

Binghamton University
The Open Repository @ Binghamton (The ORB)

English, General Literature, and Rhetoric Faculty
Scholarship

English, General Literature, and Rhetoric

Fall 9-2018

Adam Sciveyn in Cyberspace: Loss, Labour, Ideology, and Infrastructure in Interoperable Reuse of Digital Manuscript Metadata

Bridget Whearty

Binghamton University (SUNY, State University of New York), bwhearty@binghamton.edu

Follow this and additional works at: https://orb.binghamton.edu/english_fac



Part of the [English Language and Literature Commons](#)

Recommended Citation

Whearty, Bridget, "Adam Sciveyn in Cyberspace: Loss, Labour, Ideology, and Infrastructure in Interoperable Reuse of Digital Manuscript Metadata" (2018). *English, General Literature, and Rhetoric Faculty Scholarship*. 5.
https://orb.binghamton.edu/english_fac/5

This Book Chapter is brought to you for free and open access by the English, General Literature, and Rhetoric at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in English, General Literature, and Rhetoric Faculty Scholarship by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

Edited by **MATTHEW EVAN DAVIS,**
TAMSYN MAHONEY-STEEL,
and **ECE TURNATOR**

ARC HUMANITIES PRESS

MEETING THE MEDIEVAL IN A DIGITAL WORLD



ARC HUMANITIES PRESS



MEDIEVAL MEDIA CULTURES

Medieval Media Cultures offers analyses of how individuals interacted with written, visual, dramatic, and material media in medieval and early modern cultures, as well as how modern scholars interact with the remnants of medieval and early modern cultures via written, material, and now digital and electronic media.

This new series in media literacy welcomes proposals for monographs and essay collections in the fields of digital humanities, mapping, digital text analysis, games and gaming studies, literacy studies, and text production and interaction. We are especially interested in projects that demonstrate how digital methods and tools for research, preservation, and presentation influence the ways in which we interact with and understand these texts and media.

Series Editors

Toby Burrows, *University of Oxford*

Dorothy Kim, *Vassar College*

Richard Utz, *Georgia Institute of Technology*

Acquisitions Editor

Ilse Schweitzer VanDonkelaar, *East Lansing*

MEETING THE MEDIEVAL IN A DIGITAL WORLD

EDITED BY MATTHEW EVAN DAVIS,
TAMSYN MAHONEY-STEEL, AND ECE TURNATOR

ARC HUMANITIES PRESS

ARC HUMANITIES PRESS



British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

© 2018, Arc Humanities Press, Leeds



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence.

The authors assert their moral right to be identified as the authors of their part of this work.

Permission to use brief excerpts from this work in scholarly and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is an exception or limitation covered by Article 5 of the European Union's Copyright Directive (2001/29/EC) or would be determined to be "fair use" under Section 107 of the U.S. Copyright Act September 2010 Page 2 or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17 USC §108, as revised by P.L. 94-553) does not require the Publisher's permission.

FOR PRIVATE AND
NON-COMMERCIAL
USE ONLY

ISBN: 9781641891929
e-ISBN: 9781641891936

<http://arc-humanities.org>

Printed and bound by CPI Group (UK) Ltd, Croydon, CR0 4YY

CONTENTS

| | |
|--|-----|
| List of Illustrations | vii |
| List of Contributors | xi |
| Introduction | 1 |
| MATTHEW EVAN DAVIS, TAMSYN MAHONEY-STEEL, AND ECE TURNATOR | |
| 1 Statistical Analysis and the Boundaries of the Genre of Old English Prayer. | 11 |
| WILLIAM H. SMITH AND CHARLES L. BUTLER | |
| 2 If (not “Quantize, Click, and Conclude”) {Digital Methods in Medieval Studies} | 27 |
| KATAYOUN TORABI | |
| 3 Project Paradise: A Geo-Temporal Exhibit of the Hereford Map and <i>The Book of John Mandeville</i> | 45 |
| ALEXANDRA BOLINTINEANU | |
| 4 Ghastly Vignettes: <i>Pierce the Ploughman’s Crede</i> , the Ghost of Shakespeare’s Blackfriars, and the Future of the Digital Past. | 69 |
| JIM KNOWLES | |
| 5 Content is not Context: Radical Transparency and the Acknowledgement of Informational Palimpsests in Online Display | 93 |
| MATTHEW EVAN DAVIS | |
| 6 Encoding and Decoding Machaut | 123 |
| TAMSYN MAHONEY-STEEL | |
| 7 Of Dinosaurs and Dwarves: Moving on from <i>Mouvance</i> in Digital Editions | 143 |
| TIMOTHY L. STINSON | |

| | |
|---|-----|
| 8 Adam Scriveyn in Cyberspace: Loss, Labour, Ideology, and Infrastructure in Interoperable Reuse of Digital Manuscript Metadata. | 157 |
| BRIDGET WHEARTY | |
| 9 Digital Representations of the Provenance of Medieval Manuscripts. | 203 |
| TOBY BURROWS | |
| 10 Bridging the Gap: Managing a Digital Medieval Initiative Across Disciplines and Institutions | 223 |
| JOSEPH KOIVISTO, LILLA KOPÁR, AND NANCY L. WICKER | |
| Index | 241 |

LIST OF ILLUSTRATIONS

Figures

- Figure 1.1. Dendrogram showing fingerprint word distribution in Old English poetry and prose “sampler” files. 19
- Figure 1.2. Prayers and confessional materials using a combined 146-word fingerprint. 21
- Figure 1.3. Prayers and confessional materials using the thirty most frequent words in each set. 22
- Figure 1.4. Prayers, confessional materials, and the four “Confessional Prayers,” using a combined 146-word fingerprint. 23
- Figure 1.5. Prayers, confessional materials, and the four “Confessional Prayers,” using the thirty most frequent words. 24
- Figure 1.6. Prayers, confessional materials, and the four “Confessional Prayers,” using the ten most frequent words. 25
- Figure 2.1. Combined *Beowulf* and *Blickling Homily XVII* 500-word chunks (Lexomics Tools). 31
- Figure 2.2. Dendrogram—*Canterbury Tales: General Prologue* portraits. 37
- Figure 2.3. Wife of Bath *General Prologue*—Word Cloud and Word Frequency List (generated using Voyant). 38
- Figure 2.4. Prioress *General Prologue*—Word Cloud and Word Frequency List (Voyant). 38
- Figure 2.5. Combined Wife of Bath and Prioress Portraits with function words present (Voyant Word Collocation tool). 38
- Figure 2.6. Combined Wife of Bath and Prioress Portraits excluding function words (Voyant Word Collocation tool). 39
- Figures 3.1 and 3.2. Hereford map, Earthly Paradise (covered by large circle) with rivers (indicated by small dots within the larger circle). 53
- Figure 3.3. Hereford map, River from Paradise bears jewels into Prester John’s kingdom (triangle). 55

| | |
|---|-----|
| Figure 3.4. Hereford map: Within the circle of Paradise, centre, flow its four rivers. | 55 |
| Figure 3.5. Hereford map: At the top, circled: Paradise; below, just above the lower border: the walled city of Jerusalem; above it, Christ's cross.. | 56 |
| Figure 3.6. The Hereford map: Within the Earthly Paradise (encircled, above), the rivers stop before the wall. Outside the Earthly Paradise, the four rivers (from the top: Indus (Ganges), Tigris, Euphrates, and Nile) are not visually connected to Paradise. | 58 |
| Figure 3.7. The Hereford map: The triangle-studded line from the Cross (below) to Paradise and the angel (above) to the company of the blessed in the afterlife (top left). | 61 |
| Figure 4.1. Greyfriars second chapel, Phase I, 1244–48.. | 83 |
| Figure 4.2. Greyfriars, Phase II, 1260–70. | 84 |
| Figure 4.3. Greyfriars, Phase III. Blackfriars is visible in the background. | 84 |
| Figure 4.4. Greyfriars, Phase IV, after 1270. | 85 |
| Figure 4.5. Greyfriars, Phase V, late fourteenth century. | 86 |
| Figure 4.6. Blackfriars second site, Phase I, 1236–45. | 87 |
| Figure 4.7. Blackfriars, Phase II (foreground). Blackfriars, Phase V (background). | 88 |
| Figure 4.8. The underlying image is a digital copy of the Ralph Agas 1578 map of Oxford, a copy of which is in the Bodleian Museum. | 88 |
| Figure 5.1. The triadic relationship of Peirce's semiotic theory. | 101 |
| Figure 5.2. Infinite Semiosis—the Interpretant of the first sign becomes the Representamen of the next sign in the chain. | 103 |
| Figure 5.3. A constructed image intended to represent the opening showing fol. 300v–301r of the “Queen Mary Psalter.” | 104 |
| Figure 5.4. Fol. 300v and 301r as they appear on the British Library's Digitised Manuscripts site. | 105 |
| Figure 5.5. SQL Schema for the Minor Works of Lydgate website. | 112 |

| | |
|--|-----|
| Figure 5.6. The <sourceDoc>-based encoding for the version of Lydgate's "Quis Dabit Meo Capiti Fontem Lacrimarum" in Jesus College Q.G.8, Jesus College, Cambridge, as rendered via the oXygen XML Editor..... | 115 |
| Figure 5.7. The author's master transcription file for the witness of the <i>Testament of John Lydgate</i> in Huntington HM140. | 116 |
| Figure 5.8. Model of the Clopton chantry chapel, Holy Trinity, Long Melford. | 117 |
| Figure 5.9. Force-directed model indicating the relationship between the three texts at Holy Trinity, Long Melford, in all witnesses..... | 118 |
| Figure 6.1. Screenshot of jechante.exeter.ac.uk showing the triplum text of the motet <i>Quant en moy / Amour et biauté / Amara valde</i> using the "Edition" tab and "Standard" view. | 126 |
| Figure 6.2. TEI P5 conformant XML encoding of stanza one of "Helas! pour quoy virent onques mi oueil" (Lo53)..... | 126 |
| Figure 6.3. Stanza one of "Helas! pour quoy virent onques mi oueil" (Lo53) as it appears in MS C. | 127 |
| Figure 6.4. Text and translation of stanza one "Helas! pour quoy virent onques mi oueil" (Lo53) without mark-up. | 127 |
| Figure 6.5. Example Exclusion File from the Roland Park Company Papers, Special Collections and Archives, Sheridans Libraries, Johns Hopkins University..... | 130 |
| Figure 6.6. Encoding with elements each having equal hierarchy. | 131 |
| Figure 6.7. Encoding with details nested under <person>..... | 131 |
| Figure 6.8. Rubric from the index of MS BnF fr 1584, f. Av. | 134 |
| Figure 7.1. Four views of the text available on the <i>Siege of Jerusalem</i> Electronic Archive..... | 152 |
| Figure 7.2. Apparatus of Bx, <i>Piers Plowman</i> Electronic Archive. | 152 |
| Figure 9.1. Example of multiple <provenance> elements in the TEI Guidelines. | 206 |
| Figure 9.2. Amphora of Tuthmosis III—an event-centric description (without aggregation and related digital representation entities). | 214 |

Tables

| | | |
|------------|---|-----|
| Table 1.1. | Most frequent words in the <i>DOE</i> “Prayer” category (B12.4)..... | 15 |
| Table 1.2. | Most frequent words in the <i>DOE</i> “Confessional Materials” category (B11). | 15 |
| Table 1.3. | Most frequent words in the <i>DOE</i> “Prayer” category with stop words removed..... | 16 |
| Table 1.4. | Most frequent words in the <i>DOE</i> “Confessional Materials” section with stop words removed..... | 16 |
| Table 1.5. | Most frequent words in the four “Confessional Prayers” (B11.9.1, B11.9.3.1, B11.9.3.2, and B11.9.4)..... | 17 |
| Table 8.1. | Comparison of XML-encoded metadata schemas employed by participating institutions..... | 177 |
| Table 8.2. | Sample crosswalking of how the three example participating institutions mark up a work’s creator(s). | 178 |
| Table 8.3. | Sample crosswalk of Walters TEI to Stanford MODS..... | 181 |
| Table 8.4. | Descriptive metadata on the decorations and embellishments of Corpus Christi College Cambridge, Parker Library MS 304..... | 189 |
| Table 8.5. | Mock-up of what that macaronic description, drawing together English-language OAI:DC and French-language TEI, might look like.... | 193 |
| Table 9.1. | The definition of the “Event” class in the Europeana Data Model..... | 213 |
| Table 9.2. | The components of a property graph for the purchase of Phillipps MS 16402..... | 217 |
| Table 9.3. | Sample nodes and relationships for Neo4j property graph. | 217 |
| Table 9.4. | Sub-objects of the “Manuscript” Object in the nodegoat framework.... | 217 |
| Table 9.5. | A partial history of Phillipps MS 12264 modelled in nodegoat..... | 218 |

LIST OF CONTRIBUTORS

ALEXANDRA BOLINTINEANU (alexandra.bolintineanu@utoronto.ca) is an Assistant Professor, Teaching Stream, in Medieval Digital Studies, at the University of Toronto. She is cross-appointed to the Centre for Medieval Studies and Woodsworth College. Her research interests include literary wonders in Old and Middle English; poetics of wonder in oral, literate, and digital environments; imaginary geographies; medieval and digital mapping; and digital humanities pedagogy.

TOBY BURROWS (toby.burrows@uwa.edu.au) is a Senior Researcher at the Oxford e-Research Centre in the University of Oxford, and a Senior Honorary Research Fellow in Humanities at the University of Western Australia. Current projects include “Mapping Manuscript Migrations” (funded by the Digging into Data Challenge) and “Collecting the West” (funded by the Australian Research Council). Between 2014 and 2016 he held a European Union Marie Curie International Incoming Fellowship in the Department of Digital Humanities at King’s College London. Publications include *Collecting the Past* (edited with Cynthia Johnston), forthcoming from Routledge.

CHARLES LEE BUTLER is currently the Director of Institutional Research and Assessment at Weatherford College. Trained in Statistics in the Sociology Department at Baylor University, his interest in the application of statistical methods has grown and diversified, ranging from text analysis in a variety of areas to Bayesian statistics in Educational Research. He continues to explore the power of R statistical programming language through research into variables associated with student success. Outside of his professional obligations, he explores applications of data analysis to the Digital Humanities.

MATTHEW EVAN DAVIS (matthew@matthewedavis.net) currently serves as a Postdoctoral Fellow with the Lewis and Ruth Sherman Centre for Digital Scholarship at McMaster University. Prior to this he was a Lindsey Young Visiting Faculty Fellow at the University of Tennessee’s Marco Institute and served as the Council for Library and Information Resources/Mellon Postdoctoral Fellow in Data Curation for Medieval Studies at North Carolina State University. He has published articles on medieval drama in *Theatre Notebook* and on the Lydgate verses at the Clopton chantry chapel in the *Journal of Medieval Religious Cultures*. He is the editor of the *Minor Works of John Lydgate* virtual archive (www.minorworksoflydgate.net) and is also working on a database charting the network of spiritual obligation captured in fifteenth-century wills and inventories.

JIM KNOWLES (jrknowle@ncsu.edu) is Teaching Associate Professor of English at North Carolina State University. He is Managing Editor of the *Piers Plowman Electronic Archive* and co-creator (with Michal Koszycki) of the *Oxford Friars Project*. He has

published articles on medieval literature and theology in the *Journal of Medieval and Early Modern Studies* and the *Yearbook of Langland Studies*. Together with Timothy Stinson, he is working on a hybrid print and electronic edition of Robert Crowley's three 1550 print versions of *Piers Plowman*.

LILLA KOPÁR (kopar@cua.edu) is Associate Professor of English and Director of the Center for Medieval and Byzantine Studies at The Catholic University of America in Washington, DC. She is an interdisciplinary scholar of the early Middle Ages with a special interest in the intersections of visual art and literature, myth and religion in the Viking Age, epigraphy and runology, and the art of commemoration. She has published on the iconography of Anglo-Saxon stone sculptures, on representations of Germanic myths and legends in early medieval northern Europe, and on runes and inscribed objects. Kopár is also co-director of the NEH-funded digital humanities initiative Project Andvari.

JOSEPH KOIVISTO (jkoivist@umd.edu) (MSLIS) is a systems librarian at the University of Maryland, serving the University System of Maryland and Affiliated Institutions (USMAI) consortium. Joseph earned his BA from the University of Scranton in 2009 and his MSLIS from the Catholic University of America in 2014. Additionally, he has worked for the District of Columbia Public Libraries and the Library of Congress. He has served as project assistant and project manager for Project Andvari, an NEH-funded research initiative, since 2014. Joseph's research interests include digital humanities work in libraries, critical perspectives on library technical infrastructure, crowdsourcing, and authority creation.

