




2017

Will that Surrogate Do?: Reflections on Material Manuscript Literacy in the Digital Environment from Islamic Manuscripts at the University of Michigan Library

Eryn C. Kropf
ekropf@umich.edu

Follow this and additional works at: https://repository.upenn.edu/mss_sims

 Part of the [Digital Humanities Commons](#), [Medieval Studies Commons](#), [Near Eastern Languages and Societies Commons](#), and the [Scholarly Communication Commons](#)

Recommended Citation

Kropf, Eryn C. (2017) "Will that Surrogate Do?: Reflections on Material Manuscript Literacy in the Digital Environment from Islamic Manuscripts at the University of Michigan Library," *Manuscript Studies*: Vol. 1 : Iss. 1 , Article 4.

Available at: https://repository.upenn.edu/mss_sims/vol1/iss1/4

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/mss_sims/vol1/iss1/4
For more information, please contact repository@pobox.upenn.edu.

Will that Surrogate Do?: Reflections on Material Manuscript Literacy in the Digital Environment from Islamic Manuscripts at the University of Michigan Library

Abstract

The widespread dissemination of digital surrogates for Islamic manuscripts certainly has the potential to impact scholarship both positively and negatively. Realizing a positive impact is contingent on the accessibility and quality of the digital surrogates and the training extended to the scholars working with them. Indeed, while manuscript digital surrogates have the potential to enhance access for those scholars who might otherwise neglect manuscript evidence, they may also enable neglect of material qualities and with them the essential historical context for the content of a codex. This is particularly concerning for the field of Islamic manuscript studies for which so much codicological and palaeographical groundwork is yet to be conducted and ample training in material manuscript literacy is still lacking. Perhaps surprisingly, our experience with manuscript digital surrogates at the University of Michigan has demonstrated that even “materially distant” digital surrogates can actually enhance understanding of manuscript features, including appreciation for material aspects, and help advance the codicological projects of Islamic manuscript studies. The key is introducing basic material manuscript literacy via exposure to physical artifacts and relying on the surrogates as tools for descriptive training.

Keywords

Islamic manuscripts, manuscript studies, bibliography, training, digital surrogates, codicology, palaeography, material features, material manuscript literacy, digital humanities, digitization, manuscript cataloging, information science

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

VOLUME 1, NUMBER 1

(Spring 2016)

Manuscript Studies (ISSN 2381-5329) is published semiannually
by the University of Pennsylvania Press



The Schoenberg Institute
for Manuscript Studies

UNIVERSITY of PENNSYLVANIA LIBRARIES

MANUSCRIPT STUDIES

VOLUME 1, NUMBER 1

(Spring 2016)

ISSN 2381-5329

Copyright © 2016 University of Pennsylvania Libraries
and University of Pennsylvania Press. All rights reserved.

Published by the University of Pennsylvania Press,
3905 Spruce Street, Philadelphia, PA 19104.

Printed in the U.S.A. on acid-free paper.

Manuscript Studies brings together scholarship from around the world and across disciplines related to the study of premodern manuscript books and documents, with a special emphasis on the role of digital technologies in advancing manuscript research. Articles for submission should be prepared according to the *Chicago Manual of Style*, 16th edition, and follow the style guidelines found at <http://mss.pennpress.org>.

None of the contents of this journal may be reproduced without prior written consent of the University of Pennsylvania Press. Authorization to photocopy is granted by the University of Pennsylvania Press for libraries or other users registered with Copyright Clearance Center (CCC) Transaction Reporting Service, provided that all required fees are verified with CCC and paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. This consent does not extend to other kinds of copying for general distribution, for advertising or promotional purposes, for creating new collective works, for database retrieval, or for resale.

2016 SUBSCRIPTION INFORMATION:

Single issues: \$30

Print and online subscriptions: Individuals: \$40; Institutions: \$90; Full-time Students: \$30

International subscribers, please add \$18 per year for shipping.

Online-only subscriptions: Individuals: \$32; Institutions: \$78

Please direct all subscription orders, inquiries, requests for single issues, address changes, and other business communications to Penn Press Journals, 3905 Spruce Street, Philadelphia, PA 19104. Phone: 215-573-1295. Fax: 215-746-3636. Email: journals@pobox.upenn.edu. Prepayment is required. Orders may be charged to MasterCard, Visa, and American Express credit cards. Checks and money orders should be made payable to "University of Pennsylvania Press" and sent to the address printed directly above.

One-year subscriptions are valid January 1 through December 31. Subscriptions received after October 31 in any year become effective the following January 1. Subscribers joining midyear receive immediately copies of all issues of *Manuscript Studies* already in print for that year.

Postmaster: send address changes to Penn Press Journals, 3905 Spruce Street, Philadelphia, PA 19104.

Visit *Manuscript Studies* on the web at mss.pennpress.org.

Will That Surrogate Do?

*Reflections on Material Manuscript Literacy
in the Digital Environment from Islamic Manuscripts
at the University of Michigan Library*

EVYN KROPF
University of Michigan

THINKING OUTSIDE THE CODEX to its surrogate, “Will that surrogate do?” The complexities of this question and its multifaceted, even elusive answers have already been acknowledged, on multiple occasions, by Marilyn Deegan¹ and numerous other theorists and practitioners concerned with digital preservation and access. Certainly both question and answer are contextual, and the appropriate response is perhaps another question: “Well, what surrogate, for what purpose, and for whom?”

