Commonplace

Remarking on Annotation

Remi Kalir

Published on: Jun 30, 2020

DOI: 10.21428/6ffd8432.8b15fafc

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

“Annotation,” as [John Seely Brown and Paul Duguid discuss](#) in their analysis of the social life of documents, “is a rich cultural practice.” Hundreds of years ago, medieval scribes annotated manuscripts to translate words, illuminate passages, and correct errors. Today, middle school students annotate their school books to assist reading comprehension, language acquisition, or to simply doodle. Genomes, computer code, and chess games are all annotated. Artificial intelligence is trained using annotated data, just as our use of social media involves the annotation of posts, tweets, and viral videos. Across texts and time, culture and context, annotation is an everyday and social activity that provides information, shares commentary, sparks conversation, expresses power, and aids learning.

Annotation may be defined as a note added to a text. This is an intentionally flexible and inviting definition that my colleague Antero Garcia and I present in our book [Annotation](#), a volume in The MIT Press Essential Knowledge series. As with our book, this reading list introduces readers to the literary, scholarly, civic, and everyday significance of annotation across historical and contemporary contexts. Like the many examples in *Annotation*, this list demonstrates how and why annotation may be usefully understood as a genre—that is, as a synthesis of reading, thinking, writing, and communication.

From advances in natural language processing to online learning, and from studies of book culture to commentary about journalism and religious texts, the entries on this reading list represent various disciplines, traditions, and perspectives. The entries were selected for a variety of reasons; some provide useful background information, while others are provocative and usefully push forward new ideas about what counts as annotation and why such notes are consequential. Some entries are articles, whereas others are projects related to—or that make extensive use of—annotation. In total, this reading list features 19 entries about annotation. And, of course, many readings and resources have not been included (though I do suggest further recommendations via commentary, which I describe next). Nonetheless, omissions reflect my own editorial decisions or blindspots; accordingly, I encourage readers to share via discussion their favorite examples of annotation, too.

For most entries on this reading list, I have also included some corresponding annotation. My initial annotations are intended as more than introductory or expository commentary. Moreover, given the social qualities of PubPub, I presume additional annotations and replies from readers will appear over time. Many of my

annotations include linked resources as recommended further readings. By including links both within and also as my annotation content, this corpus of annotations functions as a means of connecting bits of knowledge, disparate perspectives, notable people, and ideas together. These annotations-as-connections are the “associative trails” described by Vannevar Bush in his essay “As we May Think” (and yes, this piece is included in the list below). Annotation—as illustrated in the reading list, and also as evidenced by my annotations with linked readings—functions as the connective tissue binding together a dynamic body of knowledge.

As requested, please add your commentary and questions about annotation via PubPub discussion. You are also invited to share your thoughts, additional readings, and related resources via social media using the hashtag #AnnoConvo (an abbreviation of “annotation conversation”). In addition to providing an introduction to the genre of annotation, I hope this reading list provokes further conversation about the importance of notes added to texts.

“What I Really Want Is Someone Rolling Around in the Text”

By Sam Anderson, *The New York Times Magazine*

“I want, in short, marginalia, everywhere, all the time.”

“It quickly began to feel, for me, like something more intense: a way to not just passively read but to fully enter a text, to collaborate with it, to mingle with an author on some kind of primary textual plane.”

“Twitter is basically electronic marginalia on everything in the world: jokes, sports, revolutions.”

This “Riff” is Anderson’s personal history with and appreciation for annotation. It is a lively essay that provides useful history and compelling commentary about annotation in the context of literature, everyday reading, and the social life of books. Written nearly a decade ago, Anderson provides a prescient appraisal of digital annotation technologies and everyday social annotation practices (as with Twitter) that are now taken-for-granted.

“Book Traces”

University of Virginia

“Readers wrote in their books, and left pictures, letters, flowers, locks of hair, and other things between their pages. We need your help identifying them in the stacks of academic libraries. Together we can find out more about what books were and how they were used by their original owners.”

This project from the University of Virginia captures through crowdsourced submissions and archival methods the mutability of books as physical objects. Annotation is one means by which books are changed through interaction with readers, the environment, and other materials. The project, which now features over 3,000 submissions from libraries across the United States, is specifically focused on “unique” copies of books from the nineteenth- and early twentieth-centuries. Explore the project archive and [submit your book!](#)

“Can Deep Reporting Answer the Ultimate Coronavirus Question: How Will it End?”