TAMSYN MAHONEY-STEEL (tamsyn@jhu.edu) is the Digital Scholarship Specialist for the Sheridan Libraries at Johns Hopkins University. She has a PhD in medieval French music and literature, and has published on medieval motets, linked open data, text mining the poetry of Guillaume de Machaut, library design, and collaborative scholarship. Her forthcoming monograph explores the scholarly benefits of using annotation tools on digitized medieval texts and manuscripts. Tamsyn teaches courses on digital humanities, using digital tools to explore medieval authors, and information literacy. At JHU she has instituted a series of seminars exploring the intersection of digital scholarship and issues of diversity and inclusion, and she has created workshops to introduce graduate students to text mining tools and methodologies.

WILLIAM SMITH (wsmith@wc.edu) is a Professor of English at Weatherford College, where he has taught classes in British Literature, Medieval World Literature, and Science Fiction Literature, among others. His research interests focus on Old English literature, especially the literature of learning and personal devotion, and on the application of statistical textual analysis methods to examinations of literary style.

TIMOTHY STINSON (tstinso@ncsu.edu) is associate professor of English at North Carolina State University. He has published articles on the Alliterative Revival, printing history, codicology, manuscript illumination, and the application of genetic analysis to the study of medieval parchment. He is editor of the *Siege of Jerusalem Electronic*

Archive, is co-founder and co-director of the Medieval Electronic Scholarship Alliance (MESA), associate director of the Advanced Research Consortium (ARC), co-director of the *Piers Plowman Electronic Archive*, and director of the Society for Early English and Norse Electronic Texts (SEENET). His research has received funding from the National Endowment for the Humanities, the Andrew W. Mellon Foundation, the Bibliographical Society of America, and the Council on Library and Information Resources.

KATAYOUN TORABI (torabik@tamu.edu) an instructional assistant professor in the Department of English at Texas A&M University. She is a specialist in Old and Middle English Literature, with strong research and teaching interests in the Digital Humanities. Her research, which incorporates both digital and traditional methodologies, investigates the connection between purgation and incarceration in pre- and post-mortem punitive spaces in Old and Middle English literature. Her work has been published in *Essays in Medieval Studies* and the *Association for Computing Machinery Press*.

ECE TURNATOR (turnator@mit.edu) received her PhD in medieval (Byzantine) history from Harvard University in 2013. Her dissertation is an interpretation of thirteenth-century Byzantine economy through an analysis of archaeological (coins and ceramics) and textual evidence. Between 2013 and 2016, she was a CLIR postdoctoral fellow at The University of Texas at Austin Libraries where she worked as coordinator for the Global Middle Ages Project, specialized in medieval data curation, studied and wrote about digital humanities and best practices for data curation and visualization, and taught in her areas of expertise. Currently, she is the humanities and digital scholarship librarian, and the liaison for history and linguistics departments, at the MIT Libraries.

BRIDGET WHEARTY (bwhearty@binghamton.edu) an Assistant Professor in English and Medieval Studies at Binghamton University (SUNY). She researches and publishes on the digitization of medieval manuscripts and on later medieval English literature and culture, especially the poetry of Geoffrey Chaucer, Thomas Hoccleve, and John Lydgate. From 2013–2015 she was a Council on Libraries and Information Resources (CLIR) Postdoctoral Fellow in Data Curation for Medieval Studies at Stanford University, where she served as data curator for a multi-institution digital manuscript index. She is co-editor of a special issue of *Archive Journal* dedicated to Digital Medieval Manuscript Cultures and is currently completing her first monograph—*The Digital Resurrection of Medieval Manuscripts*.

NANCY L. WICKER (nwicker@olemiss.edu) is Professor of Art History at The University of Mississippi. She has also been a Visiting Professor at Uppsala University and was elected to membership in the Royal Society of Humanities at Uppsala. Her research focuses on Scandinavian art during the Early Medieval Period, from the fifth century through the Viking Age, ca. 750–1100. She has had the support of fellowships from the National Humanities Center, the National Endowment for the Humanities, and the American Council of Learned Societies, as well as numerous grants. Wicker has served as Co-Director of Project Andvari, funded by NEH Digital Humanities Grants.

Chapter 8

ADAM SCRIVEYN IN CYBERSPACE: LOSS, LABOUR, IDEOLOGY, AND INFRASTRUCTURE IN INTEROPERABLE REUSE OF DIGITAL MANUSCRIPT METADATA¹

BRIDGET WHEARTY

And for ther is so gret diversite
In Englissh and in writyng of oure tonge,
So prey I God that non myswrite the,
Ne the mysmetre for defaute of tonge;
And red wherso thow be, or elles songe,
That thow be understonde, God I biseche!
But yet to purpos of my rather speche

—Geoffrey Chaucer, *Troilus and Criseyde*²

Once you start to aggregate these resources and combine them in a new context and for a new purpose, you find out, in practical terms, what it means to say that their creators really only envisioned them being processed in their original context.... It's as though the data has suddenly found itself in Union Station in its pajamas: it is not properly dressed for its new environment.

—John Unsworth, "Digital Humanities Centers as Cyberinfrastructure"³

In recent years, the digitization of medieval manuscripts has grown from a few intrepid trickles to a global flood. Driven by twin commitments to preservation and access, major

¹ This chapter is indebted to the Council on Libraries and Information Resources (CLIR) and the many members of Stanford University Libraries' Department of Digital Library Systems and Services who welcomed and mentored me throughout my postdoc there. I am grateful, especially, to Tom Cramer, Tony Navarrete, Bess Sadler, and Laney McGlohon; and to Greta de Groat, Lynn McRae, and Laura Wilsey who generously shared expertise in metadata and crosswalking. Throughout my postdoc, Benjamin Albritton was an invaluable sounding board, patient teacher, and enthusiastic collaborator. While I am deeply grateful to all at CLIR, SUL, and DLSS, my gratitude to him knows no bounds.

² V.1793–1799. All quotations from Chaucer, unless otherwise noted, come from Benson, ed., *The Riverside Chaucer*.

³ Unsworth, "Digital Humanities Centers as Cyberinfrastructure," quoted in Unsworth "Computational Work with Very Large Collections," 4.

digitization efforts dedicated to medieval books have arisen across the planet.⁴ This tide has also given rise to important warnings that digitization is fundamentally transforming scholarly work upon medieval books.⁵ Although “the physical” and “the digital” are frequently presented in opposition, beneath significant disagreements all sides share core convictions. Namely, that medieval books matter, that they must be cared for, preserved, read, and maintained, that these objects’ physical forms—their words, miniatures, margins, fore-edges and bindings—are vitally important to uncovering complex textual meanings, and to recovering the identities, concerns, and desires of the people who made and read these books centuries before us.

Atop these core convictions, there are significant differences between the work that can be done on physical, *analogue* codices and that which can be done using digital medieval manuscripts. In photographs, much of the sensual experience of the codex goes missing: one cannot touch it, hear it, smell the book; one cannot see the particular movement of these specific leaves as they respond to the exact humidity of the weather on the day of the reader’s visit. Although the language of these objections foregrounds an undeniable pleasure in touching the physical object, lack of physical access raises serious scholarly concerns that cannot be dismissed as the “simple” thrill of touching a medieval book.⁶ Collation formulae, for example, are difficult to determine, or verify, when photographs do not fully show an opening’s gutter, and when the only images of the head and tail show the covers pressed firmly closed. (Moreover, there are valid and real concerns that digital manuscripts’ existence might be used to bar desirous researchers from laying hands on the beloved object.⁷) And yet, despite the undeniable benefits of working directly with the physical book, there is also something equally powerful

4 Among many remarkable projects, the work by the Hill Museum and Manuscript Library stands out: both in creating digital preservation copies of the medieval manuscripts of Timbuktu and in ongoing efforts in Iraq and Syria to create access copies of priceless cultural heritage items, many of which are, at the time of writing, unreachable (and increasingly endangered) due to ongoing political campaigns carried out by the Islamic State. It is also worth noting that while medieval manuscripts are not generally physically destroyed in the process of digitizing, more modern texts have been—and in some cases, continue to be—cut apart in order to aid the speed at which they can be digitized. See, for example, Harvard’s “Free the Law” Project, which—treating books of case law as content vessels rather than objects with their own complex histories—slices books’s pages from their spines and covers in order to achieve the admirable goal of a universal database of American law. In these more modern examples, unlike the medieval ones this chapter is dedicated to discussing, the commitments of preservation and access may sometimes be less “twin goals” and more competing teams.

5 Edwards, “Back to the Real?” remains a key thought-piece in these debates. Similar concerns about the ways that digitized objects fail to convey the full materiality of the physical object are raised in Treharne, “Fleshing Out the Text”; Rudy, “Dirty Books”; Echard, “Containing the Book: The Institutional Afterlives of Medieval Manuscripts,” and her coda to *Printing the Middle Ages*.

6 Not that these thrills are ever simple. Nor should that very real visceral pleasure be dismissed.

7 On the practice of providing access to digital images of manuscripts rather than to manuscripts themselves, see Nikolova-Houston and Houston, “Building the Virtual Scriptorium,” 232. On the impact of digitization on reading room use, see Rudy, “Dirty Books,” 30.

about possessing the ability to call up images of the texts we study whenever, and wherever, we wish. Sitting at home in upstate New York, I can pour a glass of wine, or nibble cheese, as I read Huntington Library, HM 286, a deluxe copy of John Lydgate's *Fall of Princes* (ca. 1440–1460) held in southern California—approximately 2700 miles/2445 km from my home. Although I cannot touch the flesh or smell the leaves of the book, the experience of this kind of slow reading is still strange and lovely: the manuscript and I commune together, through a kind of modern image-magic of the glowing screen. And, significantly, our connection cannot be severed by limited research funding or special collections' reading room hours. I stop reading when I want to, and not before.⁸

This sense of digital access as an intimate, one-to-one communion of lone researcher with distant book is founded on a great deal of invisible and undervalued labour. A vast community of curators, photographers, and metadata and remediation specialists make possible my communion with HM 286, even if—as part of their labour—they erase all trace of their interventions. And this community grows infinitely larger, and even more invisible, when it comes to the labour of making digital manuscripts *interoperable*, by which I mean discoverable, shareable, and useable between and across institutions. The longed-for quest object of digital manuscripts may be a single hub where a researcher can search and view the manuscripts of the world, but this marvellous hub cannot exist without serious work on digital manuscripts' descriptive metadata—the machine-readable descriptions that make it possible for researchers to search by writing support, language, contents, subject, author, scribe, artist, location of origin, current repository, etc. Perhaps because we are significantly less interested in the living than the dead, medievalists rarely inquire into the labour of making and maintaining digital manuscripts. But that is our error. For digital books are, undeniably, still objects in their own rights, albeit objects of a very different sort than the “analogue” book.⁹ Furthermore, the frequently invisible work that goes into their creation and maintenance has a profound impact on the images we see on our screens, as well as those that exist but we cannot find.

This chapter explores how invisible decisions governing the creation and sharing of digitized medieval manuscripts exert a profound, albeit unacknowledged, impact on the kind of research manuscript scholars are able to do. In particular, I focus on the use and reuse of manuscripts' descriptions that I participated in as a CLIR Postdoctoral Fellow in Data Curation for Medieval Studies at Stanford University, working as data curator on the Digital Manuscript Index (hereafter DMS-Index), a multi-institution

8 I acknowledge also that this digital access is a privileged and precarious state, dependent on access to electricity, an up-to-date computer, and a stable and powerful internet connection on my end, as well as everything working smoothly on the home institution's end—and this does not always happen.

9 I am indebted to my colleague Kristen Gallant for this reminder, which has long been accepted common sense among librarians and has shaped library-based practices in producing objects' records. This careful, longstanding differentiation between “analogue” books and digital manuscripts in library communities can provide an important model for manuscript scholars seeking language to differentiate between the two, extremely different objects.

experiment working toward that dream of a single search portal through which the digitized manuscripts of the world may be seen.¹⁰ Because this chapter exists at the intersection of library, medievalist, digitization, and digital humanities (DH) cultures, I begin by defining a broad set of key tools and concepts. Next, using the sharing of manuscripts from the Walters Art Museum with DMS-Index as my case study, I explore how the processes of copying and transformation, inevitably, create loss and change. Then, I move on to the Parker on the Web app and the challenges of reusing older digital projects' data made according to earlier best practices, which might no longer rule. Finally, drawing on my experiences with metadata from e-codices, the digital manuscript library of Switzerland, I consider how cataloguing languages and decisions feed into debates over outreach, access, and the problematic monolingualism of digital humanities.

Throughout, this chapter will argue that manuscript scholars and medievalists must take a more book historical approach to digitized texts. In other words, I seek to promote a codicology of the *digital* medieval book. Rather than treating the digital book as a not-quite-good-enough window to the distant physical book upon which one might like to do codicological study, I argue that manuscript scholars and medievalists must engage in rigorous analysis of the digital book as a unique object. We need to uncover the historical, political, social, and economic pressures that shape digitization and the building of digital repositories, because these forces exert as much influence on the shape of digital medieval texts as they did, centuries ago, in the shaping of physical texts. Framed via a rather different terminology, this chapter seeks to promote a deeper sort of information literacy for medievalists and manuscript scholars, by revealing how the information we see when we work with digital manuscripts is structured by all kinds of outside social forces. I do so because, whether approached via codicology or information literacy, it is imperative for medievalists and manuscript scholars to not lose sight of the fact that all of this information and these manuscript descriptions are always based on human decision-making—and are therefore never neutral.

In particular, this chapter seeks to make visible some of digital projects' invisible labour by sharing my narrative of growth as a programmer and digital humanist in the making of the second iteration of DMS-Index. This chapter is in many ways my conversion narrative, or *bildungsroman*, as a digital humanist. When my work on this project began, I was a freshly minted PhD in English, specializing in fifteenth-century English poetry. I had no digital humanities experience, and so—as is often the case for both library work and digital humanities—much of my learning occurred on the job. Thus, part of the argument of this chapter will be to trace a narrative of skills growth. Similar narratives of growth, I contend, are part of the underlying, hidden story of DH projects and this unseen growth can exert a profound, often unacknowledged impact on the shape these final projects can take. Because this is a *bildungsroman*, there will be readers of this chapter who came of age earlier and who therefore are more advanced

¹⁰ On DMS-Index, see Foy, "Media Archaeology and the Digital Incunable," 134.

coders than I was at the moment I began my postdoc. I expect these readers will come up with coding solutions that I did not in my first years of DH practice. But that is perhaps the least interesting kind of response I seek to foster.¹¹ A much larger part of this article's argument is that there is value in revealing this inner working—how what was built looks the way it does *because* I was a digital humanist in formation. A medievalist perspective on this digital work reveals that these processes of self-formation are part of a much longer genealogy of text-technologies and transformation than is often acknowledged.¹² Rather than obscure the learning that took place as I copied, or the ways that my improved coding changed what I was able to do with different metadata sets as the project progressed, I seek to reclaim how copyists, in the constant state of forming their own best practices, have always and will always be the story of medieval texts—in digital archives and aggregators, in earlier projects that those aggregators reuse and curate, all the way back to medieval manuscript makers, who were also involved in the work of copying under social, technological, educational, and economic pressures.

Laying Foundations: Definitions for “Metadata” and “Interoperability”

Metadata

A widely used (if maddeningly circular) definition of *metadata* is “data about data,” or “data that describes or gives information about other data,” definitions that frankly do little to clarify what metadata actually *is* and *does*.¹³ More focused communities offer more specific and useful definitions. In academic libraries, the term is “commonly used for any formal scheme of resource description” for all library holdings that must be catalogued and described—both physical books and digital objects.¹⁴ As the National Information Standards Organization (NISO) puts it, metadata is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.”¹⁵ There are three distinct categories of metadata: administrative

11 Adeline Koh's treatment of how certain class, gender, geographical, racial, and ethnic backgrounds support or inhibit early technological uptake in those who go on to become digital humanists is informative for my understanding here, as is work by Eszter Hargittai, Moya Z. Bailey, Chris Bourg, and Bess Sadler. See Koh, “Niceness, Building, and Opening the Genealogy of the Digital Humanities”; Hargittai, “Digital Natives or Digital Naives?”; Bailey, “All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave”; and Sadler and Bourg, “Feminism and the Future of Library Discovery.”

12 Bredehøft's *The Visible Text: Textual Production and Reproduction from Beowulf to Maus* is exemplary in its working putting medieval texts into a much longer history of text and textual transformation, and has strongly influenced my thinking.

