

From Manuscript to Megabyte: Building Sustainable and Open Futures in the Digital Humanities

James B. Harr III¹ , Emma Stanley², Jenna Rinalducci³ , and Donna Kain⁴

¹ Department of Data Science, The College of William and Mary, Williamsburg, VA, USA

² North Carolina State University, Raleigh, NC, USA

³ University of North Carolina at Charlotte, Charlotte, NC, USA

⁴ East Carolina University, Greenville, NC, USA

Abstract

We present and introduce the proceedings from the 2026 Digital Humanities Institute on NC State's Centennial Campus. The gathering's theme, *From Manuscript to Megabyte: Building Sustainable and Open Futures in the Digital Humanities*, asked scholars, librarians, archivists, and practitioners from across North Carolina and beyond to consider a question that has grown only more urgent as public investment in higher education and the humanities continues to contract: how do we build systems of knowledge—technical, legal, institutional, and communal—that are not only innovative but durable, accessible, and just? We describe the content of the included papers and end with reflections on what happens when we move from manuscript to megabyte: who is included in the process, what is deliberately preserved and what is allowed to end, and whose knowledge counts as evidence in the first place.

Keywords: digital humanities, sustainability, digital preservation, cultural heritage, libraries

On May 27–28, 2026, the Digital Humanities Collaborative of North Carolina (DHC-NC), in partnership with the North Carolina State University Libraries, convened the 2026 Digital Humanities Institute on NC State's Centennial Campus. The gathering's theme, *From Manuscript to Megabyte: Building Sustainable and Open Futures in the Digital Humanities*, asked scholars, librarians, archivists, and practitioners from across North Carolina and beyond to consider a question that has grown only more urgent as public investment in higher education and the humanities continues to contract: how do we build systems of knowledge—technical, legal, institutional, and communal—that are not only innovative but durable, accessible, and just? The call for papers framed this question capaciously. It invited work on open education and OER, on the role of libraries and archivists in digital preservation and access, on the copyright and policy frameworks that govern digitization, on participatory and community-engaged models of digital editing, on the material and sensory dimensions of historical media, and on the promises and risks of artificial intelligence for transcription, restoration, and interpretation. Underlying all of these threads was a shared premise: digital tools are no longer peripheral aids to humanistic inquiry but have become foundational infrastructure for preserving cultural memory and sustaining scholarly ecosystems, precisely at a moment when that infrastructure can least be taken for granted.

The nine articles collected in this special issue, developed from presentations delivered across the Institute's five panels, take up that premise from remarkably different vantage points—biocodicology, handwritten text recognition, semantic web ontologies, immersive exhibition design, community archiving, accessibility law, library administration, and survey research on the field itself. Read together, they trace the arc the conference theme names: the movement of cultural material from manuscript to megabyte, and the sustained, often unglamorous labor

James B. Harr III, Emma Stanley, Jenna Rinalducci, and Donna Kain. "From Manuscript to Megabyte: Building Sustainable and Open Futures in the Digital Humanities." In: *From Manuscript to Megabyte: Building Sustainable and Open Futures in the Digital Humanities*, ed. by James B. Harr III, Emma Stanley, Jenna Rinalducci, and Donna Kain. Vol. 5. Anthology of Computers and the Humanities. 2025, 1–5. <https://doi.org/10.63744/XQGnb3aSKb2D>.

© 2025 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

required to make sure that movement serves rather than erodes the communities whose histories are at stake.

1 Manuscripts as Data: Computation, Uncertainty, and the Limits of Machine Reading

Three contributions take up the literal premise of the conference theme by asking what happens when historical documents are rendered as computational data, and what is gained and lost in the translation.

Guoying Yang's "Between the Lines: A Comparative Analysis of AI vs. Human Transcription of Lin Yutang's Bilingual Handwritten Notebooks" pursues a parallel question in the domain of handwritten text recognition (HTR). Through a controlled comparison of three HTR systems against three expert human transcribers working on Lin Yutang's bilingual Chinese-English notebooks, Yang finds that automated transcription systems perform reasonably well on mechanical accuracy but falter badly on precisely the passages that carry the most interpretive weight—code-switching boundaries and Lin's culturally specific neologisms, such as "Christo-Confucian." Yang's proposed distinction between "semantic-blind" machine errors and "mechanical" human errors, along with an "Interpretive Fitness Audit" framework, offers a methodological contribution well beyond this single case: a vocabulary for evaluating AI-assisted transcription that goes beyond character error rate to ask what is actually lost when a neologism is silently flattened into a familiar word. The essay's proposed hybrid workflow (HTR paired with targeted human correction) models exactly the kind of sustainable, resource-conscious practice the conference theme called for.