Corresponding article: “How the Pandemic will End,” by Ed Yong, *The Atlantic*

Interview and Annotations by Monique Brouillette, *NeimanStoryboard*

“Obviously you can’t be sure how the future will play out. So maybe the title should be ‘How the pandemic could end.’ But we wanted to convey a sense of strength. We are making a powerful assertion with this.”

Ed Yong, a journalist and staff writer for *The Atlantic*, published “[How the Pandemic Will End](#)” in late March, 2020, as COVID-19 spread with devastating consequence around the world and throughout the United States. About a month later, *NeimanStoryboard*—a publication of the Nieman Foundation for Journalism at Harvard University that “showcases exceptional narrative journalism and explores the future of nonfiction storytelling”—published an annotated version of the article. The *NeimanStoryboard* post includes three sections: First, an introduction with background information about Yong and the context of his piece; second, a brief interview with Yong about his practices as a journalist (“Despite describing a very dire and grim reality here, you decided to end the piece on an upbeat note. Can you tell us about that decision?”); and third, a version of Yong’s article with custom and interactive questions-as-annotations by Brouillette (color-coded beige) with responses by Yong (shaded light blue). The “[Story Annotation](#)” archive (which *NeimanStoryboard* previously referred to as the “Annotation Tuesday!” series) includes over a hundred entries.

“As We May Think”

By Vannevar Bush, *The Atlantic*

“Thus he builds a trail of his interest through the maze of materials available to him.”

“A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory.”

This seminal essay is considered by many the source text about annotation and the relationships among information architecture, networked knowledge, and human wisdom. Writing at the end of WWII, Bush—who, at the time, was the Director of the Office of Scientific Research and Development—proposed the creation of new technology and processes to make knowledge more accessible and actionable for both scientific study and everyday knowledge work. To do so, Bush introduced a fictitious and annotation-powered memory extension tool called the memex.

“Next-generation Genome Annotation: We Still Struggle to Get it Right”

By Steven Salzberg, *Genome Biology*

“The core feature of genome annotation is still the gene list, particularly the protein-coding genes. With hundreds of eukaryotic genomes and well over 100,000 bacterial genomes now residing in GenBank, and many thousands more soon to come, annotation is a critical element to help us understand the biology of genomes.”

It has been a quarter century since *Haemophilus influenzae*, the first complete bacterial genome, was sequenced and annotated. Genome annotation, as Salzberg explains at the beginning of his brief but illuminating paper, is “the process of decorating the genome with information about where the genes are and what those genes (might) do.” In other words, a sequenced genome can be considered a “text;” in order for scientists and researchers to make sense of the sequenced genome, the addition of notes (or annotation) is necessary to understand specific structural and functional elements. The genomes of [mice](#), [chimpanzees](#), the [common fruit fly](#), and—of course—us [humans](#) have now been sequenced and annotated. Various tools and

techniques support genome annotation, with some scientists even suggesting the need for “democratizing genome annotation.” Salzberg’s paper carefully considers implications for automation, error and correction, and the future of genome annotation.

“Data for: Debating Algorithmic Fairness”

By Melissa Hamilton, *Qualitative Data Repository*

Corresponding research article: “[Debating Algorithmic Fairness](#),” By Melissa Hamilton, *UC Davis Law Review Online*

“Annotation offered opportunities to incorporate dozens of findings, explanations, and supporting materials that otherwise would have been redacted. Readers wishing to learn more than within the four corners of the official, published article can review these supplementary offerings through the links.”

Hamilton’s article is a “story about perceptions of fairness in criminal justice decisions” that also functions as an innovative model for more transparent scholarship and publication. The study is associated with a [Qualitative Data Repository \(QDR\)](#) initiative called [Annotation for Transparent Inquiry \(ATI\)](#) that has advanced approaches “to achieving transparency in qualitative and multi-method research.” As with Hamilton’s research, the published version of studies that employ ATI include a layer of public Hypothesis annotation; elements of ATI annotation may include a full citation of underlying data, an excerpt from the data source, an analytic note about how data were generated or analyzed to support empirical claims, and a link to the data source housed in QDR databases. Hamilton’s publication in the *UC Davis Law Review Online* features 30 extensive ATI annotations, and her corresponding Data Overview provides a description of multiple data sources, analytic methods, and a “Logic of Annotation” that specifies how “Annotations were employed to fulfill a variety of functions, including supplementing the main text with context, observations, counter-points, analysis, and source attributions.”