13 OED “metadata” (n.); Schmitt, “Towards an Interoperable Scholarly Edition,” 9.

14 NISO, “Understanding Metadata,” 1.

15 NISO, “Understanding Metadata,” 1. See also Hodge, “Metadata Made Simpler.”

metadata,¹⁶ structural metadata, and *descriptive metadata*.¹⁷ This last category is, arguably, most important for researchers of medieval manuscripts and is the focus of this chapter.

A manuscript's bibliographic record and/or catalogue description—which outlines its contents, physical make-up, provenance, etc.—is its descriptive metadata. It is this descriptive metadata that computers query when researchers seek out manuscripts containing works by John Lydgate, or manuscripts in original fifteenth-century limp bindings, or containing miniatures of St. Margaret of Antioch. For a computer to successfully find manuscripts that fit these categories, this information about the book must be not just represented in the digital images of the book but also encoded in the digital manuscript's descriptive metadata.¹⁸

In addition to essential search-and-discovery functionality, descriptive metadata plays an increasingly important role filling the experiential gaps between what can be learned studying the physical book and what can be perceived working with a digital manuscript via a modern screen. As Mohammed Ourabah Souala and Mohamed Hassoun put it:

simply providing images of the manuscript [via digitization] is not sufficient as the images do not convey many important aspects of the manuscript including: codicological description (codex, binding, etc.); paleographic description (handwriting, etc.); manuscript transmission history.¹⁹

Until the time when providing full text versions of the manuscripts is feasible, manuscript cataloguing²⁰ proves to be a realistic solution for providing access.²¹

It is widely recognized that digitized manuscripts are fundamentally lacking when it comes to fully representing the rich complexity of a physical codex. A. S. G. Edwards highlights several particularly important aspects that are difficult to discern via digital images:

it is often difficult for viewers to take in the scale of the object being presented. It is also difficult to discern distinctions between materials such as parchment

16 This in turn can be separated further into rights management metadata and preservation metadata.

17 NISO, "Understanding Metadata," 1.

18 For more on metadata, and how the term's meaning shifts, Proteus-like, from context to context, see Gilliland, *Introduction to Metadata*, and Schmitt, 9.

19 Soualah and Hassoun, "A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts," 1.

20 The use of "cataloguing" here as a synonym for "metadata" highlights how much descriptive metadata is intimately connected to earlier physical records. For an exploration of how older data continues to shape readers' understanding of the books they see on their screens, see Echard, "Containing the Book," 104–105.

21 Soualah and Hassoun, "A TEI P5 Manuscript Description Adaptation," 2.

and paper, and between different textures of ink. Often we can't tell what the overall structure of the work is like, how many leaves it has and whether it contains any cancel leaves.²²

Elaine Treharne describes challenging her graduate students to determine the size of Cambridge, Corpus Christi College, Matthew Parker Library, MS 391 based solely on digital images from the Parker on the Web.²³ Her students' widely different suggestions (ranging from 15 cm x 7 cm all the way up to 76 cm x 45.7 cm) demonstrate how images alone are insufficient for assessing a book's size and function.²⁴ This is shared as evidence of the limitations of the digital image, all that it cannot successfully convey.

Yet this is also what descriptive metadata is for. Students in my course "Medieval Books in the Digital Age" (Binghamton University, Fall 2015) repeatedly emphasized that metadata helped them engage more deeply with medieval books. Working with original fragments, modern facsimiles, and digitized manuscripts, they came to argue that, in any medium, metadata is a book's voice—the tool by which that book tells users about itself, helping beginners understand and value what they see, helping experts uncover the complex web of connections that is the foundation for advanced research. While reading someone else's measurements of height and width may not help physically remote readers achieve what Treharne calls "optimal interpretative potential," recording this data shares important insights into those key details that cannot be discerned looking at pictures on screens.²⁵ Well before the rise of the so-called "digital age," print publications like J. J. G. Alexander's magisterial series *A Survey of Manuscripts Illuminated in the British Isles* gave manuscript scholars vitally important information, like the physical measurements, for distant manuscripts they also could neither see nor touch. In digital viewing environments, descriptive metadata fulfills a similarly important experiential function, helping bridge some of the inevitable gaps between eyes, mind, fingers, and screen.

Interoperability

"Interoperability" both in definition and in practice is as highly sought-after and elusive as a treasured relic in an Arthurian quest. The term's origins are rooted the mid-twentieth-century military discourse (in roughly the same decades that medievalists—valued for their practice "drawing conclusions from fragmentary evidence"—were

²² Edwards, "Back to the Real?" para. 6. The question of scale is further complicated by the fact that viewers may well use different sized screens to access the same digital copy of the same medieval book: smart phone, laptop, external monitor, tablet—all change the sense of how big the object might be and the viewer's bodily relation to it.

²³ Treharne, "Fleshing out the Text," 475–76.

²⁴ Treharne, "Fleshing out the Text," 476.

²⁵ Treharne, "Fleshing out the Text," 470.

entering the ranks of government programs like the Central Intelligence Agency).²⁶ Moving beyond Cold War military applications, the term was taken up in computing to describe “the ability of two or more computer systems or pieces of software to exchange and subsequently make use of data.”²⁷ In the context of information systems, the term now “denotes the ability of a system to work with or use parts of other systems.”²⁸ In digital libraries, “interoperability” has come to mean both the broader “potential for metadata to cross boundaries between different information contexts” and the much more specific ability “to exchange metadata between two or more systems without or with minimal loss of information and without any special effort on either system.”²⁹ Martin Foys defines “interoperability” as digital resources being “open in their data and connect[ing] to larger networks.”³⁰ To put it in more medievalist-friendly terms: consider the literary trope of the “envoy” in which the poet sends his or her creation to its intended audience, often by telling it, “go, little book.” *Interoperability* is the underlying philosophy—and hard work—that makes it possible for the (digital) book, once sent to a new home, to actually be read and understood with only minimal loss and transformation.

“Ther is so gret diversite in wrytyng of oure tonge,” Part I: Our Unalterable Inheritance of Codicological Variance

One of the largest challenges facing interoperability, and the creation of a single, universal digital manuscript hub, lies not in new technology but in old practice. For hundreds of years, the discipline of codicology has been characterized by widely divergent practices, with different schools and scholars developing extremely different ways of describing the same features in medieval manuscripts.

There is, for example, a centuries-long disagreement over what to call the animal-based membrane upon which a manuscript may be written. William Horman’s 1519 Latin-English phrasebook offers a range of terms one might choose to employ to describe this substance, all of which—significantly—he presents as equally valid:

That stouffe that we wrytte vpon: and is made of beestis skynnes: is somtyme called p[er]chement / somtyme velem / somtyme abortyue / somtyme

²⁶ On the military roots of “interoperability,” see John Unsworth, “Computational Work,” 1. See also, *OED* “interoperability” (n.1). On medievalists in the CIA, see Cantor, *Inventing the Middle Ages*, 261–62.

²⁷ *OED* “interoperability” n. 2.

²⁸ Haslhofer and Klas, “A Survey of Techniques,” 13.

²⁹ Haslhofer and Klas, “A Survey of Techniques,” 1, 8. See also Chan and Zeng, “Metadata Interoperability and Standardization” and Elings and Weibel, “Metadata for All.”

³⁰ Foys, 133. Foys briefly mentions metadata’s role in interoperability, but focuses his comments on new protocols like RDF and the creation of new metadata. How to continue to use the massive amount of already created metadata has yet to be raised in a sustained fashion by medievalists, even though the ability to continue to use important early digitization projects depends on engagement with legacy data as well as the latest database protocols.

me[m]bran. Parchement of the cyte: where it was first made. Velem / bycause it is made of a caluys skynne. Abortyue / bycause the beest was scante perfecte. Membraan / bycause it was pulled of [off] by hyldynge [skinning, or playing] fro the beestis lymmes.³¹

Some modern scholars share Horman's stance and insist that all these terms may be used interchangeably.³² Some use "parchment" as a generic term for any writing support made of animal membrane.³³ Others contend that "vellum" ought to be reserved for writing support made of calf skin, and use "parchment" to describe all other animal-skin writing support.³⁴ Still others draw a stricter distinction, reserving vellum for calf-skin and parchment only for goat or sheep.³⁵ In addition to disagreement between scholars, different libraries follow different cataloguing rules as they describe animal-skin writing supports.³⁶ As Christopher de Hamel notes, "In the manuscripts department of the Bodleian Library in Oxford the house usage today is to refer to the material consistently as parchment; in the British Library in London, the same substance is standardly called vellum."³⁷

Similar disagreements over practice occur in the proper methods for measuring leaves: millimetres, centimetres, or inches? The proper method for recording manuscript dates is also the subject of considerable variation. Ought palaeographers to use the traditional Latin abbreviations and roman numerals, like "s. XI"? Or Arabic numerals: "ca. 1250"? These different ways of recording manuscripts' dates vary between palaeographical schools. Sometimes, date usages also vary *within* the same palaeographical

31 Horman, *Vulgaria*, f.80v, via *Early English Books Online's* digital reproduction, of a microfilm reproduction, of the original print copy held in the Henry E. Huntington Library and Art Gallery.

32 See, for example, de Hamel, *Medieval Craftsmen*, 8.

33 See Clarkson, "Rediscovering Parchment," 5. The Library of Congress's *Thesaurus for Graphic Materials*, a key reference guide for metadata librarians, suggests "parchment" as the proper term for "skins and hides" and presents it as synonymous with, but preferable to, "vellum." Similar use of the term "parchment" for the entire genre of animal membrane can be found throughout da Rold, "Materials," and in Holsinger, "Of Pigs and Parchment."

34 See, for instance, the distinction drawn by Gillespie in "Bookbinding," 165n53.

35 On the terminology debates, see Ryder, "Parchment: Its History, Manufacture, and Composition," 392–93. Clemens and Graham also present an excellent concise overview of the terminology debates in *Introduction to Manuscript Studies*, 9–10.

36 Recently, shared standards for describing manuscripts have been created. However, these standards do not offer guidelines on the "parchment" versus "vellum" debate, preferring to leave that choice to local institutions. See *Descriptive Cataloging of Rare Materials (Manuscripts)*, 86. <http://rbms.info/files/dcrm/dcrmmss/DCRMMSS.pdf>. I am indebted to my colleague, and Binghamton University metadata librarian, Laura Evans for this reference. It is worth noting that although the standards do not differentiate between animal types for "parchment" or "vellum" or attempt to mandate what institutions ought to use, they appear to prefer "parchment" as their general term for animal-skin writing support. See, for example, the definitions of "membrane," "parchment," and "vellum," on pages 144, 145, and 147 respectively.

37 de Hamel, *Medieval Craftsmen*, 8.

school, even the same library, and the same book. For instance, both examples just cited—s. XI¹ and c. 1250—come from *A Catalogue of the Pre-1500 Western Manuscript Books at the Newberry Library*.³⁸

This variation can be problematic even when manuscripts are studied within their home institutions. It becomes even more problematic when these descriptions—as descriptive metadata—are brought together in a single digital manuscript aggregator. Aggregators like DMS-Index let researchers view side-by-side different manuscripts held at different institutions, and each of these manuscripts is described by its home institution's preferred codicological vocabulary. One knows, perhaps, at the Bodleian what “parchment” means, and at the British Library what “vellum” means—and, importantly, if one doesn't know one can ask a librarian. But what happens when users view these manuscripts from different institutions side-by-side in cyberspace, without the guiding wisdom of a human librarian well-versed in local cataloguing practice?

Picture a researcher uninterested in the long history of terminology variance in codicology. What this researcher is interested in, instead, is using the vast swathes of data now available online to analyze the regional differences in the use of calf and other animal skins in codex making. If she views side-by-side manuscripts from Harvard, the Huntington, the Bodleian, Cambridge University Library, the British Library, the Vatican Library, and the Bibliothèque Nationale de France, will she know what each of these libraries mean when they say “vellum” or “parchment”? Or, because she believes in the training she received at her particular palaeographical school, *and* because these institutions do not embed detailed explanations of their unique local practices within every single manuscript description they produce (because no one does), will she risk believing that “vellum” means something stable and dependable, across all institutions and manuscript she sees? Might she go on to trace animal-use variation based on the apparent difference highlighted in her manuscripts' metadata, when that variation, in fact, reveals nothing clear about the animals—but a great deal about curators and codicologists? And if she does go on to do this, who will be blamed when it is revealed that her data are irrevocably flawed? Certainly the researcher, but it seems quite likely that such research errors might also be used to single out digital manuscripts themselves as deceptive and flawed—when the problem lies much earlier, is in fact intertwined with and inseparable from the field.

Furthermore, more problems than just “parchment” versus “vellum” arise when combining manuscript descriptions made by different teams, at different moments, with different end goals in mind. The first iteration of DMS-Index successfully demonstrated interoperability by drawing together manuscript descriptions for medieval texts digitized by seven different partner institutions. But because of these partners' different local

³⁸ Example dates come from, respectively, the description for MS 1, a French book of homilies from the first quarter of the eleventh century, and the description for MS 19, a Franciscan Bible, also from France. Saenger, *A Catalogue of the Pre-1500 Western Manuscript Books at the Newberry Library*, 3, 35.

practices, this early DMS-Index retrieved records for 118 manuscripts on “parchment,” 14 more on “Parchment,” and 394 on “parch,” as well as 392 manuscripts on “Vellum” and 2 on “vellum”—rather than 920 manuscripts on the same animal-substrate³⁹ (if we allow ourselves to believe that “vellum” in this case means the same thing as “parchment,” which of course we should not). This wide range of search results occurred because computers cannot, without human help, intuit that “parchment” (lower-case p) is the same thing as “Parchment” (with an upper-case P)—let alone that both “parchment” and “Parchment” are the same thing as the convenient cataloguing shorthands “parch” and “perg.”⁴⁰ In their home environments, all of these terms are correct—ready, as it were, to go to the ball. But combined in an interoperable manuscript hub like DMS-Index, it is as though all of this metadata has, as John Unsworth puts it, “suddenly found itself in Union Station in its pajamas: it is not properly dressed for its new environment.”⁴¹

This data normalization issue is a familiar one to libraries and museums. To combat this almost gravitational pull toward variation, libraries and museums use “controlled vocabularies”: “an organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching.”⁴² Professional cataloguers use these specific, authoritative, controlled vocabularies to make sure that their terminology fits their community’s accepted professional standards, which in turn allows their data to be findable across all other institutions that share the same vocabulary.⁴³ However, not all of the creators and manipulators of metadata share this training. (I did not, when I began my postdoc.) As digital humanists and other scholars cross into record creation and cataloguing, both the quality—and any deliberate or fetishized uniqueness—of the metadata we create can and will affect interoperability.

This variation in terminology can be addressed in a number of ways in digitization projects. First, scholars and digital humanists can use established controlled vocabularies in all new metadata we create. Second, all creators of new metadata—in digital projects, as well as in libraries and museums—can minimize future fragmenting of terminology by not resorting to shorthand like “parch” or “perg” for “parchment” (although the underlying problem of “parchment” and “vellum” will remain). Third, data curators can expect

39 This specific example of how terminological choices can limit broader interoperability comes from conversations with Albritton early in my postdoctoral work.

40 A similar issue appears in searching “language” across different institutions’ metadata. In the earlier iteration of *DMS*, researchers find 326 manuscripts by searching “Latin.” and 48 if one searches “Latin” without the final period. For the computers we humans partner with in our research, “Latin and English.” is not the same as “English and Latin.”—and neither are the same as “Latin and English” (without the final period), “Middle English with occasional Latin phrases,” “Latin, English.” or the machine-readable, iso-639-2 language codes “lat” and “eng.”

41 Unsworth, “Digital Humanities Centers as Cyberinfrastructure,” quoted in Unsworth, “Computational Work with Very Large Collections,” 4.

42 Harpring, *Introduction to Controlled Vocabularies*, 12.

43 Museums tend to rely on the *United List of Artists Names* and the *Thesaurus of Geographical Names*. Libraries turn to the *Library of Congress Subject Headings* and *Thesaurus for Graphic Materials*.

and uncover these (and other) data normalization problems and then build solutions into the code they write to transform manuscript descriptions for interoperable aggregators. Fourth, aggregators might move partner institutions' metadata through programs like "Open Refine," where manuscript records containing Parchment (with a capital "P") can be transformed *en masse* into "parchment" (with a lower-case "p") and "eng" and "lat" can become "English" and "Latin." Fifth, inter-institution projects can develop shared nomenclature, and projects that hope to share their manuscripts via aggregators like DMS-Index might take on the terminology already in use for the manuscript aggregator they hope to join (although that runs the risk of bending their immediate project *toward* future goals and away from *immediate* research needs and/or funders' requirements). Sixth, researchers can carve out part of their research time to look into terminology variation between the institutions they work with and make sure to do a version of data normalization in their own research. The list of potential fixes could likely be expanded almost infinitely.

