



Community feedback on scholarly content: why it is important and why it should be preserved

The provision of community feedback on the exploration of science is as old as the quest itself. As the publication process has evolved and collaboration technology has adapted along with it, feedback has moved from letters to listservs to preprints to online commenting and annotation. The February 2017 approval of open standards for web annotation provides the infrastructure for an interoperable collaborative annotation layer that will make conversations over scientific content ubiquitous and standard. How is annotation different from current commenting tools? What happens when websites discontinue support for comments, as happened in February 2018 when the National Center for Biotechnology Information announced the end of PubMed Commons? Learn how this community feedback was preserved in the form of annotations that support the FAIR data principles that they be findable, accessible, interoperable and reusable.

Keywords

Annotation; comments; community; feedback; collaboration; FAIR; preservation; Hypothesis

On 1 February 2018 the National Center for Biotechnology Information (NCBI) announced that PubMed Commons, which allowed researchers to add comments to article abstracts, would soon be discontinued and that the existing comments would cease to be visible shortly thereafter.¹ The reason cited for this change was low participation in relation to investment and the existence of alternative fora for commenting elsewhere on the web. The NCBI is not the only organization to discontinue commenting functionality on their website, so one might well ask about the value of community feedback on scholarly content.² As we outline below, feedback is integral to scientific communication and scientists and scholars go to great lengths to attend conferences and communicate with peers. How have community feedback options changed over time? How do annotation tools fit in? How were PubMed Commons comments turned into annotations? Where do we go from here?



HEATHER STAINES

Director of
Partnerships
Hypothesis



MARYANN E
MARTONE

Director of
Biosciences
Hypothesis

The history of community feedback

As long as there has been scholarly content, there has been community feedback, either in the form of researchers in similar disciplines gathering together face to face, formally or casually, to discuss findings and theories or through personal correspondence or trusted gatekeepers.³ With the launch of scholarly journals came letters to the editor. It is unclear whether the first issue of *Philosophical Transactions* included letters to the editor in the way that we understand them today, but many of the submissions came in the form of letters to the Royal Society's first Secretary, Henry Oldenburg, who had previously managed group correspondence for members.⁴ The first issue of *Nature*, published on 4 November 1869, contained letters to the editor⁵ and the first issue of *Science* to publish such a letter was the 14 January 1882 issue.⁶

In addition to sharing correspondence about their work, scientific communities would often share early versions of articles, a process made easier by carbon paper after the turn of the 19th century and widespread photocopying from the late 1950s. Pre-internet preprint experiments

'feedback is
integral to scientific
communication'

2 date back to the 1960s when the National Institutes of Health (NIH) started Information Exchange Groups to share biology findings, but the experiments ultimately ended when scholarly journals saw these preprints as a threat to their business model.⁷ Not long after the NIH project began, there was a proposal for a central registry in particle physics. This too was controversial and ultimately not pursued.⁸ Another increase in the sharing of articles with colleagues for feedback came in the 1980s with the development of e-mail technology. When in the early 1990s these e-mailed articles threatened to fill up available disk space of recipients, a repository and alerting system was created. In 1998 it would be renamed arXiv.⁹

As online tools matured, new forms of feedback, most notably the blog, started to gain traction. The first blog, Links.net, was created by Swarthmore College student Justin Hall in 1994, although the term 'weblog' would not come along for another three years.¹⁰ In the late 1990s blogging platforms were released, enabling those with fewer programming skills to participate. Podcasting and video blogging would begin in 2003 and 2004.¹¹ Many scholarly publishers now offer blogs in addition to their journal and e-book content.

When journals moved online, so too did the letters to the editor. Many publications also added commenting functionality which made it simpler for researchers to share their opinions on articles and even to reply to the opinions of others. By 2011 three quarters of sites that used an external commenting tool used Disqus, launched some four years earlier.¹² About five years ago many mainstream websites began removing commenting sections, citing the incivility of the participants and the prevalence of alternative areas for discussion like Twitter and Facebook.¹³ While incivility may not be as much of a problem in the scholarly community, most journals have not had much uptake of public comments, as illustrated by the NCBI experience.

