

Title

Wikis in scholarly publishing

Authors

Daniel Mietchen[1], Gregor Hagedorn[2], Konrad U. Förstner[3][4], M Fabiana Kubke[5], Claudia Koltzenburg[6][7], Mark Hahnel[8] and Lyubomir Penev[9].

1. ↑ [Science 3.0](#)
2. ↑ [Julius Kühn-Institut, Königin-Luise-Straße 19, 14195 Berlin, Germany](#)
3. ↑ [Institute for Molecular Infection Biology, University of Würzburg, D-97080 Würzburg, Germany](#)
4. ↑ [Research Centre for Infectious Diseases, University of Würzburg, D-97080 Würzburg, Germany](#)
5. ↑ [School of Medical Sciences, University of Auckland, New Zealand](#)
6. ↑ [Cellular Therapy and Transplantation](#)
7. ↑ [Clinic for Stem Cell Transplantation, University Medical Center Hamburg-Eppendorf, Germany](#)
8. ↑ [Science 3.0](#)
9. ↑ [Pensoft Publishers, 13a Geo Milev Street, 1111 Sofia, Bulgaria](#)

Abstract

Scientific research is a process concerned with the creation, collective accumulation, contextualization, updating and maintenance of knowledge. Wikis provide an environment that allows to collectively accumulate, contextualize, update and maintain knowledge in a coherent and transparent fashion. Here, we examine the potential of wikis as platforms for scholarly publishing. In the hope to stimulate further discussion, the article itself was drafted on Species-ID - <http://species-id.net>; a wiki that hosts a prototype for wiki-based scholarly publishing - where it can be updated, expanded or otherwise improved.

Introduction

"Science is based on building on, reusing and openly criticising the published body of scientific knowledge." --Panton Principles [1]

At its inception, a research project is typically fueled by at least one of: new data, new methodological approaches, or new hypotheses. Each of these can contribute either to small incremental steps that further specific knowledge or have more profound effects on the theoretical framework of a field. The generation or interpretation of data is therefore constantly changing.

Despite the tools and methods made possible by the web, the scientific literature continues to consist mainly of journal articles, book chapters, monographs, preprints and conference proceedings, arranged in a semi-hierarchical structure that varies considerably across fields. While such persistent, reliably unchanging knowledge expressions are an essential ingredient of the process of science, the dynamic nature of science in general calls for dynamic publishing formats that can be updated in light of new research results.

A side effect of static publishing is the fragmentation of the scientific landscape, in itself a major access barrier to accessing scientific content on the web. Other important access barriers result from pay-walls erected around the documents containing research reports or the works cited therein. The considerable momentum reached by the Open Access movement - in both its Green and Gold shades - is likely to alleviate this problem significantly over the next few years. However, even if the whole scientific



literature were to become freely accessible, a number of traditions would continue to stand in the way of efficient scholarly communication.

The current publishing standard centered around the scientific 'paper' (with introduction, materials and methods, results and discussion sections) leads to a high level of redundancy between papers on related topics. A large portion of most publications (typically the introduction and parts of the discussion) serve to "recapitulate existing knowledge". These recapitulations tend to be very fragmentary within individual articles, yet highly repetitive when aggregated across articles. For instance, a Google Scholar search for the phrase "Magnetic Resonance Imaging is a"^[2] returned about 2500 results at the time of writing, and multiple other phrases are in use to explain the very same concept to similar audiences. It would it be less tiring for authors and reviewers - and more convenient to readers - to have fewer of these introductions, while at the same time have better collaborative introductions that are under ongoing revision. Review articles already provide these features, and if made available via Open Access, they could be efficiently exploited to avoid redundancies and to facilitate critical appraisal and replication instead.

The traditional scientific paper is designed to put together a series or collection of findings in order to tell a more complete story. This structure interferes with the ability to efficiently refer to (and thus, later, to retrieve) each of those individual findings.

For example, citations typically go to a whole publication, an entire monograph, a whole chapter or page, instead of to a more precise point of interest within the larger publication, that is actually the source for a particular claim. With the increase of multi-author publications, each contributing to small portions of a complete report, such a type of referencing results in inadequate attribution to the creator/s of the knowledge. In addition, the current citation system in which references are listed as endnotes or footnotes instead of being directly linked from the citing passages, becomes a burden to the reader and discourages the reading of the original literature. Microcitation specific to individual figures, data sets or paragraphs are generally discouraged by printed media on the grounds that they increase printing costs. The increased efficiency of linking to small grained information bits, however, has been proven on the web.