At the University of Michigan Library, we recently completed creation and dissemination of full color digital surrogates for 1,029 codex manuscripts from our Islamic Manuscripts Collection. The surrogates are held in the HathiTrust Digital Library and linked from catalogue records in our library’s online catalogue.² Scholars (emerging and established) have been

1 See Marilyn Deegan and Simon Tanner, *Digital Futures: Strategies for the Information Age* (New York: Neal-Schuman, 2002), 183–89; and Deegan, “Surrogacy and the Artifact,” in *Preservation Management for Libraries, Archives, and Museums*, ed. G. E. Gorman and Sydney J. Shep (London: Facet, 2006), 54–63.

2 See <http://www.hathitrust.org/> and <http://mirlyn.lib.umich.edu/Search/Home?filter%5B%5D=location%3ASPEC%20ISLM&page=1&sort=yearup>.

enthusiastically working with these surrogates, as evidenced by our web analytics and recent user study. We have also leveraged these surrogates to engage scholars with descriptive tasks in the course of a collaborative cataloguing project. Our experience engaging scholars in work with these surrogates has revealed a need for enhanced training in material manuscript literacy while simultaneously establishing the valuable role that informed work with these surrogates can play in developing appreciation for the physical aspects of manuscripts. In this essay, I will reflect on scholarly work with manuscript digital surrogates and the need for fundamental training in description of physical artifacts suggested by our experience at Michigan.

Digital Surrogates and Scholarship

The qualities of a surrogate determine the scholarly uses it can afford. Hence, surrogate design is in a position to directly impact scholarship, even privilege certain forms of scholarship over others. In this sense, surrogates are “boundary objects” integral to the informatic needs and scholarly work processes of multiple user communities. While we might endeavor to design a surrogate that responds to the diverse needs of every scholarly community, limitations in technological and fiscal resources may force us to privilege the needs of one community over another. Geoffrey Yeo has likened this challenge to the dilemmas we face with collecting and describing material in the first place—the difficulties of anticipating what some unknown future user might find worthy of study, and by what method.³

Whether users, scholars in particular, can take advantage of the affor-

3 “Even when the digital world offers new techniques that promise responsiveness to the needs of diverse user communities, the costs of exploiting their full potential may be prohibitive. If we cannot meet everyone’s needs we may have no choice but to privilege certain features of records, and thus the needs of certain communities, in the actions we take. Our recognition that records are boundary objects, and our sense of obligation to multiple user communities that this recognition entails, cannot easily prevail over practical limitations imposed by restricted resources and finite budgets”; Geoffrey Yeo, “Nothing is the Same as Something Else: Significant Properties and Notions of Identity and Originality,” *Archival Science* 10 (2010): 85–116, quote at 111.

dances of a particular surrogate, intentioned or not, is another, equally important question. We as academic manuscript specialists and curators have an instructive role to play in mediating access to collections by addressing such needs.

All the while, the dissemination of digital surrogates is influencing scholarship. We see in older⁴ and more recent studies—such as the study recently completed on behalf of the Project Bamboo Initiative and presented in Iowa at Digital Humanities 2013⁵—that humanities scholars tend to use digital surrogates in complement with originals, but with the digital comprising more than half of the material consulted in their research.

Consulting both original and surrogate seems appropriate, as there may be some qualities of the original that a surrogate does not mediate well, while the surrogate can mediate other aspects much more safely and efficiently. The question is whether users can recognize these “limits of remediation” if they exist and whether the original is even accessible for them to consult.⁶ The ability to recognize and appreciate these limits is a matter of bibliographic literacy.

Anecdotally, we see similar trends for scholars working with Islamic codex manuscripts. A recent surge of digital initiatives for preservation and access, among them numerous online cataloguing and digitization projects, has scholars flocking to manuscript digital surrogates. I can attest to this

4 E.g., Jon Rimmer, Claire Warwick, Ann Blandford, Jeremy Gow and George Buchanan, “An Examination of the Physical and the Digital Qualities of Humanities Research,” *Information Processing & Management* 44 (2008): 1374–92; and Neal Audenaert and Richard Furuta, “What Humanists Want: How Scholars Use Source Materials,” in *JCDL'10 Proceedings of the 10th Annual Joint Conference on Digital Libraries* (New York: ACM, 2010): 283–92.

5 Harriett Green and Angela Courtney, “Beyond the Scanned Image: A Needs Assessment of Faculty Users of Digital Collections,” presentation given at the Digital Humanities 2013 conference, posted online 19 July 2013 at <http://www.slideshare.net/green19/dh2013-slides-revised>. For more on the now-concluded Project Bamboo, see the archived pages at <http://www.projectbamboo.org/> and the ongoing Bamboo DiRT, a registry of digital research tools, at <http://dirt.projectbamboo.org/>.

6 See Diana Kichuk, “Metamorphosis: Remediation in Early English Books Online (EEBO),” *Literary and Linguistic Computing* 22 (2007): 291–303.

as the curator of our “Islamic Manuscripts (Michigan)” collection in the HathiTrust Digital Library,⁷ which is consistently the most popular HathiTrust collection in terms of page views.