'most journals have not had much uptake of public comments'

The history of annotation

The history of another mechanism for feedback, annotation, plays out a bit differently. Sometimes called marginalia, annotation can include highlights, underlining, tags, private notes or notations for collaborative editing, connected to a specific portion of text. Early indications of textual annotation, made by scribes or readers on hand-copied manuscripts, go back at least as far as 1000 AD.¹⁴ The rise of the printing press and increase in access to individual copies of a book made annotation more of a private activity.¹⁵

Fast-forward to the 20th century: from the earliest thinking about the communication technology that would eventually become the internet we know today, there were plans to include annotation. In 1945 Vannevar Bush published a piece in *The Atlantic* envisioning a 'mechanized file and library' for storing and organizing 'records, books, and communications'. A reader would pull up the desired book with a code and project it on a screen, with the ability to display multiple items at a time, and would then be able to 'add marginal notes and comments' to documents.¹⁶ Marc Andreessen would touch upon this idea again in his web browser Mosaic with an aim of restoring 'the big missing feature' from the web.¹⁷ Technological limitations of the day unfortunately left the promise of annotation unrealized, and it would be another 20 years before a community would form to push forward open annotation as a web standard.

On 23 February 2017 the W3C (World Wide Web Consortium) standards body for the web approved annotation as a web standard.¹⁸ This development paved the way for annotation clients to be built natively into web browsers, so that a user can set preferences in the same way that preferred search engines are indicated today. More importantly, when multiple annotation services all follow the standard, researchers will be able to interact with each other even when not using the same service – much like e-mail works today. The Annotating All Knowledge Coalition, formed in 2015, promotes collaboration between member universities, publishers and technology partners, and is also free to join for universities, publishers, platform hosts, and technology companies.¹⁹

What tools are available and how do they work?

The first generation of web-based annotation tools began to appear around 2012, with tools like Rap Genius (later Genius) and Hypothesis. These tools provided specialized extensions that could be installed by individual users within their web browsers, allowing them to annotate almost any web page. Thus, users were generally in control of what content they wished to annotate rather than requiring content providers to make annotation function available.

Web annotation tools have continued to proliferate. When considering annotation tools, researchers can look at a variety of differentiating factors. How widely can the tool be used? Some tools can be used anywhere and do not require the publisher or site owner to embed anything to make that possible. Does a user control their own annotations? Some tools have APIs so that users do not lose their annotations if they choose to switch services. What are the collaboration options? Some tools enable private collaboration groups but no means to make a public or world-readable annotation. Where does the collaboration happen? Some tools overlay annotations on the version of record and others are utilized by making a copy of the content and loading that copy onto another site.

'Does a user control their own annotations?'

PaperHive, launched in mid-2016, enables researchers to copy PDF content from sites where a publisher has embedded a link and deposit it into their hive for collaboration. This workflow serves a similar function to the scholarly collaboration networks (SCNs) referred to below. Users can make collaboration groups, but there is no way to make annotations world-readable. There is an API, so users have access to their annotations for other uses.

Hypothesis is an open source tool that researchers can use anywhere across the web with no site-owner action required. Annotations anchor to the version of record and typically work across format and across aggregations. In addition to private collaboration groups, users can annotate in the public channel. Publishers can enable the freely available version of Hypothesis to make these public annotations visible to anyone or they can support their own branded and moderated layer. With the Hypothesis API, researchers and others can access their own or publicly visible annotations.

Remarq by Redlink, launched in spring 2017, is the newest participant in the space, focusing on creating a social network with article sharing and annotation capabilities upon content where the publisher has embedded the tool. Remarq Lite enables personal notes but no world-readable annotations. Remarks on publisher sites can be read by anyone, but annotators are vetted by Remarq staff before their public annotations can appear. There is no API.

There are other proprietary tools that can be used on publisher-enabled content that offer annotation, including Digital Science's ReadCube and colwiz by Taylor & Francis. There are also additional open source tools like Pundit and BibSonomy.

Commenting vs. annotation

With the decreasing availability of commenting systems, one might wonder whether things will be any different when it comes to the latest generation of annotation technology. As long as tools meet their needs, researchers may not care whether they are making 'comments' or 'annotations'. However, clear differences emerge when you look at the degree of collaboration possibilities, portability of data and range of use.

'Commenting tools typically have only one type of visibility: public'

Commenting tools typically have only one type of visibility: public. There is no way to use the tool to make private notes or to put together a collaboration group. These collaboration groups travel across the web with the researchers to fit with their research workflows. Sponsoring organizations can also offer branded and moderated groups visible by default

- 4 on their content, with options for general discussion or layers restricted to annotations by authors or reviewers. Multiple layers can enable different conversations to happen on the same document simultaneously.

Statistics from Hypothesis show that these different modes are all used. Hypothesis users have created more than 2.8 million annotations, one quarter of which are open for the public to read, with another quarter completely private. The rest, approximately 1.4 million annotations, are in private collaboration groups. To date, more than 22,000 private groups have been created. With publishers enabling annotation tools to be visible by default, the visibility in the community and potential for uptake will increase.

