Introduction

This book is for humanities students or scholars who are classically trained in handling manuscript materials and wish to take advantage of the incredible computing power at their fingertips but are at a loss where to begin. Some of the more technical parts of the book could be challenging, but a little persistence and practice, over time, should more than suffice.

I also hope to reach those who are more skeptical; who would agree with Paul Eggert, a book historian of modern English literature, when he wrote: "Speaking as a humanities scholar who lacks programming skills and ongoing access to a funded computing laboratory, the assumed advantage of the electronic environment is far less clear."[1] My response to this challenge is twofold. First, as colleagues around you introduce computer-supported solutions into their workflow, they will gain an edge over you. In fact, as digital methods gain wider currency, digitally restyling parts of our workflow will become the norm, and you may well be left behind if you decide not to do likewise. Second, and more importantly, I will argue that using computers for your erstwhile methodology and workflow requires some adaptions. This is chiefly because digital photos impose specific limitations: their resolutions might be very low, their colors might not be true to life, or they can only be accessed from the museum website. Knowing how to spot and judge these limitations will be most useful to work efficiently and accurately, and this requires a little bit of knowledge about what digital photos of manuscripts are.

In my experience, the most daunting aspect of applying computer-supported methodologies in one's work is its demand for a life-changing choice. Because, in this day and age, although technology is capable of doing more and more, students and seasoned scholars alike have had little or no exposure to it during their training. The result is that the so-called 'digital humanities' (DH) pose a real conundrum: either one pretends it does not exist, or one takes it as a specialization at the graduate level. In the second case, one stops becoming a historian of, for example, ancient Greece, medieval Islam, or the long eighteenth century, and is now on the path to becoming something else—a 'Digital Humanist,' where the 'digital' part dominates over one's original expertise. As a consequence, such a researcher tends to be restricted to communicating primarily with other 'Digital Humanists' and to publishing in their own specialized journals. The first group, meanwhile, does not invest significant time in

1 Eggert, P. "The Book, the E-Text and the 'Work-Site.'" pp. 63–82 in *Text Editing, Print and the Digital World*, edited by M. Deegan and K. Sutherland. Surrey: Ashgate, 2009, p. 63.

learning how digital tools can be used to make our work easier and better. In the rush of assignments, teaching duties, administrative burdens, conference preparations, brushing up language skills, and keeping up with developments in the field, it is one task too many to become acquainted with computer-supported solutions. This leads to an almost perfect disconnect between manuscript studies and 'digital humanities.' On the one hand, introductions to manuscript studies seem to overlook the part of our work that happens on computers, assuming that we have access to the material manuscripts.[2] On the other hand, introductions to digital humanities tend to suppose that the researcher already has the text in digital format.[3] The gap between the two is not seriously addressed.[4]

Another major hurdle in engaging with computer-supported solutions is the perceived cost of both time and funds. In digital humanities, research, all too often, is conducted using team-based projects funded by generous grants. The rationale behind this is that expertise in digital humanities can best be partitioned into two groups: those specializing in the 'digital' aspect and those who focus on the 'humanities' aspect. At the simplest level, this would result in teams of two experts, the one being a humanities scholar and the other a technician or a developer. While the scholar would develop the research questions and the conceptual path to a solution, the technician would make it happen. Experience shows, however, that what Snow calls 'the two cultures problem'[5] becomes almost insurmountable. If the scholarly problem becomes

2 Notably, even a team working specifically on 'digital editing of medieval manuscripts' produced a guide that assumes there are only material manuscripts, see Haltrich, M., E. Kapeller, and J.A. Schön, eds. *From Sheep to Shelf: An Illustrated Guide to Medieval Manuscripts for Students.* DEMM, 2017.

3 Blackwell's companion, arguably the field's flagship introduction, fails to mention IIIF and does not have a snippet of TEI to show what it is like, see Schreibman, S., R. Siemens, and J. Unsworth, eds. *A New Companion to Digital Humanities.* Oxford: Wiley-Blackwell, 2016.