The attraction is understandable. Access to the more than 3 million extant manuscripts in the Arabic script remains problematic due to their geographic and political distribution.⁸ Manuscript digital surrogates enhance access for those who might otherwise neglect manuscript evidence owing to impossibilities of travel and material fragility. These surrogates can facilitate—in remarkable ways—time-intensive, detailed study of both content and carrier within a single manuscript, and—as Nir Shafir has recently highlighted—efficient comparative study across a large corpus of manuscript witnesses.⁹

However, the voracious appetite for manuscript digital surrogates has aroused some misgivings. Dagmar Riedel has expressed concern for the impact of digitization on the ongoing, essential project of manuscript cataloguing.¹⁰ She argues that digitization projects will overtake cataloguing projects in the competition for scarce resources, diminishing the value of careful, detailed description.

My own concern is whether material features—informing dating, recognition of later intervention, and other aspects of the essential physical context in which any study of manuscript content should be grounded—are more easily neglected in the context of a flood of readily available digital surrogates. This comes down to surrogate design and more significantly to training in an element of bibliographic literacy that I will call “material

7 See <http://babel.hathitrust.org/cgi/mb?a=listis;c=1961411403>.

8 The figure of more than 3 million is Geoffery Roper’s estimate; see “The History of the Book in the Muslim World,” in *The Oxford Companion to the Book*, ed. Michael F. Suarez and H. R. Woudhuysen (Oxford: Oxford University Press, 2010), 321–39 at 323.

9 See Nir Shafir, “How Digitization Has Transformed Manuscript Research: New Methods for Early Modern Islamic Intellectual History,” *HAZĪNE*, 8 November 2013, http://hazine.info/2013/11/08/digitized_manuscript_libraries/.

10 See Dagmar Riedel, “How Digitization Has Changed the Cataloging of Islamic Books,” *Islamic Books: A Research Blog about Manuscripts, Printed Books, and Ephemera in Arabic Script*, posted 14 August 2012, updated 13 January 2013, <https://researchblogs.cul.columbia.edu/islamicbooks/2012/08/14/digitalsurrogates/>.

manuscript literacy,” that is, the ability to recognize the significance of the physical features of manuscripts and to discern and study them. Seeing and evaluating these features in the digital environment requires distinct skill and is affected by the ability of the digital surrogate to mediate the features adequately or inadequately.

This is not to say that all scholars will be assessing material features. Many scholarly projects will build on the material groundwork that other scholars are laying. The issue is that with Islamic manuscript studies so much of this groundwork remains to be laid and training in material manuscript literacy is at such a deficit.

Recent decades have seen remarkable trends in the advancement of Islamic manuscript studies. Nevertheless, by contrast with the strides made in Western manuscript studies, the field remains in its early stages of development with hundreds of thousands if not millions of manuscripts yet to be explored. A vast amount of codicological and palaeographical data for various regions and periods has yet to be collected, analyzed, and contextualized within the larger projects of manuscript cultural history, intellectual history, social history, art history, and philology for the various locales, cultures, and literary outputs.

Further, we anecdotally see a dearth of manuscript studies training within most North American Middle Eastern studies graduate programs. Regular courses in the relevant fields of palaeography, manuscript culture, book history, and codicology are quite rare. More typical offerings are occasional workshops given by experts from outside of the university.¹¹ The alternative is training in Europe or the Middle East where such short courses, workshops, lectures, and even graduate programs are much more widely available. When courses are offered in North America the response is phenomenal. An Islamic codicology short course offered at Stanford during the summer of 2013 invited applications for a limited number of places in a hands-on-plus-lectures track, and the applications received numbered five

11 Adam Gacek, being based in North America, is the expert most often called upon. Courses with François Déroche and Jan Just Witkam have also been offered in North America, but are more often offered in Europe.

times the available spots.¹² Additional places were offered for a lectures-only track, which was also enrolled to capacity.

Nevertheless, while deficits in material manuscript literacy raise concerns that widespread dissemination of digital surrogates for Islamic manuscripts is enabling the neglect of material qualities and with them the essential historical context for the content of a codex, I would argue that these same digital surrogates are no threat.¹³ Rather, these surrogates can actually help us advance the codicological projects of Islamic manuscript studies, and indeed of material manuscript literacy, as tools for training and engagement via descriptive tasks.

Digital Surrogate Design (or “Give Us Another Affordance”)

Properly designed and deployed digital surrogates—even careful use of imperfectly designed surrogates—can thus actually serve the project of increasing material manuscript literacy. In order to consider how this might be accomplished, we can first consider some existing Islamic manuscript digital surrogates and their affordances for mediating elements of materiality and physicality.

Enhanced image color and resolution certainly improves mediation of a manuscript’s physical attributes. HathiTrustDigital Library page images for Michigan’s Islamic manuscripts were captured at 400 ppi, 24-bit color.¹⁴ Composition and completeness are also important. Including a color bar or

12 I extend my thanks to John Eilts, my colleague at Stanford University Libraries who was instrumental in organizing the short course with the sponsorship of The Islamic Manuscript Association, for supplying these data.

13 Nevertheless, it must be acknowledged that some libraries do use the availability of digital surrogates as a pretext for limiting access to the physical manuscripts, irrespective of physical condition or value (which can certainly be legitimate concerns) and of the quality of the digital surrogates provided.

