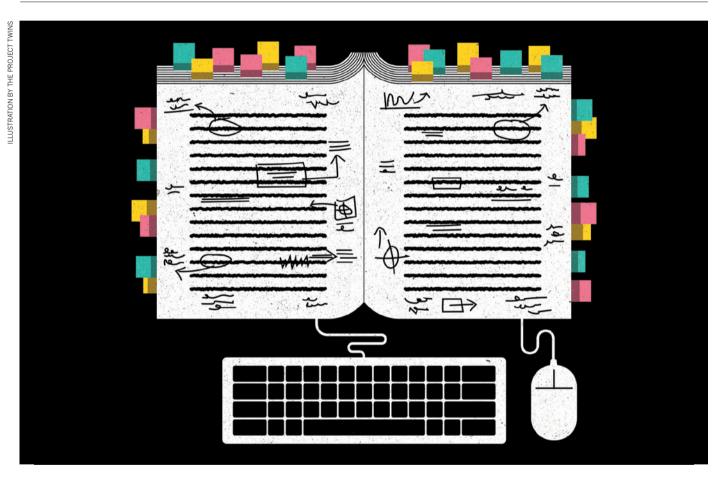# ANNOTATING THE SCHOLARLY WEB

Scientific publishers are forging links with an organization that wants scientists to scribble comments over online research papers.

BY JEFFREY M. PERKEL

Would researchers scrawl notes, critiques and comments across online research papers if software made the annotation easy for them? Dan Whaley, founder of the non-profit organization Hypothes.is, certainly thinks so.

Whaley's start-up company has built an open-source software platform for web annotations that allows users to highlight text or to comment on any web page or PDF file. And on 1 December, Hypothes.is announced partnerships with more than 40 publishers, technology firms and scholarly websites, including Wiley, CrossRef, PLOS, Project Jupyter, HighWire and arXiv.

Whaley hopes that the partnerships will encourage researchers to start annotating the world's online scholarship. Scientists could scribble comments on research papers and share them publicly or privately, and educators could use annotation to build interactive classroom lessons, he says. If the idea takes off, some enthusiasts suggest that the ability to annotate research papers online might even change the way that papers are written, peer reviewed and published.

Hypothes.is, which was founded in 2011 in San Francisco, California, and is supported by philanthropic grants, has a bold mission: "To enable conversations over the world's knowledge." But the concept it implements, online annotation, is as old as the web itself. The idea of permitting readers of web pages to annotate them dates back to 1993; an early version of the Mosaic web browser had this functionality. Yet the feature was ultimately discarded. A few websites today have inserted code that allows annotations to be made on their pages by default, including the blog platform Medium, the scholarly reference-management system F1000 Workspace and the news site Quartz. However, annotations are visible only to users on those sites. Other ▶

▶ annotation services, such as A.nnotate or Google Docs, require users to upload documents to cloud-computing servers to make shared annotations and comments on them.

Hypothes.is is not the only service that wants to make it easy for users to leave annotations across the entire web. A competing offering is a web annotation service from Genius, a start-up firm that began as a site for annotating rap lyrics. In April, it launched services such as browser plugins to help users to annotate any web page. But unlike Hypothes.is, the Genius code is not open-source, its service doesn't work on PDFs, and it is not working with the scholarly community. On the scholarly side, the reference-management tool ReadCube makes it possible for users to annotate PDFs of papers viewed on a ReadCube web reader — but that software is proprietary. (ReadCube is owned by Digital Science, a firm operated by the Holtzbrinck Publishing Group, which also has a share in *Nature*'s publisher.)

By contrast, the open-source nature of the Hypothes.is platform means that anyone could use it to create their own annotation reader or writer — just as anyone can create their own web browser using standards-based technology. The company is also a member of a working group within the World Wide Web Consortium, the standards body for the web, which is developing a universal standard for annotations and how they are communicated. The hope is that web pages that allow annotations would all adopt the same underlying code and protocols (as they do with hyperlinks, for example), making the function easier to use and interact with. The working group has released a draft version of its standard, which is expected to be finalized by the end of 2016.

HOW IT WORKS
For now, Hypothes.is users have several options for creating and viewing annotations. These include bookmarklets (a simple program within a browser bookmark), browser plugins or adding 'via.hypothes.is/' to the start of any URL.

When a Hypothes.is user opens a page — a scholarly article, for instance — the web browser shows any annotations to which the user has access. These appear as highlighted words and comments on top of the text, like an overlaid transparency. Users can then add their own comments, similar to a student marking up a textbook. These are public by default but can be made private, and, following an update added on 3 November, annotations can be shared with private groups. That should enable the tool to be used for journal clubs, classroom exercises and even peer review.

