The Digital Orientalist:
Scholarship is Becoming Bigger and Better

L. W. C. (ERIC) VAN LIT
YALE UNIVERSITY

Introduction

Over the last few years, I have been writing about how computer technology can be implemented in day-to-day scholarly activities on my weblog, www.digitalorientalist.com. That name, *The Digital Orientalist,* was chosen to reflect the very question I wished to explore: can computer technology of today and tomorrow be fruitfully put to use in a field like Islamic Studies which is borne out of old-fashioned philological work? Is there such a species as a "digital orientalist," or is that a contradiction in terms?

I soon realized that the question is actually of a slightly different order. It is not a question of possibility, but rather of necessity. The shift from print to digital comes with a whole slew of ramifications that either have already taken place or are in the process of coming about. Students and scholars operate in a drastically different way than ten to twenty years ago. But at the same time, there is an astonishing lack of reflection on this shift and just as little technical know-how among said students and scholars.

In this essay I shall describe some of those drastically different ways in which scholarly activities are executed, especially those that have in one way or the other to do with libraries. My description is largely based on my own experience, and this piece may therefore be best read as a user case. I think libraries and librarians could (and already do) play a vital role in integrating computer technology into day-to-day academic activities, in a flourishing and sustainable way. I hope that this piece inspires a continued dialogue between librarians and patrons on such issues.

Growth of Private Libraries

My visits to libraries are currently limited to 30 minutes, sometimes less. I have now become accustomed to scan the parts of the book that

I need at the library and give it back immediately. No matter what the book is, I have one approach: I look at the book at the library and I decide if I need to scan it or not. I am currently at a university that allows me to have library staff bring books downstairs even if they are just on the shelf. This means that no matter if a book is on the shelf, in storage, or has to come from another institution, I always pick it up downstairs, where it is sitting on a shelf with my name on it, and I only come when I have received an e-mail confirming it is sitting there. This makes my trip to the library extremely efficient, and that small investment of time behind a scanner is an easy trade off against having that source with me for the rest of my life.

The biggest drawback of such short visits is that I am no longer browsing the shelves. For a while I was at McGill University in Montreal, and their library is entirely open-stack. At the dedicated library for Islamic Studies, you get this wonderful spatial feeling for the collection of 150k+ volumes and simply walking through it has made me discover interesting books I had never heard of before. However, I soon realized that this is still not a guarantee of getting the entire picture, as a certain percentage of the books are loaned out—usually exactly those books that are the most interesting. Other books are simply not in the collection and therefore also fall out of bounds. Yet I find it to be better than nothing, as I have not found a comfortable way of browsing the shelves digitally. If such a thing would be developed, it would still depend on how rich the metadata of the catalogue is. In other words, as the distance between patron and collection widens, cataloguing becomes more important to usefully direct the patron to the right book. How wonderful it would be if I could just look in every book's table of contents while browsing the online catalogue.

Until then, all I can do is page any book I may think is of interest and have a look myself. The need to scan is mostly an educated guess about the need for this text in the future. I am quite generous in this guess; sometimes it seems more about peace of mind than actual need. I would rather have a PDF that I never touch than somewhere down the line, whether it is weeks or years from now, wishing I could take a quick look at it and not being able to do so. For an often-used book it certainly is great to have a hard copy, but print copies of books can still be a scarce good. As an individual, it is sometimes not possible to find a copy for sale, especially books published years ago with a small print run. Even if it is for sale, books of academic interest can

sometimes be very costly. Out-of-print academic books, sold through an antiquarian, rarely go for less than their original value. New copies only seem to become more expensive, as academic publishers rely on selling to institutions and can therefore ask more for a book than they would if they had targeted the consumer market. It is not only more cost-effective to scan, but requires far less time than obtaining a private copy. One popular scanner, the Scannx 6167, allows me to do about a hundred pages in ten minutes. A Xerox machine, like the WorkCentre 5665, is even faster. My educated guess is thus a generous one, and invariably includes the title page and the table of contents.

My compulsive hoarding and paranoia about books going missing have left me with hundreds of gigabytes of data, mostly PDFs.

I especially like PDFs that are a collection of images of a printed book (like a scan) because it is a bits and bytes representation of print material, allowing me to see how the text looks on the page and on which page exactly it is. A misprint, for example, seems more authoritative in print than in digital format. Even for fully digital documents PDF is preferable, because it cannot be altered (easily) and is therefore more stable and will probably still be able to be opened many years from now. I would like to think that PDF is here to stay, but at the same time I know that cars first looked like horse carriages but not anymore, and keyboards first looked like typewriters and do not anymore. In fact, printed books looked at first a lot like manuscripts. Equally, digital copies of texts in this first stage look a lot like printed books, but perhaps they will not twenty years from now. What I am betting on is that someday OCR technology will be good enough for Arabic to simply turn my PDFs into fully digital, searchable texts, so that I can export them in whatever the standard file format will be.