14 E.g., see a zoomed image of the upper cover of Isl. MS 1016 at <http://babel.hathitrust.org/cgi/pt?id=mdp.39015079132000;view=1up;seq=1;size=400>.

ruler—as Cambridge University Digital Library surrogates do¹⁵—helps to mediate color and size. Including views of the spine, fore-edge, head-edge, tail-edge, flap (where present), and other housing—as do most surrogates of the Bibliothèque nationale de France’s Gallica Digital Library¹⁶—helps mediate aspects of the binding and attachment (including endbands), presence or absence of edge inscriptions (often titles or shelfmarks), edge decoration, presence or absence of spine text or decoration, and evidence of repair.

The chosen access model (i.e., whether free or at cost, accessible to particular affiliates or not, requiring log in or not, etc.) is also important, and file sizes, file types, and delivery units (i.e., image by image and/or an entire surrogate package of images) will influence users’ long-term examination. Methods of image display (i.e., image by image, thumbnails, scrolling, flip, zooming, full screen, simultaneous views, etc.) as well as display direction and whether it is possible to switch according to reading direction will have a more immediate impact. The Bibliothèque nationale de France’s Gallica includes a “switch reading direction” link,¹⁷ while the Bibliotheca Alexandrina’s Digital Assets Repository (DAR) viewer allows a user to change the display direction by selecting one of several language icons.¹⁸

As Cécile Bonmariage has argued,¹⁹ entry points that allow one to navigate directly to features of interest such as illuminated headpieces, centers of gathering, cover views, and so on could greatly enhance study. This might be achieved via tags, fully labeled navigation bars, or descriptions with page numbers. The “Penn in Hand” viewer provides an example of this for both

15 E.g., see the ruler visible with images of MS Mm.4.15 at <http://cudl.lib.cam.ac.uk/view/MS-MM-00004-00015/18>.

16 E.g., see the views of spine, head-edge, fore-edge, and tail-edge for Bibliothèque nationale de France, MS Supplément turc 317 at <http://gallica.bnf.fr/ark:/12148/bt-v1b8436331b/f836.planchecontact.rtl.r=recueil%20turc-oriental.langFR>.

17 Visible just below the page navigation bar, see again the previous link for MS Supplément turc 317.

18 E.g., see the icons visible in the upper navigation bar for <http://dar.bibalex.org/web-pages/mainpage.jsf?PID=DAF-Job:100459>.

19 Cécile Bonmariage, “Access and Rights,” unpublished paper delivered at The Islamic Manuscript Association’s Fifth Islamic Manuscript Conference, 24–26 July 2009, Christ’s College, University of Cambridge: <http://www.islamicmanuscript.org/biennialconference/pastconferences/2009conference.aspx>.

contents and illuminated headpieces,²⁰ as do the Harvard Digital Repository Service (DRS) viewer²¹ and Princeton University Digital Library viewer.²²

Another important aspect is the location of the textual description with respect to the images. Rather than requiring navigation to a different page (as with HathiTrust) or scrolling away from the image (as with Gallica), the Cambridge viewer nicely allows for display of the textual description in a separate frame that allows the user to scroll through while the image view remains fixed.²³ Most of these delivery features are of course determined by the repository, which may be an online database devoted exclusively to the collection, a larger-scale digital library, an independent photo-sharing site such as Flickr, or some combination thereof as used by the Walters Art Museum²⁴ and Kislak Center for Special Collections, Rare Books, and Manuscripts.²⁵

What Can We Get Away With? (or “A Little Goes a Long Way”)

These digital surrogates show tremendous promise for relaying features that would support codicological tasks, but how can we as curators of these

20 E.g., see the “Contents” and “Illustrations” drop-down menus in the case of Penn Museum NEP27 at <http://hdl.library.upenn.edu/1017/d/medren/5829382>.

21 E.g., see the left navigation frame visible at <http://pds.lib.harvard.edu/pds/view/10662513> for MS Arab 9 (Houghton Library, Harvard University), served via the Digital Repository Service (DRS) for the Islamic Heritage Project (<http://ocp.hul.harvard.edu/dl/ihp/>).

22 E.g., see the navigation panel visible with Garrett no. 968Y <http://pudl.princeton.edu/viewer.php?obj=m326m1793>.

23 See again the view of Cambridge University Library MS Mm.4.15 <http://cudl.lib.cam.ac.uk/view/MS-MM-00004-00015/18>.

24 Walters Museum manuscript images are freely available via Flickr (<https://www.flickr.com/people/medmss/>) and the Digital Walters (http://www.thedigitalwalters.org/01_ACCESS_WALTERS_MANUSCRIPTS.html) as well as their Works of Art site (<http://art.thewalters.org/browse/category/manuscript-and-rare-books/>).

25 Images freely available via OPenn (<http://openn.library.upenn.edu/>, also featured at <https://www.flickr.com/photos/130418531@N02/sets/72157651660303132/>) as well as the Penn in Hand site (<http://dla.library.upenn.edu/dla/medren/index.html>).

surrogates suggest such affordances to scholars? Why not attempt to engage them in tasks that reference material features, such as aspects of descriptive and analytical bibliography?

