Transformation of Conventional Research Environments and Publication Forms

Clifford Anderson

Abstract In this paper, I explore how and to what extent transformations in scholarly research and publication have effected changes in scholarship itself. Taking my theoretical cues from media studies, I use the dual lenses of media displacement and media saturation theory to analyze the alteration in practice that has occurred as scholars shift from analog to digital forms of research and writing. We see that this shift was not itself binary but exists along a continuum. For instance, the digitization of primary and secondary sources promised to unlock new methods of digital research, but the often poor quality of optical character recognition impedes the application of those methods. Word processing promised to speed up scholarly production and, in some senses, succeeded but also managed to occlude the digital texts, making them harder to aggregate and repurpose. Web annotations aimed to fulfill the vision of a distributed web of critical commentary but the scale of the internet makes achieving such dreams hard to pull off. Digital tools for enumerative bibliography have largely automated the formulation of citations, though they have not yet broken with the form itself. Finally, digital publishing still relies, by and large, on interfaces that mimic their analog counterparts. In short, we find that digital tools and methods are not displacing the analog but supplementing them. Is this a sign of an ongoing and incomplete digital revolution or a stable and enduring scholarly synthesis?

Keywords Scientific Culture, Media Studies, Digitization, Librarianship, Text Processing, Digital Humanities Pedagogy

There is widespread agreement that Digital Humanities has transformed conventional research and publication. The emergence of blog posts, digital editions, code notebooks, and data repositories alongside the conference talks, articles, and monographs has enriched the range of scholarly outputs. How has this expanded scope of academic outputs changed the nature of research itself? At its root, this is an empirical question that should be addressed through mixed methods research into the changing habits of religious studies scholars. In this contribution, I prepare the ground for such a study by examining the transformations of scholarly research environments in theology and religious studies through the lens of media studies.[1] To analyze these

1 Editor's note: The author uses the terms *theology* and *religious studies* according to the Anglo-Saxon understanding, where religious studies refer to all research that has religion as its object of study (cf. the contribution by Ch. A. Nunn in this volume, p. 71).

Clifford Anderson: Transformation of Conventional Research Environments and Publication Forms, in: Christopher A. Nunn and Frederike van Oorschot (eds.): Compendium of Computational Theology 1. Introducing Digital Humanities to Theology, Heidelberg: heiBOOKS, 2024, S. 443–462.
DOI: https://doi.org/10.11588/heibooks.1521.c21964

transformations, I draw on contested concepts in media studies, namely media displacement theory and the notion of media saturation (Newell et al. 2008). To what extent have new digital media supplanted our conventional research methods? To what degree has our media ecology become inundated with digital technologies?

What is media displacement theory? The core idea is straightforward. We have limited time at our disposal to consume media. If we begin to engage with new forms of media, we need to find the time somewhere in our day. As Bryant & Fondren (2009, 505) remark, "In displacement theory, the core assertion is that media consumption will displace some other activity or activities, such as exercise or social interaction, or even shift time from one medium to another."

Scholarship differs from media consumption, of course, by its focus on production. The purpose of research and publication is to produce and communicate new knowledge. The scholarly activities that support these activities have changed considerably as humanists engage with new forms of digital media. Analogous questions arise about potential displacements. Do scholars spend more time reading ebooks or perusing PDFs online than pulling monographs from library shelves or consulting offprints of journal articles? Has the recent availability of audiobooks from university presses impacted the time spent reading?

The question of saturation is correlated closely with displacement theory. In media studies, saturation indicates the limits of media consumption. As such, the concept resembles the notion of information overload. On its own, information overload is nothing new. As Blair (2011) has argued in *Too Much to Know: Managing Scholarly Information before the Modern Age,* scholars since antiquity have faced the challenge of having too many sources and have designed scholarly technologies, from indexes to florilegia and commonplace books, to mitigate the flow of information. Media saturation theory sharpens the problem of information overload by underscoring that scholarly communication flows through proliferating forms of media. As scholars, we must consider whether or how to combine traditional channels of research – archival manuscripts, journal articles, monographs – with data flowing through newspapers, podcasts, preprint servers, radio, social media, television shows, webinars, websites, YouTube, and now generative artificial intelligence. "The saturation of media tends to result in increasing fragmentation of information," remarks Wasiak (2008, 113), "as one navigates space and media sources." As media proliferates, we cannot connect the threads, tracing conversations from TikTok to blogs to scholarly articles and then back again. As new channels (*Clubhouse,* anyone?) pop up, we may just tune them out because our media environment is already suffused with competitors for our attention. As we shall see, a strong trend in the Digital Humanities resists the proliferation of new media by reducing all media, as far as possible, to data and using command line utilities (Bash, ZSh, etc.) to manipulate those data.

Adopting a media studies lens also helps us to look more holistically at the digital turn in religious studies and theology. On the one hand, theologians and scholars of religion continue to carry out their research and publication in ways similar to

fifty years ago. While computation has wholly transformed the sciences, requiring graduate students and postdocs to develop skills in data analysis and large-scale computing to carry out their research, the same cannot be said of theologians. In *The Place of Computation in the Study of Culture,* Daniel Allington distinguishes, in the line of C.P. Snow, two cultures of the academy: the hermeneutical and the empirical. Among what he terms the "*essentially* hermeneutic," he places the discipline of theology (Allington 2022, 374). "And computation has almost nothing to contribute under such a paradigm," he insists, "unless we mean those computations which go on, unnoticed, in the background, incessantly, so that emails can take the place of the postal service and a word processor can take the place of a typewriter. But almost nobody cares about those (ibid., 373)." From a media studies perspective, we may come to care about these unheralded shifts in research and publication practices because, collectively, they add up to qualitative differences in our scholarly media ecology.

1. The Analog to Digital Shift

When we consider the transformation of conventional research environments, the critical driver that comes to mind is the shift from analog to digital research methods and publication forms. At the center of contemporary scholarly communications stands the networked computer, replacing nearly all previous forms of academic technology. Or so the presumption goes.