Ultimately, however, as useful as these and other fixes will be for future digital manuscript work, the problem of human-introduced variation is likely to continue, rearing its head—Hydra-like—in new and interesting ways just when we believe we have solved it for good. This is in part because humans are prone to diversification, and humanists are especially fond of coining neologisms to fit the specific thing—or problem—we are grappling with in that moment. (Take, for example, the evolving debates over what one ought to call digital photographs of medieval books—are they digital manuscripts? Digital facsimiles? Digital surrogates? Avatars? Metaobjects?) Manuscript aggregators, thus, are a constant negotiation between the needs of humans, who as academics might build their careers on the coining of neologisms and new terminology, and the needs of machines, which require precision and repetition. Human researchers get far more from phrases like "Middle English with occasional Latin phrases,"⁴⁴ "Latin with some later scribbles in Middle English,"⁴⁵ "Latin with some English added on folios at the end,"⁴⁶ "Latin with Middle English verse in margin of F. 107v," "Latin and Old French with some English on the flyleaf under paste-down," than we do "Latin" and "English." But the computers on which we rely for searching can use the latter, and not the former.⁴⁷

Making a manuscript aggregator, thus, is an endless negotiation between the living and the dead, between human and machine, between different palaeographical camps and different cataloguers working in different decades, between digital humanists/scholars rewarded for uniqueness and information professionals and librarians

44 "Manuscript Description," Corpus Christi College, Parker Library, CCCC MS 282.

45 "Manuscript Description," Corpus Christi College, Parker Library, CCCC MS 21.

46 "Manuscript Description," Corpus Christi College, Parker Library, CCCC MS 112.

47 There is an important distinction between what is produced by a targeted search using the "Language" facet versus what the broader net of a keyword search might generate. A keyword search would catch manuscripts with language descriptions like "Latin and Old French with some English on the flyleaf under paste-down," while a targeted language-search would not. At the same time, this sort of broad keyword search also fails to distinguish between "Chaucer" as author/creator and "Chaucer" as subject, or "Chaucer" as largely irrelevant reference buried deep in the bibliography.

rewarded for regularity, between what we have inherited, and what we now want and need, and what we might guess our heirs in the future might want and need too. But that negotiation happens, strangely, off stage for most manuscript scholars. Despite the fact that we are trained to see these negotiations between past, present, and future as they play out in physical medieval books, we do not tend to seek them out as they pertain to new digital copies of old physical books. Despite our expertise in changes of older text technologies, these newer technological changes are invisible to us, even though our work is facilitated, and perhaps sometimes curtailed, by the invisible labour of data curation happening off screen.

One fairly simple intervention in the data problems produced by inevitable human variations is for all digitization projects to openly share all their underlying, messy metadata—as well as the cleaned-up, human-readable display seen in search screens.⁴⁸ But as important as that sharing is, it lays bare the problem. It does not fix it. Another, very different, intervention can be done by medievalists. Even as we seek to maximize data normalization in and between our digital manuscript projects, a medievalist perspective can also offer an important layer of practical resignation to the soaring ambitions of much digital work. As Boethius might note, only God is perfect and changeless, above the mutability of worldly things. The state of humans, by contrast, is constant change. Whether or not one agrees with Boethius about God, his point about the unavoidable changefulness of human work is useful to recall in digital projects. Remembering human mutability, or—to offer a different analogy, also connected to medievalists—accepting that there is no “one true ring to bind them all” and there never will be, can keep us focused on what we are actually doing and can do. We cannot revert to some state of prelapsarian, pre-Tower-of-Babel, never-existed-in-the-first-place grace where manuscript scholars always used the precise same terminology in precisely the same way. Instead, we must learn to respectfully reuse each other’s data as they are and not as we might idealistically wish them to be. To return to Boethius and Lady Philosophy’s final lesson: in medieval digital humanities we must always strive to do good, but we must understand that—in large and small ways—we will always fail.

“Ther is so gret diversite in writyng of oure tonge,” Part 2: The Unalterable Fact of Different Metadata Standards

There is another cause of variation that has to be understood when dealing with digital manuscripts. To search and discover digital manuscripts, object descriptions must be wrapped in XML tags that are shaped according to metadata standards that libraries, museums, and other cultural heritage institutions use to encode descriptive metadata. Strictly speaking, these metadata standards are *encoding* standards, which is to say, they are guidelines that determine what type of descriptive (and administrative and structural) metadata can be recorded in each file and where in each object description

⁴⁸ The Digital Walters and OPenn are important pioneers for this kind of open sharing of raw data.

this information must appear.⁴⁹ The best-known of these metadata standards include the Text Encoding Initiative (TEI), Dublin Core (DC), the Metadata Object Description Schema of the Library of Congress (MODS), Encoded Archival Description (EAD), and the Metadata Encoding and Transmission Standard (METS). This chapter focuses on the first three—TEI, DC, and MODS—because these are the standards I worked with closely in my time on DMS-Index. However, the work of metadata sharing across standards/schemas and my larger arguments are applicable to EAD, METS, and other forms of metadata and mark-up.

The foundational truth that humanities scholars must grasp about metadata is that even though there are established standards and controlled vocabularies and we would do well to follow them, there is no “one best metadata standard for all digital medieval projects”—and there never will be. As of 2010, there were “over a hundred community-recognized metadata standards from which to choose” and “additional metadata standards will continue to emerge as new classes of digital resources are created and as new user needs are identified.”⁵⁰ Therefore, this chapter vigorously refuses to endorse a single metadata standard for all digitized medieval manuscripts. While TEI may be the most familiar XML-metadata standard for digital humanists and has long played an important role in the marking up and sharing of medieval manuscripts, it is by no means the only way to render medieval manuscripts’ descriptive metadata.⁵¹ Museums, libraries, and archives have long had their own cataloguing methods, standards, and needs for medieval manuscripts. As all these libraries, digital humanities projects, and cultural heritage institutions have moved into manuscript digitization, they have carried their cataloguing practices and metadata schemas with them.

Perhaps the second most important thing for medievalists and digital humanists to understand about metadata standards is that each of these standards was developed at a specific moment in history, in response to a particular community’s concerns and specific pressures, in order to serve that community’s larger needs and goals. Furthermore, as communities and needs have evolved, so too have the standards. Understanding—and sharing—digital manuscripts requires historicizing digitization projects *not* in the time that the manuscripts were created, but within the specific historical moment of the digital project’s creation. That moment, that community, and those goals all determine not just the

49 In other words, encoding standards will not tell you *how* that metadata ought to be encoded. TEI and DC do not tell you what to call parchment or whether you can abbreviate to “parch” or “perg” without inhibiting interoperability. They simply provide you with a field or a tag where you can enter the information as you (ideally with recourse to your field’s controlled vocabularies) see fit.

50 Cole and Han, *XML for Catalogers and Metadata Librarians*, 96. Digital Librarian Jenn Riley has created a remarkable visualization of the 105 most-discussed and best-known metadata standards used by cultural heritage institutions for different objects and different communities: “Seeing Standards: A Visualization of the Metadata Universe.”

51 The most recent major revision of the TEI guidelines includes a chapter dedicated to the best practices, as defined by the TEI consortium, for describing medieval manuscripts. P5, first released November 2007; at time of writing most recently updated March 3, 2016.

method of digitization but the form that the metadata takes. And the form that metadata takes determines how shareable and discoverable these digital resources really are.

Writing the full history of each schema/community is beyond the scope of this chapter. However, some basic knowledge of who these communities were (and still are), as well as the problems each metadata standard exists to help fix, reveals how impractical, impossible, and frankly undesirable it would be to attempt to select one universal metadata standard. Rather than engaging in a quixotic quest to erase the inevitable variations of individual practice, a better solution to interoperability begins with understanding the diverse communities of practice that have come together in shared digital manuscript environments, because this confluence of communities, goals, metadata standards, and values continues to shape—visibly and invisibly—the digital objects that increasingly drive our research.

TEI

Perhaps the most well-known of these XML-based metadata standards within digital humanities and literary studies is the Text Encoding Initiative (TEI). TEI is an internationally recognized set of guidelines developed for encoding electronic texts, such as novels, plays, and poetry: it functions primarily in support of humanities research.⁵² TEI was not created specifically for medieval manuscripts, nor for encoding their physical descriptions—although the fifth iteration of the TEI guidelines includes an extensive chapter dedicated to methods for encoding the physical descriptions of medieval manuscripts.⁵³ Since TEI was not created for medieval books, projects that use TEI for medieval manuscript metadata adapt techniques for encoding literary content to encode bibliographic content.

TEI was established in 1987, in response to the rapid growth of a wide variety of digital technologies “dominated by mutually incompatible formats.”⁵⁴ These incompatible formats imposed “serious technical obstacles even to the simple transfer of data files from one machine to another, to say nothing of the difficulties posed by mutually incompatible and proprietary file formats.”⁵⁵ Against this bubbling sea of one-off programs, TEI was developed to aid mutual comprehension, information sharing, and the building of tools that could be used on more than one humanities computing project. TEI’s founders came from a variety of academic backgrounds.⁵⁶ Their ambitious goal was to create “practical recommendations as to how an extensible set of guidelines consistent with the goal of a universal text-encoding scheme might be achieved.”⁵⁷ Such ambition is very much in keeping with late 1980s computing, in and beyond humanities disciplines.

52 NISO, “Understanding Metadata,” 4.

53 TEI: Text Encoding Initiative. “Chapter 10: Manuscript Description.”

54 Burnard, “The Evolution of the Text Encoding Initiative,” 3.

55 Burnard, “The Evolution of the Text Encoding Initiative,” 3.

56 Burnard, “The Evolution of the Text Encoding Initiative,” 2–3.

57 Burnard, “The Evolution of the Text Encoding Initiative,” 3.

As visionary as the creators of TEI were, they were not the only people in the late 1980s with “an expertise in the creation and management of digital text” concerned by the ways that a widespread lack of standardization was inhibiting the universal sharing of text and data.⁵⁸ Nor were they the only community seeking to aid mutual intelligibility and information sharing through metadata standardization across projects and institutions.

DC

Twenty years earlier, in 1967, the presidents and library directors of colleges and universities in Ohio created a nonprofit dedicated to “develop[ing] a computerized system in which the libraries of Ohio academic institutions could share resources and reduce costs.”⁵⁹ Dubbed the Ohio College Library Center, its shared cataloguing system debuted in 1971 and revolutionized library cataloguing and academic research, shepherding in (among other things) Interlibrary Loan as we know it today.⁶⁰ By the early 1990s, the Ohio College Library Center had expanded to all fifty states and had a new name: the Online Computer Library Center (OCLC), which is perhaps best known today as the organization behind the WorldCat global library catalogue.⁶¹ It is important to realize that the information professionals of the OCLC were not unaware of the standardization that troubled the founders of the TEI, but—for their community—that was not the real problem. Well before the digital revolutions that convulsed the 1980s, they had already developed a standard system for data sharing.

The problem that librarians faced, instead, was the creation and growth of the World Wide Web. By 1994 approximately half a million addressable objects were online, but they were not yet mutually comprehensible and discoverable.⁶² To solve this problem, in 1995 the OCLC and the National Center for Supercomputing Applications (NCSA) held a joint workshop in Dublin, Ohio.⁶³ Participants there developed a concise “standard set of 15 interoperable metadata elements designed to facilitate the description and recovery of document-like resources in a *networked* environment,” which they

⁵⁸ Burnard, “The Evolution of the Text Encoding Initiative,” 3.

⁵⁹ Rosenheck, “OCLC: From an Historical Perspective,” 2.

⁶⁰ This revolutionized the role of computers in academic libraries in two important ways: first, it “enabled libraries to rapidly and efficiently catalog books” because each library no longer had to catalogue by hand every book they received. If the book was already in the shared catalogue, they could use existing cataloging information. If it was not in the shared catalogue, they could catalogue it so that other partner libraries would be able to save time by reusing that information. Second, it “provided location information for the materials listed in the catalog by participating libraries:” users were suddenly able to easily find and request materials held at another partner library. Rosenheck, “OCLC: From an Historical Perspective,” 2.

⁶¹ For views on how OCLC and WorldCat transformed twentieth-century librarianship, see *What the OCLC Online Union Catalog Means to Me*.

⁶² “History of the Dublin Core Metadata Initiative,” Dublin Core Metadata Initiative.

⁶³ “History of the Dublin Core Metadata Initiative,” Dublin Core Metadata Initiative.

named Dublin Core (DC) after their meeting place.⁶⁴ Today, DC is “by far the most used descriptive metadata standard in the library domain,”⁶⁵ and in 2012 expanded to include a total of fifty-five elements and element refinements, in order to provide more, and more specific, metadata. Again, this metadata standard was not invented with medieval manuscripts in mind. But, because—when digitized and put online—manuscripts are, in fact, “document-like resources in a networked environment,” DC is an important player in digital manuscript metadata. This importance is also due to the fact that the Open Archives Initiatives Protocol for Metadata Harvesting (OAI-PMH), “which supports metadata sharing and interoperability,” “mandates the use of Simple Dublin Core as its lingua franca, or lowest-common-denominator XML metadata grammar.”⁶⁶ In other words, if a library or institution believes in the broader mission of the Open Archives Initiative, and wants all content in their repository to be interoperable with the enormous network of repositories that are part of the OAI, they must do at least part of their metadata according to the constraints of Simple DC.

Despite its deep roots in data sharing between academic libraries, DC is not *the* default metadata standard for all academic libraries, let alone all libraries, museums, and cultural heritage institutions that hold, digitize, and display content online. This is, in part, because within library communities “there has historically been some tension between supporters of a minimalist view” of metadata creation, “who emphasize the need to keep the elements to a minimum and the semantics and syntax simple,” and “supporters of a structuralist view, who argue for finer semantic distinctions and more extensibility for particular communities.”⁶⁷ The same concision that makes DC such a useful tool for sharing basic metadata across institutions limits the complexity of descriptions fit into DC. What this means is that DC, either in its original “Simple” 15-element form or in its expanded “Qualified” 55-element form, is an excellent solution for some kinds of metadata problems and goals, but not others.

MODS

In 2002, the Library of Congress developed the Metadata Object Description Schema (MODS) as a middle path between the forceful simplicity of DC and the complexities that many cataloguers and metadata librarians still desired.⁶⁸ As the National Information Standards Organization puts it,

Rich description of electronic resources is a particular focus of MODS, which provides some advantages over other metadata schemes. MODS elements are

64 “Dublin Core (DC),” *Dictionary for Library and Information Science*, 234, my italics. See also, “Understanding Metadata,” 3.

65 Following the descriptive metadata system known as MARC. Cole and Han, *XML for Catalogers and Metadata Librarians*, 98.

66 Cole and Han, *XML for Catalogers and Metadata Librarians*, 101.

67 NISO, “Understanding Metadata,” 3.

68 The history and use of MARCXML, while not part of this chapter, is covered extensively in Cole and Han.

richer than the Dublin Core; its elements are more compatible with library data than ... Dublin Core standards; and it is simpler to apply than the full MARC 21 bibliographic format.⁶⁹

MODS, that is, helped update and streamline earlier, complicated bibliographic formats while simultaneously giving libraries a richer set of elements and attributes for describing their holdings than the fifteen elements of Simplified DC.⁷⁰ Once again, MODS was not created with medieval manuscripts in mind, but digitized manuscripts are, in fact, “electronic resources,” and it makes sense for the creators and maintainers of library catalogues who use MODS to make their digital manuscript holdings discoverable alongside their other physical and electronic holdings.

If TEI was created to fit the needs of humanists and academic researchers and DC was created to solve the problems of information professionals seeking to harness the networking power of the internet, MODS was created to serve the needs of individual libraries, which need to be able to catalogue diverse resources throughout their collections in ways that facilitate searching while not flattening out the complexities of vastly different objects. Like TEI and DC, MODS was not created specifically for describing medieval manuscripts, but as digitized manuscripts enter library catalogues—if those catalogues are already filled with thousands, or millions, of items described according to the local interpretation of MODS—then digitized manuscripts’ metadata will likely be slotted into MODS elements, so that patrons can find these digital resources when they search the library’s catalogue.