4 Notably, the manuscript specialists from Hamburg include a little about digitized manuscripts in Bausi, A. (et al), ed. *Comparative Oriental Manuscript Studies: An Introduction*. Hamburg: COMSt, 2015. Likewise, the DH specialists Kurz and Jannidis devote a short section in their books to images: Jannidis, F., H. Kohle, and M. Rehbein. *Digital Humanities: Eine Einführung*. Stuttgart: J.B. Metzler, 2017; Kurz, S. *Digital Humanities: Grundlagen und Technologien für die Praxis*. Wiesbaden: Springer Vieweg, 2015. Two resources in English that come close but are a bit too specialized to serve as introductions are the book: Andrews, T., and C. Macé, eds. *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*. Turnhout: Brepols, 2014; And the journal issue: "The Digital Middle Ages." *Speculum* 92, no. S1 (2017). See also the series of collected volumes *Kodikologie und Paläographie im digitalen Zeitalter* (5 volumes so far).

5 Snow, C.P. *The Two Cultures and the Scientific Revolution*. Cambridge: Cambridge University Press, 1959.

more intricate and advanced, it becomes more likely that the technician will fail to provide a solution that truly encapsulates the problem. And when the technology becomes more advanced, it becomes more likely that the scholar will fail to understand how it can be improved to meet their requirements. If we scale back our ambitions towards using and modifying the existing technology, it is possible to have scholars operate on their own, as is customary in the humanities. The tools discussed in this book require no generous grant—they are mostly open source; they are free to download, use, and adapt. Such small adaptions can then be shared back to the community, fostering the organic growth of our toolbox.

Finally, another issue for manuscript studies is that when it comes to digitized manuscripts, there is no prior agreed upon conceptual framework, nor an acknowledged, basic skill set. This book addresses this matter by providing a conceptual and practical toolbox. If something does not have a name, we literally cannot speak or think clearly about it,[6] hence working out a conceptual framework is essential. Similarly, without discussing what we can do with digital tools at the beginner and intermediate level, we cannot properly determine the skills we must develop as a normal part of our methodological toolbox.

Two things should be noted. First, this book is not about digitization itself. Parts of this book, such as Chapter Six, discuss work that leads towards selecting artifacts to be digitized, but for the most part, the assumption is that your manuscripts of interest have already been digitized. If you wish to digitize artifacts yourself, or if you are a professional in this regard, you will profit most from Chapters Two and Three.

Second, neither is this book about plain-text analysis of repositories—such as *Index Thomisticus*,[7] *Thesaurus Linguae Graeca*,[8] *Library of Latin Texts*, *al-Maktaba al-shāmila*,[9] *Chinese Text Project*,[10] *Perseus Digital Library*, or *Project Gutenberg*, to name a few of the largest full-text databases from various fields. Such resources are a major advancement in the humanities, capable of yielding

6 Carroll, L. *Through the Looking-Glass, and What Alice Found There*. London: MacMillan, 1872, pp. 61–64.

7 This resource is generally considered to be the first 'digital humanities' project. It was started by Fr. Roberto Busa SJ in 1949 with the aim to have the entire corpus of Thomas Aquinas in electronic full-text format, searchable, indexed, and eventually syntactically analyzed.

8 The TLG has been a powerhouse since the 70's. I do find it odd that its title is written in Latin.

9 The name literally means "The All-Comprehending Library" (المكتبة الشاملة).

10 The Chinese name literally means "The Chinese Philosophical Book Digitization Project" (中國哲學書電子化計劃).

fantastic new avenues of research if we exploit them by automated mining, using a distant reading methodology.[11] But whereas digitization itself is presupposed, such plain text processing is a stage beyond the subject topic of this book. If what you have are images of texts, this is the book for you.