James B. Harr III's "The Manuscript as Dataset: Biological Evidence and Computational Analysis in Medieval Palimpsests" pushes the "manuscript as data" metaphor into genuinely new territory, treating parchment not only as a textual surface but as a biological archive. Working with a fourteenth-century codex from the University of Pennsylvania's Kislak Center, Harr extracted mitochondrial DNA from parchment leaves to test whether the chemical and mechanical processes of palimpsesting—scraping and washing a page for reuse—leave a measurable biological trace. His machine learning classifiers ultimately struggled to distinguish palimpsested from single-use parchment on the basis of mtDNA alone, a result he treats not as a dead end but as an honest accounting of what computational methods can and cannot recover from the material past. The essay is a useful reminder that "sustainability" in digital humanities work includes epistemic sustainability: knowing the limits of one's data is as important as knowing its possibilities.

Yu Lee An's *Reconstructing Historical Events with CIDOC CRM* turns from transcription to temporal representation, asking how historians can encode the pervasive uncertainty of archival dating—circa dates, bounded ranges, fuzzy eras—into machine-interpretable form without flattening the humanistic nuance those uncertainties carry. Using the professional networks of the Georgian and Victorian British music trade as a case study, An demonstrates how CIDOC CRM's temporal boundary properties can hold multiple degrees of precision within a single semantic graph. Read alongside Harr's and Yang's contributions, An's essay rounds out a shared argument across this cluster of papers: computational formalization is not the enemy of interpretive complexity, but working with machines responsibly requires building that complexity into the data model itself, rather than assuming the machine will supply it.

2 Infrastructure and Access: Libraries as the Connective Tissue of Digital Humanities

A second cluster of essays turns from manuscripts themselves to the institutional scaffolding, chiefly libraries, that makes sustained digital humanities work possible at all, a concern squarely at

the center of the conference's call for attention to "libraries, archives, and access" and "sustainable systems of knowledge sharing."

Allison Kaeffring, Jennifer Daugherty, and Jeanne Hoover's *Opening Digital Humanities to All: Approaches to Supporting Accessible Omeka Projects* addresses a deadline that has quietly reshaped digital humanities practice at public institutions: the Department of Justice's 2024 Title II regulations requiring WCAG 2.1 AA compliance for government digital content. Drawing on East Carolina University Libraries' audit of its own Omeka S projects, the authors distinguish between accessibility failures built into third-party software and those introduced when well-meaning practitioners—increasingly aided by AI "vibe coding"—edit underlying code without accessibility expertise. Their essay also surfaces a genuinely unresolved policy question that will be familiar to many DH practitioners: whether a finished student capstone project, orphaned when its creator graduates, should be treated as "archived" and thus exempt, or as an ongoing public resource still owed full compliance. The piece is a valuable, practical companion to the conference's calls for engagement with "copyright, licensing, and policy" beyond the more familiar terrain of intellectual property.

Christin Lampkowski, Savannah Lake, and Jenna Rinalducci's *From Working Group to Campus Resource: Strengthening Digital Humanities through the Library* documents a complementary, more organizational form of infrastructure-building: the decade-long evolution of an informal librarians' discussion group at UNC Charlotte into a cross-departmental Digital Humanities Working Group. Their environmental scan of comparable institutions, from richly staffed centers like San Diego State's to purely distributed, "ad hoc" models, situates UNC Charlotte's approach within a broader landscape of decentralized DH support, where libraries function less as a physical center than as connective tissue among faculty, students, and departments who might otherwise never find one another. The essay's honest accounting of what sporadic, under-resourced programming can and cannot sustain speaks directly to the conference's interest in the "structural" conditions shaping how knowledge circulates.

Kathryn Wymer's *Understanding Ephemerality and Loss in Digital Humanities Projects: What Surveys of the Field Can Reveal* takes up sustainability's harder edge: what happens when digital humanities projects end. Drawing on the 2009 and 2022 iterations of the Graceful Degradation survey and situating them alongside the University of Victoria's Endings Project, Wymer argues that digital humanities can usefully be understood as "research on fire", necessarily fast-moving and combustible in ways traditional scholarship is not and poorly served by an academic reward system still oriented toward permanence. Her call to distinguish the parts of a project worth preserving (often its underlying data) from the parts that can be gracefully retired (often its interactive interface) offers a frame that resonates across nearly every other essay in this issue, each of which is, in its own way, a record of infrastructure built under conditions of uncertainty about how long that infrastructure will last.

