A Digital Orientalist

How can we hold the digital world in one hand and the manuscript world in the other? From 2013 onwards, all my activities involved in putting computers to the service of humanities studies were grouped under the project name 'The Digital Orientalist.' 'Digital' referred to anything researchers can achieve by means of computers, especially the seemingly futuristic and limitless potential of obtaining ever larger data sets and manipulating them in whatever way they think is meaningful. The term 'Orientalist' is, of course, not aiming at the Orientalism that Edward Said so profoundly dismantled[1] but to an archaic grouping of fields of studies, among them my own Islamic studies, which have the usage of ancient manuscripts in common, as their most important source for the study of civilizations to the East of Europe.[2] At first glance, the combination of these two terms seems to be an oxymoron. However, by now it has become more and more evident that the supposed contradiction between what is 'Digital' and what is 'Orientalist' is, in fact, a fusion that leads to a great increase in our research capabilities.

This state of affairs makes for exciting times for classical fields of the humanities. Our fields often involve the painstaking collection of large amounts of snippets of evidence from disparate and scattered sources as well as close reading of texts to tease out its many layers of meanings, allusions, cross-references, and intertextuality. Much of these activities can be done cheaper, faster, and better by using digitized materials in a digital workflow. We can give a computer the boring stuff that test our patience and abilities, such as remembering where we kept every note we made or noticing a similarity between a new note and one we made a long while ago. As for us, we only need to focus on the interesting stuff, such as deciding whether such a similarity is significant or not. In addition, computers add entirely new tools to our toolbox, such as image manipulation and distant reading. 'Classical' humanities will thrive because of 'digital' humanities.

---

1 Said, E. *Orientalism*. New York: Pantheon Books, 1978.

2 As still witnessed by learned societies such as the American Oriental Society or the Deutsche Morgenländische Gesellschaft, and also still present in the name of renowned institutions such as Chicago's Oriental Institute, London's School of African and Oriental Studies, and Paris's Institut national des langues et civilisations orientales.

This new wave comes with an important challenge, namely, that we will have to invest ourselves in making these digital tools our own. This is only logical since our fieldwork, or rather the soil of our fields, has changed dramatically. To get to our manuscripts, we no longer go to dusty libraries but turn to buzzing computers, and if we do go to a library, we often come back with digital photos. Not so long ago, it was not expected of classically trained students and scholars to have computer skills.[3] As it becomes more and more obvious that a 'Digital Orientalist' is not an oxymoron, it will become increasingly necessary and expected that all researchers possess some understanding of and skills in computer technology.

What has reading this book been good for? We have gone through extensive conceptual and theoretical discussions and worked our way through a diverse palette of technical and practical skills. The theoretical discussions attempted to foster an awareness of the changing nature of our fieldwork through which we can avoid mistakes that would arise from approaching our digital workflow with a manuscript or print mindset. Two big mistakes highlighted in this book are (1) to disregard the digital materiality of digitized manuscripts, and (2) to limit oneself to print-like publications.

Issues surrounding the first mistake are addressed in Chapters Two and Three. From Chapter Three, we learned that not all digitized manuscripts are created equally. In other chapters, we prepared ourselves for this by learning skills to change digitized manuscripts to a more usable shape. In Chapter Five, we found out that this itself can introduce differences. For instance, we noticed that using different methods of converting a file from one format to another will yield different outcomes. In Chapter Two, we found the right vocabulary to describe this change and difference and introduced ten notions that can help us describe the digitized manuscript as a file and its *Sitz im Leben*. From now on, when we use a digital surrogate in our research, we are able to notify the reader of this in our publications by including a sentence that says something along the lines of 'The digital images used in this work are $X_{px} \times Y_{px}$ showing N pages, at about z kb. The cut is tight and color balance is fine.' The X and Y values are the pixel **dimensions** of the actual images, N is usually one or two depending if the image is of one **page or a page-spread**, z is the **file size**, the **cut** is a description of how much is visible around the page, and the **color balance** is a judgment on whether the colors in the image are true to life.

The implications arising from the second mistake are not simply a plea for digital editions. As addressed at the end of Chapter Five, born-digital

---

3   Babeu, A. *"Rome Wasn't Digitized in a Day": Building a Cyberinfrastructure for Digital Classics*. Washington: Council on Library and Information Resources, 2011, p. 139.

publications could be the right outcome of your project but are not necessarily the most useful ones. More importantly, the print-publication mindset can limit us even far earlier in the creative process and this would be a serious mistake. In Chapters Four and Seven, we saw that digital surrogates can be manipulated in ways virtually unimaginable in the print world. While the first of the two above-mentioned issues chartered the limits of using digital surrogates, this second one highlights the incredible potential. Concurrently, it should be noted that this potential does not impact our tried and tested methodologies and project management: older approaches are still valid, as we are merely superimposing specific computer-supported methods. This is important to note since there is a mindset that has portrayed digital humanities as possibly antagonistic towards the fundamental methodologies of classical humanities. This mindset encouraged the shift towards big-grant team projects. In this book, I have forcefully argued against this shift. Such projects want to split technical and scholarly tasks and assign them to different people. In my opinion, we ought to combine in our own selves the technical know-how and the field-specific expertise. This can take many shapes; instead of thinking of 'digital humanities' as in binary opposition to 'classical humanities,' I advocate for seeing digital humanities as a spectrum. Only if we manage to fill that entire spectrum can we make the most of computer technology, and then each of our humanities disciplines can gradually progress in building an ecosystem of standards, tools, and technologies, and possibly repositories, that we actually use. Of course, this does not exclude the possibility to work within or to start a big-grant project. In fact, having verifiable experience in both technical and field-specific aspects will make you a better candidate for that.

