(DLH) The Journal for Digital Legal History Open Access

Publishing license: https://creativecommons.org/licenses/by/4.0/ CC-BY 4.0

ETitle:	Editorial: *Is Sharing Research Materials Worth the Effort?*
Author(s) and contributors:	Gijs van Dijck
Issue:	1
Year:	2025
DOI:	*10.21825/dlh.99809*
Keywords:	*Open Science, FAIR, digital legal history*
Abstract:	This editorial discusses the considerations for sharing research materials in the domain of Digital Legal History.

(JLH) The Journal for Digital Legal History Open Access

Is Sharing Research Materials Worth the Effort?

Gijs van Dijck

Maastricht Law and Tech Lab, Maastricht University

In an editorial for another journal, I discussed the advantages and disadvantages of Open Access (OA) publishing in the legal domain (Van Dijck, 2025). OA publishing concerns not only books and articles but also data and research materials. In this editorial, I will explore the benefits and drawbacks of publishing and sharing research materials, particularly in the domain of (digital) legal history.

I tend to publish and share my data as much as possible. This, I believe, makes sense when working in a domain where data is scarce, data collection is expensive, and many other researchers can and do reuse the data. At the same time, releasing the data takes time and effort. Is it worth the time and effort to prepare and share data in digital legal history? I started wondering about this question when co-supervising a Ph.D. student who doctrinally and empirically analyzes 18th-century case law from the Netherlands. To my knowledge, there are no other researchers in the Netherlands who specialize in this domain. It might, therefore, well be that the Ph.D. student and the other supervisor are the only two researchers in this universe who, at least in the foreseeable future, study this type of data. Is sharing 18th-century Dutch case law data and their materials worth the investment?

The benefits of digitizing historical legal datasets are evident. It preserves fragile archival material, reduces the effort required for transcribing texts, and enables analyses across time, comparisons across jurisdictions, and integration with other contextual information to better understand certain practices. As in other domains, AI-assisted tools enable the detection of patterns and trends that might otherwise go unnoticed, support collaboration across disciplines, and prevent repetitive, labor-intensive data collection efforts. Immediate demand may be limited, but this may change over time. Such collections have the potential for educational benefits, scientific advances, and the transformation of archival materials into accessible cultural and historical exhibits.

But what about annotated data, notebooks developed over the course of the research, and other materials that capture the insights, decisions, and workflows of the study? There is significant documentation on how digital and computational methods have shaped *how* researchers can or should make their materials available (Burrows, 2023). There are also guidelines available for doing this (e.g., Black, 2021; Blaney et al., 2021). Is it worthwhile to share these as well? A brief cost-benefit analysis:

The biggest drawback is time. Producing a shareable annotated dataset, for example, can be almost as labor-intensive as the research itself. This particularly applies to historical legal documents, which can be messy, may require OCR scanning and review, and can pose difficulties when converting them to commonly used formats such as HTML, TEI, or CSV. There is also considerable 'overhead', as README files must be prepared, dependency lists created, and code comments clarified. In this sense, materials can be technically "open" but practically unusable. Also, digital solutions age quickly. Without maintenance and

nurturing, what is open and available today can become closed tomorrow. And what becomes open runs the risk of being crawled and used as training data for LLMs.[1]

Furthermore, research materials are created with the research questions the researcher aims to answer in mind. The network the researcher built and analyzed, the entities they were interested in, or the annotations they created may not align with what others find relevant or necessary. The materials, such as text annotations or citations identified in the text, may also be subject to bias, particularly in a domain with few researchers to verify the data and test interrater reliability. Moreover, sharing materials is only useful if others can understand them. Notebooks often lack explanations and structure, making them difficult to rerun (Rule et al., 2019).

An important reason *for* sharing materials is that they can serve as a catalyst for future research. Document annotation enhances machine-readability and improves the corpus's retrieval and searchability for legal scholars, computer scientists, historians, and others. Off-the-shelf NLP models often underperform on historical texts as well as on legal language. Consequently, community-shared gold-standard annotations are invaluable. In this way, notebooks may at a minimum serve as tutorials or templates for other researchers. A Jupyter notebook that walks the reader through creating a citation network or training a classifier to detect lawyers, locations, names, and dates can serve as inspiration for others, not only for the specific corpus but also for other corpora.

A second important reason concerns interoperability. Annotated collections can be linked to other resources to support richer analyses. For instance, the Italian Digesto project aims to convert Justinian's Corpus Juris into TEI and Linked Open Data formats, enabling researchers to query and compare texts across languages (Cinini et al., 2025). Thirdly, shared notebooks and other research materials can serve as educational resources for students and researchers (novice or experienced). A well-documented Jupyter notebook can become an interactive environment, or even a narrative for exploring or replicating the data, or a way to make complex workflows accessible (Rule et al., 2019). Research materials can also serve as teaching benchmarks, providing examples for students. They can help lower the barriers to learn adopting digital methods in legal history. Finally, for the researcher, sharing materials enhances or ensures reproducibility, which both strengthens the credibility of the research and is necessary for others to reproduce the findings.

Where does this leave digital legal history researchers? This is a question to which I do not have a concrete answer. More than in contemporary research, where it is more common for multiple researchers or groups of researchers to reuse data and to verify results, including through checks of interrater reliability, the costs and benefits of sharing materials in digital legal history are more difficult to balance; the decision to share is therefore less straightforward than in other domains. Hence, doing so deserves only praise.

Bibliography

Burrows, 2023

Burrows, T. (2023). Reproducibility, verifiability, and computational historical research. International Journal of Digital Humanities, 5, 283–298. https://doi.org/10.1007/s42803-023-00068-9

[1] Although this can be prevented by requiring approval before sharing the data.

Black, 2023

Black, J. W. (2023). Creating specialized corpora from digitized historical newspaper archives: An iterative bootstrapping approach. Digital Scholarship in the Humanities, 38(2), 779–797. https://doi.org/10.1093/llc/fqac079

Blaney et al, 2021

Blaney, J., Milligan, S., Steer, M., & Winters, J. (2021). Doing digital history: A beginner's guide to working with text as data. Manchester University Press.

Cinini, 2025

Cinini, A., Marongiu, P., Sassolini, E., & Monachini, M. (2025). Preserving and enhancing cultural heritage: The Digest project. In S. Rebora, M. Rospocher, & S. Bazzaco (Eds.), Diversità, equità e inclusione: Sfide e opportunità per l'informatica umanistica nell'era dell'intelligenza artificiale (pp. 660–667). AIUCD. https://doi.org/10.6092/unibo/amsacta/8380

van Dijck, 2025

van Dijck, G. (2025). To Open Access or Not to Open Access? Maastricht Journal of European and Comparative Law, 32(1), 3-7. https://doi.org/10.1177/1023263X251338075

Rule et al., 2019

Rule, A., Birmingham, A., Zuniga, C., Altintas, I., Huang, S. C., Knight, R., Moshiri, N., Nguyen, M. H., Rosenthal, S. B., Pérez, F., & Rose, P. W. (2019). Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks. PLoS computational biology, 15(7), e1007007. https://doi.org/10.1371/journal.pcbi.1007007