3 Recovering and Contesting Cultural Memory

A third cluster of essays turns to the conference's abiding concern with cultural memory: who gets to preserve it, through what forms, and with what consequences when digital platforms mediate its circulation.

Shaun Bennett, Franziska Bickel, Kevin Oliver, and Angela Wiseman's *Bringing Cultural Heritage to Campus through Innovative Technical Platforms and Physical Spaces* documents the Russell School Digital Preservation Project, centered on the last remaining Rosenwald School in Durham County, North Carolina, a two-teacher schoolhouse built through Black community fundraising, county support, and the Rosenwald Fund in 1926. The authors trace how a Matterport 360-degree virtual tour, developed in partnership with the Friends of Russell Rosenwald School, was repurposed across NC State Libraries' VR studio, projection gallery, and interactive exhibit

spaces to create a public, community-anchored event that paired scholarly context with the testimony of a Russell School alumna. The project stands as one of the clearest realizations in this issue of the conference's call for "digital editions and community engagement": a case where digital preservation was undertaken not on behalf of a community but visibly with and for it.

Rhonda D. Jones and Maurika Smutherman's *More Than Just Planting Pansies: Black Women's Garden Clubs and the Neo-Archive in North Carolina's Digital Landscape* poses a related but distinct archival problem: how to recover the history of civic institutions (here, African American women's garden clubs across the Jim Crow South) whose records were never centralized in the first place, surviving instead across newspapers, scrapbooks, photographs, oral histories, and living landscapes. Jones and Smutherman's proposed concept of the "neo-archive" reframes gardens, seed-saving, and beautification practices themselves as legitimate historical evidence, and their trauma-informed reading of archival absence—asking not why the records are missing but how communities preserved knowledge despite that absence—offers digital humanities practitioners a methodological model for community-engaged, Omeka-based reconstruction that extends well beyond this particular case.

Samone Jacobs's *Google Map Reviews: A Window into Plantation Narratives* turns the same attentiveness to whose stories get told toward a distinctly contemporary and contested site: the online reviews visitors leave for plantation museums along Louisiana's Great Mississippi River Road. Using computational text analysis to examine how eight plantation sites' divergent institutional narratives, ranging from Whitney Plantation's explicit focus on the history of slavery to sites organized around weddings and ghost tours, are reflected, reproduced, or challenged in visitor reviews, Jacobs extends scholarship on plantation "narrative economies" into networked digital space. The essay is a pointed reminder that the circulation the conference theme names is not always benign: platforms can just as easily launder or amplify a site's chosen mythology as they can democratize access to a fuller history, and Jacobs's attention to electronic word-of-mouth offers a template for tracking which outcome is occurring, and where.

4 Reading the Issue as a Whole

Individually, these nine essays answer to different corners of the original call for papers: biocodology and machine learning, transcription ethics, semantic modeling, accessibility law, library organizational history, project sustainability, community-engaged exhibition design, archival theory, and platform-mediated public memory. Read together, they resist any single, tidy account of what "digital humanities" now means in North Carolina and its collaborating institutions, and that resistance is itself a finding worth naming. Several through-lines nonetheless emerge. Machines are treated throughout not as neutral instruments but as interpretive actors whose errors and silences require the same critical scrutiny long applied to human sources, whether the machine in question is a supervised learning classifier, an HTR engine, or a large language model coding plantation reviews. Libraries and librarians appear again and again not as background infrastructure but as the field's actual connective tissue, quietly making possible the community partnerships, accessibility compliance, and organizational continuity that individual projects depend on and rarely credit sufficiently. And sustainability is treated consistently as a condition to be actively built and maintained, whether that means designing for a project's eventual end, auditing a decade of Omeka sites for accessibility, or reckoning honestly with what a biological or textual dataset cannot tell us.

The editors are grateful to the authors for the generosity and rigor with which they revised their conference presentations into the articles collected here, to the panel moderators and the Institute's keynote speaker, Dr. Jeffrey Bardzell, Peterson Family Dean of the School of Arts and Sciences at the Worcester Polytechnic Institute, for framing the conversations from which this volume grew, and to the NC State University Libraries and the Digital Humanities Collaborative

of North Carolina for the institutional partnership that made the 2026 Institute, and this issue, possible. Finally, we want to acknowledge the contributions made by Dr. Kathryn Wymer from the UNC-Chapel Hill School of Data and Information Sciences, for her work soliciting articles from panelists and reviewing/editing submissions for this volume. If the conference theme asked what it means to move from manuscript to megabyte, the essays that follow answer, collectively, that the move is never simply technical. It is always also a question of who is included in the process, what is deliberately preserved and what is allowed to end, and whose knowledge counts as evidence in the first place.