Open Arabic Periodical Editions

A framework for bootstrapped scholarly editions outside the global north

Till Grallert, Orient-Institut Beirut

2020-07-10

# Introduction

Early Arabic periodicals published across the Eastern Mediterranean from the mid-nineteenth to the early twentieth centuries were at the core of social developments and formative discourses of modern(ising) societies that carry continued prominence across the region and beyond: the Arabic (cultural) renaissance (*nahḍa*), Arab nationalism or the Islamic reform movement (*ṣalafiyya*), to name just a few. Yet, the vast majority of journals and newspapers remain obscure and understudied beyond a few well-known titles from Beirut and Cairo. As physical artefacts—frequently printed with cheap ink on cheap paper and handled without care in transit—they are vulnerable to neglect and active destruction. Neoliberal defunding of cultural heritage institutions as a global phenomenon is compounded by an onslaught of iconoclasts, failing institutions, and wars ravaging Syria, Yemen, and Iraq.

Turning to Gil and Ortega’s question of “What do we need?”[[1]](#footnote-1), the answer, therefore is simple: preservation and access! Societies of the Global South have a right to unhampered access to their own cultural record—a cultural record that is frequently held by institutions of the Global North. Access would also allow the scholarly communities concerned with the histories of Middle Eastern societies to study them through their own cultural production.

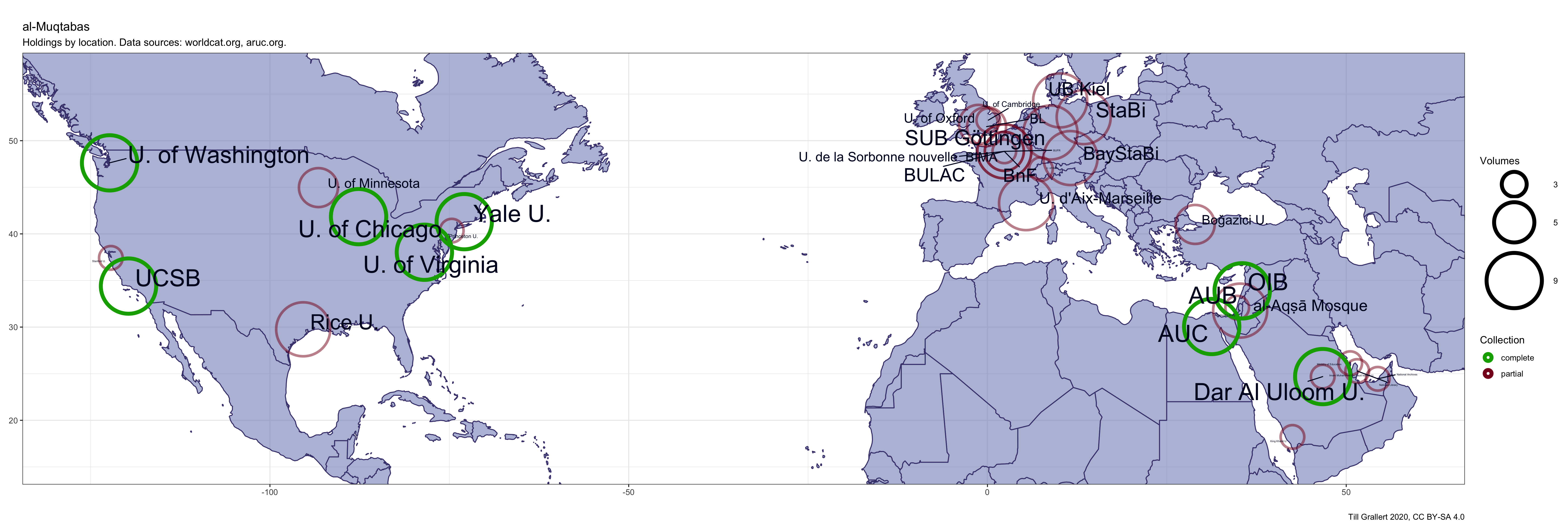
The following paper consists of two parts. I begin with challenging the equation of “digitisation = access” by outlining the layers of inaccessibility inherent to existing digitisation efforts of cultural artefacts from the late Ottoman Eastern Mediterranean—historically multilingual and multiscriptoral societies, whose material heritage has been and continuous to be looted, destroyed and neglected; cultural artefacts, which resist digisation by being dependent on scripts other than Latin (such as Arabic, Armenian, Syriac and Hebrew scripts) and calendars—and corresponding conceptions of time—other than the Gregorian (such as the Ottoman fiscal, or *mālī*, calendar, the reformed Julian calendar or the Islamic *hijrī*); and contemporary societies, who cannot draw on the on the vast resources in wealth and socio-technical infrastructures of the Global North. The second part introduces the project [Open Arabic Periodical Editions (OpenArabicPE)](https://openarabicpe.github.io) as an attempt to address some of these layers of inaccessibility through bootstrapping existing data, tools, and infrastructures into a framework to produce and distribute open scholarly periodical editions under the affordances of the Global South. I argue that OpenArabicPE proves the feasability of our bootstrapping approach by, inter alia, having published five full-text editions of Arabic journals originally published in Baghdad, Cairo, and Damascus between 1892 and 1918 with a total of 20 volumes comprising 246 issues, 12.830 pages that link to digital facsimiles, and 3.166.783 words, without any resources beyond voluntary labour and private laptops.