If a page has been altered since an annotation was made, the software uses 'fuzzy' logic to map annotations to their approximate original location. The system can also map annotations from HTML to PDF and back again (for instance, if a user annotates the web version of an article and subsequently views a PDF of the same document).

Annotations are stored on a dedicated Hypothes.is server, which Whaley says looks set to log around 250,000 comments from some 10,000 users in 2015. For instance, after Hurricane Patricia in October, climate scientists left comments and highlighted text on a widely shared mashable.com article (see go.nature.com/rcsesf). But publishers that wish to host annotations for their own content, or companies that want to annotate corporate documents behind a firewall, could run their own server using the same software platform, Whaley adds.

PUBLISHER PARTNERSHIPS
A Hypothes.is user can already annotate any web page — including research papers and pay-to-view articles to which they have access. But the formal partnership announced this week sees some publishers working harder to encourage annotation, including tackling content that annotation systems stumble over, such as page frames and embedded page readers.

The digital library JSTOR, for example, is developing a custom Hypothes.is tool for its educational project with the Poetry Foundation, a literary organization and publisher in Chicago, Illinois. Alex Humphreys, who is director of JSTOR Labs in New York City, says that teachers will be able to use the tool to annotate poems with their classes. An instructor selects the poem to be annotated, sets up a dedicated page with a copy of it, and restricts access to their class only. Students can then create personal notes or share them with the group; an extra annotation layer finds the scholarly resources in JSTOR that quote each line of poetry and provides links out to those resources. The tool is slated to launch in mid-December, Humphreys says.

"You can think of this as a fabric that allows those comments to move freely in time and [across] versions."

The scientific publisher eLife in Cambridge, UK, has been testing the feasibility of using Hypothes.is to replace its peer-review commenting system, says Ian Mulvany, who heads technology at the firm. The publisher plans to incorporate the annotation platform in a site redesign instead of its current commenting system, Disqus. At a minimum, says Mulvany, Hypothes.is provides a mechanism for more-targeted commentary — the equivalent of moving comments up from the bottom of a web page into the main body of the article itself.

Another partner, the arXiv preprint service run by Cornell University Library in Ithaca, New York, has been working on making annotations flow across multiple article versions, says information scientist Simeon Warner, who leads technology development for arXiv. To jump-start interest in the annotation program, arXiv has been converting mentions of its articles in external blog posts (called trackbacks) into annotations that are visible on an article's abstract page when using Hypothes.is.

NOT JUST GRAFFITI
Hypothes.is plans improvements to its platform that include a way to validate the identities of commenters, by incorporating researchers' unique ORCID digital profiles. That could go a long way towards improving adoption of the system among scholars, by facilitating expert commentary on published works and filtering out unwanted marginalia, says Paul Ginsparg, the founder of arXiv and a physicist at Cornell University. "If people start looking at articles and they see the equivalent of graffiti, then people will turn off the comments and the experiment will fail," he says.

If it takes off, online annotation could represent a fundamental shift in the way scholarly communication is done, adds Cameron Neylon, part of the research team at the Centre for Culture and Technology at Curtin University in Perth, Australia, who formerly worked at PLOS.

At the moment, Neylon explains, the scholarly publishing process involves ferrying a document from place to place. Researchers prepare manuscripts, share them with colleagues, fold in comments and submit them to journals. Journal editors send copies to peer reviewers, returning their comments to the author, who goes back and forth with the editor to finalize the text. After publication, readers weigh in with commentary of their own.

With an open-source annotation platform, Neylon says, the document is the centre of attention. Different contributors act on the content simply by changing who has access to it and its comments, with the document becoming richer over time. "You can think of this as a fabric that allows those comments to move freely both in time and [across] versions in a way that we've never been able to do before," he says.

But as Ginsparg points out, it is not clear that researchers — who have proved reluctant in repeated trials to comment on published articles — will take to annotation, even if they can share their comments privately. "There's no incentive structure for people to comment extensively, because it can take time to write a thoughtful comment, and one currently doesn't get credit for it," he says. "But it's an experiment that needs to be done." ∎

Jeffrey M. Perkel *is a writer based in Pocatello, Idaho.*

CORRECTION
The table in the Toolbox article 'Eight ways to clean a digital library' (*Nature* **527,** 123–124; 2015) wrongly stated that ReadCube runs on only desktop and web platforms. In fact, it also runs on mobile platforms.