Open annotation tools also offer APIs that make researcher annotations portable, in the event that the researcher wants to move them to another platform. Researchers do not want to invest time creating notes in proprietary tools which could be discontinued or changed without notice. There is no way to get your comments out of a commenting system for use elsewhere.

Commenting tools can only be used on sites where a publisher has chosen to offer them. A researcher cannot take these tools with them wherever they are doing research, which is typically across publisher platforms and content types. This is one reason for the popularity of SCNs.

Commenting tools keep all comments in silos, providing researchers with no way to see all of their comments in one place or to search or organize them. With annotation, however, profile pages and search features can enable researchers to access all of their annotations, links to the documents they annotated upon and any tags they may have made. Standards-based annotation enables cross-format (HTML, PDF, EPUB) and cross-platform visibility. Researchers need not worry that they will miss important information because they are working on one format or if the publisher also hosts content on aggregator sites like PubMed Central or Project MUSE.

'Commenting tools can only be used on sites where a publisher has chosen to offer them'

What kinds of communities offer feedback?

Much collaboration occurs today in SCNs like Mendeley or ResearchGate well away from the publisher's version of record. Readers move a copy of the article into these private platforms because the tools they need to collaborate are not available on the publishers' platforms. They also wish to collaborate across content from multiple publishers. This fractures conversations and hides significant usage that librarians and publishers depend upon to understand how content is being used.

'Much collaboration occurs today in SCNs'

Another growing avenue for collaboration is around preprints posted on discipline-specific servers like arXiv, bioRxiv, Center for Open Science servers, American Geophysical Union's Earth and Space Sciences Online Archive (ESSOAr), Social Science Research Network (SSRN) and more. Researchers post these early versions of their results with the hope that reader input will improve or inform them, setting up an ideal opportunity for community interaction.

Through annotation, limited communities or private groups can collaborate on top of the documents themselves, whether they are hosted at a publisher site, an aggregator site, a preprint server or an SCN.

'Publishers are enabling branded annotation layers'

Publishers are enabling branded annotation layers, visible by default on their content, to keep discussion closer to the content itself. Layers can be limited for specific types of community feedback, separating general annotation from peer-review or society-member activity. With the support of preprint servers, who have indicated their willingness to integrate publisher-branded and -moderated layers, publishers can expand their community to make their branded layers visible on those sites as well.

5 The teaching and learning space is yet another forum for community feedback. Students and teachers can collaboratively annotate content assigned via the LMS environment as well as on the open web in a private group setting that meets university privacy needs. Scholars of education have taken an interest in classroom use of collaborative annotation and have begun researching and publishing about its efficacy. Evidence suggests web annotation helps improve students' reading comprehension and critical thinking, development of domain-specific knowledge, and collaboration.^{20,21,22,23}

Communities of experts can band together to fact-check content across the web, using annotation to correct inaccurate or missing information and to provide links to other relevant content. One example of this is Climate Feedback, a global group of more than 200 climate scientists, annotating in the public interest. Publications, either mainstream or scholarly, are distributed to the relevant experts who grade them on accuracy, produce 'feedbacks' for each, and link to their public annotations on the articles themselves. Their action has led publishers to issue corrections for misleading statements within articles.²⁴

'Communities of experts can band together to fact-check content'

With so many potential communities in play, it is important to consider how they will be organized to enable readers to determine who is annotating where and for what purpose. Researchers will also need the ability to discover annotation groups and integrate them into their workflow. The Annotating All Knowledge Coalition is considering both discovery and display of annotation layers to meet multiple needs.

PubMed Commons: from comments to annotations

Almost as soon as the NCBI revealed that commenting would be removed, researchers took to Twitter, asking what could be done to preserve the content and the functionality. Some asked if Hypothesis, a non-profit and open source technology company started in 2011, might be able to offer a solution. An immediate effort began, and eight days later we announced the preservation of more than 7,000 comments, including replies, as Hypothesis page notes (annotations connected to a document but not a specific line in that text). You can learn more about that initiative in this blog post.²⁵

The exercise was about more than just moving comments from one server to another. Care was taken to display and link the licences for these PubMed Commons archive annotations. In addition, utilizing a mapping file supplied by Europe PMC, digital object identifiers (DOIs) were matched with many of the articles to which the comments referred. Where previously these comments were siloed on PubMed Commons, now they are visible as annotations there and on the original publications. The entire corpus of annotations can be visited and explored through the Hypothesis search page. PMID tags – unique identifiers used in PubMed – enable readers to filter comments on an individual abstract. In taking these additional measures, Hypothesis was practising FAIR data principles to make the resulting annotations findable, accessible, interoperable and reusable.²⁶