This book represents a case study of how the different aspects of digital manuscript studies may be integrated into the work of one person. This means that most chapters draw heavily from examples of my own work in Islamic studies. The non-Islamicists, I hope, will forgive me for my field-specific examples and find that the problems (and solutions) encountered are universal in nature, applicable to any other field involving digitized manuscripts. Similarly, my focus lies on manuscript and archival materials. Ancient texts or inscriptions, I am aware, may also appear on papyrus, stone, pottery, textile, coins, bone, and other materials. However, for reasons of space and my lack of experience with these materials, I do not address the peculiarities of engaging with such texts through a digital surrogate.[12]

The seven chapters between this introduction and the conclusion are split into two parts. The first three cover the conceptual and theoretical framework of thinking about digital surrogates. The next four explain the practical and technical skills.

Chapter One is theory-laden and, as such, comes logically prior to the rest of the book. Nevertheless, it could very well be read after the other conceptual chapters. In it, a framework is developed to examine the relationships between a material manuscript, print publication, and digital document. Here, I introduce the concepts 'manuscript world,' 'print world,' and 'digital world,' and I discuss how our work can be explained through the different relationships between these worlds. The manuscript world is a realm in which participants use and produce texts by writing them with ink by hand, on parchment, or paper. The print world is a world in which texts are machine-printed on mass-produced paper. Last, the digital world comes into existence when you type

11 This term was popularized by Franco Moretti as an opposite to 'close reading.' It has found diverse adaption in different fields of the humanities. For my own interpretation, see Lit, L.W.C. van. "Commentary and Commentary Tradition: The Basic Terms for Understanding Islamic Intellectual History." *MIDEO* 32 (2017): 3–26. Lit, L.W.C. van. *The World of Image in Islamic Philosophy: Ibn Sīnā, Suhrawardī, Shahrazūrī, and Beyond*. Edinburgh: Edinburgh University Press, 2017.

12 That is not to say that there is no interesting work done on them. See e.g. the successful markup standard for epigraphy called EpiDoc: Bodard, G. "EpiDoc: Epigraphic Documents in XML for Publication and Interchange." pp. 101–118 in *Latin on Stone: Epigraphic Research and Electronic Archives*, edited by F. Feraudi-Gruénais. Lanham: Lexington Books, 2010.

on a computer keyboard and see your input appear on an electronic screen. When we edit, we base our work on artifacts from the manuscript world. We work, meanwhile, on a computer, that is, in the digital world. Our final product, however, is often times a printed book, part of the print world.

Chapter Two forms the book's conceptual core. I discuss the perception scholars have of digitized manuscripts as 'larger-than-life' objects, emphasizing the ability to zoom in and make the tiny details invisible to the naked eye visible. I discuss how this perception rests on larger trends of thinking about mechanical reproduction and digital surrogates, in effect viewing digitized manuscripts as though they are a window through which one can look at the material manuscript. As such, it is unsurprising that scholars cite the material manuscript when, in reality, they make use of a digital surrogate. I also respond to the opposite view; that digitized manuscripts destroy the pure experience of handling the material manuscript itself. I argue that both views result from ignoring the 'digital materiality' of digital photos, which in turn occurs because we do not have a vocabulary to describe its features. I propose ten aspects by which to evaluate a digitized manuscript and its repository: (1) size of the collection; (2) online availability; (3) ability to download; (4) the portal through which the repository is accessed; (5) the viewer; (6) indication of page numbers; (7) image resolution; (8) color balance; (9) lighting; and (10) how the image is cut.

Chapter Three shows how these ten aspects can be used to evaluate twenty repositories, which are chosen to give a representative picture of the state of digitization of Islamic manuscripts worldwide. Since many of these libraries also host manuscripts of other disciplines, readers from beyond Islamic Studies should still find this informative of the general state of digitization. The result is that quality and usability varies wildly. Not all manuscripts are downloadable, and the legal restrictions applied to them are often ambiguous. I end this chapter by speculating on the future of these repositories.