Looking over this book, you will have found very few technologies that were specifically developed for the humanities (TEI and IIIF are the exceptions: see Chapter Five.) There are many technologies from other industries that we can profitably adapt and use. In Chapter Four, we cherry-picked tools from the graphic design industry; in Chapter Six, we reaped the low-hanging fruits of the web development industry; in Chapter Seven, we made thankful use of the tools developed by the industry that needs automated object detection on video footage. The tech industry today has an extraordinary 'sharing is caring' mentality, which means that many powerful tools are free to use and free to learn. In particular, the world of web development is easy to penetrate with its very large pool of learning materials. The goal, then, of this book is not so much to outfit you with the tools you need but, rather, after you have acquired the foundations, to motivate you to keep learning. In each of the above-mentioned technologies, changes are bound to happen. But with proper knowledge of the foundations, adjusting and adapting new software or new programming

libraries will only be natural, as you continue to incorporate into your personal toolbox the latest, most interesting technology that is relevant to you.

The specific examples used in the chapters can be merged into powerful combinations. For example, you can draw vector-based SVG-images using what you learned in Chapter Four, then mass-manipulate them with Python using Chapter Seven, then present them through an interface using Chapter Six, keeping in mind the standards discussed in Chapter Five. Even if you intend to remain a passive consumer of technology, there is great merit in looking over these practical chapters. This will yield a sense of mastery, of being in control and, by extension, it will instill a creative liberty providing you with the confidence to handle digitized manuscripts and do with them whatever you want—nothing is impossible. Moreover, you will know which technical limits have been reached and can, consequently, temper the exaggerated expectations that some may have.

What can we expect in the future? Here is a sweeping statement by Jerome McGann from 2005: "In the coming decades—the process has already begun—the entirety of our cultural inheritance will be transformed and re-edited in digital forms."[4] Our own fields and their unstoppable wave of digitizing manuscripts are proof that this is true, so much so that we are already leaving behind the era in which it is a novel request to ask for digital images of a manuscript. Instead, we can begin to transform our workflow into digital forms. An excellent sign of this is will be the inclusion of digital methods in our teaching. As this progresses, the work now known as 'digital humanities' will become increasingly integrated into every field of the humanities in their own particular ways. Its application will become normal, even expected, and so the epithet 'digital' will lose significance.[5] Again, we should look out for the moment when specific training in 'digital methods' quiets down and inclusion in regular courses becomes expected.[6] Digital humanities as a field surely still has a future of its own, perhaps similar to Statistics which is used throughout the sciences but knows its own specialists with their own discourse and manifests itself as a department in its own right here and there. Nonetheless, it is the institutional embedding of this field that may need the greatest change. Already now, DH

---

4 McGann, J. "Information Technology and the Troubled Humanities." pp. 105–21 in *TEXT Technology* 14, no. 2 (2005), p. 109.

5 Cf. Zorich, D.M. "Digital Humanities Centers: Loci for Digital Scholarship." pp. 70–78 in *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*. Washington, D.C.: Council on Library and Information Resources, 2009, p. 77.

6 Note that this book's chapters refer to 'Codicology', 'Paleography,' and 'Philology' and not 'Digital Codicology' and so forth.

centers are criticized for being "silos of activity and redundant resources,"[7] and the people therein are accused of being "inward focused."[8] These characterizations are especially painful since the most recurrent values mentioned in digital humanities literature talk about desilofication, reuse, and expansion of our horizons. Part of the problem, as I see it, is that DH centers are almost always financed by generating income through grants. Such grants can only be obtained by proposing big projects. Then, digital humanities is locked into doing big projects that invariably foster an in group–out group culture and produce technology that is created and stored at a center, to the detriment of the periphery. To counter this situation, decentralization will hopefully become a key theme; decentralization of things such as repositories, tools, and standards but also of labor such as programming and encoding. In short, decentralization of responsibility. This would foster a positive stance towards open source and open access and encourage everybody to contribute with small additions. We know that in a manuscript culture, readers cast votes by choosing texts they like to write out and make new copies of. This way, texts either become popular or die out. Similarly, faced with the new challenges presented by the digital world, we do not have to create the perfect end solution in a single effort, nor expect our solution to be universally accepted. We can, instead, use a collection of small tools and standards, swap out one thing for another if something else works better, and if we add something of our own, then share it back with the rest of the world. Gradually, but perceptibly, our fields will turn digital. It has never been a better time to be studying ancient manuscripts.

---

7   Zorich, p. 71.
8   Prescott, A. "Consumers, Creators or Commentators? Problems of Audience and Mission in the Digital Humanities." pp. 61–75 in *Arts & Humanities in Higher Education* 11, no. 1–2 (2011), p. 67.