Beginning in 2009, we at the University of Michigan Library tried just that. Our daring project²⁶ attempted digitally mediated, collaborative cataloguing for our Islamic Manuscripts Collection, a dauntingly diverse array of volumes—chiefly in codex format but including a few rolls and single leaves—representing manuscript production from the eighth to the twentieth century CE across the historically Islamicate cultural areas and carrying texts mainly in Arabic, Persian, and Ottoman Turkish.²⁷

To define the descriptive tasks, I (as project cataloguer) developed a rich scheme—following the *AMREMM* cataloguing standard with modifications inspired by Adam Gacek and other scholars such as Jan Just Witkam—to suit the palaeographic and codicological features of the manuscripts.²⁸ The intended outcome was a searchable database of transcriptions,

26 For more on the original project concept and early implementation, see Eryn Kropf and Jonathan Rodgers, “Collaboration in Cataloguing: Islamic Manuscripts at Michigan,” *MELA Notes* 82 (2009): 17–29 and at <http://www.lib.umich.edu/special-collections-library/collaboration-cataloging-islamic-manuscripts-michigan>.

27 Together these volumes, rolls, and leaves carry nearly 1,800 titles. For more detail on the collection size, scope, history, and organization, see “About the Collection” on my collection research guide: Eryn Kropf, “Islamic Manuscripts Collection,” <http://guides.lib.umich.edu/islamicmss>.

28 *AMREMM* = Ancient, Medieval, Renaissance, and Early Modern Manuscripts; see Gregory Pass, *Descriptive Cataloging of Ancient, Medieval, Renaissance, and Early Modern Manuscripts* (Chicago: Association of College and Research Libraries, 2003); Jan Just Witkam, “Aims and Methods of Cataloguing Manuscripts of the Middle East,” in *Les manuscrits du Moyen-Orient: Essais de codicologie et paléographie, Actes du Colloque d’Istanbul (Istanbul, 26–29 mai 1986)*, ed. François Déroche (Paris: Institut français d’études anatoliennes et Bibliothèque nationale, 1989), 1–5; see the contribution of Annie Berthier and Marie Geneviève Guesdon, “Codicology and the History of Collections,” in François Déroche et al., *Islamic Codicology: An Introduction to the Study of Manuscripts in Arabic Script*, ed. Muhammad Isa Waley (London: Al-Furqān Islamic Heritage Foundation, 2006), 345–60; and Adam Gacek, “Appendix V: Describing the Manuscript,” in Gacek, *Arabic Manuscripts: A Vademecum for Readers* (Leiden: Brill, 2009): 333–38. The work completed at Princeton University Library for many of their Islamic manuscripts was also an inspiration.

notes, and headings that would represent not only the contents of the codices—text, paratext, image, and ornament—but also their material form via further notes addressing the script and hand, structure (i.e., composition of gatherings, sewing, cover, and attachment), dimensions, writing material, layout, and other material aspects.

The project concept centered on the creation and deployment (in the HathiTrust Digital Library) of digital surrogates that contributors could examine to ascertain and provide the requested descriptive details for the manuscripts as comments on existing preliminary descriptions posted to a project site. The primary motivation in adopting this “crowdsourcing” approach was to assemble a capable “community of expertise”—chiefly for knowledge sourcing—and to get a large number of scholars to engage with this still poorly-known collection by involving them in its cataloguing via surrogates.²⁹ As with the best-laid schemes, things did not quite proceed as expected. Suffice it to say that, as the project carried on, we found our approach serving knowledge creation more than knowledge sourcing and engaging more emerging scholars than established ones.³⁰

29 “Crowdsourcing” projects are now rampant in library circles with mixed results that depend a great deal on interpretation in light of project aims. See Rose Holley, “Crowdsourcing: How and Why Should Libraries Do It?” *D-Lib Magazine*, 16.3–4 (2010); Christine Madsen, “Will 2011 Be the Year of Crowdsourcing in Libraries?,” posted 30 December 2011, <http://christinemadsen.com/2011/will-2012-be-the-year-of-crowdsourcing-in-libraries/>; Johan Oomen and Lora Aroyo, “Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges,” in *Proceedings of the 5th International Conference on Communities and Technologies, Brisbane, QLD, Australia, June 29–July 02* (New York: ACM, 2011): 138–49; Ben Brumfield, “Lessons from Small Crowdsourcing Projects,” talk at IMLS WebWise 2012, transcribed in a blog post dated 17 April 2012, “Crowdsourcing at IMLS WebWise 2012,” <http://manuscripttranscription.blogspot.com/2012/03/crowdsourcing-at-impls-webwise-2012.html>; a series of blog posts by Trevor Owens posted in 2012, cf. <http://www.trevorowens.org/tag/crowdsourcing/>; and a more recent case study by Zoë D’Arcy, “‘The Hive’: Crowdsourcing the Description of Collections,” in *Description: Innovative Practices for Archives and Special Collections*, ed. Kate Theimer (Lanham, MD: Roman & Littlefield, 2014), 1–15.

30 I have discussed crowdsourcing challenges and other aspects of the project outcomes in greater detail elsewhere; see Eryn Kropf, “Collaboration in Cataloging: Sourcing Knowledge from Near and Far for a Challenging Collection,” in *Description: Innovative Practices for Archives and Special Collections*, ed. Kate Theimer (Lanham, MD: Roman & Littlefield, 2014), 99–114.

First, we saw greater engagement than ever anticipated with our “local community” of emerging scholars on the project staff (myself and a cadre of graduate student assistants), chiefly due to unforeseen challenges, that led to greater opportunities for developing our own codicological and palaeographical skills. We came onto the project having had training in the fundamentals of Arabic manuscript studies with Adam Gacek and in book structures, condition, etc. with Julia Miller. Descriptive work with digital surrogates enhanced our basic knowledge and increased our level of material manuscript literacy, essentially moving us beyond the fundamentals while cementing concepts and enhancing skill. Altogether our local community catalogued all but 8 of the 880 manuscript volumes fully described in the course of the project (including 696 digitized manuscripts and the rest not digitized), with the graduate student assistants providing virtually full descriptions for 86 volumes, about 10 percent of the total catalogued. They contributed full physical descriptions for another 7 volumes, and collation, pagination and dimension details for still another 56 volumes.