In the standard publishing environment, updating any information requires the generation of a new article. Given the growing number of research groups and the rapid production of new information, static articles can quickly become outdated. Review articles may already be outdated by the time they finally get published. In the case of print copies, updating information is rarely economically viable for academic publications other than text books. For text books, ironically however, the process of updating is especially lengthy, resulting in a long delay and often large portions of outdated materials. In an online publishing context, producing an update can be more easily implemented and such a step comes at a lower cost. This makes the web a medium that is more aligned with the dynamic requirements of scientific communication. Further, online publishing offers tools for version control and the ability to display, compare and cite different versions of a given document. Reuse-friendly open content licenses can be used to lower sharing barriers enabling knowledge producers and users to handle cross-document versioning in an efficient manner.

Good version control tools support (and are likely to promote) collaborative editing and updating. Collaboration on science reporting might also be promoted if new types of author attribution are put in place that link not only large chunks but also small-size contributions to individual authors. New findings can potentially be added as 'stand alone' sections within the existing dynamic entity, or optionally with links to other pages with more detailed information (such as electronic lab notebooks). While radically different from the way research reporting is being performed in most fields today, a collaborative and dynamic way to rapidly update existing knowledge is better suited to the needs and time scales of knowledge generation within a typical research project. An updated perspective on scientific output reporting schemes would mean that "writing" (and hence "publishing") should actually consist of collaborative updating.

"Although the potential of the Internet for improving the dissemination of information is now

taken for granted, it would seem that the attitudes of those involved in sharing this information have not kept pace with the technology." --The PLoS Medicine Editors (2010)[3]

Wikis as dynamic publishing environments

The issues outlined above have prompted the suggestion of a set of "Criteria for the information scientific journal of the future[4]". Therein, wikis are explicitly mentioned as "working platforms of the future". Consequently, it appears logical that the criteria are now collaboratively refined on a wiki environment[5].

Wikis behave differently than more conventional publishing platforms with respect to the following criteria: they are versioned, their contents is often available under reuse-friendly licenses, they allow recapitulation to be replaced by contextualization through hyperlinks to background articles, they support fine-grained referencing and attribution, and are well suited for transparent review and timely revisions. Moreover, they are "web-native" and thus fully compatible with semantic enhancements and the inclusion of multimedia.

With respect to dynamic publishing models, three approaches go in this direction:

1. Extending traditional publishing by providing the article in an editable format in addition to the standard non-editable ones,
2. Extending the traditional repository approach by supplying previously published Open Access articles with an editable version, and
3. Moving the entire publishing process to a wiki platform.

Virtually all scholarly articles published today are available in PDF format. In addition to that, HTML versions are available for most, and XML versions for some journals. Experiments with other formats - including wikis - continue (e. g., videos at the Journal of Visualized Experiments[6]).

The main purpose of providing a wiki publishing model is to allow published articles to become editable, so as to allow them to stay current and be expanded as knowledge evolves. Rendering a published article editable requires legal permissions for copying and modification. Under the prevailing copyright arrangements, that give publishers a long term (typically around 100 years) monopoly on the scientific articles, editable wiki versions of entire articles published by the majority of contemporary scholarly journals are prohibited by copyright law.

While editable versions of articles behind paywalls are certainly of limited use, making some aspects of the article available under a reuse-friendly license may well be worth considering as a dissemination strategy. Given that some methodologies and data cannot be copyrighted under most jurisdictions, those sections of the article could be released under more open licenses. A related approach is taken by the journal RNA Biology[7], which requires[8] article submissions regarding new RNA families to be accompanied by the draft of a corresponding Wikipedia entry (that would then cite the original paper). This new policy has been well received by the community, to the extent that a number of authors have since submitted drafts of Wikipedia articles on topics for which the journal does not require that[9].

Open Access journals have a wider range of options to consider when rendering their publications editable. They could, for instance, actively upload the images and other non-text media to collaborative media repositories like Wikimedia Commons, place the entire text or parts thereof - or a synopsis, if available - in a wiki environment. If they do not wish to engage in these activities by themselves, the open licenses enable others to do part or all of this instead. However, the handling by the journals itself can be more efficient and automated.

This system works best if the wiki environment does not yet have an entry on the topic in question, as is the case with novel RNA families in the example cited above. The description of taxonomically new species is another such situation, and we have now implemented a workflow that allows the descriptions

of new animal species published in the journal ZooKeys[10] to be uploaded onto Species ID[11], a wiki dedicated to taxonomy and host to the wiki version of the article you are reading.