“An Eyesore in Washington Becomes an Icon”

By Mark Leibovich, *The New York Times*

“Other protesters were perusing the chain-linked collage of signs, messages and artwork that covered nearly every part of the 8-foot barrier across Lafayette Park-- a kind of chaotic bulletin board cluttered with ‘Black Lives Matter’ logos,

renderings of Mr. Floyd and statements of ridicule, often profane, aimed at the president who lived inside the barricades.”

“The fence had become a must-see attraction for them, a monument to how random citizens can reclaim a democratic space.”

Is annotation graffiti? Whether written inside a book or on a wall, it is useful to contemplate the ways in which an expansive interpretation of notes added to texts mimics forms of graffiti. In 1517, an Augustinian monk named Martin Luther nailed his “Ninety-Five Theses” to the door of All Saints’ Church. In 1978, the human rights activist Wei Jingsheng wrote an essay titled “The Fifth Modernization” and [posted it to the Democracy Wall in Beijing](#). From Luther to Jingsheng to the Berlin Wall to the White House fence, annotation is not only a means of marking up books but also the built environment—and often so as to speak truth to power.

“The Best of Doctors Go to Hell’: How an Ancient Talmudic Aphorism Can Inform the Study and Practice of Medicine”

By Andrea Wershof Schwartz, *AMA Journal of Ethics*

“The ancient text of the Mishna, dating from around the second century of the common era, sits in the center of the page, followed by later rabbinic commentary, and then surrounded by marginalia, font size growing ever smaller, added over the centuries, with modern editions continuing to ring the conversation.”

“Just as the words of these texts convey subtle layers of meaning and can be understood in multiple different ways, so too can our patients’ stories and symptoms.”

The word Talmud is derived from the Hebrew root *lamad*, meaning “to teach” or “to learn.” As Schwartz notes, the Talmud is a collection of complex rabbinic teachings that is architected, in both print and purpose, by annotation; it is an amalgamated manuscript comprised of a source text (the Mishnah), commentaries (like the Gemara), and layers of annotation and further commentary-upon-commentary added by many influential rabbis and scholars over many centuries. Schwartz’s “personal narrative” presents a single passage—translated as “The best of doctors go to hell”—and describes how the power of the Talmud’s commentary and ongoing, contemporary interpretation can provide new perspective and professional meaning.

“Marks in Books”

Curated by Pablo Alvarez, University of Michigan

“These marks are extraordinary witnesses offering unique information on various aspects of book history such as production, textual transmission, reception, and provenance history.”

“During the first decades of the fifteenth century, printers assumed that some sections of their books would be completed by hand, including marks such as rubrication, underlying of the table of contents, foliation, marginal index-letters, and decorated, or even historiated, initials.”

Annotation in illuminated medieval manuscripts. Annotation as [rubrication](#). Annotation by scribes for translation and correction. Organized into eight sections, this digital exhibit is a visually-rich and comprehensive historical survey of the many ways in which manuscripts and early print books were marked both intentionally and incidentally, revealing all manner of social, cultural, religious, literary, and scholarly activities preserved over time.

“Inmates Annotate an Archive of Prison Photos With Their Own Stories”

By Bridget Quinn, *Hyperallergic*

“In 2012, Poor began teaching with the prison photos, asking students to ‘map’ their thoughts, analyses, descriptions, and interpretations directly onto the pictures. The results are arresting, as the writers, who are also men in prison, make anonymous images their own, speaking out of their own experiences, bringing insights and empathy that no outside critic or art historian could.”

Nigel Poor is well-known as a co-creator of the popular podcast *Ear Hustle*. Prior to that creative endeavor, Poor taught photography at San Quentin State Prison near San Francisco. This article recalls Poor’s teaching methods, describes how inmates authored annotations to accompany a trove of photographs from the prison archive, and provides information about various exhibitions of the annotated images.

“Annotating Austen: English 581: San Francisco State University”

By William Christmas and students, San Francisco State University

“Annotating Austen is an ongoing digital humanities project that aims to create multi-media annotated electronic editions of Jane Austen’s six published novels. The project engages undergraduate students in researching and writing scholarly explanatory annotations using the web annotation tool Hypothesis.”

All 61 chapters from Austen’s *Pride and Prejudice*, and the 50 chapters of *Sense and Sensibility*, as well as the 31 chapters comprising *Northanger Abbey* are—thanks to the “Annotating Austen” project—all openly accessible and publicly annotated by Prof. Christmas and his SFSU English students. Hundreds of open Hypothesis annotations provide readers with definitions, exposition, links to related resources, photographs and illustrations, as well as literary analysis and critique. As with related efforts to annotate famous literature, this project provides a rich corpus of detailed notes that is a useful resource for both students and everyday readers.