Second, we received only meager response to our wider “crowdsourcing” approach. While by project’s end we had received over 284 descriptive contributions from 34 scholars in Belgium, Egypt, France, Germany, Iran, Israel, Lebanon, Russia, Turkey, the UK, and the US via our project site—with the vast majority of the quality contributions being provided by two or three individuals³¹—we certainly could not rely on those contributions to generate virtually complete descriptions.³² Interestingly, these remote contributors tended not to cover the full range of descriptive elements, but instead focused on bibliographic details pertaining to textual contents. Usually they provided only a single title, citation, or transcription of incipit and explicit. Features such as visual content, decoration, layout, writing

31 As expected given the typical power distribution seen in small crowdsourcing projects; again, see Brumfield, “Lessons from Small Crowdsourcing Projects,” and blog post of 17 April 2012, “Crowdsourcing at IMLS WebWise 2012,” <http://manuscripttranscription.blogspot.com/2012/03/crowdsourcing-at-impls-webwise-2012.html>.

32 Still, we did receive a couple of contributions that made a significant impact. Specifically, for a manuscript carrying Punjabi texts (outside my expertise and the expertise of our student assistants), remote scholars were able to assist with the identification.

material, and binding were almost completely neglected, although aspects of those features were ascertainable via the surrogates provided.

In the end, the remote contributions amounted to nearly complete descriptions for only 8 of the 696 digitized manuscripts. The student assistants worked on another 28 of the digitized manuscripts, with my review and contributions. All others I catalogued, working first from their digital surrogates. In fact, via the enhanced affordances of the full-color, cover-to-cover digital surrogates (features including zooming and simultaneous views of images—even for comparison across different manuscripts), text editors, reference databases, and electronic catalogues—I was able to catalogue (to a nearly complete level) close to 550 volumes in a single year.

What made our project staff so productive in terms of valuable contributions? I would argue that the basic material manuscript literacy afforded by our training in the fundamentals of Arabic manuscript studies had the greatest impact, followed by motivating factors such as pay and dissertation projects that ensured our ongoing investment in the descriptive tasks and subsequent increase in skill. The digital surrogates were crucial to the efficiencies we were able to achieve, as well as to a better appreciation of the physical aspects of the manuscripts.

A descriptive exercise will illustrate this point. Aiming toward a full codicological description, consider Isl. MS 1016 from our Islamic Manuscripts Collection (fig. 1).³³ The digital surrogate³⁴ provides a clear view of the cover. Even at a distance (and certainly after zooming in), it is possible to see the texture in what appears to be brown cloth covering the spine and board corners. The cover image also reveals what appears to be a cream paper block-printed with reptilian design in blue-green. Continuing on through the images, it is also possible to see the turn-ins of the board coverings through the creamy pastedowns, which (looking at the image of the flyleaf) appear to be of woven paper. Inscriptions including a hint at a title and a book price are clearly visible on the “title page” along with a few stains and what may be a repair to the leaf. Turning to the incipit page, the

33 Cf. the description of this manuscript can be found at http://mirlyn-classic.lib.umich.edu/F/?func=direct&doc_number=006834164&local_base=AA_PUB.

34 Available at <http://hdl.handle.net/2027/mdp.39015079132000>.



FIGURE 1. Viewing the cover of Isl. Ms. 1016 (Ann Arbor, University of Michigan, Special Collections Library), via the digital surrogate in the HathiTrust Digital Library.

opening of the text clearly appears, more than adequate for transcription, and thereby for analysis to identify the text. A characterization of the hand is also possible. A seal impression is also visible on the incipit page and zooming in reveals the numerals of a date in the seal. The writing surface again appears to be wove paper as suggested by the texture and appearance of the zoomed image. Continuing on, the presence of catchwords, page numbers penciled in Western numerals, the number of lines per page, the absence thus far of any rubricated keywords or headings, and the persistence of the same hand are all evident. Skipping all the way to the last image of the lower cover reveals the edges of a label on the spine. The sequence number of this image suggests something of the extent. Continuing in reverse order reveals a blank but once folded scrap of paper that has been paginated and scanned, as well as several blank leaves at the close of the codex following the close of the text and brief colophon. This colophon provides the copyist's name but no date of transcription. The seal impression and its date allow for only a very rough *terminus ante quem* (since the seal could have been used long after its production) and the unfortunate positioning of the seal's ground decoration makes two date readings likely (though one perhaps more likely).

Obviously what has been compiled thus far represents a substantial description—nearly complete, but not fully complete—drawn from the digital surrogate for this manuscript without need to consult the physical codex. Work with the digital surrogate makes assessing these features quite quick and efficient, as long as one knows what to look for. After all, the surrogate does not suggest these features by providing direct links to them.