Chapter Four starts off the practical part of the book with two topics. First, I discuss how manuscript research specifically concerned with computer-supported solutions has evolved into team projects supported by big grants. I wish to highlight, in particular, when such big projects work well and when they seem to fall short of expectations. From this, we learn that we cannot rely on such teams to produce technical solutions to our problems. Instead, we need to take matters in our own hands. As the first step in this direction, I provide a practical example of how a tablet and free drawing software can be used to perform simple yet effective paleographic work. This part of the chapter is an extended and more in-depth version of an article I previously published,

which discussed three glyphs that appear in a text by twelfth-century philosopher Suhrawardī, who claims that only the initiate will understand how these symbols represent the essence of his philosophy. By (literally) drawing from several medieval manuscripts and combining different versions of the glyphs, I arrive at the interpretation that the symbols are constructed from Arabic letters.

Chapter Five discusses the workflow of digital editing. The particular software one uses will change over time, but many of the technical standards on which digital documents rest will remain the same. For this reason, it is essential to know these standards, such as Unicode, TEI, and IIIF. To lower the barrier for working in these standards, I shall provide some pointers on how to set up your computer—for example, how to create your own keyboard layout. The chapter finishes with a discussion of the pros and cons of what is called a 'digital edition': a publication that does not appear in print and, therefore, can take on digital forms unimaginable in the print world.

For digital publishing of any kind, knowing web development technology is a terrific asset. In Chapter Six, I explain the entire process of creating an online catalog of a hitherto uncatalogued collection. We first look at how computers and smartphones can be helpful for fieldwork in an archive and then turn to create an interactive website to make the catalog openly available to others, using HTML, CSS, JSON, and JavaScript. We finish by looking at how those same technologies can help in creating attractive and insightful diagrams.

Chapter Seven adopts a notably technical character. I explain how one might use a simple programming language such as Python and a well-known function library called OpenCV to analyze the covers of several thousand digitized manuscripts. The method is automated image recognition, and it aims to say meaningful things about the shape of the codex. All the while, the core skills for programming that are explained can be applied to any other use case. As such, the chapter revolves around introducing programming in general and Python in particular.

Chapter Eight, the conclusion, is divided into three sections. The first stresses the point that 'digital' humanities and 'classical' humanities are not contradictions; that the latter will benefit from working in the digital world. There has, for this reason, never been a better time to conduct classical philological, codicological, or paleographical studies. The second section synthesizes the lessons learned from the book. In the third section, I offer my perspective on the future of the ongoing incorporation of digital assets and tools within humanities.

This book has two additional parts. As a postscript, I include a few stories about my experience in handling digitized manuscripts in a style similar to

Ignaz Kratchkovsky's memoir *Among Arabic Manuscripts*.[13] With these stories I wish to show that the experience of reading a material manuscript may be destroyed by using a digital surrogate, but other experiences of equally personal and emotional quality come about. It also gives additional insight into my daily practice concerning obtaining, storing, and retrieving digitized manuscripts.

The last part of this book is not actually in this volume, but is a companion website, a digital appendix. You can access it through the URL right below. There you will find images, code, data and other relevant digital documents. In this book itself you will find no URLs or DOIs: they have all been moved to the digital appendix. In fact, many more online resources, technologies, and tools are listed there to give you the opportunity to explore specialized or more advanced options after you have acquired the foundation that this book offers.

For the digital appendix go to GitHub.com/Among

For the stable, citable repository go to Zenodo.org/record/3371200 and use this DOI: DOI 10.5281/zenodo.3371200

Additionally, you will find a QR-code for every chapter. It will take you to the correct folder within the digital appendix.

13 Kratchkovsky, I.Y. *Among Arabic Manuscripts*. Translated by T. Minorsky. Leiden: Brill, 1953 [Reprinted 2016].