To reframe these brief histories: it bears repeating that *none* of these standards were developed for medieval manuscripts, nor medievalists, nor manuscript scholars. MODS was developed by librarians and cataloguers for librarians and cataloguers. It offers more opportunity for granularity and fine-grained details than DC, but that granularity comes at a cost: even as best practices and rules for implementers proliferate, “MODS records from different institutions and different projects tend to be less consistent at this time.”⁷¹ By contrast, DC was also developed by librarians, in partnership with information technologists in academia, government, and industry. DC creates opportunities for impressive regularity across wildly diverse digital collections held at different institutions and repositories. This cross-collection regularity, however, comes at the cost of the granularity in individual description. TEI, by contrast, was developed largely outside of academic libraries—it is primarily by (humanities) researchers for (humanities) researchers. Its flexibility makes it endlessly adaptable to particular projects’ and researchers’ needs, but often has little to do with libraries’ and/or cataloguers’ needs.

⁶⁹ NISO, “Understanding Metadata,” 5–6.

⁷⁰ In the increasingly complicated timeline of evolving metadata standards, this places the creation of MODS before the expansion of DC to “Qualified Dublin Core,” with its fifty-five elements. Qualified DC was created after MODS, and fulfills some, but not all, of the same granular functionality of MODS.

⁷¹ Cole and Han, *XML for Catalogers and Metadata Librarians*, 111. For the MODS best practices document I consulted, see the Library of Congress, “DLC/Aquifer Summary of MODS Requirements and Recommendations Table.”

As data curator for DMS-Index, my job was to aggregate metadata from different collections, marked according to different metadata standards, which were always already shaped by very different communities, goals, and needs (see, for example, Table 8.1). The Walters Art Museum marked up its descriptive metadata according to TEI P5 in order “to ensure broad access to [their] manuscript descriptions” and ensure “that [their] manuscript descriptions be available to a wide range of users and can be rendered in multiple output formats (such as PDFs, HTML web pages, and page turning applications).”⁷² However, because Stanford Library’s Digital Object Registry represents its descriptive metadata in MODS, and because Stanford and the Walters entered into a data sharing agreement shortly before my postdoc began which allowed Walters manuscripts to be discoverable through Stanford’s catalogue, I needed to turn the Walters Art Museum’s TEI into Stanford’s local implementation of MODS. Because Stanford Libraries was also home to the first iteration of DMS-Index, there were good institutional reasons for continuing to have MODS be the metadata standard underlying DMS-Index. Parker on the Web, by contrast, was significantly older than the other digital manuscript repositories this chapter covers. It predated the TEI P5 guidelines used by Walters and e-codices, and—adhering to accepted TEI standards when the Parker Project began—encoded existing manuscript catalogues in the same way other digital humanities projects at the same time were encoding novels, plays, and poems. Because ecodices is an inter-library digital repository, its needs as a collaboration of more than forty libraries across Switzerland made DC a logical metadata choice. At the same time, a profound need for highly detailed manuscript descriptions inspired the additional use of TEI P5 as a second, more granular metadata standard. Reading these repositories’ medieval manuscript metadata in this way, with and against each other, and with profound interest in the logic behind each repository’s choice of schema, demonstrates, as Johanna Drucker has eloquently argued, that

metadata schemes must be read as models of knowledge, as discursive instruments that bring the object of their inquiry into being, shaping the fields in which they operate ... Analysis of metadata and content models, then, is an essential part of the critical apparatus of digital humanities.⁷³

Crosswalks and Transforms

Because there never has been, and never will be, a single metadata standard appropriate to the needs of all projects, objects, and stakeholders, libraries have developed terminologies and best practices for taking information expressed in one metadata standard and remaking that same information according to the rules and constraints of another metadata standard. The first step is creating “a crosswalk.” A crosswalk, as defined by the National Information Standards Organization (NISO), is

a mapping of the elements, semantics, and syntax from one metadata scheme to those of another. A crosswalk allows metadata created by one community

⁷² “Describing Manuscripts with TEI,” The Digital Walters.

⁷³ Drucker, *SpecLab*, 11.

to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme.⁷⁴

Also sometimes called “field mapping” or “metadata mapping,” a crosswalk is generally represented as a table, or a chart, showing “equivalent or nearly equivalent metadata elements or groups of metadata elements within different metadata schemas.”⁷⁵ The phrase—“nearly equivalent metadata elements”—might seem to suggest a close, or close-enough, fit. But in reality, different metadata standards can present the same information in markedly different ways (see Table 8.1, row 3).

Moreover, crosswalk building is rarely as simple as these basic descriptions might suggest: metadata is created at different times, by different institutions, according to different standards, to serve different needs. Thus, different sets of metadata rarely fit together seamlessly or, in some cases, without a great deal of shoving. A particular challenge for medieval manuscript aggregators is the fact that, as NISO notes “the mapping of schemes with fewer elements (less granularity) to those with more elements (more granularity) is problematic.”⁷⁶ For example, the DC element <creator> is “an entity primarily responsible for making the resource.”⁷⁷ In a single-author illustrated manuscript, <creator> encodes not just author, but also translator(s), artist(s), and/or scribe(s) since they are also, according to long-standing custom in manuscript studies, also acknowledged as “entities primarily responsible for making” the manuscript. In manuscripts that compile works by different authors, or that were copied by several different scribes, or that contain work by more than one artist, the list of “entities primarily responsible for making” tagged by <creator> explodes. By contrast, in TEI’s most recent guidelines for manuscript descriptions there is a much more targeted element for the specific creator-figure <author>. Since authors are not the only humans responsible for the making of medieval books, institutions like the Walters Art Museum have developed ways of creatively adapting the TEI element <respStmnt> to acknowledge the unique contributions of manuscripts’ scribes, translators, and artists. E-codices uses a different TEI element—<persName role= “author”>—to achieve the same precision. And, of course, all of these contributors are acknowledged in a different way in MODS (see Table 8.2).

The crosswalk, or mapping, is only first step for sharing metadata and manuscripts between institutions. After a crosswalk has been written showing how the data ought to be transformed, the data still has to actually *be* transformed. In my work on DMS-Index, this was done by writing a custom XQuery script (also called a “transform”) in the program oXygen for each institution’s digital manuscript collection. For each institution I created one script, which would take tens, hundreds, sometimes more than

⁷⁴ NISO, “Understanding Metadata,” 11.

⁷⁵ See “crosswalk” and “metadata mapping” in “Glossary,” Gilliland, *Introduction to Metadata*.

⁷⁶ NISO, “Understanding Metadata,” 11.

⁷⁷ “Term name: creator,” dublincore.org.

Table 8.1. Comparison of XML-encoded metadata schemas employed by participating institutions, along with a sample of how each institution represents the concept/field of the maker of the intellectual work in their local interpretation of their preferred metadata schema(s).

| Institution | Stanford University | Walters Art Museum | Parker on the Web | e-codices |
|---------------------|--|-------------------------------|-------------------------------|--|
| XML Metadata Schema | MODS | TEI P5 | TEI, P4 | OAI:DC TEI P5 |
| Sample crosswalk | <name> <namePart>Boethius </namePart> <role> <roleTerm>author </roleTerm> </role> </name> | <author>Boethius </author> | <author>Boethius </author> | <creator>\Boethius<creator> AND ALSO <author>Boethius</author> |

Table 8.2. Sample crosswalking of how the three example participating institutions mark up a work’s creator(s), according to each institution’s local interpretations of their chosen metadata schema(s).

| Metadata Standard | Walters TEI (P5) | Stanford MODS | e-codices Simplified DC | e-codices TEI, P5 |
|--|---------------------------------|--|-------------------------|--|
| Crosswalk for the concept “author” | <author> | <name type=“personal”> <namePart> <role> <roleTerm>author | <creator> | <author> AND ALSO <persName role=“author”> |
| Crosswalk for the concept “scribe” | <respStmnt> <resp>scribe | <name type=“personal”> <namePart> <role> <roleTerm>scribe | <creator> | <persName role=“scribe”> |
| Crosswalk for the concept “artist” | <respStmnt> <resp>artist | <name type=“personal”> <namePart> <role> <roleTerm>artist | <creator> | <persName type=“scribe”> |
| Crosswalk for the concept “translator” | <respStmnt> <resp>translator | <name type=“personal”> <namePart> <role> <roleTerm>translator | <creator> | <persName role=“translator”> |

1,000 individual metadata documents, and make new versions of those descriptions in new metadata schema. This step is known as “transforming” the metadata. For those concerned about digital *mouvance* run amok, it is important to note that this process does not fundamentally change the parent record. It is creating a new copy of the manuscript description, in the new schema, that—ideally—loses or changes as little of the original content as possible. However, as will become clear in my discussion of transforming the Walters Art Museum digital manuscript metadata, some data loss is almost always inevitable in metadata transformations.

Sympathy for Adam Scriveyn: Transforming the Walters Metadata

As a medievalist specializing in English literature, I work in a field haunted by lost books: from the Reformation reconfigurations and destruction of medieval libraries, to the Cotton fire of 1731, to the mid-century socialist biblioclasm of Otto Ege and the ongoing, present-day commercial destruction of books like the Hungerford Hours.⁷⁸ And, despite my straightforward pronouncements above on inevitability of loss and change, when I began curating manuscript metadata for DMS-Index I felt the weight of centuries of destruction bearing down on me. Although I knew, intellectually, that the loss of descriptive metadata crosswalked into a new digital environment was *not* actual physical destruction enacted on medieval books, I still felt I had to preserve all that was known about each digital medieval manuscript assigned to my care. If I failed, it seemed a betrayal of my professional, and ethical, duty to serve and preserve the past and the dead.⁷⁹

My preservation principles were not merely driven by dread of professional failure. They were also influenced by what can only be described as a kinship I was beginning to feel with medieval scribes—particularly those cursed by medieval authors and modern editors for their failures to properly copy the text. I knew that, at least in some circles, my failures in metadata curation might be cursed as quickly and smoothly as Chaucer (if in fact he wrote the lyric) curses his scribe in “Adam Scriveyn.”⁸⁰ The final line of the short poem describes the scribe’s poor copying as the result of the Middle English

78 On the Renaissance reconfiguration of medieval libraries, see Summit, *Memory’s Library*. On the Cotton fire, see Prescott, “Their Present Miserable State of Cremation.” On Ege, see Gwara, *Otto Ege’s Manuscripts*. See also Benjamin Albritton’s work digitally reconstructing Ege’s dismembered books and Morcos, “Piecing Together the Puzzle of the Hungerford Hours.”

79 Kiss et al. touch on similar dilemmas and fears of data loss faced in textual editing as it moves from print to digital modes of production. See “Old Light on New Media,” 20. Searle’s “Possible History” was a touchstone text, discussing medievalists’ duty to the past and the dead. See especially Searle, “Possible History,” 779.

80 By and large, the fifteenth-century scribe John Shirley’s attribution of the poem to Chaucer has been accepted in modern scholarship. For a provocative case against attributing the poem to Chaucer, see Edwards, “Chaucer and ‘Adam Scriveyn,’” 135–39.

word “rape,” which can mean either “haste” or “sexual assault and violence.” In my first weeks transforming manuscript metadata, I dreaded being indicted for enacting similar violations on texts given to my care.

Beyond lingering fears of being cursed by my colleagues, my developing curatorial principles were also influenced by my research in fifteenth-century English poetry and poetics, in which there is a strong precedent for expansive principles of preservation. Writing on John Lydgate, David Lawton offers backhanded, yet entirely accurate, praise: “this is one of his most impressive characteristics, that he is no trimmer.”⁸¹ Although praise of Lydgate’s expansive principles of inclusion tends to be shadowed with irony, my engagement with Lydgate’s curatorial practice was decidedly straightforward. As a translator and transmitter of inherited continental literary traditions, Lydgate was no trimmer; as a data curator, it turns out, neither was I. Thus, with Adam Scriveyn as my negative exemplar and John Lydgate as my positive model, I sought to reproduce the entirety of the original metadata entrusted to my care.

At times, when the different elements used by the Walters Art Museum’s particular dialect of TEI and those favoured by Stanford’s local interpretation of MODS matched closely, it was easy to adhere to this idealistic vision of data curator-as-perfect-scribe. For instance, the Walters’ TEI element `//msContent/msItem` may be crosswalked directly to the Stanford MODS element `//relatedItem type=“constituent.”`⁸² Although the match is not quite as seamless between the ways the different metadata standards denote authors (see Table 8.2, above), it is still straightforward and fairly easy to carry over that information from the Walters TEI to Stanford MODS.

However, as the granularity and richness of the TEI metadata increases, the ability of MODS to contain that data lessens. The problem is, from the standpoint of the crosswalk builder, that there are no matching elements in MODS for TEI’s specialized elements for describing writing support, dimensions for the average leaf size and the written space, foliation, collation, signatures, catchwords, layout, hand description, and decorations. Nor does MODS support the kind of nesting of items within items within items that TEI allows. Because it was created to serve broader collection needs than just those of medieval manuscripts, MODS does not have pre-existing elements or attributes to be pulled into service here. Instead, recognized MODS elements had to be bent and stretched into scarcely recognizable shapes to fit my exacting curatorial principles. My commitment to aggressive preservation paired with the formal constraints of MODS, fostered within my Walters TEI-Stanford MODS crosswalks an increasingly long and baggy collection of `<note>` elements, all modified by the attribute `@displayLabel`, in order to contain all possible manuscript descriptors (see Table 8.3).

Readers might be tempted here to dismiss MODS as a useful tool for manuscript metadata, but that would miss my point. A rigorous codicology of the digital manuscript

⁸¹ Lawton, “Dullness and the Fifteenth Century,” 779.

⁸² Here, for simplicity’s sake, I show the XPaths for the particular example lines, as they appear in the Walters and Stanford’s XML metadata documents. This is not how it is actually encoded in the XML itself in either institution.

Table 8.3. Sample crosswalk of Walters TEI to Stanford MODS demonstrating how heavily the @displayLabel attribute was worked to hold important metadata that had no other place in the MODS.

| Walters TEI | Stanford MODS |
|--|---|
| <msContents> <textLang> | <note displayLabel="Language Note"> |
| <msContents> <mslItem> <locus> | <note displayLabel="Locus"> |
| AND <msContents> <mslItem> <mslItem> <locus> | |
| <msContents> <mslItem> <rubric> | <note displayLabel="Rubric"> |
| AND <msContents> <mslItem> <mslItem> <rubric> | |
| <msContents> <mslItem> <incipit> | <note displayLabel="Incipit"> |
| AND <msContents> <mslItem> <mslItem> <incipit> | |
| <physDesc> <objectDesc> <supportDesc> <extent> | <note displayLabel="Foliation Statement"> |
| <physDesc> <objectDesc> <supportDesc> <extent> <dimensions type="leaves"> <height unit="cm"> <width unit="cm"> | <note displayLabel="Dimensions"> |

(continued)

Table 8.3 (Cont.)

| Walters TEI | Stanford MODS |
|--|--|
| <code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><extent></code> <code><dimensions type="written"></code> <code><height unit="cm"></code> <code><width unit="cm"></code> | <code><note displayLabel="Writing Block"></code> |
| <code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><collation></code> <code><p></code> <code><formula></code> | <code><note displayLabel="Collation: Formula"></code> |
| <code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><collation></code> <code><p></code> | <code><note displayLabel="Collation: Notes"></code> |
| <code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><signatures></code> | <code><note displayLabel="Signatures"></code> |
| <code><physDesc></code> <code><objectDesc></code> <code><supportDesc></code> <code><collation></code> <code><p></code> <code><catchwords></code> | <code><note displayLabel="Catchwords"></code> |
| <code><physDesc></code> <code><objectDesc></code> <code><layoutDesc></code> <code><layout @columns></code> | <code><note displayLabel="Layout: Columns"></code> |
| <code><physDesc></code> <code><objectDesc></code> <code><layoutDesc></code> <code><layout @ruledLines></code> | <code><note displayLabel="Layout: Ruled Lines"></code> |

(continued)

Table 8.3 (Cont.)

| Walters TEI | Stanford MODS |
|--|--|
| <code><physDesc></code> <code><objectDesc></code> <code><layoutDesc></code> <code><layout></code> | <code><note displayLabel="Layout: Notes"></code> |
| <code><physDesc></code> <code><bindingDesc></code> <code><binding></code> | <code><note displayLabel="Binding"></code> |
| <code><physDesc></code> <code><decoDesc></code> <code><decoNote></code> | <code><note displayLabel="Decoration Note"></code> |
| <code><history></code> <code><origin @notBefore @notAfter></code> <code><origDate></code> | <code><note displayLabel="Date Note"></code> |
| <code><history></code> <code><provenance></code> | <code><note displayLabel="Provenance"></code> |
| <code><history></code> <code><acquisition></code> | <code><note displayLabel="Acquisition"></code> |

does not retrospectively judge *whether* MODS was “the right choice” (whatever that might mean), but rather seeks to understand *why* MODS was chosen. What larger social and institutional pressures made MODS the correct choice for these project-builders in this moment? And what might that choice reveal about the larger structures in place that support the creation of digital manuscripts? And, perhaps most pressing, what does this metadata standard reveal about the nameless, unseen humans behind the digital manuscript we see hovering on our screens?