For example, the wiki entry on the diving beetle *Neobidessodes darwiniensis* sp. n.[12] (Coleoptera, Dytiscidae, Bidessini) from northern Australia was set up the same day that the original description[13] had appeared in ZooKeys, and four wiki contributors have since edited it to improve and enhance its presentation. While this page was started manually, an automated workflow to import taxon treatments from future ZooKeys manuscripts into the wiki is currently being finalized (Penev et al., ZooKeys, upcoming).

Wiki entries can also be set up retroactively for suitably licensed articles already published in other formats. In taxonomy, this would be especially interesting for legacy taxon treatments from dozens or hundreds of years ago that are often barely accessible. Other scenarios are possible, for instance, an article on Thirdhand Smoke Exposure Concerns[14] in the journal Environmental Health Perspectives[15] has been adapted into a corresponding wiki entry at the Encyclopedia of Earth[16].

Technically, there are few limitations to importing scholarly contents into wiki repositories, so it could be envisaged that wiki versions might one day complement other forms of scholarly communication. Prototypes for such scenarios already exist - some science blogs[17] run on wiki software, and a number of electronic lab notebooks are being kept public by way of wiki platforms, e.g. at OpenWetWare[18] or Useful Chemistry[19].

With over 60 000 CC-BY-licensed articles currently available from PubMed Central in XML format, a systematic approach to importing such content into a wiki stands excellent chances of seeding a useful tool for literature searches, especially if combined with semantic enhancements (see below).

The examples mentioned above concerned the integration of wikis into publishing workflows. It is also possible to handle the entire journal workflow on a wiki, as exemplified by the review journal Scholarpedia[20]. Such an approach naturally fits with public peer review (i.e. the reviews are being posted in public), and the reviewers have (unless the authors exclude that) the possibility to edit the submitted text by themselves rather than writing a report about what should be changed. This adds a new quality to the peer review process. The idea of a wiki journal has also surfaced from time to time at Wikipedia and other wiki projects, but no decisive action has been taken yet.

As in the case of the Wiki Repository, there is no inherent reason to believe that such an approach would only work for formal publications (or review articles even) - this wiki-style peer review by direct editing could in principle be implemented for any step of the research cycle, and be it lab notebooks or grant proposals.

Social knowledge organisation on wikis mirrors to a large extent the processes of scientific publications. On platforms like Wikipedia, dedicated (voluntary) editors process the incoming materials, similar to how it is done in managed journals. Material that is undesirable or clearly below quality standards is removed often within minutes (Quick deletion workflow). Other substandard material is submitted to a one-week process of pointing out errors and requesting improvements and revisions (Requests for deletion workflow). To some extent, this process involves expert review, provided experts have organized themselves appropriately (e.g. "Redaktionen" on German Wikipedia - "Editorial teams" are under discussion[21] for the English version, too). While the expert review in Wikipedia may occasionally be substandard, the mass review is highly transparent, and all improvements resulting from it immediately enhance the quality of the product.

Crucial elements of site policies for wiki-based publishing are guidelines on which kinds of content should be uploaded or modified and by whom, who is admitted to the platform, what are the norms of interaction, and, for instance, to what extent automated approaches will be permitted. For instance, WikiGenes[22] monitors PubMed for new gene-related content and posts extracts thereof on the wiki, whereas Gene Wiki[23] harvests molecular biological databases to keep infoboxes in Wikipedia articles updated.

Despite their great potential to radically improve the searchability as well as the automated analysis and processing of the corpus of scientific literature, the uptake of semantic technologies in scholarly publishing is slow, and for the few journals that already do provide semantically tagged content, possible reuse scenarios have yet to be explored. Manual annotation is labour-intensive, and automated annotation not yet reliable outside well-circumscribed areas. Again, taxonomy is in a privileged position here, and both ZooKeys as well as Species ID can handle tags obeying the TaxPub[24] extension to the National Library of Medicine of the USA (NLM) Document Type Definitions (DTD)[25] standard.

Outlook

A system of the kind described above would gravitate towards a smaller number of more meaningful and useful resources on any given topic. As readers, scientists would be able to focus on a smaller number of publications, trusting that they contain the most up-to-date information, that efficiently link to the relevant data, processes, and references. As writers, scientists could choose to contribute to the collaborative efforts reducing the time spent on writing long repetitive manuscripts that distract them from the task of generating new knowledge.