The transition from analog to digital has eventuated in both subtle and profound transformations in research and publication patterns. Analog methods have not simply given way to digital successors. Pace the so-called *media displacement* theory, analog and digital methods flourish side by side. From an anecdotal perspective, scholars take notes about the book they are reading, underling or highlighting critical passages in pencil; they also mark up articles in PDF format, storing their digital annotations in tools like *Papers* or *Zotero.*

Robert Hassan, Professor of Media and Communication at the University of Melbourne, argues in *Analog* that we should not consider analog and digital as antipodes. The expanse of the analog world is greater and more connected with our personal and cultural histories; analog habits of research have become second nature, or parts of "our extended mind" (to use the language of Clark & Chalmers 1998). Given the millennia of entanglement of our minds with analog tools of scholarship, it should come as no surprise that these habits did not immediately give way to digital equivalents with the advent of *Unix* time on 1 January 1970. "We need to remind ourselves that to write and read is to interact with a technology," contends Hassan (2023, 132). "At a deeper level we need to remind ourselves also that the technology is analog, in that it corresponds, symbolically, to speaking, to hearing the voice and, with it, the mind's thoughts." Hassan acknowledges that just as Friedrich Nietzsche's purchase of

an early typewriter (a Hansen Writing Ball) in 1882 may have pushed him toward his late aphoristic style, so may the digitization of our research environments shape the contours of our research. "At a deeper philosophical level, screen-reading suggests a new relationship with knowledge, which is to say, a new interface in the constituting of the reality of the world (ibid., 139)." As we shall see, the interface question looms large over the new world of digital research and publication.

2. Digitization

If you found yourself teleported into a library from fifty years ago, things at first inspection would look generally the same. The same row on row of books ordered (at least in the United States) on shelves according to the Library of Congress classification system would span the floors. You would find students in the reading rooms and carrels, pouring over books and bound periodicals. The differences would strike you when you sought a book on the shelf. To carry out that task, you would need to consult the catalog, held in drawer after drawer of cards that might span an entire room. In the library of the 1970s, you would also find technology that has since disappeared or radically diminished. In the interlibrary loan department, alongside analog phones and Xerox 4000 photocopy machines, for example, you would find a teletype machine used to communicate loan requests from peer libraries.[2] Off the side of the reading room, you would find metal cabinets full of microfilm and microfiche along with the requisite readers; would you know how to thread the spool and advance the reader to the beginning of the article you hoped to read?[3] As for computers, you might find mainframes here and there in the technical services units of large university libraries. Still, you wouldn't find them anywhere in the public areas. What you might notice at the reference desk is a so-called "dumb terminal," that is, a system like the DEC VT52 terminal, which connected to the DIALOG Online Search System or a competitor; using a compact query language, the librarian would search multiple databases, carefully avoiding returning too many results as search results were metered (Schatz 1997).

From a phenomenological perspective, the most significant change you would experience relates to the level of intermediation. In the library of the past, librarians played a central role in assisting patrons with navigating to sources of information. The placement of the reference desk at the center of the library both reflected and reinforced the librarian's mediating function. The past fifty years have seen tremendous

2 See https://www.facebook.com/pasadenalibrary/posts/whats-that-machine-its-a-teletype-machine-this-is-a-teletype-model-28-which-had-/10157896789598049 (Accessed 26 June 2024).

3 See https://edtechmagazine.com/higher/article/2017/05/microfiche-was-dawn-multimedia-research (Accessed 26 June 2024).

disintermediation of research environments, primarily removing librarians from the day-to-day research process.

The first wave of disintermediation made metadata directly available to the public. Librarians began to migrate cataloging records from print to digital format in the 1960s and 1970s. By the 1980s, OPACs *(Online Public Access Catalogs)* had become standard in academic libraries in North America, at first coexisting with and then functionally displacing card catalogs. Partnerships like the *Research Libraries Group* (RLG) and the *Ohio College Library Center* (OCLC)[4] enabled researchers to discover literature in peer library collections; these days, scholars can use OCLC's *WorldCat* both to search for relevant items and to identify libraries that hold those items. The same applies to periodical sources. Readers can readily find metadata about articles of interest through *Google Scholar*, though they may encounter paywalls when seeking to download the PDFs.

The second wave of disintermediation resulted from the digitization of collections. In *Along Came Google: A History of Library Digitization,* Marcum & Schonfeld (2021) document the initiatives to digitize library collections before Google. From the mid-1990s, librarians began forming coalitions to make library collections available online. However, the scale of these ambitions, disagreements about directions, and common pool resources dilemmas inhibited the success of these projects. Starting in 2004, however, the entry of Google changed the equation drastically; the *Google Books* project revived hopes that all books could be made accessible online. Google's scanning process introduced errors with fingers and hands appearing in images occasionally but proved efficient (James 2010), all too efficient for publishers. A flurry of lawsuits led to the retrenchment of Google's planned universal library. As Marcum & Schonfeld (2021, 188 f.) remark, "Rather than a universal digital library, we have a potpourri of digital collections, with greater or lesser access, as well as libraries that have individually become digital, more or less." While the prospects of realizing a universal digital library have diminished, scholars can still find monographs online between these sources, reducing their dependence on the library as a physical center of information.

From a Digital Humanities perspective, digitizing these volumes provides an incredible start but also falls far short of the goal. To conduct data-driven research, scholars need direct access to the underlying data and metadata. In many cases, the intellectual property restrictions bar access to the data in raw form. When it is possible to download or scrape the data, scholars frequently find, to their dismay, that the optical character recognition software has produced a nonsensical mess. Quoting from a recent document, for example, we see this sentence, which is no worse or better than the ones surrounding it: "After a week' s lness, E5= abeth, eldet. dn. of Mr. H. W. By* 1s Mfay. m e Fhebs wif of A. C. 1e_, esq. d f of h" e Be,. Ta. s forrlyo em of s Rea f Uonl." Good luck trying to carry out textual analysis on documents this garbled.

4 RLG merged with OCLC in 2006.

3. Word Processing

As Allington remarked, the most profound shifts from analog to digital research may simultaneously be regarded as the most trivial. To wit, the shift in writing practices. Fifty years ago, theologians generally wrote their texts by hand, then typed them up (or sent their manuscripts by interoffice mail to department secretaries to be typed) for submission to a publisher. These days, nearly everyone writes using word processing software and emails files (in Microsoft Word format) directly to editors.

In *Track Changes. A Literary History of Word Processing*, Matthew Kirschenbaum examines the effects of the switch to word processing on writers and writing through the lens of media studies. He demonstrates that what feels inevitable now struck writers of the period as alien, exciting, and unsettling in equal parts during the 1980s. As writers experimented with word processing systems, they regularly remarked about how these systems would transform literary production. Some, like Stephen King, experimented with word processors early on in the hope that they would accelerate their already prolific output. Others, like Gore Vidal, bemoaned that "the idea of literature is being erased by the word processor (Kirschenbaum 2016, 43)." Reflecting from a contemporary vantage point, Kirschenbaum pronounces a tempered judgment. "*Pace* Gore Vidal, word processing did not erase literature, not in any sense I can fathom. Neither, of course, did it perfect literature," he writes, commenting, "But like the typewriter before it, word processing changed the face of literary culture and our imagination of literary authorship (ibid., 243)." That is, the switch to the word processor altered our image of the author; today, we picture writers (and, by extension, scholars) huddled over laptops in coffee shops rather than pounding out prose on manual typewriters or composing essays by candlelight with quill pins. "See what big letters I make as I write to you with my own hand!" Paul exclaimed in Galatians 6:11 as he stopped dictating to his amanuensis. These days, our handwriting has become so squiggly due to our focus on typing that students prefer faculty to proffer feedback in their learning management systems than to scrawl comments in red on their papers.

Will the current ubiquity of Microsoft Word persist among scholars in the long term? As anyone who has used Word for a decade or longer knows, the project managers at Microsoft did not have researchers in view when designing their word processor. For years, academics struggled with footnotes that appeared on the wrong page and other annoying glitches. While these challenges to using Word for academic writing have generally been overcome, there remains a significant mismatch between Word and the Digital Humanities. The "What You See Is What You Get" (WYSIWYG) model of word processing, which sought to replicate the polished layout of the page even in the phase of composition, impedes the perspective of text as data. Why is this so? Word obfuscates the structure of documents by focusing on the presentation of the text on the screen (and, hence, on the printed page). There are tools in Word to mark out the structure of documents, like primary and secondary headings, etc.

However, authors typically ignore such features when formatting their documents and use visual clues like bolding, italics, and larger fonts to indicate textual features.

Another challenge with using Word arises from file formats. Before 2007, versions of Microsoft Word used a proprietary binary format *(.doc)* to store Word files. While Word could export documents to many other formats, including plain text and HTML, the exporting process typically stripped away many formatting codes, effectively diminishing detail about the structure of documents. In the early 2000s, Microsoft began planning to use XML to store documents. The goal was to provide an open, documented standard for transparency and interoperability. The history of the development of what became the *Office Open XML standard (.docx)*[5] and its relationship to a previous competing XML-based standard, *Open Document Format* (ODF),[6] need not detain us here. While the move to these open formats certainly did advance the goals of interoperability, the hope that they would render the technical markup of word-processing documents legible to nonspecialists did not materialize. Few users, even among digital humanists, have dared to unpack these compressed files to inspect their arcane contents. The byzantine structure of these XML formats stymies data extraction from them.

In the end, how many authors care about the format of their electronic texts? The telos of WYSIWYG word processing is to produce a document that resembles the printed page.[7] Where XML-based formats behind the scenes have faltered in their original promise, the *Portable Document Format* (PDF) has solved the right side of the WYSIWYG equation. The PDF's success stems from its capacity to mimic the printed page. "Whether they render digitized text or text that has been born digital, as it were, pdf's present what are called page images," explains Lisa Gitelman. "They look something like pictures of pages produced by one printing process or another, or by word processing (Gittelman 2014, 115)." Given their success at this task, PDFs have become the central format of scholarly exchange, from Interlibrary Loan programs to open access repositories like *ArXiv* to "shadow libraries" such as *SciHub.* The technology backing implementations of the PDF standard belies its simplicity of presentation; the ISO standard for PDF 2.0 runs 986 pages.[8] Vanishingly few users tinker with the internals of the PDF format. Instead, they collect PDF documents in file folders, reference managers, and file-sharing servers, building their libraries of articles to read and cite.

The Digital Humanities community has taken different approaches to aligning digital writing with digital research. But breaking away from the spectral grasp of

5 ECMA-376 2021; see https://www.ecma-international.org/publications-and-standards/standards/ecma-376 (Accessed: 26 June 2024).

6 See https://www.oasis-open.org/2021/06/16/opendocument-v1-3-oasis-standard-published (Accessed: 26 June 2024).

7 As Gitelman (2014, 123) expresses the point, "For wysiwyg to work, there had to be continuity across screens (wys) and the pages printed out (wyg)."

8 See https://www.iso.org/standard/75839.html (Accessed: 26 June 2024).

paper-based formats proves challenging, if only because of disagreements about the path forward.

On the one hand, entrepreneurs have seized on the limitations of Microsoft Word to create niche products for authors and academics. First off the mark was *Nota Bene,* a word processor developed to format papers according to academic style guides such as the *Chicago Manual of Style* or the *American Psychological Association.* Steven Siebert, a doctoral candidate at Yale University studying philosophy and religion under Hans Frei, created *Note Bene* in the early 1980s out of frustration with the limitations of existing word processing software and has continued to market it successfully to academics for forty years.[9] More recently, *Scrivener*[10] has made headway among academics because of its nonlinear, associative approach to organizing long-form writing projects.

On the other hand, a series of alternatives to traditional word processors has emerged to render texts more tractable as data. An early alternative to the WYSIWYG paradigm was Donald Knuth's TeX (a play on τέχνη + X), which he developed in the late 1970s to facilitate the typesetting of his *The Art of Computer Programming.* In the 1980s, computer scientist Leslie Lamport developed macros to automate the features of TeX, making it more accessible to a broader group of users. "LaTeX is not a word processor!" notes the website for the LaTeX project. "Instead, LaTeX encourages authors not to worry too much about the appearance of their documents but to concentrate on getting the right content."[11] Of course, this gap between word processing and formatting with textual macros opened up a potential market. *Overleaf* has emerged as a productive compromise, providing a cloud-based authoring environment for LaTeX with conversion to PDF for visualization.[12] *Overleaf* also offers tools like collaborative editing and revision tracking. These days, a significant amount of scientific publishing takes place in LaTex; the American Mathematical Society "strongly encourage[s]" authors to use LaTeX because of its compatibility with its production systems and, presumably, its fine-grained handling of mathematical symbols.[13] LaTeX has made fewer inroads among digital humanists but may become more prominent as interdisciplinary work between theology and the sciences grows.

Markup languages push the distinction between content and appearance further by separating them nearly completely. By interweaving text and markup, markup languages make documents into data structures; unlike the .docx standard, the goal

9 In 1986, Kevin P. Roddy, a medievalist and early digital humanist (back when it was known as *Humanities Computing*) at the University of California, Davis, remarked, "Steven Siebert, the author of *Nota Bene,* had not at last report finished his dissertation in Philosophy at Yale. I hope that he has now returned to it, and left Version 3 to someone else. In America and elsewhere, we need as many philosophers as we do programmers (Roddy 1986, 95)."

10 See https://www.literatureandlatte.com/scrivener (Accessed: 26 June 2024).

11 See https://www.latex-project.org/about (Accessed: 26 June 2024).

12 See https://www.overleaf.com (Accessed: 26 June 2024).

13 See http://www.ams.org/publications/authors/tex/latexbenefits (Accessed: 26 June 2024).

of markup languages like *Text Encoding Initiative* (TEI) is to make XML documents equally human- and machine-readable. The TEI community has promoted the advantages of semi-structured formats for humanities since the 1980s; the commercial roots of markup languages like GML and SGML go back to the 1960s. In the late 1990s, the *eXtensible Markup Language* (XML) emerged as the dominant standard for creating markup languages.

Humanists have exercised an outsized influence on the XML community, co-developing the standards that emerged from SGML. The XML toolchain contains sophisticated tools for the entire lifecycle of documents, from definition to publication. The availability of open-source XML databases like *BaseX* and *eXist* has enabled scholars to publish their digital editions online. In contrast to relational databases, XML databases use *XQuery* as their combined application and query language.

Creating an XML-based digital edition has become significantly easier in recent years. *TEI Publisher* is a rapid application development environment built on *eXist* that provides the essential functionality for digital editions, including stylesheets for reading onscreen and in print, facetted search, and ready web hosting. *CETEIcean* takes a leaner approach, focusing on web publication of TEI documents using web components to integrate TEI elements into HTML.

XML and its related technology stack fell from favor in the 2010s. Just as XML reached a fever pitch of hype with industry leaders like Microsoft and IBM pivoting entire product lines to XML-based formats, the web community rebelled against XML. Douglas Crockford introduced JSON *(JavaScript Object Notation)* as a lightweight alternative to XML-based message-passing protocols like SOAP *(Simple Object Access Protocol).* JSON was not only more straightforward and, arguably, less verbose than XML, but it was also data-centric rather than document-centric.

The blowback against XML technologies in the early 2000s caused significant collateral damage in the Digital Humanities community. The community found itself isolated to some extent from the mainstream of technological development. XML-focused presentations began disappearing from the agendas of major tech events, though conferences like *Balisage* bound the XML community together.

To make XML more palatable to the workaday Digital Humanities scholar, *Performant Software Solutions* has released a TEI-based word processor called *FairCopy.*[14] *FairCopy* provides a graphic user interface for the TEI, allowing users to compose documents without having to wrangle angle brackets. Such graphic overlays over XML were already available in tools like *oXygen.* Still, *FairCopy* cleverly addresses the complexity of the TEI with its myriad elements and attributes by bundling relevant structures together. Still, you need to know some TEI to write effectively in *FairCopy.*

In the early 2000s, *Markdown* emerged as a stripped-down alternative to WYSIWYG systems and LaTeX and XML-based markup systems. John Gruber, the developer of *Markdown,* created the system out of frustration with online writing using

14 See https://faircopyeditor.com (Accessed: 26 June 2024).

HTML (Lockridge 2020). Flavors of *Markdown* have since evolved into the syntax of web-based writing, from *Discord* to *Github.* As Mailund (2019) remarks, "With *Markdown* you don't have quite the same power to control your formatting as you do in a language like LaTeX, but the simplicity of *Markdown* more than makes up for it."

But what good is *Markdown* for scholarly publishing in conventional contexts? After all, few scholarly publishers (with notable exceptions, such as *PubPub*[15]) accept manuscripts authored in *Markdown.* To facilitate its interoperability with other document formats, a command line program called *Pandoc* has become an essential companion to markdown in the scholarly writing workflow. John McFarlane, Professor of Philosophy at the University of California Berkeley, developed *Pandoc* and, not coincidentally, contributed to the standardization of *Markdown,*[16] to make switching between document formats straightforward.

The combination of *Markdown* and *Pandoc* has made it feasible to dispense with the WYSIWYG paradigm. As I type these words, for example, I am writing in *Visual Studio Code,* Microsoft's open-source code editor. My editing environment is more austere and componentized than Microsoft Word. A system of extensions allows me to add the features I want to the editor, including syntax highlighting for *Markdown,* a spellchecker, and a word counter. When I am finished writing, I use *Pandoc* to convert this document into Microsoft Word, PDF, or any required format. In a strange inversion, Microsoft Word .docx becomes an output file; in the final stage of the writing workflow, I will use *Pandoc* to convert the *Markdown* document(s) (along with accompanying references in BibTeX) into .docx for delivery to my editors.

4. Annotation

The problem of information overload is not a new issue for scholars. In *Too Much to Know,* Ann Blair explores pre-modern scholars' methods of organizing their research. Most techniques she surveys from the ancient and medieval world remain recognizable to contemporary researchers, though some have fallen out of favor. The act of marking up, taking notes, or otherwise annotating books has long served the practice of memory. The purpose of creating such annotations varied. "Annotations might make corrections to the text, add cross-references to similar material in the same or difficult texts, or include occasional words of praise or criticism," explains Blair (2011, 71), "but predominantly they flagged passages of interest […]." The tools have changed, but the motivations stay constant.

As a librarian, I have a professional duty to warn researchers against taking notes in borrowed books. The temptation to mark significant passages with light

15 See https://www.pubpub.org (Accessed: 26 June 2024).
16 See https://commonmark.org (Accessed: 26 June 2024).

pencil marks or, worse, a felt-tipped pen can prove strong. As a reader, I vacillate between abhorrence of texts defaced by yellow highlighting pen and amusement at witty comments or retorts sketched in the margins. In the contemporary world of digital scholarship, the restrictions on marking up documents have been lifted. When the bulk of your sources exist as PDFs or ebooks, analog forms of annotation no longer apply unless you print everything out. The developers of ebook software have sought to create digital analogs to marginal notes, allowing readers to mark up PDF documents with virtual highlights or put digital sticky notes in electronic books.

A new paradigm of social annotation has pushed beyond these simple replacements to instaurate marginalia for digital documents. Social annotation allows readers to mark up what they read online and, if they wish, to share their musings with private circles of others or the public at large. The W3C *(World Wide Web Consortium),* which creates standards for web-based protocols, has formulated a *Web Annotation Data Model* that specifies the conceptual underpinnings for web annotations. The definition of an annotation the working group provides is considerably more abstract than a pencil scrawl in the margins of a codex: "An annotation is considered to be a set of connected resources, typically including a body and target, and conveys that the body is related to the target."[17] This recondite language serves the ambitions of the web annotation; the goal is to provide a framework for the annotation of the entire web, including PDFs and audio/video resources. The motivating ideas behind web annotation harken back to the origins of the web. As Tim Berners-Lee envisioned the *Hypertext Transfer Protocol* (HTTP), the specification contained verbs to get web pages and post, put, and patch them. In other words, he imagined users not only reading and sharing information with web pages but updating them with new information. This utopian vision of a readable and writable web did not survive commercialization, but web annotation brings it back in revised form.

Commercial and open-source options now exist for sharing these digital marginalia. The leading open-source project is *Hypothesis.*[18] During the past decade, the *Mellon Foundation* provided grants to support the *Hypothesis* project and foster academic interest in web annotation. If you install the *hypothes.is* plugin in your browser, you can see where annotators have spoken back to web pages. Seeking annotations on the open web proves a hit-or-miss affair in practice. The stray comments you come across resemble graffiti rather than scholarly marginalia: On 16 October 2022, user jacobknight annotated the front page of Wikipedia with "GREEAATTTTT" and another user remarked a few days later in reply, "I know right?" While occasionally humorous, such scattered exchanges hardly constitute the thick web of scholarly annotation the fiscal sponsors had in view.

In contemporary humanities research, spreadsheets have become an essential companion to the word processor as the locus of note-keeping. The spreadsheet has

17 See https://www.w3.org/TR/annotation-model (Accessed: 26 June 2024).

18 See https://web.hypothes.is (Accessed: 26 June 2024).

taken over the role the box of note cards formerly played in scholarly research. Scholars in the humanities have become accustomed to populating spreadsheets with row after row of data as they read through ship manifests, colonial accounting books, or sermon texts. The spreadsheet serves as an ersatz database without the fuss of normalizing data. The ability to export from Microsoft Word or Google Sheets to *Comma Separated Values* (CSV), a loosely defined but ubiquitous format for sharing tabular data, makes moving data from spreadsheets to databases, code notebooks, and other data-driven software straightforward. Conferences like *csv,conf*[19] have emerged to facilitate the movement of data across these contexts. Once again, academics are converting office technologies to scholarly purposes rather than adopting unfamiliar and more pertinent tools such as web annotations.

5. Bibliography

Acknowledging and citing sources remains an essential element of academic discourse. A key mark of popular scholarship is its dispensation of a bibliographic apparatus. From academics crossing over to trade nonfiction, the transition can be jarring. In a tongue-in-cheek encomium titled *A Man without Footnotes,* Nathan Glazer remarks on what he terms Irving Kristol's *no footnotes* approach (Glazer 1995, 6)," a style that academics "consider a sign of arrogance (ibid.)" and "which has led to much agony among contributors, particularly since most of them are academics (ibid., 7)."

A significant focus of library instruction has been the teaching of bibliographic styles. At an undergraduate level, students trip over bibliographic rules through little fault of their own. "As college students go from class to class, they are often asked by their professors to use different citation styles, thereby preventing the students from becoming familiar with one particular style and all its nuances," writes Pfitzinger (2011, 28). "Since, for the foreseeable future, those in academia will continue to be forced to juggle multiple citation styles, familiarity with any one style manual will be difficult to achieve." An inventory of citation styles shows that more than ten thousand exist. However, many represent variations on the leading guidelines from the *American Philosophical Association* (APA), *Chicago Manual of Style, Modern Language Association* (MLA), and the *Institute of Electrical and Electronics Engineers* (IEEE).[20] Given the success of *Uniform Resource Locators* (URLs) for identifying webpages, we might imagine that a movement toward replacing bibliographic references with semantic identifiers would have occurred by now. While initiatives such as *Citation Identifiers* (CIDs), which condense bibliographic references into numeric facets, have

19 See https://csvconf.com (Accessed: 26 June 2024).

20 See https://citationstyles.org (Accessed: 26 June 2024).

gained traction in the sciences,[21] plans to replace bibliographic references with identifiers have faltered in the humanities.

What has sprung into the gap are bibliographic reference tools like *Zotero* (open source), *Endnote* and *Mendeley* (proprietary), along with a markup language called the *Citation Style Language*, which provides machine-readable descriptions of nearly all citation formats. Browser plug-ins allow readers to click on online articles and import the corresponding bibliographic data and the full text, if available, into their libraries. Using another plug-in, writers can use these citation managers to punctuate their manuscripts with dynamic links that generate the corresponding citations in the correct format on the fly. While maintaining a dynamic connection between the document and bibliographic manager proves handy when writing, authors must remember to *flatten* these references, that is, to break the links with their *Zotero* installation, before submitting their documents. Otherwise, their editors will receive a Word document with broken *Zotero* codes rather than a properly formatted bibliography.

6. Digital Publishing

To remark that the advent of the internet from the early 1970s and, in particular, the popularization of the *World Wide Web* from 1993 revolutionized scholarly communications is both a truism and easily overlooked. Before the advent of the *World Wide Web*, scholarly communications primarily took the form of sharing offprints. If your academic library did not subscribe to a journal, you could write to the author of the article, requesting an offprint be sent to you, or request a copy be sent to your library via interlibrary loan.

The desire to break away from the limitations of the print era has given rise to digital editions. Early in the era of personal computing, Ted Nelson explored the affordances of digital publication, dreaming of the *Xanadu Hypertext System*. "Its unique facilities of backtrack, linkage and windowing will allow the creation of new forms of multi-level, explorable collections and collages of material," he opined, "without losing the well-defined authorship and ownership of all parts (Nelson 1981, chapt. 3,5)." Like Ted Nelson's concept of hypertext, digital editions present scholarly information in nonlinear forms.

While the web's openness allows freedom to experiment with nearly infinite varieties of forms with a decent knowledge of HTML and a pinch of *JavaScript*, scholars have gravitated toward web frameworks for good reason. Web frameworks not only provide the technical backbone of digital editions, connecting frontend user interfaces with backend databases but also encourage the development of third-party

21 See https://www.asme.org/publications-submissions/journals/administration/citation-identifiers-(cids) (Accessed: 26 June 2024).

ecosystems offering open source or commercial plugs-in or *skins.* The most popular frameworks among scholars are *WordPress* and *Drupal,* both open-source solutions that serve commercial audiences beyond the academy. In Digital Humanities circles, entrepreneurial scholars have responded to the mismatch of these web frameworks with scholarly ends by introducing academic content management systems, chief among them *Omeka*[22] and *Scalar.*[23] In the XML world, *TEI Publisher*[24] provides a *rapid application development* (RAD) environment for TEI texts. Suppose you have created a corpus of TEI documents using a customized ODD *(One Document Does It All),* that is, a specification that describes the semantics of your documents. In that case, *TEI Publisher* allows you to quickly create an online edition with the primary functions you would expect, including browsing, searching, and reading texts.

A counter-movement against database-backed content management systems has emerged in recent years under the rubric of *minimal computing.* By slicing through the complexity of web frameworks, *minimal computing* aims to streamline the online publication process and, at least in theory, to render the creation and consultation of digital editions more accessible. "While those who doubt their ability to learn how to code see the use of GUI-driven platforms as the key to access," write Risam & Gil (2022, 16) in their preface to a special issue on *minimal computing,* "often these systems foreclose more control over the production of knowledge, and by extension, participation that is more meaningful to those who seek access." Of course, they acknowledge the trade-offs when moving away from databases and graphic user interfaces (ibid.). Chief among those questions is that trading away user interfaces for code and data demands a higher level of computing knowledge.

The movement toward minimalism reaches its apex in the contemporary debate over data and interface. To what extent is an interface truly a requirement for sharing scholarly data on the web? In 2016, the Karl-Franzens-University in Graz hosted a conference titled *Digital Scholarly Editions as Interfaces.*[25] Dot Porter, Curator of Digital Research Services in the Schoenberg Institute for Manuscript Studies at the University of Pennsylvania and co-creator of the *OPenn* project, advocated prioritizing data over interface in digital editions. "Excellent, robust data with no interface isn't easily usable (although a creative person will always find a way)," noted Porter in her keynote, "but an excellent interface with terrible data or no data at all is useless as anything other than a show piece."[26] This call to prioritize data in digital editions has the potential to make academic projects far more interoperable, but is the academy ready to recognize these stripped-down digital editions as credible scholarly outputs?

22 See https://omeka.org (Accessed: 26 June 2024).
23 See https://scalar.me/anvc/scalar (Accessed: 26 June 2024).
24 See https://teipublisher.com (Accessed: 26 June 2024).
25 See https://informationsmodellierung.uni-graz.at/en/department/events/archive/digital-scholarly-editions-as-interfaces (Accessed: 26 June 2024).
26 See http://www.dotporterdigital.org/what-is-an-edition-anyway-my-keynote-for-the-digital-scholarly-editions-as-interfaces-conference-university-of-graz (Accessed: 26 June 2024).

The Digital Humanities community has pressed for recognition of formats beyond the article and the book as legitimate forms of scholarship. During the golden age of blogging in the aughts, scholars debated what weight to give blog writing in scholarly assessment and tenure review.[27] For the past decade, microblogging has taken over from blogging as an informal but essential running dialogue within the *Digital Humanities* research community. For better or worse, Twitter (now X) has served as the primary locus of this conversation. Ernesto Priego, Senior Lecturer in the Department of Computer Science at University College London, has analyzed tweets associated with the major Digital Humanities conferences since 2010.[28] As with many disciplines these days, following the right hashtags keeps you current with cutting-edge scholarship. But does it also promote a scholarly superficiality that favors, in the words of Lewis (2014), "the New New Thing?"

7. Literate Programming

The central idea behind literate programming is simple: optimize code for readability rather than conciseness. In contemporary Computational Humanities, literate programming frequently takes the form of code notebooks. A code notebook is a device that holds explanatory narrative and executable code together. A code notebook offers two kinds of cells: code cells and text cells. The code cells provide an environment for executing code and, if relevant, for displaying the output of the computation. The text cells, by contrast, contain non-executable information, generally a narrative about how the code blocks function or the purpose of their computation.

The most popular form of code notebook is called a *Jupyter notebook.* Since these notebooks were developed in the *Python* programming community,[29] *Python* is the default choice of kernel, or programming environment, for *Jupyter.* But *Jupyter* is fully extensible, and kernels exist for many other languages, including *R* and *SQL,* among many others. Google has popularized *Jupyter* by offering its *Colaboratory* (or *"Colab"*) service;[30] *Colab,* which offers both free and paid service, connects to cloud servers on the *Google Cloud Platform* (GCP), eliminating the need to set up hosting either locally or remotely for *Jupyter. Colab* has proved immensely popular among artificial intelligence specialists. These days, the "Open in *Colab*" button regularly appears alongside

27 See https://www.science.org/content/article/science-blogging-and-tenure (Accessed: 26 June 2024).

28 See https://ernestopriego.com/2019/07/15/dh2018-and-dh2019-twitter-archive-counts-a-comparison (Accessed: 26 June 2024).

29 See https://cs.lbl.gov/news-media/news/2021/project-jupyter-a-computer-code-that-transformed-science (Accessed: 26 June 2024).

30 See https://colab.research.google.com (Accessed: 26 June 2024).

papers, encouraging readers to give the code a spin and try their new machine-learning techniques for themselves.

Not all code notebooks work the same way. *Observable*, for instance, uses *JavaScript* rather than *Python* as its primary coding language.[31] Whereas *Jupyter notebooks* execute code sequentially from top to bottom, *Observable notebooks* execute code topologically; that is, code cells update whenever any of the cells they depend on change.[32] Technically, this paradigm is termed functional reactive programming, but the way *Observable* works in practice is similar to a spreadsheet. For example, if you change a cell in a spreadsheet, your calculations that depend on that cell will update accordingly. *Observable* works similarly, responding to updates in both code and data dynamically. Given digital humanists' ubiquitous use of spreadsheets, the *Observable* paradigm may feel more familiar than the top-to-bottom execution style of *Jupyter notebooks.*

The move toward literate programming with code notebooks has not gone without criticism. While notebooks provide an ideal environment for exploratory data analysis, they suffer from bloat and complexity as they increase in size. From a pedagogical standpoint, students may pick up problematic coding habits from notebooks (Johnson 2020). But, despite the criticisms, code notebooks have found their place as essential complements to scholarly papers in computer science and specific sectors of the Digital Humanities.

8. Quantification of Theological Knowledge

The transformation from analog to digital research methods inevitably disrupts our approach to scholarly analysis. To play on Wittgenstein's adage, Digital Humanities "does not leave everything as it is." The shift to computational methods in theology carries along a host of other techniques, including programming, data management, and statistical modeling. A trope about data science is that it exists at the intersection of computer science, statistics, and disciplinary knowledge. The problem, to return to Daniel Allington, is that almost nothing of computer science or quantitative methods is taught during seminary or Ph.D. programs in the humanities, leading to his sweeping judgment, "there is scant prospect for the development of quantitative methods in primarily hermeneutic disciplines such as literature, philosophy, or theology (Allington 2022, 381)." As a generalization, this is essentially correct, but there are exceptions, both at the curricular and the (sub-)disciplinary level.

What is the solution to this lack of background knowledge among theologians and scholars of religious studies?

31 See https://observablehq.com (Accessed: 26 June 2024).

32 See https://observablehq.com/observablehq/how-observable-runs (Accessed: 26 June 2024).

On the one hand, programs like *Software Carpentry*[33] and *Data Carpentry*[34] have emerged as boot camps to bring graduate students and scholars up to speed with data-intensive programming. These carpentries provide crash courses on *git,* the *Unix shell, Python,* and *R* programming. In Digital Humanities circles, intensive summer programs, like the *Digital Humanities Summer Institute* (DHSI) at the University of Victoria, have trained cohorts since 2001 in techniques ranging from textual encoding to network analysis. These boot camps evince mixed success. While they orient newcomers to the toolsets of the Digital Humanities, they exude a similar misleading appeal as language courses pitched to business travelers: promising functional proficiency after only a couple of weeks of study. Your actual mileage may vary. On the statistics side, there are fewer opportunities to develop fluency. Arnold & Tilton (2019, 293) have purported to "show how statistics – the organization, analysis, interpretation, and presentation of data – is a fundamental interlocutor and methodological approach for the Digital Humanities." But who will teach faculty the requisite mathematics and train them in data analysis?

The solution to this quandary is generally not for religious studies scholars to develop computer science and statistics expertise. The path toward gaining facility in these disciplines is long and inevitably involves passage through mathematical fields like calculus and linear algebra. It is safe to say that we will not see these subjects appearing in seminary curricula in the foreseeable future.

> No one person can be expected to understand such a wide range of disciplines brought together in the digital humanities in the depth required to develop the innovative insights and methods that are the promise of the field. Rather, the digital humanities should welcome statistics to the table, and this begins by better acknowledging the critical role of statistics in the field (ibid., 298).

The same goes for software engineers, project managers, metadata specialists, and DevOps experts. Building a Digital Humanities project requires a team or, to use the nomenclature of the natural sciences, a lab. We see movements in this direction in the Digital Humanities, but the organizational apparatus remains challenging to muster. As we shift from conventional research environments to Computational Humanities (and theology), the emergence of these labs may be the most tangible sign of the transformed research environment.

33 See https://software-carpentry.org (Accessed: 26 June 2024).

34 See https://datacarpentry.org (Accessed: 26 June 2024).

9. From Saturation to Datafication

In this essay, we surveyed a series of digital surrogates for analog research and publication methods. While we have covered much territory, we've only just explored the landscape. If you are feeling overwhelmed by the options, you are not alone. But how do digital humanists deal with media saturation?

The shift in emphasis from interface to data points toward one path forward. In *On the Existence of Digital Objects,* Yan Hui explores the being of digital objects in dialogue with Husserlian phenomenology and post-Husserlian ontology. As he remarks near the outset of his study, "the term 'digital object' remains ambiguous here, because the vast quantity of digital objects are comparable in breadth and diversity to the vast array of animal species (Hui 2016, 48)." We interact with increasing numbers of digital objects, running the gamut from Instagram posts, tweets, Google Docs, file folders, Android or iPhone apps, and so on. These objects multiply every time a new medium comes into existence, inflating our digital ontologies. Putting the point more familiarly, whenever we download a new app on our phones or try out a new educational technology, we need to familiarize ourselves with its way of cutting up its digital domain. As the channels proliferate, we tend to experience a diminution of interest. "How do we share a post in Mastodon again? How can I change the order of my feed in Threads?" In the language of media studies, our digital media environment has become saturated; adding novel technologies to our academic workflows threatens to sap energy rather than accelerate our scholarship.

In the Digital Humanities, scholars have responded to the saturation of our academic media landscape by moving away from digital interfaces toward data, metadata, and code. There remains tremendous diversity at this level of scholarly computing as well. But, as Yuk Hui suggests, the operating ontologies of Computational Humanities – and, by extension, Computational Theology – become perspicuous when we strip away the interfaces and examine the code, data, and metadata that animate them behind the scenes (ibid., 26). Still, the shift to the command line or code notebook threatens to exclude scholars who lack the time, resources, or inclination to explore the arcane languages and protocols that animate it. As a remedy to media saturation, this kind of computational reductionism works, but it may also narrow the field. How many digital humanists have the capacity, or inclination, to retool completely their scholarship around the software engineering toolchain?

In the near term, most theological scholars will simply pick and choose academic tools from conventional (analog) and new (digital) media. They will continue to borrow books from the library while downloading PDFs online. They will write marginal comments in their texts while keeping track of key dates in Excel documents. And they will publish monographs with university presses while showcasing their projects in multimedia digital editions. In other words, the digital will not displace the analog but supplement it. The question is whether this mixed media approach will persist. Will we look back on this era with nostalgia for the books, pencils, file cards,

and notebooks that have since vanished from our scholarly environment? Or will we find some efficient equilibrium that blends the best of analog and digital into a perduring scholarly synthesis?

References

Allington, D. (2022). The Place of Computation in the Study of Culture. In J. O'Sullivan (Ed.), *The Bloomsbury Handbook to the Digital Humanities* (pp. 373–384). London: Bloomsbury Publishing.

Arnold, T., & Tilton, L. (2019). New Data? The Role of Statistics in DH. In M.K. Gold & L. Klein (Eds.), *Debates in the Digital Humanities* (pp. 293–299). Minneapolis: University of Minnesota Press.

Blair, A.M. (2011) *Too Much to Know. Managing Scholarly Information Before the Modern Age.* New Haven: Yale University Press.

Bryant, J., & Fondren, W. (2009). Displacement Effects. In R.L. Nabi & M.B. Oliver (Eds.), *The Sage Handbook of Media Processes and Effects* (pp. 505–517). Los Angeles: Sage.

Clark, A., & Chalmers, D. (1998). The Extended Mind, *Analysis,* 58(1), 7–19. DOI: https://doi.org/10.1093/analys/58.1.7 (Accessed: 26 June 2024).

Gitelman, L. (2014). *Paper Knowledge. Toward a Media History of Documents.* Durham: Duke University Press Books.

Glazer, N. (1995). A Man Without Footnotes. In Ch. DeMuth & W. Kristol (Eds.), *Neoconservative Imagination. Essays in Honor of Irving Kristol* (pp. 3–12). Washington, D.C: AEI Press.

Hassan, R. (2023). *Analog.* Cambridge: The MIT Press [= *The MIT Press Essential Knowledge Series*]. DOI: https://doi.org/10.7551/mitpress/13644.001.0001 (Accessed: 26 June 2024).

Hui, Y. (2016). *On the Existence of Digital Objects.* Minneapolis: University of Minnesota Press.

James, R. (2010). An Assessment of the Legibility of Google Books, *Journal of Access Services,* 7(4), 223–228. DOI: https://doi.org/10.1080/15367967.2010.503486 (Accessed: 26 June 2024).

Johnson, J.W. (2020). Benefits and Pitfalls of Jupyter Notebooks in the Classroom. In *Proceedings of the 21st Annual Conference on Information Technology Education* (pp. 32–37). New York: Association for Computing Machinery. DOI: https://doi.org/10.1145/3368308.3415397 (Accessed: 26 June 2024).

Kirschenbaum, M.G. (2016). *Track Changes. A Literary History of Word Processing.* Cambridge: Belknap Press.

Lewis, M. (2014). *The New New Thing. A Silicon Valley Story.* New York: W.W. Norton & Company.

Lockridge, T. (2020). *Writing Workflows. Beyond Word Processing.* Michigan: University of Michigan Press.

Mailund, Th. (2019). *Introducing Markdown and Pandoc. Using Markup Language and Document Converter.* Berkeley, CA: Apress. DOI: https://doi.org/10.1007/978-1-4842-5149-2 (Accessed: 26 June 2024).

Marcum, D., & Schonfeld, R.C. (2021). *Along Came Google. A History of Library Digitization.* Princeton: Princeton University Press.

Nelson, T. (1981). *Literary Machines.* Sausalito: Mindful Press.

Newell, J., Pilotta, J.J., & Thomas, J.C. (2008). Mass Media Displacement and Saturation, *International Journal on Media Management,* 10(4), 131–138. DOI: https://doi.org/10.1080/14241270802426600 (Accessed: 26 June 2024).

Pfitzinger, S.G. (2011). The UPs and Downs of Citation Titles. What Librarians Should Know When Helping Students, *Indiana Libraries,* 30(2), 25–29.

Risam, R., & Gil, A. (2022). Introduction. The Questions of Minimal Computing, *Digital Humanities Quarterly,* 16(2), 1–34. URL: http://www.digitalhumanities.org/dhq/vol/16/2/000646/000646.html (Accessed: 26 June 2024).

Roddy, K.P. (1986). Once More into the Breech. Computer Literacy and the Humanities, *Historical Social Research,* 40, 91–95. URL: https://www.jstor.org/stable/20755058 (Accessed: 26 June 2024).

Schatz, B.R. (1997). Information Retrieval in Digital Libraries. Bringing Search to the Net, *Science,* 275(5298), 327–334. DOI: https://doi.org/10.1126/science.275.5298.327 (Accessed: 26 June 2024).

Wasiak, J. (2008). Epistemological Implications of Media Saturation, *The International Journal of Technology, Knowledge, and Society,* 4(2), 113–118. DOI: https://doi.org/10.18848/1832-3669/CGP/v04i02/55862 (Accessed: 26 June 2024).