One notable discovery, once one digs into the underlying philosophies of information organization revealed by an institution or project’s chosen metadata schema, is that TEI manuscript-description may ultimately suggest collections of like-objects grouped with like. Nineteenth-century novels read against other nineteenth-century novels. Medieval manuscripts grouped with other manuscripts. These TEI descriptions are rich and lovely, but may also be somewhat cut off from the larger transhistorical movements of text and text technologies by virtue of that same delicious precision. By contrast, because MODS is for describing libraries’ collections—across time and media—use of this metadata schema for medieval manuscripts highlights similarities between medieval manuscripts and other, non-medieval, non-manuscript items. TEI celebrates manuscripts’ difference from all other forms of text. MODS highlights the similarities. The choice of one schema tunnels users ever deeper into the minutiae of medieval books, the other moves users along a much broader and more diverse continuum of text. Both have their gifts, as well as their costs. But rather than making a swift comparison and then dismissing MODS as

the wrong tool for the wrong job, a codicology of digital medieval books along the lines of what I advocate considers all of these things.

Even with my heavy reliance on MODS `<note @displayLabel>` to contain the information tagged by the Walters in separate TEI elements, the full granularity and precision of the Walters' metadata could not always be preserved. A particular challenge lay in the `<note>` elements nested inside `<msItem>` elements (which appeared alongside other nested elements like `<incipit>`, `<rubric>`, and so on). The Walters cataloguers and metadata creators drew distinctions between different types of notes: `<note type="content">`, `<note type="text">`, and `<note type="work">`.⁸³ MODS has a `<note>` element, which I was already using enthusiastically for things like `<note displayLabel="rubric">`, so in theory I could easily turn the Walters' ever-more granular `<note>` elements into matching MODS `<note>` elements. However, if I were to transfer into MODS the Walters' careful distinctions between `@type="text"` and `@type="content"`, I would need to create separate `<note>` elements for each type: the Walters `<note type="content">` *could* be matched to a Stanford `<note displayLabel="Note(s) about Manuscript's Content">`, the Walters `<note type="text">` to a Stanford `<note displayLabel="Notes about Textual Variation">`. But should it? While this would preserve the Walters' elegant precision, at a certain point I risked capsizing the metadata's usefulness under the weight of an expanding series of notes about notes about notes about notes.

If this project was merely about one institution's data, such a solution might be feasible. However, because my early work was laying a foundation for a much larger interoperable environment, I could not design with just the Walters in mind. In my work with the Walters, I was also designing what might be considered the general "metadata envelopes" that I would go on to reuse for all subsequent institutions who shared their manuscripts and metadata via DMS-Index. While I included a statement encouraging users of DMS-Index to seek out further information with the original holding institution, the fact remained that the majority of readers in DMS would not know, at a glance, what the Walters' distinctions between "text," "content," and "work" meant when read alongside manuscripts from Oxford, Cambridge, and the Bibliothèque Nationale de France, or any other institutions who might not draw the same distinctions in their metadata.⁸⁴

Consider the profoundly distorting effect on readers if they were viewing a Walters manuscript alongside one from the Lund University Library in Sweden. The Walters uses meticulous sub-types of notes within notes to highlight different information. To achieve a similar effect, the Lund Library uses a single `<note>` element with each `<msItem>`, and

83 See The Digital Walters, "Describing Manuscripts with TEI."

84 The obvious response and solution here is project documentation, but this is not always enough. The Walters is a model in meticulous documentation. This documentation stands outside the individual metadata documents I was processing. Therefore, because it functions as a kind of digital paratext, standing outside the text of the metadata proper, this documentation would not automatically be seen in DMS. While I included links in each transformed metadata set back to the original metadata record, these links only mean that readers of Walters metadata in DMS would have the opportunity to follow links back to the original metadata in its original environment and from there might dig out the original project documentation.

within that <note> parent element employs the child element <p> to indicate a break. Both institutions have similar goals, but use different XML tagging solutions to achieve them. But the formal properties of the XML are not commensurate. Imagine if I preserved all the Walters' separate sub-types of <note> and simply matched them to the Lund <note>. An unintended effect of my structural preservation of the Walters' original form might be that the Lund's cataloguing would look less meticulous and rigorous in comparison to the Walters because the Walters had many categories to the Lund's one overarching information unit. Alternately, the effect might also be that the Walters could appear curiously fragmented, and therefore lessened, against the Lund's unitary vision of what the <note> should be and do. Neither case would be true: my metadata forms would be participating in a kind of suggestive untruth, pushing readers to see difference rather than showcasing coherence across continents, countries, and libraries. Faced with the need to create metadata that would serve across standards and across different institutions' cataloguing and XML approaches, I created a single MODS <note displayLabel="Catalog Note(s)"> to contain all the Walters' TEI <note> elements. I did so to create a standard metadata form that would serve all institutions who chose to share their manuscripts and metadata. But I also knew that in so doing I was failing to create a perfect copy of the metadata entrusted to my care. To serve the greater project, my zeal for perfect changeless preservation had to begin to crack.

Similar challenges emerged when dealing with metadata concerning the Walters' manuscripts' languages. The Walters systematically seeks to serve the needs of human researchers and computer systems by pairing machine-readable information with human-friendly nuanced text.

`<textLang otherLangs="dut" xml:id="n496.203230" mainLang="lat">The primary language in this manuscript is Latin. The secondary language of this manuscript is Dutch; Flemish.</textLang>`⁸⁵

It is simple to write an XQuery script crosswalking the information enclosed by the TEI element <textLang> tag into similar tags using DMS-Index-style MODS. However, machine-comprehensible information about manuscripts' languages was locked within the @otherLangs and @mainLang attributes inside the Walters <textLang> element. The easiest way to move this information into MODS would have been to command oXygen to directly copy the three-letter language codes ("dut" and "lat") from the @mainLang and @otherLangs attributes. But this would serve neither the needs of humans, nor computers. After all, computers do not know that "lat" means "Latin" unless humans tell them. So, "the easy fix" would result in "lat" and "dut" being listed as languages, instead of "Latin" and "Dutch" or "Flemish," thus repeating and exacerbating all the problems of "parchment" versus "parch" described earlier.

Even more than being a good, dutiful scribe and "no trimmer," my goal as data curator was to begin to help undo part of the digital Tower of Babel that develops when different institutions' metadata jostles together in interoperable settings like DMS. Therefore,

85 W.169, my italics.

I wrote a series of *if-then-else* statements in my XQuery script that ran through every possible three-letter language code in the Walters digitized manuscripts, in both @mainLang and @otherLangs attributes. (Doing so added more than 100 lines to my script.) Each time my program saw “lat,” “dut,” “gre,” or any other three letter language code in the inherited, “parent” metadata, its job was not to copy that language code into MODS element but to write in a new line of metadata. When the code saw “lat” in Walters TEI, it wrote in Stanford MODS

```
<language>
    <languageTerm>Latin</languageTerm>
</language>
```

When it saw “arm” in Walters TEI, it wrote

```
<language>
    <languageTerm>Armenian</languageTerm>
</language>
```

and so on. In short, to preserve the metadata, I had to change the metadata. *Mouvance* was inevitable. Despite my deep desire to dutifully copy my exemplars, even when those exemplars were characterized by beautifully organized XPaths, interoperability mandated that I could not *not* be Adam Scriveyn.

Curation is Not the Same as Hoarding: Lessons Learned from Parker on the Web

Unlike the Walters, the descriptive metadata for manuscripts accessible via Parker on the Web is not produced according to TEI P5, Chapter 10 guidelines. This is because the Parker on the Web app predates the creation of those TEI P5 guidelines on how to describe medieval manuscripts. Instead, the Parker team drew on what was considered best practice at the time the project was being designed, using then-current TEI P4 guidelines to encode M. R. James’s print catalogue (1909–1912), with some changes, additions, and customization.⁸⁶ What this means is that the formal structures of the

⁸⁶ “Manuscript Descriptions,” in “About the Project: Project Technology,” Parker Library On the Web: About the Projects. The decision has been the subject of pointed critique, most notably in Ken Pennington and A. S. G. Edwards’ 2013 *TLS* exchange (Edwards’ piece can be found at: www.the-tls.co.uk/articles/private/better-on-the-continent/. Pennington’s response is at: www.the-tls.co.uk/articles/private/lets-get-digital-2/. Edwards’ reply is at: www.the-tls.co.uk/articles/private/digital-texts/.) For a reasoned consideration of the social and economic pressures that influenced this decision by the Parker team, see Rundle, “Virtual Manuscripts and the Real World. Part I.”

Parker library metadata are very different from those of the Walters Art Museum's metadata. While the Walters presents its metadata as short phrases nested along regular XPaths, and these XPaths are consistent across manuscript descriptions, Parker on the Web marks up prose. In the Parker on the Web metadata, long paragraphs of description are punctuated with TEI elements bracketing important search information.⁸⁷

Due to the funding structures and project mission for my position, I wrote only one transform per repository, which meant that my transforms were limited to copying data that occurred in predictable XPaths across the majority of manuscripts in that repository. For instance, James's manuscript descriptions often include a heading for Decoration Notes, which, when it is used, appears consistently along the XPath

```
/TEI/text/body/msDesc/p[@function="Decoration"]/decoNote
```

Using this XPath, my XQuery transform could pick out all information found between the <decoNote> tags at this path and then copy that same information into the appropriate MODS elements using the following command:

```
{for $f in $y/text/body/msDesc/p[@function="Decoration"]
return
<note displayLabel="Decoration Note">{data($f/fn:normalize-space())}</note>}
```

which, essentially, tells oXygen to take whatever information it finds along the XPath `TEI/text/body/msDesc/p[@function="Decoration"]` and copy it into the space between the start-tag `<note displayLabel="Decoration Note">` and the end-tag `</note>`. However, a number of manuscripts in Parker on the Web do not have the element `<p function="Decoration">`: from them, the command would collect no information along this path. Some of these manuscripts without the element `<p function="Decoration">` contain decoration notes elsewhere in their metadata. For example, Corpus Christi Collect Cambridge, Parker Library MS 153 has no `<p function="Decoration">` but has two `<decoNote>` elements retrievable at the XPath `/TEI/text/body/msDesc/p/decoNote`.⁸⁸ Decoration notes for Corpus Christi College Cambridge, Parker Library, MS 304 appear not at `/TEI/text/body/msDesc/p[@function="Decoration"]` but instead at `/TEI/text/body/msDesc/p[@function="Research"]/q/decoNote`. For the data curator tasked with writing a single transform to work for an entire collection, such variation in XPaths is a significant challenge, because it means that similar information is stored at very different places, all of which need to be taken into account.

ARC HUMANITIES PRESS

87 Paratextual elements, including paragraph breaks and running headnotes of the pages of the James catalog, are also all preserved in Parker's XML tags.

88 See lines 96 and 98 in the descriptive metadata.

Furthermore, taking these other paths “into account” is not just a matter of obsessively close reading metadata patterns until all possible options emerge.⁸⁹ In the Parker on the Web metadata, some of the information framed by granular tags like the <decoNote> mentioned above would be of extremely limited—or no—value once it has been extracted from its original setting and served out in an interoperable manuscript hub. For example, the descriptive metadata for Corpus Christi College Cambridge, Matthew Parker Library MS 304, as it is marked up in the Parker on the Web’s original P4 TEI, has eight extractable <decoNote> elements along the XPath /TEI/text/body/msDesc/p[@function=“Research”]/q/decoNote. But while the first is of undeniable value for researchers encountering the manuscripts’ metadata in an external manuscript hub like DMS-Index, the latter seven contain content of limited value for reuse outside Parker on the Web (see Table 8.4).

Again, I do not want to be misunderstood. This is not a critique of what the Parker team did or did not do in marking up their metadata. Although those critiques have been, and continue to be, levelled (and may be of some value for new projects seeking models), such critiques are not particularly useful for anyone seeking to use the metadata of the Parker Project manuscripts *as it exists* today. One may dream of an ideal world in which every digitization and digital humanities project meticulously designs its metadata to simultaneously serve the needs of the project at hand, as well as all possible needs of all possible future inheritors of that data. One may also dream of an ideal world in which legacy projects, like the Parker, are given so many financial resources that they can hire a host of workers to retype the entire Parker Library catalogue according to the current TEI P5 manuscript description guidelines. While one is dreaming, one might also dream up a related endowment so that the Parker, and all libraries around the world like it who were early innovators in digitization, can be rewritten every time metadata standards evolve or the TEI changes. In the world in which we actually live and work, however, projects come into being in their own historical moment; their design is driven by that project’s particular needs and goals; and their metadata is shaped accordingly.

As a data curator, and as a literary scholar interested in the long evolution of text technologies, I have no interest in critiquing this project or arguing for that new method or standardization. Instead, I am interested in the concrete facts that underscore all digital work, and which therefore underscore the possibilities of that universal portal to all manuscripts that medievalists speak of in hushed and excited tones. Not all data can be reused. The painful truth is that, for various reasons, there will always be some data that cannot be translated from one project to another. And, in fact, in the case of <decoNote> here, translating that metadata into the new schema would result in a less useful final product for researchers using an aggregator like DMS-Index, a manuscript description riddled with enigmatic fragments that served former project needs but make little sense in the new system. As I worked with the metadata of the Parker Library, I set aside my motto of “be no trimmer” and—in collaboration with my project manager Benjamin

⁸⁹ Taken to the extreme and committing to doing this to all 500+ manuscripts currently available in Parker on the Web, this approach would be extremely time-consuming, and still would not guarantee perfect extraction of the metadata.

Table 8.4. Descriptive metadata on the decorations and embellishments of Corpus Christi College Cambridge, Parker Library MS 304 as that information appears in the Parker on the Web TEI compared to what information would appear once this metadata had been extracted and transformed to DMS-Index MODS. Emphasis in TEI column added.

| Decoration Notes as they appear in Parker on the Web TEI (surrounding context included) | Extractable information as it would appear in DMS-Index MODS |
|--|---|
| <i><decoNote>The three preliminary leaves are ornamentally written in large capitals (11 lines to a page), within borders, the first and last pages being enclosed in an arch, the rest within rectangular borders</decoNote>.</i> | <i><note displayLabel="Decoration Note"> The three preliminary leaves are ornamentally written in large capitals (11 lines to a page), within borders, the first and last pages being enclosed in an arch, the rest within rectangular borders</note></i> |
| on the opposite page the text of <name type="author">Iuuencus</name> begins without any rubric, the first preface being written <i><decoNote>in the same fancy capitals</decoNote></i> as the verses of <name type="person">St Isidore</name> | <i><note displayLabel="Decoration Note">in the same fancy capitals</note></i> |
| <i><locus>4v</locus> <name type="person">Iohannis</name> fremit ñ misteria uitae caluetii: aquilini: siue iuueni: so far <decoNote>in large fancy capitals</decoNote>, then immediately in ordinary uncials: Immortale nihil ñ</i> | <i><note displayLabel="Decoration Note">in large fancy capitals</note></i> |
| then at once in larger capitals: sapientissimi uiri Iuuenci: xpiani: euangeliorum liber primus: explicit Incipit liber secundus caluetii ~ aquilini ~ Inde philippus ait >>> these last words are <i><decoNote>in red</decoNote></i> but in ordinary uncials | <i><note displayLabel="Decoration Note">in red</note></i> |
| the 2nd book ends, and the 3rd begins thus (67<abbr rend="superscript">a</abbr><locus>67r</locus>): turbasque reliquit. Explt Incpt liber tertius (<i><decoNote>in red</decoNote></i>) | <i><note displayLabel="Decoration Note">in red</note></i> |
| the 3rd book ends, and the 4th book begins (93<abbr rend="superscript">b</abbr><locus>93v</locus>): hominum seletio fiet: Euangeliorum liber tertius explt (<i><decoNote>in red</decoNote></i>) | <i><note displayLabel="Decoration Note">in red</note></i> |
| Incpt eiusdem liber ...(<i><decoNote>in black</decoNote></i>) | <i><note displayLabel="Decoration Note">in black</note></i> |
| quartus? feliciter? (<i><decoNote>in red</decoNote></i>). Talia ñ The MS. breaks off book IV, verse 733 | <i><note displayLabel="Decoration Note">in red</note></i> |

Albritton—developed a new motto: “curation is not the same as hoarding.” Everything cannot be kept. Users of digital manuscripts and their metadata would do well to recall Ariel’s song from *The Tempest*:

Full fathom five thy father lies;
Of his bones are coral made;
Those are pearls that were his eyes:
Nothing of him that doth fade,
But doth suffer a sea-change
Into something rich and strange.⁹⁰

Creating new copies, physical or digital, is always a process of loss and change.