As for the size of my library, I need not worry. In fact, the physical size is only shrinking: the most precious part of my private library sits on a 128Gb Micro SD Card, the size of which is smaller than a postage stamp. This costed me $55. My entire library sits on an external hard drive of 2Tb, which costed me $88. Enclosed in its protective case, it is smaller than one book. And still the industry is able to cram more gigabytes onto a smaller surface, for a lower price. These are consumer goods that are readily available to anyone, and at such price points should be within reach for anyone even remotely serious about their studies.

This changing nature of the private library changes not only the opportunities, but equally the threats. It is easy enough to destroy an external hard drive, and once that one tiny device is destroyed, so are all its contents. We all too eagerly mix private equipment with institutional equipment. We have no problem sticking foreign USB sticks into our computer. And we happily click and download whatever we can find on the internet. All of these reckless actions pose threats to lose everything we have in a matter of seconds. One thing I would like to point out in particular is the case of collateral damage. Sony Entertainment was hacked in November 2014, or at least that is when it became clear it was hacked. It seems most likely that the only thing the hackers were after was retaliation for Sony's production of a satire about North Korea.[1] But instead of targeting specific individuals connected with that production, the hackers simply tried to erase any and all content they could find on the internal network of Sony. What this means is that even if your own project seems innocent enough, if somebody within your organization does do something that may irk the wrong person, your data is in jeopardy as well. I have described on my website the back-up strategy that I currently employ, and it is still undergoing changes.[2] What I consider most important is to have an offsite, offline backup.

The end result is a private library of a magnitude never seen before. What I have described here pertains to myself, and even though I like to think I am ahead of the wave, it should be understood that many students and scholars work with similar principles if only subconsciously. The question "does anyone have a PDF of …" has become ubiquitous, and is preferred to going to a library to pick up a hard copy. It is only a matter of time before digital library curation will become better understood and practiced, though technical support in this area will always be much welcomed.

The logical conclusion of all of this is that the status quo in knowledge dissemination—the complex of publishing, distributing, and collecting, and the laws to protect the interests of all parties

---

1. One insightful account is P. Elkind, "Sony Pictures: Inside the Hack of the Century," *Fortune* (July 1 2015), available at http://fortune.com/sony-hack-part-1/

2. L. W. C. van Lit, "Al-Ghazālī on Archiving and Backups," *The Digital Orientalist* (May 19 2015), http://digitalorientalist.com/2015/05/19/al-ghazali-on-archiving-and-backups/

involved—is headed towards a major disruption. At the moment it seems that those with the biggest economic interests, commercial publishers and distributors, are doing their best to nip any deviation from the old model in the bud. Well-known incidents that indicate this are Elsevier's changed policy for sharing academic work published through them (unusually restrictive),[3] JSTOR's lawsuit against a student who had a computer automatically download extensive parts of JSTOR's collection (surprisingly hard-hitting: the student committed suicide while on trial),[4] and the cease-and-desist letter of a consortium of publishers that killed Gigapedia (also known as library.nu).[5] But I think those are just symptoms of a much larger force stirring, one that goes bottom-up. To navigate through the minefield of legal issues, individuals will likely act increasingly amongst themselves and not act with or through an institution.

Diminishing Demand for Scarcity

Obviously, reproduction in a digital era in which the internet connects everything is no longer something to which you can attach exclusivity. It simply does not work that way anymore. This combination of virtually unlimited reproduction at almost no cost with the fast and easy transportation of such files over the entire planet has a big impact on the way scholarship proceeds.

Most importantly, completeness is fast becoming a standard one needs to meet rather than a happy luxury. The argument "I have not been able to inspect this text" would in the past satisfy a reader. It signals that the author is aware of the importance of this or that book or article, but was unable to get it in his hands to read it and incorporate it in his analysis. Especially in a field like Islamic Studies, books of interest can be produced in faraway places, with only a limited print

---

[3] A. Wise, "Unleashing the power of academic sharing," *Elsevier* (April 30 2015), http://www.elsevier.com/connect/elsevier-updates-its-policies-perspectives-and-services-on-article-sharing

[4] J. Schwartz, "Internet Activist, a Creator of RSS, Is Dead at 26, Apparently a Suicide," *The New York Times* (January 12 2013), available at http://www.nytimes.com/2013/01/13/technology/aaron-swartz-internet-activist-dies-at-26.html

[5] C. Kelty, "The disappearing virtual library," *Al-Jazeera* (March 1 2012), http://www.aljazeera.com/indepth/opinion/2012/02/2012227143813304790.html

run, and perhaps decades ago. The reader, undergoing the same trouble as the author in obtaining certain texts, knows how hard it can be sometimes and can easily forgive the author for not including one or another text in his study. But currently this argument sounds more like laziness than honesty. "You are telling me you could not get your hands on that early twentieth-century lithograph print from Calcutta? I find that hard to believe!" If a scholar wishes to write about a certain topic, they are now responsible for collecting every and all texts that are pertinent to that topic.