Still, an attempt at completing the description would swiftly reveal the limitations of this particular surrogate and the value of consulting the original. For example, it would be desirable to check the edges of the text block and the spine (including that label) for titles, but alas, no views are provided. The sewing and the gatherings should be assessed, and expecting quaternions or quinions gives a good idea of where in the surrogate to look, but no good views of the sewing are provided. How then to complete the collation? It would also be desirable to check each page for marginalia. Some of the views provided allow for counting and skimming each page, as would downloading the manuscript in PDF (eventually, given load times), but without good views of the sewing it would still not be possible to complete a collation statement. And what if some pages were inadvertently skipped by the digitizers? And what about the dimensions of the cover, the page, and the written area? And the issue of the dating? A better look at the paper would be incredibly valuable.

Examining the physical codex would allow for assessing many of these features in just seconds, and, with a careful collation, allow for evaluating the composition of the gatherings and the sewing (including any endbands) while simultaneously checking for any lacunae, quire numbering or middle-of-the-quire marks, repairs, marginalia (including seal impressions), illustrations, mistakes in the pagination, changes in hand, and paper characteristics.

In this case, the physical examination would reveal an embossed watermark, virtually impossible to catch in the current digital surrogate given its lighting, even if one were to spend the time carefully loading and skimming each zoomed image. With a bit of further research, this embossed watermark can be associated with a particular maker and with other dated examples, allowing for refinement of the manuscript dating.

This exercise demonstrates both the richness and the limitations of the

surrogates with which we were working, as well as the vital importance of training to make the most of what the surrogates afford (or do not afford). Approaching these imperfect surrogates with even rudimentary codicological training (basic “material manuscript literacy” if you will) made description to a certain level realizable at exceptional efficiency. We nevertheless clearly saw the limitations of the surrogates and found that it was necessary to consult the physical manuscripts in order to complete our codicological descriptions efficiently. Now, had the surrogates been more carefully designed for work with codex manuscripts, we might have been able to achieve more without physical examination. I have already discussed a few examples of surrogates that could have performed better.³⁵ Still, regardless of improved surrogate design, it is inevitable that examining the physical codex will remain more efficient, if not absolutely essential, for certain types of investigation: e.g., spectroscopy, tactile investigation, densitometry, etc.

Here Come the Users (or “Take Your Partner by the Hand”)

As I mentioned earlier, a vast number of users have worked with the digital surrogates in HathiTrust, often without our knowledge.³⁶ In late 2012,

35 Still another solution might be creation and deployment of multiple surrogates (such as the outputs of spectral imaging, transmitted light imaging, raking light imaging, etc.), each serving a different purpose and perhaps deployed relationally to form a single composite surrogate. I will leave those nuances of complex surrogate design to experts like our “SharedCanvas” colleagues; cf. Robert Sanderson, Benjamin Albritton, Rafael Schwemmer, and Herbert Van de Sompel, “SharedCanvas: A Collaborative Model for Medieval Manuscript Layout Dissemination,” in *JCDL’11 Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* (New York: ACM, 2011), 175–84.

36 Google Analytics figures reveal that a somewhat random sampling of 25 percent of the digitized manuscripts saw a total of almost 47,000 page views between 1 January 2012 and 8 August 2013, with most views (in descending order) for this selection of manuscripts coming from users in Turkey, Azerbaijan, Albania, Iraq, Hungary, Kuwait, Israel, India, Pennsylvania, Liberia, Iran, and New York. Most often users are arriving at the digital surrogates having searched (in Google or HathiTrust) for specific titles, authors, or keywords such as “münşeat,” “mecmua,” etc.

an Iranian researcher published an article on a text appearing in one of our manuscripts³⁷ relying entirely on the digital surrogate available to him through HathiTrust. The researcher submitted no requests for permission, nor did he send word of the publication. I learned of his research via another colleague and watched for the appearance of the publication for months. So, now the world is our reading room, it would seem.

Again, this interest is no surprise given the extraordinary level of access which these digital surrogates provide. That being said, if some basic level of material manuscript literacy is essential to engage with manuscript physicality via digital surrogates, are these users approaching our surrogates with such fundamentals? If not, are they somehow still managing to appreciate material qualities? If not, could the textual description and its treatment of physical characteristics suggest as much to them?

In 2013, a colleague and I conducted a user study to address these and many of the other questions I have raised in this essay.³⁸ Using Qualtrics, I created a survey in separate but equivalent English and Arabic versions, which we deployed at the page image and collection level in the HathiTrust user interface. The survey was active almost continuously from 1 July 2013 to 31 December 2013. The preliminary results are quite fascinating.

As of 13 November 2013, we had seen thirteen English version surveys completed and eighteen Arabic version surveys completed.³⁹ Respondents connected from Algeria, Egypt, France, India, Iraq, Kuwait, Libya, Morocco, Pakistan, Saudi Arabia, Singapore, Switzerland, Tunisia, Turkey, UAE, and USA. Most respondents indicated that the manuscript would be a primary source for their textual study, even critical edition. One respondent indicated that their purpose was to search for an image in a text. An-

37 Isl. MS 264, http://mirlyn-classic.lib.umich.edu/F/?func=direct&doc_number=006804519&local_base=AA_PUB, cf. ‘Alī Ṣafarī Āq’al’ah [= Ali Safari AghGhaleh], “Ḥayrat al-fuṣṣahā wa khajlat al-fuṣṣahā [= Hayrat al-Fuṣṣahā wa Khijlat al-Fuṣṣahā] matn-i fiqhī Fārisī nivishtah-yi 695 q. az ‘Alā’ Muftī Bukhārī,” *Guzārish-i Mīrās* 52–53 (2012): 87–94.