However, while some loss is inevitable, there are data that cannot be lost: for instance, manuscripts’ dates of creation. M. R. James dates his manuscripts in charmingly idiosyncratic phrases—and in roman numerals. In Parker on the Web these date notes are tagged in XML, using the same method as the <decoNote> tagging, which helps human readers but does not help a computer tasked with finding manuscripts from a certain century. By this time in my term as data curator, I was working with XQuery functions, which allowed me to help machine applications focus on standardized data within idiosyncratic human phrasing. In the Parker transform, I had hoped to use fn:contains to pull the date’s data from within the shifting human words and recopy it into machine-friendly form. I found, however, that the underlying logic of roman numerals stood in the way of my proposed solution. Since “xiii” contains “xii,” “xi,” and “x,” the new digital humanities skills I had been honing did not help me with this part of the Parker metadata (although they were exceedingly useful elsewhere). In projects like these, when one works on grant money for the set term of two years, time spent learning new skills to pursue the particular challenges of one institution’s metadata set is useful to a point: beyond that, it is a field of diminishing returns. In time-sensitive, grant-funded work, excessive time expended on one data set is time that cannot be given to other libraries’ metadata. In the case of the Parker on the Web, rather than continue to spend more time and grant money learning new techniques (say, a reverse case statement, which would have helped here) in hopes that they *might* work, after a certain period of time I moved forward with the skills I had. With Albritton’s support, I ceased trying to pull dates from the XML marked-up James’s catalogue and instead pulled the Parker manuscripts’ dates from the Parker on the Web’s .html (which had the added benefit of containing revised dates based on new research and better information than what James used).

Significantly, our solution echoes medieval copying practices of combining flawed or fragmentary exemplars to create a better copy. For example, midway through Thomas Hoccleve’s *Series* (early 1420s), the narrator Thomas shows a character known as “the Friend” the new English translation of a story from the *Gesta Romanorum* that the Friend asked Thomas to create.⁹¹ Thomas tells the Friend that the poem is complete. However, the Friend objects that Thomas’s new translation lacks the moralization that the Friend treasures in his personal copy of the *Gesta*. The Friend lends this second copy

⁹⁰ Shakespeare, *The Tempest*, I.ii. 474–79.

⁹¹ See “Fabula de Quadam Imperatrice Romana” in *My Compleinte’ and Other Poems*, lines 953–80.

to Thomas, whereupon Thomas copies the missing portion into his book. From two Latin exemplars arises Thomas's single English text. Neither Thomas nor the Friend are troubled by Thomas's failure to make a perfect copy of either exemplar. Precise copying of one parent manuscript is not the point for them. Combining the best parts of both to create a newer and more useful copy is.

Similarly, in drawing dates from the Parker .html and everything else from the Parker's TEI-marked-up James's catalogue, I found myself making a "better" object by copying parts of two separate documents. There are perhaps no downsides here. The dates are more correct. The manuscripts more discoverable. But in taking the best data from two different copies I had to let go of what I had seen as a scribe-like duty to make a *precise* copy in order to make a more *useful* copy. Over the months of my postdoc work, both my skills as a copyist and my understanding of the kind of copying I was supposed to be doing were evolving. I no longer saw data curation as a duty to strictly preserve inherited documents. Instead, my duty was to create documents that could be *used*.

To put it less like a data curator and more like a medievalist, in working with the Parker on the Web metadata for DMS-Index, I moved past the association of copyist-induced change with betrayal, loss, and violence that I had drawn from the Chaucerian lyric "Adam Scriveyn." Instead, my models became Chaucer's heirs: Thomas Hoccleve and John Lydgate. In the prologue to *The Fall of Princes*, Lydgate (translating and transforming Laurent de Premierfait, who in turn is translating and transforming Boccaccio) describes the positive reasons that copyists change old books as they make them anew:

Artificeres hauyng exercise
May chaunge and turne bi good discrecioun
Shappis, formys, and newli hem deuyse,
Make and vnmake in many sondry wyse,
As potteres, which to that craft entende,
Breke and renewe ther vesselis to a-mende.

Thus men off crafft may off due riht,
That been inuentiff & han experience,
Fantasien in ther inward siht
Deuises newe thoruh ther excellence;
Expert maistres han therto licence
Fro good to bettir for to chaunge a thyng,
And semblabli these clerkis in writyng,

Thyng that was maad of auctours hem befor,
Thei may off newe fynde and fantasie,
Out of old chaff trie out ful cleene corn,
Make it more fressh and lusti to the eie,
Ther subtil witt and ther labour applie,
With ther colours agreable off hewe,
Make olde thynges for to seeme newe.

[Craftsmen, having practical experience,
May change and turn by good judgment

Shapes, forms, and newly construct them
 Make and undo in many different ways—
 Like potters, who are inclined to that craft,
 Break and revise their vessels to emend.

Thus men of craft may of due right
 (That are inventive and have experience)
 May think up in their inward sight
 New designs through their excellence.
 Experienced masters have the right
 From good to better to change a thing—
 And similarly, these clerks in writing

A thing that was made by earlier authors/authorities
 The clerks may, through new discoveries and invention,
 Out of old chaff produce very pure corn,
 Make it more vivid and pleasing to the eye
 Their clever judgment/understanding and their labor apply
 With their colours of pleasing hue
 Make old things appear new.]⁹²

Like Lydgate's "artificeres," "men off crafft," and "clerkis in writyng" who "make and unmake in many sondry wyse," curating manuscript metadata for reuse in interoperable environments requires—and likely will always require—a certain amount of breaking, renewing, and amending. Rather than see that breaking and renewal as a betrayal of the medieval past, I propose that closer attention to the hidden labour of metadata transformation underlying digital manuscripts reveals just how deeply digital medieval book culture is rooted in copying practices akin to those practiced by medieval writers, scribes, and bookmakers.

The Tower of Babel in the Catalogue: e-codices and the Place of English in Digital Manuscript Studies

E-codices, the digital medieval manuscript library of Switzerland, collects the digitized manuscripts of more than fifty-eight libraries and private collections, both in and outside of Switzerland. A basic level of manuscript metadata is shared via the Open Archive Initiative Dublin Core (OAI:DC) metadata standard: manuscript's title, short description/abstract, author or creator, publisher, date, and format, which includes material support and foliation statement. This DC-tagged information is in English, and is lean and accessible across institutions. However, because it has been carefully shaped to fit the goals and requirements of OAI:DC, this first layer of metadata does not match the scope of the descriptive metadata DMS-Index contained drawn from the Walters Art Museum and Parker on the Web. For instance, e-codices' top-layer DC metadata does not contain a full statement of contents, detailed decoration descriptions, or detailed

⁹² Lydgate, *Fall of Princes*, bk 1, lines 9–28.

Table 8.5. Mock-up of what that macaronic description, drawing together English-language OAI:DC and French-language TEI, might look like.

<abstract> This ethical work by Boccaccio, originally written between 1353 and 1356 and expanded in 1373, addresses the subject of the unevenness of fate. Manuscript copies of the work were frequently made; it was issued in print and translated into many languages. It enjoyed great popularity in Europe. The French translation by Laurent de Premierfait for Jean de Berry was equally popular, as evidenced by the 68 manuscript copies of this text still in existence....</abstract>

...

<note displayLabel="Catalog Note(s)">Traduction par Laurent de Premierfait exécutée en 1409. Ce manuscrit ne comporte pas le premier prologue de Laurent de Premierfait (dédicace de sa traduction).</note>

...

<note displayLabel="Foliation: Statement">Foliotation récente au crayon de 10 en 10. Les deux feuillets de garde, au début et à la fin, sont contemporains du manuscrit. Le feuillet 338v° est blanc.</note>

...

<note displayLabel="Hand Note">Ecriture bâtarde. Une seule main.</note>

provenance information. This information, instead, appears in a second set of metadata documents: extended, granular descriptions in TEI P5 rich enough to empower the far-reaching research that interoperable hubs like DMS-Index seek to support. Thus, by pairing two different metadata standards, to support two different kinds of interaction with their data, e-codices manages to serve the inter-repository needs of simplicity, and individual researchers' needs for scholarly depth. Using the same approach I had developed on the Parker Library metadata, I was determined to pull the best information from the DC and the TEI, seeking to make the richest possible metadata for future researchers using DMS-Index.

However, dipping into these richer TEI manuscript descriptions brought a new curation challenge. As the e-codices personnel with whom I corresponded had warned, these extended TEI descriptions were often not in English, but in German, Italian, and French. For example, Cologne, Fondation Martin Bodmer, Cod. Bodmer 174, a fifteenth-century manuscript of Laurent de Premierfait's French translation of Giovanni Boccaccio's *Des Casibus Virorum Illustrium*, has a short description in English tagged in the OAI:DC metadata schema. It also has a more extended manuscript description in TEI, which contains more granular information akin to what I transformed from the Walters Art Museum and Parker on the Web. This extended TEI manuscript description is in French. Although I was technically able to combine the <abstract> from the English-language OAI:DC with the cataloguing and content details supplied in the longer French-language TEI description, the resulting MODS metadata would have been macaronic (see Table 8.5).

For most data curators, this kind of linguistic mash-up would be immediately and obviously untenable. Yet I wondered if it could be defended. Medieval writers wrote macaronic verse. Some of the most famous and influential "English" medieval manuscripts

are, in fact, multilingual, and the recovery of medieval multilingualism is an important focus in the field.⁹³

Ultimately, however, DMS-Index was not just for experts who have the languages to tackle medieval and modern multilingualism. In helping build the next iteration of DMS-Index, I was hoping to create something also useful for undergraduates, book artists, calligraphers, and interested hobbyists. But seeking this broader audience imposed strict limits on what could be defined as “usable” metadata. What would students in a medieval literature course or an introductory book history course do when set to research using digital sources? The formal language of manuscript description already carries a high cognitive load for beginners.⁹⁴ Asking new recruits to medieval studies to also engage with formal manuscript descriptions that slide from English into Italian, German, and French would likely push that cognitive load to the breaking point. Accessibility does not just mean putting manuscripts on the web. It means surrounding those manuscripts with information that acts as a doorway in for new users. Words, as Geoffrey of Vinsauf puts it in his *Poetria Nova* (1208–1213), should work to render difficult ideas accessible.⁹⁵ His early thirteenth-century counsel on the ornaments of style offers an important critique of language used to render medieval manuscripts “accessible” in the digital age:

If, therefore, you introduce any words that are strange or recondite, you are displaying your own virtuosity thereby.... Take counsel: it may be you know all things—you are greater in others than this—still, in your mode of expression be one of those others.... Regard not your own capacities, therefore, but rather his with whom you are speaking.⁹⁶

Seeking to open up manuscripts to a wider range of possible users, I abandoned the macaronic manuscript catalogue.

This decision is not ideologically neutral. There are important, growing critiques of digital humanities that highlight its problems of monolingualism, monoculturalism, and the dominance of English as its *lingua franca*.⁹⁷ In seeking to better serve more diverse

93 See, for example, Butterfield, *The Familiar Enemy*.

94 For this point, in particular, I am indebted to the students in my “Medieval Books in the Digital Age” course, Binghamton University, Fall 2015. While I have taught manuscripts at other universities and to other students who I am certain also found manuscript descriptions difficult to parse, these students were uniquely honest about their difficulties parsing the stylized language of formal manuscript description. My thinking on language, access, and medieval studies is also informed by Carruthers, “‘Micrological Aggregates’: Is the New Chaucer Society Speaking in Tongues?”

95 Geoffrey of Vinsauf, *Poetria Nova*, 54–55.

96 Geoffrey of Vinsauf, *Poetria Nova*, 55.

97 Exploration of the ways that pervasiveness of English and Anglo-American interests deform digital humanities threads throughout Fiormonte, Numerico, and Tomasi, *The Digital Humanist*; see especially “Conclusions—DH in a Global Perspective,” 207–18. See also Fiormonte, “Toward a Cultural Critique of Digital Humanities,” 59–76; as well as his blog posts “Digital Humanities in Kerala: Some Lessons from the South” and “Towards a Monocultural (Digital) Humanities?” and Columbia’s blog post “Postcolonial Studies, Digital Humanities, and the Politics of Language.”

audiences within the English-speaking world, my curation decisions also reified in DMS-Index—at least for now—the problematic state of English as a default language of international scholarship.⁹⁸ Put differently, in seeking to solve one problem of the lack of diversity in the academy, my choices contributed to another.

Conclusion

I remain profoundly displeased by the choices that I made regarding the e-codices metadata. In the same way, I am haunted by other curation decisions I made as data curator for digital metadata from many other institutions generously sharing their manuscripts with DMS-Index. There are good reasons why I made these decisions, and in the previous pages I hope I have suggested some of those:

- The ever-present panting of “time’s winged chariot” that one lives under on a short-term project supported by external grant money and how that pressure influences project forms and timelines
- The slow-growing DH- and data curation skills that beginning digital humanists can only learn on the job through the process of trial, error, and *making*, and the impact that this learning will inevitably have on project processes and final productions
- The idea that we do not digitize these manuscripts solely for scholars, and my metadata is meaningless if it does not help recruit future generations to the study and care of medieval literature and culture.

But these good reasons do not erase the fact that I have equally good reasons to continue to be haunted.

In the end, I would argue that there is great value in publicly owning the haunting choices we make on digital projects. It is *good* for digital humanists to be haunted by the choices we make. Exorcism is not—should not—be the point: instead, we need to continue to discuss those hauntings and ethical compromises publicly. It is wrong to hide these ghosts because, in the case of my work on DMS-Index and other labour on digital manuscripts, hiding our ghosts contributes to an incorrect impression that all digital manuscripts can be all things, to all people, in the first iteration. As a corrective, those of us involved in digital projects and digitization need to carve out spaces to openly reflect on the compromises we have made, under what conditions, and for what reasons, while perhaps also laying out ways that we hope future iterations and future makers might improve on our compromises and alleviate our betrayals. When information is excised and done so consistently, that needs to be made transparent in multiple areas. It is of paramount importance to be as clear—as honest—about any modifications or transformations as possible.

By explicitly using medieval writers like Chaucer, Hoccleve, and Lydgate to theorize digitization and manuscript metadata, I hope I have also shown two additional

⁹⁸ See, for instance, Bocanegra-Valle, “English is my default academic language,” and Clavert, “The Digital Humanities Multicultural Revolution Did Not Happen Yet.”

things: first, how much of one's own world view and research digital makers bring to their projects. Just as data is not neutral, making digital things is never neutral. DH projects are shaped by who we are, and what we value and need, as we engage in the act of building. After seven years' intensive study on medieval literature and culture, I could no more stop being a medievalist as I did my digital work than I could stop breathing—and (even if those medievalist principles are not, or at least not yet, explicitly described alongside my work in DMS-Index) my deep reading in late medieval English literature acted as a lodestar for my nascent theories of digital book making.

Second, by reflecting on how much Chaucer, Hoccleve, and Lydgate shaped my curatorial work, I hope to have demonstrated more broadly just how much the cultures of pre-modern book-making have to offer contemporary digital humanities. The rhetoric of DH and digitization can slip into technological determinism and techno-utopianism. Against these impulses, medieval writers' insights into the labour, purpose, and limitations of book-making can serve as a useful corrective to modern visions of unchecked progress. Furthermore, when medieval works are marked up for further processing using modern computer languages and technologies, editors are engaging in a centuries-old practice that they may not be entirely aware of. Developing a richer, more self-conscious awareness of the long history of those practices can in turn inform and temper some of the central presumptions about digital technologies today—including any claim that the present is a break with the past simply because today's work happens to be digital. Bringing this prehistory into contemporary conversations on digitization helps shift modern discussions away from wide claims of democratization to more nuanced consideration: *democratizing for whom? To what end? At what cost—both literally, in terms of dollars or pounds expended to achieve project goals, and figuratively. Who, or what, has been treated as expendable?*

Moreover, making those inevitable compromises explicit does not just allow modern users to become more literate in digital manuscripts: it embeds the digital manuscript more overtly, and more firmly, in the historical moment of its own making, revealing the pressures under which all digital books are produced and circulated. The exploration of networks of making and circulation is, of course, a central credo of book historians. By making both our data and statements on the compromises that have shaped them openly available—participating in a self-conscious and rigorous codicology of digital medieval books—medievalists can do important work demystifying what has, too often over the last few decades, been treated as a transparent or simple process. Medieval manuscripts on screens are, fundamentally, different objects from the physical ones they seek to represent. As we leave the incunabula age of digitization, the time has come to not just say this—but to treat them as such, and to study their creation for insights into medieval labour and bookmaking, and into the invisible making that increasingly shapes our scholarship today.