One important incentive to write traditional publications is the reputation authors gain in the process. In the current system, the value of a publication is most often associated with the perceived value of the journal in which the work is published (impact factor, IF) or through the number of citations an author receives (H-factor, HF). But reputation resulting from both IF and HF fall equally on all authors of a given publication. Micro-contributions (for example, in the Gene Wiki project), when aggregated, can positively benefit the scientific community and they should too be incorporated into the reputation portfolio of scholars. ORCID[26] - a unique identifier for researchers which is under development with strong support from many scientific publishers - could eventually provide the architecture to support the accumulated measurement of a scientist's activity, taking into account both traditional publications as well as micro-contribution models.

Historically, one of the most valuable improvements provided by the traditional journal is the role played by pre-publication peer reviewers. But the peer-review system does not scale to the realities of the scientific flow of the 21st century and post-publication peer-review do not feed back into the system. Scientific reputation systems could make use of end-user rating systems (as part of their post-publication peer review) that are increasingly found in web media (e. g., Stack Overflow[27]) and that can easily reach across disciplines.

The reputation systems described above are of crucial importance, since they dominate the systems through which researchers are assessed at the time of applying for jobs or promotions, tenure, research funding and other important elements and milestones of the individual's academic career. Performance assessment systems could easily accommodate to these new formats, assessing data contributions (the currency of science) instead of 'papers' (the currency of scholarly publishers). This requires efficient interaction with data repositories (e. g. Dryad[28] or FigShare[29]).

Wiki environments offer the opportunity for commenting on and rating of the article content by means of associated "talk" or "discussion" pages. In contrast to the classical non-public anonymous pre-publication peer review currently in place at most journals, the public peer review process on wiki platforms allows reviews to be read by any interested party, and if reviewers choose to reveal their name, they can take credit for their contributions. A transition to a wiki style-based publication system is ideally aligned with the realities of scientific workflows, allowing individual contributions to transfer to scientific reputation and assessment system in a more suitable granular manner.

Conclusions

Wikis can be used for dynamic scholarly publishing in several ways, ranging from scenarios which

change existing publishing workflows in minor ways to a fundamental restructuring thereof. The wider uptake of such collaborative approaches hinges upon community acceptance and proper recognition of contributions to large-scale collaborative environments. **You can edit this document!**[\[30\]](#)

References

1. [↑ Panton Principles. Principles for Open Data in Science](#)
2. [↑ http://scholar.google.com/scholar?q=\"Magnetic+Resonance+Imaging+is+a\"](http://scholar.google.com/scholar?q=\)
3. [↑ The PLoS Medicine Editors \(2010\) Journals, Academics, and Pandemics. PLoS Med 7\(5\): e1000282. doi:10.1371/journal.pmed.1000282](#)
4. [↑ Criteria for the information scientific journal of the future](#)
5. [↑ Wikiversity:Journal of the future](#)
6. [↑ Journal of Visualized Experiments](#)
7. [↑ RNA Biology homepage](#)
8. [↑ RNA Biology Guidelines for Authors](#)
9. [↑ Plans for Rfam 2010-2011](#)
10. [↑ ZooKeys homepage](#)
11. [↑ Species ID homepage](#)
12. [↑ Neobidessodes darwiniensis entry at Species ID](#)
13. [↑ Neobidessodes darwiniensis entry at ZooKeys](#)
14. [↑ Burton A 2011. Does the Smoke Ever Really Clear? Thirdhand Smoke Exposure Raises New Concerns. Environ Health Perspect 119:a70-a74. doi:10.1289/ehp.119-a70](#)
15. [↑ Environmental Health Perspectives homepage](#)
16. [↑ Thirdhand Smoke Exposure Concerns. Encyclopedic article at the Encyclopedia of Earth](#)
17. [↑ Post by Jim R. Wilson to the MediaWiki-l mailing list, Jul 1, 2008](#)
18. [↑ OpenWetWare homepage](#)
19. [↑ Useful Chemistry homepage](#)
20. [↑ Scholarpedia homepage](#)
21. [↑ Proposal:Editorial teams \(Redaktionen\), Wikimedia Strategic Planning wiki](#)
22. [↑ WikiGenes homepage](#)
23. [↑ Gene Wiki homepage](#)
24. [↑ TaxPub page at Sourceforge](#)
25. [↑ National Library of Medicine of the USA \(NLM\) Document Type Definitions \(DTD\)](#)
26. [↑ ORCID homepage](#)
27. [↑ Stack Overflow homepage](#)
28. [↑ Dryad homepage](#)
29. [↑ FigShare homepage](#)
30. [↑ Click here to edit the document. Requires registration at Species ID.](#)