Preserving comments as part of the scholarly record

While there were not a large number of comments on PubMed Commons compared to the number of documents on the platform, many of the comments provided important information on changes or corrections to the article. Without reader analytics, it is not possible to know how many benefited from reading these comments without choosing to create any themselves. The siloed and often fleeting nature of commenting systems may discourage researchers from investing their time to create content on a platform that could go away. Without an easy way to copy comments and transfer them to another tool, users face the prospect that their remarks will disappear when a company decides to withdraw support from commenting systems. Discussions are under way with preservation initiatives at CLOCKSS and Portico to identify strategies to preserve annotations.

'The siloed and often fleeting nature of commenting systems may discourage researchers'

6 The importance of community-led and open source tools

Hypothesis was specifically created as a non-profit to ensure the continuation of an independent voice in the annotation space. It cannot be acquired by a commercial player. The technology was based on an open source code around which a community has grown. It supports a standards-based annotation ecosystem that will enable users of different services to interact with each other or move their annotations at will. Organizations who wish to run their own in-house annotation server can do so, and users will soon be able to connect the browser-based client to multiple servers as they wish. Interested parties can join the Annotating All Knowledge Coalition and participate in the annual I Annotate conference, now in its sixth year.

Where do we go from here?

The additional work that was done on the PubMed Commons annotations made them more useful as a mechanism for community feedback. As is the case with most comments on the web today, PubMed Commons comments did not comply with most FAIR data principles that would make them findable, accessible, interoperable and reusable. Visible only on the PubMed Commons site itself, they had limited use as part of the community feedback infrastructure. With no API to export them and no DOIs associated, they were not very accessible or interoperable, though they did display a clear licence on the individual annotation level to indicate how they could be reused.

The ~7,700 annotations across 5,888 articles that were collected from PubMed Commons join the three million annotations on Hypothesis.

Despite some sites removing commenting, the opportunities for different types of community feedback are expanding. The publication of the web standard will enable users to designate their default annotation client within the browser itself.

And annotations made with clients that implement the standard should be able to interact with each other, in the same way that users can use different e-mail clients today. Until then, more sites are embedding annotation technology so that their readers will not need to bring plug-ins or extensions. Annotating All Knowledge Coalition members are actively seeking to experiment around interoperability.

'opportunities for different types of community feedback are expanding'

Acknowledgements

The authors would like to thank colleagues Dan Whaley, Jon Udell, Nate Angell and the Hypothesis Engineering Team, as well as folks at PubMed and Europe PMC without whom the initiative to preserve PubMed Commons would have not been possible.

Abbreviations and Acronyms

A list of the abbreviations and acronyms used in this and other *Insights* articles can be accessed here – click on the URL below and then select the 'Abbreviations and Acronyms' link at the top of the page it directs you to: <http://www.uksg.org/publications#aa>

Competing interests

The authors have declared no competing interests.

References

- PubMed Commons to be Discontinued, 1 February 2018, *NCBI Insights Blog*: <https://ncbiinsights.ncbi.nlm.nih.gov/2018/02/01/pubmed-commons-to-be-discontinued/> (accessed 6 April 2018).
- Finley K, A Brief History of the End of Comments, *Wired*, 8 October 2015: <https://www.wired.com/2015/10/brief-history-of-the-demise-of-the-comments-timeline/> (accessed 6 April 2018).
- Fjällbrant N, Scholarly Communication – Historical Developments and New Possibilities, *Proceedings of the IATUL Conferences*, 1997, West Lafayette, IN, Purdue University Press: <https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1389&context=iatul> (accessed 6 April 2018).
- Yeo R, *Notebooks, English Virtuosi, and Early Modern Science*, 2014, Chicago, IL, University of Chicago Press; DOI: <https://doi.org/10.7208/chicago/9780226106731.001.0001>
- E-mail to Heather Staines from Allison Mitchell, Springer Nature, 5 March 2018.
- E-mail to Heather Staines from Sarah Whalen, AAAS, 5 March 2018.
- Cobb M, 16 November 2017, The prehistory of biology preprints: A forgotten experiment from the 1960s, *PLOS Biology*; DOI: <https://doi.org/10.1371/journal.pbio.2003995> (accessed 6 April 2018).