Finally, there is an important ethical aspect to fostering a richer and more rigorous curiosity into the digital labour and labourers that make and maintain digital medieval books. Digitizers and data curators are inescapably part of the present history of the digitized manuscripts upon which much of our research depends. They are also inescapably part of our world today. The photographers who helped digitize the Matthew Parker

Library at Corpus Christi College, Cambridge are Polish.⁹⁹ At the time of writing, some of them are still working as digitizers at Cambridge University Library. The “Brexit” vote of June 2016 raises questions about their future employment. It also raises questions about their safety, as anti-Polish and anti-immigrant violence in Britain surges in the wake of that referendum. As I write these lines in New York in November 2016, the United States is facing its own wave of post-election, racist, anti-immigrant, anti-Muslim, and anti-woman violence. How many medievalists and manuscript scholars know whether American-based digitizers of medieval books will find themselves targeted in the weeks and months to come? (How many of us even know these digitizers’ names?) Whether or not we pre-modern book historians and digital humanists choose to make these connections, we are implicated in the infrastructures and politics that shape the lives of these modern makers of medieval books. We dedicate our lives to studying all that can be discovered about the processes of bookmaking and the makers of medieval books more than half a millennium ago: we should be able to extend that care to the modern copyists of medieval books today.

In this call to action, I have no medieval antecedent. Medieval book history is teeming with unnamed and uncredited labourers. Alexandra Gillespie has noted that the individual agency generally attributed to single individuals—especially printers, such as “Caxton,” “de Worde,” “Grafton,” “Pynson,” etc.—is a convenient shorthand for much more complicated relationships among networks of labourers.¹⁰⁰ As Gillespie insists, whenever these names are used they “should be understood as collaborative (involving other publishers, other printers, correctors, compositors, foremen, apprentices).”¹⁰¹ Unfortunately, by-and-large they are not—Caxton is lionized as a unique genius and the workers who actually made the type and pressed the books fade into shadows from which they cannot return. But in this next push to copy medieval books into new media, it need not remain this way. We have a chance, in this next chapter of medieval book history, to rectify some of these long-standing wrongs and fill in those vexing silences with living voices. Ultimately, this is my challenge for medievalists, manuscript scholars, and digital humanists: that we expand our circle of inquiry to include living bookmakers as well as the dead, in order to write them—to invite these bookmakers to write themselves—back into the long history of medieval books.

Bibliography

- Bailey, Moya Z. “All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave.” *Journal of Digital Humanities* 1 (2011). <http://journalofdigitalhumanities.org/1-1/all-the-digital-humanists-are-white-all-the-nerds-are-men-but-some-of-us-are-brave-by-moya-z-bailey/>.
- Bocanegra-Valle, Ana. “‘English is my Default Academic Language’: Voices from LSP Scholars Publishing in a Multilingual Journal.” *Journal of English for Academic Purposes* 13 (2014). 65–77.

⁹⁹ Personal conversation, Suzanne Paul, Cambridge University Librarian, October 20, 2016.

¹⁰⁰ Gillespie, *Print Culture and the Medieval Author*, 25–26.

¹⁰¹ Gillespie, *Print Culture and the Medieval Author*, 25–26.

- Bredehoft, Thomas. *The Visible Text: Textual Production and Reproduction from Beowulf to Maus*. Oxford Textual Perspectives. Oxford: Oxford University Press, 2014.
- Burnard, Lou. "The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure." *Journal of the Text Encoding Initiative* 5 (2013). <https://jtei.revues.org/811>.
- Butterfield, Ardis. *The Familiar Enemy: Chaucer, Language, and Nation in the Hundred Years War*. Oxford: Oxford University Press, 2010.
- Cantor, Norman. *Inventing the Middle Ages: The Lives, Works, and Ideas of the Great Medievalists of the Twentieth Century*. New York: Harper Perennial, 1991.
- Carruthers, Mary. "'Micrological Aggregates': Is the New Chaucer Society Speaking in Tongues?" *Studies in the Ages of Chaucer* 21 (1999). 1–26.
- Chan, Lois Mai, and Marcia Lei Zeng. "Metadata Interoperability and Standardization: A Study of Methodology, Part I: Achieving Interoperability at the Schema Level." *D-Lib Magazine* 12 (2006). www.dlib.org/dlib/june06/chan/06chan.html.
- Chaucer, Geoffrey. *Troilus and Criseyde*. In *The Riverside Chaucer*. Edited by Larry D. Bensen. 3rd ed. Boston: Houghton Mifflin Company, 1987.
- Clarkson, Christopher. "Rediscovering Parchment: The Nature of the Beast." *The Paper Conservator* 16 (1992). 5–26.
- Clavert, Frédéric. "The Digital Humanities Multicultural Revolution Did Not Happen Yet." *L'histoire contemporaine à l'ère numérique* (blog), April 26, 2013, <http://histnum.hypotheses.org/1546>.
- Clemens, Ray, and Timothy Graham. *Introduction to Manuscript Studies*. Ithaca: Cornell University Press, 2007.
- Cole, Timothy W., and Myung-Ja K. Han. *XML for Catalogers and Metadata Librarians*. Third Millennium Cataloging Series. Santa Barbara: Libraries Unlimited, 2013.
- Corpus Christi College, "Manuscript Description: Language." CCCC MS 2I. Parker on the Web. https://parker.stanford.edu/parker/actions/manuscript_description_long_display.do?ms_no=2I.
- . "Manuscript Description: Language." CCCC MS 282. Parker on the Web. https://parker.stanford.edu/parker/actions/manuscript_description_long_display.do?ms_no=282.
- . Parker Library. "Manuscript Description: Language." CCCC MS 112. Parker on the Web. https://parker.stanford.edu/parker/actions/manuscript_description_long_display.do?ms_no=112.
- da Rold, Orietta. "Materials." In *The Production of Books in England 1350–1500*, edited by Alexandra Gillespie and Daniel Wakelin, 12–33. Cambridge: Cambridge University Press, 2011.
- de Hamel, Christopher. *Medieval Craftsmen: Scribes and Illuminators*. Toronto: University of Toronto Press, 1992.
- . *Descriptive Cataloging of Rare Materials (Manuscripts)*. Chicago: Rare Books and Manuscripts Section of the Association of College and Research Libraries, 2016.
- . *Dictionary for Library and Information Science*, edited by Joan M. Reitz. "Dublin Core (DC)," 234–35. Westport, CN: Libraries Unlimited, 2004.
- Drucker, Johanna. *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. Chicago: University of Chicago Press, 2009.

- Dublin Core Metadata Initiative. "History of the Dublin Core Metadata Initiative." <http://dublincore.org/about/history/>.
- Echard, Siân. "Coda: The Ghost in the Machine: Digital Avatars of Medieval Manuscripts." In *Printing the Middle Ages*, 198–216. Philadelphia: University of Pennsylvania Press, 2008.
- . "Containing the Book: The Institutional Afterlives of Medieval Manuscripts." In *The Medieval Manuscript Book: Cultural Approaches*, edited by Michael Johnston and Michael Van Dussen, 96–118. Cambridge: Cambridge University Press, 2015.
- Edwards, A. S. G. "Back to the Real?" *Times Literary Supplement*, June 7, 2013, www.the-tls.co.uk/articles/public/back-to-the-real/.
- . "Chaucer and 'Adam Scriveyn.'" *Medium Aevum* 81 (2012): 135–39.
- Elings, Mary, and Gunter Weibel. "Metadata for All: Descriptive Standards and Metadata Sharing across Libraries, Archives and Museums." *First Monday* 12 (2007). <http://firstmonday.org/article/view/1628/1543>.
- Fiormonte, Domenico. "Digital Humanities in Kerala. Some lessons from the South." Infolet (blog), November 25, 2015. <https://infolet.it/2015/11/25/digital-humanities-in-kerala-some-lessons-from-the-south/>.
- . "Toward a Cultural Critique of Digital Humanities." In *Controversies around the Digital Humanities*, edited by M. Thaller. Special issue, *Historical Social Research/Historische Sozialforschung* 37 (2012): 59–76.
- . "Towards a Monocultural (Digital) Humanities?" Infolet (blog), July 12, 2015. <https://infolet.it/2015/07/12/monocultural-humanities/>.
- Fiormonte, Domenico, Teresa Numerico, and Francesca Tomasi. *The Digital Humanist: A Critical Inquiry*. Translated by Desmond Schmidt and Christopher Ferguson. Brooklyn: punctum books, 2015.
- Foys, Martin. "Media Archaeology and the Digital Incunable." In *The Medieval Manuscript Book: Cultural Approaches*, edited by Michael Johnston and Michael Van Dussen, 119–39. Cambridge: Cambridge University Press, 2015.
- Geoffrey of Vinsauf. *Poetria Nova*. Translated by Margaret F. Nims. Toronto: Pontifical Institute of Medieval Studies, 1967.
- Gillespie, Alexandra. "Bookbinding." In *The Production of Books in England 1350–1500*, edited by Alexandra Gillespie and Daniel Wakelin, 150–72. Cambridge: Cambridge University Press, 2011.
- . *Print Culture and the Medieval Author: Chaucer, Lydgate, and Their Books, 1473–1557*. Oxford: Oxford University Press, 2006.
- Gilliland, Anne J. "Setting the Stage." In *Introduction to Metadata*, edited by Murtha Baca, 1–19. 2nd ed. Los Angeles: The Getty Research Institute, 2008.
- Golumbia, David. "Postcolonial Studies, Digital Humanities, and the Politics of Language." Uncomputing (blog), May 31, 2013, www.uncomputing.org/?p=241.
- Gwara, Scott. *Otto Ege's Manuscripts: A Study of Ege's Manuscript Collections, Portfolios, and Retail Trade*. Medieval Manuscript Collections and Collectors in North America. Cayce: De Brailles, 2013.
- Hargittai, Eszter. "Digital Natives or Digital Naives? The Role of Skill in Internet Use." Presentation at CNI: Coalition for Networked Information, Washington, DC, December 9–10, 2013.

- Harpring, Patricia. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Worlds*. Los Angeles: J. Paul Getty Research Institute, 2010. www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.pdf.
- Haslhofer, Bernhard, and Wolfgang Klas. "A Survey of Techniques for Achieving Metadata Interoperability." *ACM Computing Surveys* 42 (2010): 1–37.
- Hoccleve, Thomas. *'My Complainte' and Other Poems*, edited by Roger Ellis. Exeter: University of Exeter Press, 2008.
- Hodge, Gail. *Metadata Made Simpler: A Guide for Libraries*. Bethesda: NISO Press, 2001.
- Holsinger, Bruce. "Of Pigs and Parchment: Medieval Studies and the Coming of the Animal." *PMLA* 124 (2009): 616–23.
- Horman, William. *Vulgaria uiri doctissimi Guil. Hormanii Caesariburgensis*. London: Richard Pynson, 1519.
- Kiss, Farkas Gábor, Eyal Poleg, Lucie Doležalová, and Rafal Wójcik. "Old Light on New Media: Medieval Practices in the Digital Age." *Digital Philology* 2 (2013): 16–34.
- Koh, Adeline. "Niceness, Building, and Opening the Genealogy of the Digital Humanities: Beyond the Social Contract of Humanities Computing." *Differences: A Journal of Feminist Cultural Studies* 25 (2014): 93–106.
- Lawton, David. "Dullness and the Fifteenth Century." *ELH* 54 (1987): 761–99.
- Library of Congress, "DLC/Aquifer Summary of MODS Requirements and Recommendations Table." www.loc.gov/standards/mods/userguide/dlfaquifer.html.
- Lydgate, John. *The Fall of Princes*. 4 vols. Edited by Henry Bergen. Washington: Carnegie Institution of Washington, 1923.
- Morcos, Hannah. "Piecing Together the Puzzle of the Hungerford Hours." British Library Medieval Manuscripts Blog (blog), November 17, 2015. <http://blogs.bl.uk/digitisedmanuscripts/2015/11/piecing-together-the-puzzle-of-the-hungerford-hours.html>.
- National Information Standards Organization (NISO). "Understanding Metadata." Bethesda: NISO Press, 2004.
- Nikolova-Houston, Tatiana, and Ron Houston. "Building the Virtual Scriptorium." In *Digitisation Perspectives*, edited by R. Rikowski, 229–44. Rotterdam: Sense Publishers, 2011.
- Oxford English Dictionary*. "interoperability."
- Parker Library On the Web. <http://parker.stanford.edu>.
- Prescott, Andrew. "'Their Present Miserable State of Cremation': The Restoration of the Cotton Library." In *Sir Robert Cotton as Collector: Essays on an Early Stuart Courtier and His Legacy*, edited by C. J. Wright, 391–454. London: British Library Publications, 1997.
- Riley, Jenn. "Seeing Standards: A Visualization of the Metadata Universe." *Jenn Riley: Digital Librarian* (blog). <http://jennriley.com/metadatamap/>.
- Rosenheck, Donna. "OCLC: From an Historical Perspective." *The Katharine Sharpe Review* 4 (Winter 1997). <http://hdl.handle.net/2142/78252>.
- Rudy, Kathryn M. "Dirty Books: Quantifying Patterns of Use in Medieval Manuscripts Using a Densitometer." *Journal of Historians of Netherlandish Art* 2 (2010). doi:10.5092/jhna.2010.2.1.1.

- Rundle, David. "Virtual Manuscripts and the Real World. Part I." *Bonae litterae* (blog), July 8, 2013. <https://bonaelitterae.wordpress.com/2013/07/08/virtual-manuscripts-real-world-part-i/>.
- Ryder, Michael L. "Parchment: Its History, Manufacture, and Composition." *Journal of the Society of Archivists* 2 (1964): 391–99.
- Sadler, Bess, and Chris Bourg. "Feminism and the Future of Library Discovery." *Code4Lib Journal* 28 (April 2015): 1–5. <http://journal.code4lib.org/articles/10425>.
- Saenger, Paul. *A Catalogue of the Pre-1500 Western Manuscript Books at the Newberry Library*. Chicago: University of Chicago Press, 1989.
- Schmitt, Desmond. "Towards an Interoperable Scholarly Edition." *Journal of the Text Encoding Initiative* 7 (2014). <https://jtei.revues.org/979>.
- Searle, Eleanor. "Possible History." *Speculum* 61 (1986): 779–86.
- Shakespeare, William. *The Tempest*. Edited by Barbara A. Mowat and Paul Werstine. Folger Digital Texts. Washington: Folger Shakespeare Library: www.folgerdigitaltexts.org/html/Tmp.html.
- Soualah, Mohammed Ourabah, and Mohamed Hassoun. "A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts." *Journal of the Text Encoding Initiative* 2 (2012). <https://jtei.revues.org/398>.
- Summit, Jennifer. *Memory's Library: Medieval Books in Early Modern England*. Chicago: Chicago University Press, 2008.
- TEI: Text Encoding Initiative, P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.2.0. "Chapter 10: Manuscript Description." www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html.
- The Digital Walters. "Describing Manuscripts with TEI." <http://thedigitalwalters.org/Supplemental/ManuscriptDescription.html>.
- . W.169, the Hours of Pastor Denys. Walters Art Museum. www.thedigitalwalters.org/Data/WaltersManuscripts/html/W169/.
- Thesaurus for Graphic Materials*. Library of Congress. www.loc.gov/pictures/collection/tgm/.
- Treharne, Elaine. "Fleshing Out the Text: The Transcendent Manuscript in the Digital Age." *postmedieval* 4 (2013): 465–78.
- Unsworth, John. "Computational Work with Very Large Collections: Interoperability, Sustainability, and the TEI." *Journal of the Text Encoding Initiative* 1 (2011). <https://jtei.revues.org/215>.
- . "Digital Humanities Centers as Cyberinfrastructure." Presentation at Digital Humanities Centers Summit, National Endowment for the Humanities, Washington, DC, April 12, 2007.
- What the OCLC Online Union Catalog Means to Me: A Collection of Essays*. Dublin, OH: OCLC Online Computer Library Center, Inc., 1997.