# Inaccessibility of digitised Arabic periodicals

Digitisation is the obvious answer to the need of preservation and access. Arabic periodicals are indeed being digitised and made available to online access for at least 15 years by a combination of private and public, non-profit and for-profit vendors. The terms *digitised* and *access* have the power to evoke hyperbolic promises but they have always a concrete meaning bound to their particular historical context. Depending on who we are, digitised artefacts are fraught with multiple layers of *inaccessibility*.

The first layer of *inaccessibility* is rooted in the physical artefact and its history. First and foremost, we will not know what we are a looking for and where we could do so. There is a plethora of encyclopaedic works on the Arabic press, listing titles and dates of first publication, but the individual history of the vast majority of publications remains unknown—particularly for cities beyond Cairo and Beirut.[[2]](#footnote-2) If we were interested in the series of food riots that shook cities across the pre-dominantly Arabic speaking provinces of the Ottoman Empire in summer 1910, we simply would not know which titles were actually published in Homs, Hama or Gaza.

The second layer of *inaccessibility* can be summarised by survival bias, collection bias, and digitisation bias. Surviving copies of early Arabic periodicals are scattered across libraries and private collections worldwide. Known collections are frequently incomplete and predominantly located in the Global North.[[3]](#footnote-3) Many of these periodical collections came about by chance and, to a large extent, reflect local, regional and global distribution networks and individual publishers’ ability to gift library copies long after their publication.[[4]](#footnote-4) Existing catalogues are not necessarily published and with union catalogues haven fallen out of fashion,[[5]](#footnote-5) we would also be at a loss in locating a specific publication or the press output from a particular place at a given time. Digitisation and hosting digital artefacts are expensive and the cost scales almost linearly. In need to justify their expenses, curators will almost always turn to the rare and beautiful. Hundreds of thousands of periodical pages in foreign languages are frequently not considered important enough to warrant this investment for institutions in the Global North. Funds, infrastructures and access to the physical artefacts, on the other hand, are commonly not available in the Global South. In consequence, we witness a neo-colonial absence of the Global South from the digital cultural record.[[6]](#footnote-6) In the absence of aggregators or an index, specific titles are incredibly hard to find in the patchwork of existing, eclectic collections of digitised periodicals.[[7]](#footnote-7)



Map 1: Geographic distribution of library holdings of al-Muqtabas as recorded by Worldcat

The third layer of *inaccessibility* is one of socio-technical infrastructures: Digital infrastructures, despite all promises towards the opposite, are rooted in the epistemic hegemony of late twentieth-century, anglo-american neoliberal capitalism. Most *digitised* periodicals are kept in data silos without any means for interchange or interoperability in the form of APIs or the option to download data in standardised, open file formats.[[8]](#footnote-8) Reading *access* to these silos is restricted by paywalls and geo-fencing. Proprietary web-interfaces are commonly neither tailored to the display of Arabic material nor themselves available in Arabic.[[9]](#footnote-9) Downloading content in order to circumvent ill-suited interfaces is limited to individually identifiable users. Bulk download frequently violates terms of use and most vendors try to prevent this on the technical level. Finally, bibliographic metadata is regularly limited to the issue level and woefully inadequate. This is due to ambiguity of data in the physical artefact, limited knowledge about these artefacts among librarians and the contractors, who did the actual digitising work, and software stacks incapable of handling anything but hegemonic Western concepts of dates and names and Western scripts. In consequence, bibliographic metadata are recorded and shared in transcription to Latin script—of which there are as many varieties as there are languages written in Latin script.[[10]](#footnote-10)