8. Till J E, Predecessors of Preprint Servers, *Learned Publishing*, 2001, 14(1), 7–13: <https://arxiv.org/html/physics/0102004>; DOI: <https://doi.org/10.1087/09531510125100214> (accessed 6 April 2018).
9. Ginsparg P, 13 September 2011, It was twenty years ago today..., arXiv: <https://arxiv.org/pdf/1108.2700.pdf> (accessed 6 April 2018).
10. Chapman C, 14 March 2011, A Brief History of Blogging, WebDesignerDepot: <https://www.webdesignerdepot.com/2011/03/a-brief-history-of-blogging/> (accessed 6 April 2018).
11. Chapman C, ref. 10.
12. Caprio S, 2 March 2011, Lijit Study Shows Publisher Adoption of Social Media Tools Grows 80%, Lijit Blog: <https://web.archive.org/web/20130705221155/http://www.lijit.com/blog/2011/03/02/lijit-study-shows-publisher-adoption-of-social-media-tools-grows-80/> (accessed 6 April 2018).
13. Finley K, ref. 2.
14. Wolfe J L and Neuwirth C M, From the Margins to the Center: The Future of Annotation, *Journal of Business and Technical Communication*, 2001, 15(33), 333–370; DOI: <https://doi.org/10.1177/105065190101500304> (accessed 6 April 2018).
15. Wolfe J L and Neuwirth C M, ref. 14.
16. Bush V, July 1945, As We May Think, *Atlantic*: <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/> (accessed 4 April 2018).
17. Shontell A, 3 October 2012, Marc Andreessen Reveals The Feature He Almost Included In Web Browsers But Never Got Around To Adding, *Business Insider*: <http://www.businessinsider.com/theres-a-feature-that-was-supposed-be-in-web-browsers-from-the-very-beginning-but-it-was-dropped-at-the-last-minute-2012-10> (accessed 6 April 2018).
18. Cole T, 23 February 2017, Making It Easier to Share Annotations on the Web, W3C Blog: <https://www.w3.org/blog/2017/02/making-it-easier-to-share-annotations-on-the-web/> (accessed 20 March 2018).
19. Annotating All Knowledge: <https://Hypothesis/annotating-all-knowledge/> (accessed 6 April 2018).
20. Johnson T E, Archibald T N and Tenenbaum G, Individual and team annotation effects on students' reading comprehension, critical thinking, and meta-cognitive skills, *Computers in Human Behavior*, 2010, 26(6), 1496–1507; DOI: <https://doi.org/10.1016/j.chb.2010.05.014> (accessed 6 April 2018).
21. Novak E, Razzouk R and Johnson T E, The educational use of social annotation tools in higher education: A literature review, *The Internet and Higher Education*, 2012, 15(1), 39–49; DOI: <https://doi.org/10.1016/j.iheduc.2011.09.002> (accessed 6 April 2018).
22. Su A Y S, Yang S J H, Hwang W-Y and Zhang J, A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments, *Computers & Education*, 2010, 55(2), 752–76; DOI: <https://doi.org/10.1016/j.compedu.2010.03.008> (accessed 6 April 2018).
23. McNutt M, More science in the classroom, *Science*, 2014, 346(6212), 895–895; DOI: <https://doi.org/10.1126/science.aaa1728> (accessed 6 April 2018).
24. McMahon J, 27 January 2017, Climate Scientists Launch Brainy Attack on Inaccurate News, *Forbes*: <https://www.forbes.com/sites/jeffmcmahon/2017/01/27/climate-scientists-launch-brainy-war-on-fake-news/> (accessed 6 April 2018).
25. Udell J, 9 February 2018, Preserving Comments from PubMedCommons, Hypothesis Blog: <https://web.Hypothesis/blog/archiving-pmc-comments/> (accessed 6 April 2018).
26. Force11, Guiding Principles for Findable, Accessible, Interoperable, and Re-Usable Data, Force11 Blog: <https://www.force11.org/fairprinciples> (accessed 6 April 2018).

Article copyright: © 2018 Heather Staines and Maryann E Martone . This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use and distribution provided the original author and source are credited.



Corresponding author

Heather Staines

Director of Partnerships

Hypothesis

14 Raynor Avenue, Trumbull, Connecticut 06611, US

E-mail: heather.staines@gmail.com

ORCID ID: <http://orcid.org/0000-0003-3876-1182>

Maryann E Martone

ORCID ID: <http://orcid.org/0000-0002-8406-3871>

To cite this article:

Staines H R and Martone M E, Community feedback on scholarly content: why it is important and why it should be preserved, *Insights*, 2018, 31: 13, 1–8; DOI: <https://doi.org/10.1629/uksg.418>

Published by UKSG in association with Ubiquity Press on 01 May 2018