutt, R. A., Evans, A. T., Fletcher, R. H., and Fletcher, S. W. (1990). The effects of blinding on the quality of peer review. A randomized trial. *Am. Med. Assoc.* 263, 1371–1376.

- Mons, B., and Velterop, J. (2009). "Nano-publication in the e-science," in *Proceeding of Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*, (Washington, DC).
- Peters, D. P., and Ceci, S. J. (1982). Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behav. Brain. Sci.* 5, 187–195.
- Riggs, T., and Wilensky, R. (2001). "An algorithm for automated rating of reviewers," in *Proceedings of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL'01*, (New York, NY, USA: ACM), 381–387.
- Rodriguez, M. A., Bollen, J., and van de Sompel, H. (2006). The convergence of digital libraries and the peer-review process. *J. Inf. Sci.* 32, 149–159.
- Rosenblatt, B. (1997). The digital object identifier: solving the dilemma of copyright protection online. *J. Electron. Publ.* 3, 2.
- Smith, J. W. T. (1999). The deconstructed journal—a new model for academic publishing. *Learn. Publ.* 12, 79–91.
- Smith, J. W. T. (2003). "The deconstructed journal revisited—a review of developments," in *From Information to Knowledge: Proceedings of the 7th ICC/IFIP International Conference on Electronic Publishing held at the Universidade do Minho*, eds S. M. de Souza Costa, J. A. A. Carvalho, A. A. Baptista, and A. C. S. Moreira, (Portugal: ELPUB).
- Wenneras, C., and Wold, A. (1997). Nepotism and sexism in peer-review. *Nature* 387, 341–343.
- Wicherts, J. M., Kievit, R. A., Bakker, M., and Borsboom, D. (2012). Letting the daylight in: reviewing the reviewers and other ways to maximize transparency in science. *Front. Comput. Neurosci.* 6:20. doi: 10.3389/fncom.2012.00020
- A. A. Baptista, and A. C. S. Moreira, (Portugal: ELPUB).
- that could be construed as a potential conflict of interest.

Received: 16 May 2011; accepted: 21 May 2012; published online: 27 June 2012.

Citation: Walther A and van den Bosch JFF (2012) FOSE: a framework for open science evaluation. *Front. Comput. Neurosci.* 6:32. doi: 10.3389/fncom.2012.00032

Copyright © 2012 Walther and van den Bosch. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships