

The Great Digital Humanities Disconnect: The Failure of DH Publishing

Emily Öhman
Waseda University
ohman@waseda.jp

Michael Piotrowski
University of Lausanne
michael.piotrowski@unil.ch

Mika Hämäläinen
Metropolia University of Applied Sciences
mika.hamalainen@metropolia.fi

Abstract

We discuss the disconnect in interdisciplinary publishing from a disciplinary divide perspective as to how research is expected to be presented and published according to disciplinary conventions. We argue that this divide hinders interdisciplinary collaboration and even more so the dissemination of research results from interdisciplinary projects to other interdisciplinary researchers. The disconnect is not simply theoretical but also encompasses practical considerations such as manuscript creation standards. The disconnect can also be detrimental to academic careers in terms of evaluations by peers on funding and tenure committees as well as peer reviews. With this analysis, we want to foster further discussion about the state of academic publishing from a digital humanities perspective.

1 Introduction

Different academic disciplines have different cultures and traditions, notably different standards and expectations for what constitutes acceptable research within the discipline, but also how this research is to be presented and published. This, in turn, defines what constitutes a good “track record” for researchers, which determines their career prospects (hiring, tenure, and promotion) in this discipline. In fact, the definition and assurance of quality standards could be said to be the main purpose of disciplines, their *raison d’être*. These standards are enforced through institutions that are able to make decisions about funding, hiring, teaching, etc. On this level, all disciplines are effectively in competition, and cultures and traditions thus also have important social roles that may have little to do with the actual quality of research and more with staking out claims in order to influence the allocation of resources (see, e.g., [Becher and Trowler, 2001](#)).

Although the cultures and traditions of established disciplines can be arcane and difficult to navigate (absorbing the discipline’s culture is an important, though mostly implicit, part of the training in a discipline), the expectations are—in principle—known or at least knowable. However, the picture changes radically when we consider *interdisciplinary* research, and even more so if the work is not just a clearly delimited collaboration of researchers from two or more disciplines, but literally “between” disciplines, i.e., in a kind of academic no man’s land that is not ruled by any discipline.

In both cases, researchers are confronted with and need to be aware of the publishing conventions of different disciplines if they want to benefit from their work: they need to ensure that it is visible in the right community, and that is cited or otherwise acknowledged, e.g., for obtaining tenure. The phenomenon is well known as “publish or perish” ([Hammarfelt and De Rijcke, 2015](#)).

Interdisciplinary researchers are also confronted with practical issues when considering publication venues. When language technology researchers on [Reddit¹](#) were asked the question: “Would you submit a paper to a venue that did not allow LaTeX/demanded .doc?” specifying the focus as interdisciplinary research ([askja, 2023](#)) one scholar commented:

It depends on the “weight” of the journal. If that’s a journal where I’d plan to do some research work explicitly targeting that journal, then requiring .doc would be inconvenient but not that important. But if that is research work which I’d do and then, when it’s mostly done, see what’s the best place for it – then by the time I’d consider that journal, the content would be already mostly written in LaTeX and then I’d be hesitant to rewrite it in Word.

¹<https://www.reddit.com/>

The comment echoes two challenges for the dissemination of interdisciplinary research. The question of where to submit one’s work seems obvious and largely unavoidable in interdisciplinary research. The choice of the writing tool, however, seems to be merely a practical question; yet in the case of digital humanities as a field between computing and the humanities, it tends to be a question that goes far beyond personal preferences or technical merits. The question of \LaTeX vs. Microsoft Word can be considered a “fault line” in digital humanities, which aligns with numerous other divisions inside the supposed “big tent.”

“Big Tent Digital Humanities” was the theme of the DH 2011 conference, and the “big tent” metaphor has been used since to emphasize the diversity, openness, inclusiveness, and fluidity of digital humanities. While well-meaning, already at the time some scholars noted potential problems with this notion. For example, [Svensson](#) cautioned that “[e]ven if the big-tent vision of the digital humanities gives the field a sense of openness and invitation, it does not necessarily remove institutional predispositions and thresholds or make the field into a blank slate” ([Svensson, 2012, 47](#)); he remarked that, consequently, “there is a risk that a wealth of traditions and perspectives are subsumed and conflated in a tent primarily keyed to one particular tradition” ([Svensson, 2012, 45](#)).

This seems in fact to be the case—otherwise, why highlight the fact that the Computational Humanities Research conference,² established in 2019 by a group of scholars who do computationally oriented work in DH, has adopted “practices that align with norms in computer science and linguistics (e.g., submission of 6- or 12-page papers, exclusive use of \LaTeX)” ([Dombrowski, 2023, 138](#)) for their conference?

In this position paper, we will explore the disconnect between these two flavors of DH from a research dissemination perspective. In addition to discussing the difficulties in interdisciplinary research and how the lack of disciplinary coherence reflects on the evaluation of research outputs, we focus on some practical issues within the publication disconnect: what open science means for the different disciplines involved including “publish or perish”, the authorship and readership of different publication venues, document processing, and fi-

nally what the impact will be for digital humanities as a field going forward.

2 Disciplinary Backgrounds

As most publication venues are associated with a specific discipline, the work to be published will be evaluated according to the standards of this discipline, both with respect to the research *and* the formal requirements concerning its presentation. But as interdisciplinary work is not, or at least not entirely, within a single discipline, this raises significant problems for the evaluation of the work—what constitutes appropriate peer review for interdisciplinary research and who is qualified to judge it? ([Bammer, 2016](#); [McLeish and Strang, 2016](#))—and it has been shown that interdisciplinary research has consistently lower funding success ([Bromham et al., 2016](#)). Additionally, the lack of disciplinary conventions means that it can in practice be difficult to find reviewers with appropriate disciplinary knowledge of all the fields involved, which can lead to interdisciplinary research being unfairly evaluated from the perspective of only one of the involved fields, and the difficulty in recruiting appropriate reviewers can often delay the publication process further.

In traditional humanities disciplines, the main research output is journal papers of 15–25 pages (or even longer), edited volumes, and books. It takes a long time to write such manuscripts and it can take years from submission to publication. Typically, there is no document template for submission or publication; instead, there are extensive lists of margin sizes, comma use, capitalization rules, bibliographic guides, and many more submission guidelines. The submission is in the form of a .doc(x) file. Conferences in the humanities are primarily networking events, where scholars present work in progress. These conferences tend to only have abstract submissions (a typical limit is 500 words), which may or may not be published in a book of abstracts.

In language technology, natural language processing (NLP), and computational humanities, the prestige of publication venues is often reversed, with conference proceedings being the main and most prestigious venue for disseminating information; publication in journals is generally considered too slow and low-impact. There are often two categories of conference papers: short papers (4 to 6 pages) and long papers (8 to 12 pages).

²<https://computational-humanities-research.org/>

The papers are indexed and double-blind peer-reviewed. The delay between submission and publication is usually between 3 to 6 months, depending on the specific venue. In these more computational disciplines, L^AT_EX is overwhelmingly used for submissions—these days, a direct link to a collaborative template on Overleaf³ is even typically provided. Nevertheless, Word templates are often also available.

Preprints are strongly encouraged and sometimes required in computational fields, further speeding up the information dissemination process. As an example, arXiv, the main preprint service for computer science and other STEM fields received 20,170 submissions in October 2023. Another study found that of arXiv preprint papers submitted in 2017-2018 77% were later published in peer-reviewed venues (Lin et al., 2020). However, preprints in the humanities are mostly frowned upon (Laporte, 2017) and if not outright forbidden it is common that referencing preprints is discouraged in humanities journals with only 45% of humanities journals allowing preprints and nearly all computer science journals allowing them (Klebel et al., 2020).

Digital Scholarship in the Humanities and *Digital Humanities Quarterly* could be considered two of the main journals within DH. However, the articles published in them lean heavily towards digitization (roughly 80%) rather than computational humanities, and much of the actual computational content is limited to stylometrics (Roth, 2019; Piotrowski, 2020). Digitization, especially of cultural heritage, is certainly a part of digital humanities, but the imbalance is striking. Where are the computational humanities papers published, if not in these journals? Are we perhaps after all not all in the same “big tent” despite disciplinary gaps?

3 The Disconnect

We argue that there is a growing theoretical and practical disconnect in DH due to widespread hostility towards certain types of computational approaches and other customs from computational fields. As a result, current DH can to a large extent be characterized as pseudo-interdisciplinary: it is in fact largely dominated by traditional humanities practices, in particular with respect to publishing, which, whether consciously or subconsciously, tend to exclude interdisciplinary researchers from

fields such as computer science, computational linguistics, and even computational humanities.

3.1 Open Science

Most academics in any discipline would agree that *Open science* is a good thing allowing everyone access to research results and makes these results more transparent. However, for most humanities scholars open science in practice tends to be limited to paying open-access journal publication fees. Corpora and digitized datasets in the humanities are often locked behind online interfaces or even CD-ROMs, perhaps mostly due to copyright issues, but undeniably also disciplinary culture. On the other hand, in every aspect of computational research, it is highly encouraged for everything to be made publicly available from preprints, to data, and code, and this openness and accessibility is often a part of the peer-review criteria. In combination with the different approaches to the analysis of results, this leads to a situation where the more analysis-focused results rarely face methodological scrutiny from reviewers as it is not expected or possible in most cases due to lack of access to code and data.

3.2 Publish or Perish in a Faux-Interdisciplinary Context

It is unfortunate that many researcher are faced with the choice of “publish or perish,” as this leads to insular publication practices. It also means researchers feel compelled to pump out articles at a pace at which it is difficult to maintain academic rigor. “Publish or perish” is an issue in all disciplines; however, the publishing disconnect between digitized and computational humanities (i.e., within the “big tent” of digital humanities) exacerbates the problem in DH.

Universities and academic journals both contribute to the pervasive culture of “publish or perish.” Facing budgetary pressures, institutions depend on prestige to attract research funding, and one of the easiest ways to increase prestige (as measured by rankings) is to be highly visible in prestigious journals.

It is easy to dismiss “publish or perish” as an old aphorism that academics use to complain about their working conditions, but there is ample evidence (e.g., the replication crisis) that the longer this unhealthy pressure persists, the greater the risk to research integrity. As the researchers start to suffer, and the cracks begin to appear, we can see real consequences: in an attempt to increase publi-

³<https://www.overleaf.com/>

cation metrics, researchers split up project results into “minimum publishable units,” when one paper would have sufficed, join each others’ publications as co-author, publish only research with positive results, or even resort to forgery. Not all of these practices are necessarily bad as such, and splitting up papers into multiple papers can help clarify specific contributions and increase the citability of a paper, however, when viewed as a whole or as a method to “game the system” the ethics of such practices become murky.

3.3 Disciplinary and Interdisciplinary Publication Venues

Most academics are vaguely aware of the fact that publication practices vary by field. However, many are woefully ill-equipped to evaluate the publication record of someone from a different field. Computational sciences expect rapid publication of preprints on the one hand, and on the other end of the spectrum, humanities scholars are expected to take years to write whole books. When these two disciplines come together, which disciplinary background is used to evaluate research output and results?

The choice of publication venue is an important one as it decides not only where your research will be published, but also who will review and who will read about it. This means that journals that call themselves DH but only accept manuscripts structured and created using standards from one end of the interdisciplinary spectrum will be overlooked by those closer to the other end.

3.4 Document Standards as Gatekeeping and Virtue Signaling

The state of the art in scholarly publishing (even when only considering the technical aspects) is appalling. When PDF output is required, \LaTeX remains more or less the only comprehensive authoring solution. Writing a paper for, say, an ACM⁴ or ACL⁵ conference is easy: there are official document templates, you literally just have to write your paper. The point here is not that \LaTeX is “better,” but rather that there is a clearly defined path to the submission, and authors do not have to concern themselves with the formatting of the document or the references: this is all taken care of automati-

⁴<https://www.acm.org/publications/proceedings-template>

⁵https://2023.aclweb.org/calls/style_and_formatting/

cally. This also means that no conversion and no manual interventions are required, which tend to introduce errors. In addition, the system is open source and highly portable, you can use it on any platform and with any editor you want.

In the humanities, however, an expensive license for Microsoft Word is usually required and authors have to manually adjust numerous settings and manually ensure that their submission conforms to the guidelines. Thus, although \LaTeX is older, it is much better suited for modern publication practices, including automatic compilation of the final product (proceedings, journal). This is not to say that it is perfect, but it is perhaps the best we have at the moment. \LaTeX comes with its own host of issues, specifically since the end product, a PDF document, suffers from loss of semantic information that is only available in the source code (Piotrowski, 2016). It has been noted that the requirement to submit a Word document might in some cases simply be due to the editors being unaware of the possibility of \LaTeX submissions and a request to submit using \LaTeX might be granted especially since all major publishers provide \LaTeX templates including Springer, Elsevier, Wiley etc. (Jensen, 2018). Nonetheless, none of the purely DH journals, and almost none of the DH conferences offer a \LaTeX option despite this undoubtedly being an issue that has been raised by contributing authors. \LaTeX templates reduce the workload of both editors and authors so it seems strange to deny contributions written in \LaTeX .

XML formats have many potential advantages, but the use of TEI (Text Encoding Initiative) in some DH contexts seems akin to virtue signaling and is far from practical, useful, or even in the spirit of TEI: the conferences that use TEI do *not* actually accept submissions in TEI, but require authors to use the DH Convalidator tool, which converts Word documents to TEI format, so the paper must be written in Word, and it cannot contain more information than available in Word, and it is not even published as a TEI document, but as PDF. Effectively, the semantic information painstakingly extracted from graphically typeset texts (.docx) is eventually completely erased. This marks the other end of the spectrum of the art of digital publishing (Cremer, 2018). The *Other* category in table 1 almost exclusively refers to these types of submissions.

There is a widespread fear of \LaTeX in the humanities; this is to some extent understandable,

Venue	Type	LaTeX	.docx etc	Other	Abstract only
Association for Computational Linguistics (ACL)	Umbrella for NLP conferences	O	O	X	X
Digital Humanities Quarterly	DH journal	X	O	O	N/A
Digital Scholarship in the Humanities	DH journal	X	O	O	N/A
Digital Humanities	DH conference	X	O	X	O
European Association for Digital Humanities	DH conference	X	O	O	O
Digital Humanities in the Nordic and Baltic countries	DH conference	O	O	X	X/O
Computational Humanities Research	CH conference	O	X	X	X
Journal of Data Mining & Digital Humanities	DH/NLP journal	O	O	O	N/A
International Journal of Digital Humanities	DH journal	O	O	X	N/A
Humanist Studies & the Digital Age	DH journal	X	O	X	N/A
Journal of Digital History	DH Journal	X	X	O	N/A

Table 1: ACL includes all ACL-affiliated conferences such as EACL, EMNLP, CoLING, etc. as well as co-located workshops. Abstract-only means that only a book of abstracts will be published. "Other" almost exclusively refers to the use of the DH Convalidator tool, except for DHQ where direct XML/TEI submissions are possible, and the Journal of Digital History which only accepts Python Notebooks.

even though it takes no more than 30 minutes to learn the basics of \LaTeX (or a tool like Pandoc⁶), as it is not part of traditional humanities curricula and one study found that just over 20% of humanities scholars were comfortable using TeX or other markup languages including XML and TEI (Bonn and Swatscheno, 2017). In that particular instance, the finding led to a decision to only accept Word documents.

The hostility against \LaTeX (or, in fact, anything that is *not* Word) in large parts of DH—a field that commonly describes itself as located at the intersection between computer science and the humanities and that prides itself on its interdisciplinarity, inclusiveness, and progressiveness—is a different story, though. One possible explanation is that it, like the refusal to define DH, serves a gatekeeping function: as Piotrowski (2020) argues, humanities scholars that wield DH as “a term of tactical convenience” (Kirschenbaum, 2014) to obtain a vanguard status in their discipline are natural wary of potential intruders that could strip them of this status.

Figure 1 presents an overview of different digital humanities and computational humanities venues and the file types expected of submissions. All computationally oriented conferences (ACL, ACM, CHR) accept \LaTeX submissions and provide templates - most of them provide templates for Word as well. On the other hand, very few DH venues accept \LaTeX . Notable exceptions include the Journal of Data Mining and Digital Humanities which requires preprint submissions which means any submission type is acceptable - and later provide both \LaTeX and Word templates, the Digital Humanities in the Nordic and Baltic Countries conference

which accepts both and allows both abstract-only and short and long paper submissions. Noteworthy are also the Digital Humanities Quarterly which does not accept \LaTeX but accepts XML and TEI submissions **not** generated with the DH Convalidator tool, and the Journal of Digital History, which only accepts .ipynb notebooks using their template. Most DH conferences do not publish proceedings beyond a book of abstracts, whereas no computational conference allows abstracts with the exception of lightning talks at the Computational Humanities Research conference.

4 Concluding Remarks

This position paper is a call to action. We believe that there is no future for DH if this disconnect is not addressed. Instead of waiting for the “big tent” to collapse—to the detriment of everybody who is in it—we should work to establish *computational humanities* as a discipline in its own right, that sets its own standards and evaluation criteria. In the end, this will be the only way to ensure adequate recognition of computational research in the humanities. This does not have to mean the collapse of digital humanities, but instead a strengthening of the position of digital humanities as a field separate from traditional humanities with appropriately adjusted evaluation criteria and mutually agreed upon publication practices that are neither cumbersome nor slow to use or publish.

References

- askja. 2023. [Reddit post](#).
- Gabriele Bammer. 2016. [What constitutes appropriate peer review for interdisciplinary research?](#) *Palgrave Communications*, 2(1).

⁶<https://pandoc.org/>

- Tony Becher and Paul R. Trowler. 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Discipline*, 2nd edition. Open University Press, Buckingham, UK.
- Maria Bonn and Janet Swatscheno. 2017. Humanities without walls: Understanding the needs of scholars in the contemporary publishing environment.
- Lindell Bromham, Russell Dinnage, and Xia Hua. 2016. [Interdisciplinary research has consistently lower funding success](#). *Nature*, 534(7609):684–687.
- Fabian Cremer. 2018. [Nun sag, wie hältst Du es mit dem Digitalen Publizieren, Digital Humanities?](#)
- Quinn Dombrowski. 2023. Does coding matter for doing digital humanities? In James O’Sullivan, editor, *The Bloomsbury Handbook to the Digital Humanities*, chapter 13, pages 137–145. Bloomsbury, London.
- Björn Hammarfelt and Sarah De Rijcke. 2015. Accountability in context: Effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the Faculty of Arts at Uppsala University. *Research Evaluation*, 24(1):63–77.
- Lars Christian Jensen. 2018. [Guest blog post: Latex for the humanities](#).
- Matthew G. Kirschenbaum. 2014. [What is “Digital Humanities,” and why are they saying such terrible things about it?](#) *differences*, 25(1):46–63.
- Thomas Klebel, Stefan Reichmann, Jessica Polka, Gary McDowell, Naomi Penfold, Samantha Hindle, and Tony Ross-Hellauer. 2020. Peer review and preprint policies are unclear at most major journals. *PLoS One*, 15(10):e0239518.
- Steven Laporte. 2017. Preprint for the humanities—fiction or a real possibility? *Studia Historiae Scientiarum*, 16.
- Jialiang Lin, Yao Yu, Yu Zhou, Zhiyang Zhou, and Xiaodong Shi. 2020. How many preprints have actually been printed and why: a case study of computer science preprints on arxiv. *Scientometrics*, 124(1):555–574.
- Tom McLeish and Veronica Strang. 2016. Evaluating interdisciplinary research: the elephant in the peer-reviewers’ room. *Palgrave Communications*, 2(1):1–8.
- Michael Piotrowski. 2016. [Future publishing formats](#). In *Proceedings of the 2016 ACM Symposium on Document Engineering, DocEng ’16*, page 7–8, New York, NY, USA. Association for Computing Machinery.
- Michael Piotrowski. 2020. [Ain’t no way around it: why we need to be clear about what we mean by “digital humanities”](#). SocArXiv.
- Camille Roth. 2019. [Digital, digitized, and numerical humanities](#). *Digital Scholarship in the Humanities*, 34(3):616–632.
- Patrik Svensson. 2012. [Beyond the big tent](#). In Matthew K. Gold, editor, *Debates in the Digital Humanities*, pages 36–72. University of Minnesota Press.