“Tell-tail Signs of Dual Dog Domestication”

Corresponding research article: “Genomic and archaeological evidence suggest a dual origin of domestic dogs,” by Laurent Frantz et al., *Science*

Annotations by Hazel O’Connor, *Science in the Classroom*

“SitC aims to help educators, undergraduates, and advanced high school students understand the research contained in scientific primary literature by using annotations and providing accompanying teaching materials.”

In 2016, the journal *Science* published findings from an international group of geneticists and archaeologists who had analyzed the genetic sequences of ancient dogs, including the complete genome of a late Neolithic dog from nearly 5,000 years ago. The team’s analysis indicated “that dogs may have been domesticated independently in Eastern and Western Eurasia from distinct wolf populations.” Frantz and colleagues’ article is among dozens included in a public-facing educational initiative of the American Association for the Advancement of Science called Science in the Classroom (SitC). SitC demonstrates how expert annotation of primary sources can aid learners as they develop new content knowledge and familiarity with disciplinary methods. SitC makes openly accessible select research papers from the *Science* family of journals about molecular biology, chemistry, evolution, and space science, among [many other topics](#). As with Frantz and colleagues’ study, all of SitC’s curated articles are annotated with illustrative notes about vocabulary, disciplinary methods, descriptions of prior research, and explanations of major conclusions (there are, for example, 78 public Hypothesis annotations added to Frantz and colleagues’ article by

SitC annotator Hazel O'Connor). Initial research about the use SitC in undergraduate education indicates that the annotated articles productively aid students' domain-specific learning.

“Social Annotation and an Inclusive Praxis for Open Pedagogy in the College Classroom”

By Monica Brown and Benjamin Croft, *Journal of Interactive Media in Education*

“In this paper, we will explore both the opportunities for subverting traditional knowledge structures offered by open social annotation, while also bringing to the surface the critical tensions that may make engaging in social annotation more dangerous or ineffective for students from historically marginalized backgrounds.”

The relationship between annotation and student learning is complex; it may not be causal, though it is no doubt constructive. Kenneth Grahame, the Scottish author famous for writing *The Wind in the Willows*, observed in an 1892 essay the following about student [marginalia](#): “The child’s scribbling on the margin of his school-books is really worth more to him than all he gets out of them.” Research about annotation and student learning has flourished in recent decades. These studies have been driven, in part, by broad interest in the development of online education, digital pedagogy, and advances in social annotation technologies and practices. Brown and Croft’s *JIME* article is a notable addition to the literature given their explicit emphasis on equity, power, and the need for critical approaches to open learning arrangements and practices. In the COVID-era, this article is a must-read for faculty and instructional designers committed to equity-oriented digital education as mediated by open and social annotation.

“Scientists Explain What New York Magazine Article on ‘The Uninhabitable Earth’ Gets Wrong”

Edited by Emmanuel Vincent, *Climate Feedback*

“Seventeen scientists analyzed the article and estimated its overall scientific credibility to be ‘low’. A majority of reviewers tagged the article as: Alarmist, Imprecise/Unclear, Misleading.”

[Climate Feedback](#) is a collective of scientists who peer review media reports about climate change via public annotation. Since 2015, the nonpartisan organization has relied on a global community of volunteer scientists to evaluate climate change

journalism through post-publication peer review. The organization’s “[article reviews](#)” aim to hold journalists accountable for reporting factual science. For example, in July 2017 *New York Magazine* published “The Uninhabitable Earth” by David Wallace-Wells. The article went viral, alongside skepticism about the article’s claims. In response, seventeen scientists affiliated with Climate Feedback added nearly a hundred public annotations atop the first online version of the article and, collectively, judged the article’s overall scientific credibility to be low. Two days after the piece’s publication, *New York Magazine* and Wallace-Wells responded to Climate Feedback’s public peer review by [republishing a second version of the article](#) with their own custom annotation (the yellow highlighted text is hard to miss, with annotations—that include hyperlinks to supporting sources—appearing in the left margin). The revision included many clarifying contributions and fact-checking. Climate Feedback’s annotated peer review compelled the magazine’s editors, author, and fact-checkers to revise and improve their reporting.

“Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks”

By Rion Snow et al., *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*

“Human linguistic annotation is crucial for many natural language processing tasks but can be expensive and time-consuming. We explore the use of Amazon’s Mechanical Turk system, a significantly cheaper and faster method for collecting annotations from a broad base of paid non-expert contributors over the Web.”

Snow and colleagues’ 2008 paper concerns what may be perceived, today, as enduring tensions and trade-offs related to data labeling—including automation, expertise, accuracy, quality, cost, time, and labor. Data labeling is, in essence, the addition of new information to existing data. In other words, data labeling is annotation. Such processes have contributed to advances in machine learning and natural language processing, consequently aiding innovations like the ability for autonomous vehicles to navigate about a city or for speech recognition software in a smartphone app to understand a playlist request. Snow and colleagues concluded that non-experts, such as everyday individuals hired through Amazon’s Mechanical Turk service, were able to match the accuracy of expert annotation on tasks such as affect recognition and word similarity. Of course, the insight associated with this paper’s findings may be tempered by more recent—and critical—assessments of data labeling, particularly as concerns

exploitative labor practices and the ethics associated with “turking” for 97 cents an hour.

“A White-collar Sweatshop’: Google Assistant Contractors Allege Wage Theft”

By Julia Carrie Wong, *The Guardian*

“Interpreting a spoken request isn’t magic, rather it has taken a team of underpaid, subcontracted linguists to make the technology possible.”

“From the beginning, Google planned to build the team with just a handful of full-time employees while outsourcing the vast majority of the annotation work to an ‘army’ of subcontracted linguists around the world.”

Published a decade after this list’s prior entry, I recommend you read this exposé about Google and their Pygmalion team as a much-needed counter-narrative. And then read more about the invisible labor force “[working behind the AI curtain](#)” and whose annotation-as-data labeling is necessarily understood as a form of “[ghost work](#).”

“Discourse or Gimmick? Digital Marginalia in Online Scholarship”

By R Lyle Skains, *Convergence: The International Journal of Research into New Media Technologies*

“Academic discourse in the form of annotation and marginalia overlaps both editorial annotation and reader commentary, in the form of peer review.”

“It is yet to be seen whether these developments will result in either a standard academic practice of making personal annotations on digital texts or an evolution in peer review to more open discourse.”

Annotation has, for centuries, contributed to the both the material and scholarly production of knowledge. In Western thought, the so-called “Great Conversation” refers to the idea that scholars participate in an ongoing and iterative process whereby an author references another colleague, ideas are built on prior insights, and intellectual inquiry refines the cumulative progress of learned societies. As this Great Conversation continues, it propels forward an ever-expansive network of annotation—including endnotes, citations, and recent advances like hyperlinked text and open data (as with the previously described ATI)—that provides intellectual lineage to substantiate claims. Annotation is elemental to the advance of research and enables

scholarly publication. In her article, Skains focuses on the ways in which annotation contributes to open peer review processes and outcomes. She presents multiple cases by analyzing both technological affordances, participation patterns, and social resistance. Despite some promising developments, she concludes that annotation, at least in the context of open peer review, may be more of a gimmick than substantive scholarly discourse.

“Severus Snape’s Copy of Advanced Potion-Making”

By members of the *Harry Potter Wiki*

“Snape, being a talented wizard even at a young age, had made alterations to the many recipes within the book for even better effects. He also made notes in the margins of several spells that he had invented himself.”

Spoiler alert: The content of this wiki page concerns a very special annotated textbook, an item that was quite useful to Harry Potter and is familiar to readers of the beloved series. The collaborative authoring of this particular wiki page is also the result of annotation. Over 500 revisions have been made to this entry by dozens and dozens of members of the *Harry Potter Wiki*; from contributor Yatanogarasu, who created the page on April 30, 2009, through recent updates to the French translation of the entry just about one month ago (in May, 2020). Annotation—in the form of page corrections, additions, commentary in a dedicated “Talk” section (in which community members debate the accuracy of content and revisions), and page metadata (including a “Warning” about the provenance of Snape’s textbook and the page’s source material)—is essential to both the content of this wiki page and the processes by which the encyclopedia entry has been iteratively authored over time. Moreover, the entire *Harry Potter Wiki* may be perceived as a corpus of notes (a collection of wiki pages) added as explication to the broader Harry Potter “text” comprised of various media, properties, and cultural discourses.

“The Death of YouTube Annotations Marks an End for Early Interactive Web Video”

By Matt Binder, *Mashable*

“The interactivity made possible by annotations will simply no longer work.”

Video killed the radio star and, in early 2019, YouTube ended annotation.