38 Deepest thanks to my colleague Kat Hagedorn for her contributions to the survey design and deployment. Immense thanks also to Roger Espinosa for his technical assistance with distributing the survey in HathiTrust.

39 The first response was to the Arabic version.

other indicated that the manuscript would be used for scholarly research in manuscript studies and textual editing and elaborated on a need for quality digital surrogates of Arabic manuscript exemplars when not possible to view the manuscript itself. Virtually all respondents indicated that the surrogate was meeting their research needs.

While additional detail on the nature of the research being conducted should be gathered in order to better interpret the responses, an initial comparison across English and Arabic versions of the survey is particularly interesting for a number of the questions. For example, when asked whether they would try to visit the Library in order to view the physical manuscript, no respondents to the English version replied in the affirmative and 55 percent of respondents said they did not need to view the physical manuscript. However, responding to the Arabic version, 50 percent of respondents said they would try to visit to view the physical manuscript and only one respondent said that they did not need to view the physical manuscript. With regard to whether or not they had looked at some version of the description in one of the online catalogues (the question was accompanied by links and screenshots to help communicate the intent), all English version respondents said they had looked at the manuscript description in one or multiple locations with 92 percent indicating that they had seen the description in the HathiTrust catalogue. For the Arabic version only 78 percent said they had looked at the manuscript description with only 44 percent looking at the description in the HathiTrust catalogue.⁴⁰ With regard to whether physical features are of interest when working with a manuscript, only 22 percent of English version respondents replied in the affirmative, whereas 33 percent of Arabic version respondents indicated that these features are of interest. Regarding the importance of considering physical features when working with a manuscript, none of the English version respondents selected “very important” and 15 percent said “not important at all,” whereas 47 percent of the Arabic version respondents said “very important” and none said “not important at all.” When asked to elaborate, the English version re-

40 It should be noted that the language of description is chiefly English with transcriptions for titles, names, and excerpts in the original script.

spondents mentioned the primacy of “legibility,” “searchability,” and “main text and marginal notes,” whereas the Arabic survey respondents provided textbook explanations of the essential nature of physical features for dating a transcript and as objects of study in and of themselves.

This discrepancy in response could suggest a discrepancy in training, and although the data are not robust enough to fully establish a correlation, additional details in the responses at least support the notion. When asked whether they had any training in “manuscript studies (i.e., codicology and/or palaeography)” only 15 percent of English version respondents said “yes,” whereas 47 percent of Arabic version respondents said “yes.” Arabic version respondents indicated that they had received this training formally as part of their graduate work in text editing (in universities in Algeria, Morocco, etc.), in separate short courses focused on the editing of texts (a cultural heritage center in Riyadh was among those named specifically), or in training at the Institute of Arabic Manuscripts in Cairo (which confers graduate degrees in manuscript studies, as do some universities and institutes in Iran).

Overall, the tendency of users to recognize the significance of the physical attributes and feel compelled to view the physical artifact is encouraging in the case of the Arabic survey respondents and discouraging in the case of the English survey respondents.⁴¹ A lack of training in material manuscript literacy could at least be part of the explanation—Arabic version respondents clearly indicated more extensive training than English version respondents, but the limited data more likely reflect the nature of the work being conducted with these digital surrogates by this small number of respondents. Thirty-eight percent of English version respondents did at least indicate that considering physical features was “somewhat important.” Another 38 percent selected “neutral,” perhaps reflecting a situational dependency. After all, not all research projects will require new assessment of material features, and quality work can be conducted without fully taking these features into consideration. Still, the significance of these features for suggesting value and changes in ownership as well as for more thoroughly

41 At least there is some assurance that users are viewing the catalogue descriptions. Whether they are consulting them for existing information on physical characteristics is a matter for further study.

establishing the original dating and any later interventions would hopefully be recognized and articulated regardless. Unfortunately only one of the very few English version respondents that elaborated on the question mentioned the potential significance of the physical features for providing “an idea about the quality of the work and the possible economic situation of its owners/readers, etc.” On the whole the English version respondents seemed to lack training in material manuscript literacy and to lack an enthusiastic appreciation for the manuscripts’ material qualities. By contrast, Arabic version respondents with training were able to appreciate material qualities and their significance to a greater degree.

Conclusions

The flood of readily available digital surrogates for Islamic manuscripts is both thrilling and somewhat concerning in light of the codicological and palaeographical work that remains to be done in the area of Islamic manuscript studies. Although these surrogates are “materially distant” they can actually perform quite well at mediating the material qualities of the artifacts they represent, particularly if designed to suit work with manuscripts. Nevertheless, without “material manuscript literacy,” that is, the ability to recognize the significance of the physical features of manuscripts and to discern and study them, scholars working with these digital surrogates may be more inclined to neglect the significance of physical qualities and even the essential historical context of the manuscript content they are studying.

Our experience with manuscript digital surrogates at the University of Michigan has demonstrated that providing scholars with basic material manuscript literacy and then asking them to perform descriptive tasks working with both physical manuscripts and digital surrogates can impart a better understanding of manuscript features as well as instill an appreciation for physical aspects. Our responsibility then, is to invest more in Islamic manuscript studies instruction to ensure that both the value of material qualities and the affordances of digital surrogates are properly understood. Fundamental training and exposure to the physical artifacts combined with ongoing engagement via digital surrogates can help us realize this goal.