A corollary to this is that due diligence is also expected, for which an author can be held accountable. By due diligence I mean the act of inspecting non-obvious texts for any relations with the topic under investigation. In a field like Islamic Studies, pertinent texts can hide in unexpected places, and we thus need to cast our nets widely. Then we discard the majority of these texts and never mention them in our research, but it will allow us to rebuff inquiries as to whether we looked into this or that text. The answer should at the very least be "yes, I flipped through it once and I deemed it to be of no importance to my research."

Finding a source has become mandatory, but so has revisiting a source. The affordability of losing a source has dropped to practically zero. Especially at crucial stages of an academic career, such as during the writing of a Ph.D. dissertation, failure to revisit a source is not an option. This naturally encourages scholars to scan and store sources digitally, as that way it is easy to keep a large amount, have multiple copies at different locations, and look at them wherever you are. With this digital option now readily available, it is quickly turning into a requirement rather than a luxury. Here the principle is at work that if everyone else is doing it, they are not at an advantage; you are simply at a disadvantage if you are not.

In the past, research could validate its novelty through exclusivity of access to a source. With the democratization of access to knowledge, this belongs to the past. Take, for example, the case of Wilfred Cantwell Smith, the great scholar of Islam and religious studies. His doctoral dissertation is remarkably simple; a survey of the content of a dozen or so years of the journal of al-Azhar University. Clearly, the project was at the time sound, and he was granted the degree, because he was disclosing a source unavailable to most others. This is now of the past; a scholar can still partially justify his research as disclosing a source that has previously not been given the attention

it deserves, but no longer by claiming that only they can disclose it because no one else has access to it.

Similarly, sources that were normally not citable, because unpublished, are now becoming valid references. I am thinking in particular of manuscripts, which have always been the backbone of research in Islamic Studies. Digitization of manuscripts has taken off in a spectacular way, and the holdings of many collections are now easily digitally browsable and viewable. It is currently still more work to look up a manuscript than a book, but a reference to a particular folio of a particular manuscript is not as funky as it used to be.

It is my impression that these changes in possibilities also bring about changes in the scope and nature of our research. Looking again at W. C. Smith's dissertation, its scope was fixed and limited to looking at a couple of issues of one journal. Now with so many resources available, and a subset of them even digitally searchable (such as *al-Maktaba al-shāmila* or *Noorlib.ir*), scope can be fixed in other ways. For example, the scope can be an argument, such as Ibn Kammūna's novel argument for the soul's eternity *a parte ante,* which is identified throughout Ibn Kammūna's corpus and contextualized through a myriad of texts from before him.[6] I have myself fixed the scope of a chapter of my dissertation by looking at one single sentence, and tracing its trajectory as it is copied from text to text starting in the twelfth century and ending in our present time.[7] We equally see a rise of quantitative analysis, such as a study on the composition of a 14th-century encyclopedia,[8] and the construction of a social network around 18th-century scholar Murtaḍā al-Zabīdī.[9] These are some simple examples of how scholars are looking for new ways to utilize the possibilities that are at their disposal.

---

[6] L. Muehlethaler, "Ibn Kammūna (d. 683/1284) on the Eternity of the Human Soul: the three treatises on the soul and related texts" (unpublished Ph.D. thesis, Yale University, 2010).

[7] L. W. C. van Lit, "Eschatology and the World of Image in Suhrawardī and His Commentators" (unpublished Ph.D. thesis, Utrecht University, 2014).

[8] E. I. Muhanna, "Encyclopaedism in the Mamluk Period: The Composition of Shihāb al-Dīn al-Nuwayrī's (d. 1333) Nihāyat al-Arab fī Funūn al-Adab" (unpublished Ph.D. thesis, Harvard University, 2012).

[9] S. Reichmuth, *The World of Murtada al-Zabidi* (Cambridge: E. J. W. Gibb Memorial Trust, 2009).

Conclusion

The practice, the scope, and the nature of scholarship in Islamic Studies is changing due to a shift to digital media. The size of private libraries is quickly surpassing those of the past, making students and scholars to a greater extent self-sufficient. Conversion into digital format is easy enough to absorb any materials found non-digitally, making library use more sporadic and more efficient. The biggest drawbacks of this shift in practice is that random discoveries through browsing are impossible and a tighter back-up plan is needed.

These changes greatly enhance the potential productivity of scholars; however, this also means that expectations towards peers are increasing. Top scholarship demands absolute completeness in the body of evidence that is involved in the analysis. In this sense, the possibility of a more free flow of information turns into a demand for a more free flow, as one cannot afford to miss out on a certain source.

Since this demand for a more free flow will ensure an ever larger body of readily accessible texts, this in turn will influence the very nature of scholarship. We are now able to ask different questions of our sources, and this process will undoubtedly be among the most significant trends for the next few decades. Like a hermeneutic circle, the changing nature of scholarship will influence our practices, which will influence our scope, and therefore also the nature of our work.

The Pandora's box of the digital format has been opened, and it will be most exciting to see how librarians will partner with the scholarly field to accommodate and catalyze the many changes that are underway.