The fourth layer of *inaccessibility* is the digital artefact itself: *Digitised* commonly means scanned facsimiles with limited or no text layer due, inter alia, to the abysmal state of Arabic OCR.[[11]](#footnote-11) Available text layers are frequently hidden, while vendors foreground full-text “search” that won’t ever report false negatives and verifiable information on error rates is either inexistent or not published. On the other hand, informal online-libraries of Arabic literature, most prominent and popular among them [*al-Maktaba al-shāmila*](http://www.shamela.ws/) (The Comprehensive Library),[[12]](#footnote-12) provide access to a small number of full-text editions of Arabic periodicals based on the work of anonymous human transcribers. Unfortunately, they do not provide information on editorial principles, the quality of the transcription or the material artefact they were based on.[[13]](#footnote-13) In addition, such informal “editions” lack information linking the digital remediation to the original artefact, namely bibliographic metadata and page breaks, which makes them almost impossible to employ for scholarly research. In a final twist, I found that the informal editions from *al-Maktaba al-shāmila*, with all their gaps and omissions, were rendered as images with a pseudo-original layout to provide “fakesimiles” served through [*Arshīf al-majallāt al-adabiyya wa-l-thaqāfiyya al-ʿarabiyya*](http://archive.alsharekh.org/) (Archive of literary and cultural Arabic journals), the largest Arabic online platform for historical periodicals.[[14]](#footnote-14)

These layers of inaccessibility have an already crippling effect on any Arabic speaker from Syria, Jordan or Palestine with only limited knowledge of foreign languages, particularly English, and no affiliation with wealthy institutions from the Global North. They are, however, further compounded by the infrastructural dependencies of digital remediations on access to devices, internet connections, and electricity. Neither of these can be taken for granted for the societies of the Global South. In the—all too frequent—worst case, devices are old, internet connections are slow, traffic is prohibitively expensive, and electricity is perilously scarce.[[15]](#footnote-15) Let me illustrate this multi-layered “digital divide”[[16]](#footnote-16) with the Lebanese example: At least 30 per cent of the population live on less than 120 USD a month in Lebanon.[[17]](#footnote-17) Mobile internet connection costs about 40 USD for 20GB traffic or 30-days, whatever is less. People depend on mobile internet because landlords avoid paying council tax by not registering contracts with the council. Without a registered rent contract, people cannot sing-up for a landline.[[18]](#footnote-18) Electricity is another major issue. The single, state-run utility relies on insufficient power plants and imported fuel and even the capital city of Beirut suffers from regular daily power cuts of three hours.[[19]](#footnote-19) Although heavily subsidised , electricity is expensive and easily comes in at about 120 USD per month per household. There is no available data on the cost of generator subscriptions, but we pay another 80 USD in an up-market neighbourhood of Beirut. The impact of these infrastructural realities is not limited to users of digital cultural artefacts but also severely impacts institutions planning to host them.

# Bootstrapping for access

[“Open Arabic Periodical Editions” (OpenArabicPE)](https://openarabicpe.github.io) was born out of frustration in August 2015 in order to address the socio-technical layers of inaccessibility of existing digitised Arabic periodicals with the affordances of the Global South. Building on the guiding principles of *simplicity* and *credibility* for the sake of *accessibility* and *sustainability*, OpenArabicPE was inspired by Alex Gil’s talk on *minimal computing* at [Digital Humanities Institute—Beirut](https://dhibeirut.wordpress.com/) and discussions at [Digital Humanities Summer Institute](https://dhsi.org/) 2015. The idea is simple: re-use—sometimes creatively—openly available data, tools, and infrastructures to produce open, collaborative, scholarly digital editions of early Arabic periodicals. OpenArabicPE’s resources are extremely constrained: we have no funds, no staff and no equipment beyond volunteering interns. My employer certainly enjoys patting me on the shoulder for engaging the digital humanities but I still mostly work on it after hours and on my own computer. OpenArabicPE, therefore, also represents (and reflects the changing boundaries of) my own ability “to produce, disseminate, and preserve digital scholarship ourselves, without the help we can’t get, even as we fight to build the infrastructures we need at the intersection of, with, and beyond institutional libraries and schools.”[[20]](#footnote-20) As such, OpenArabicPE resembles an attempt at what Valeria Graziano, Marcell Mars and Tomislav Meda have recently called “pirate care”.[[21]](#footnote-21)

## The data layer

On the data level, OpenArabicPE combines the virtues of immensely popular, but non-academic informal online libraries of volunteers on the one hand with academic/institutional scanning efforts as well as editorial expertise on the other. We transformed the digital text (EPub, HTML) of six Arabic periodicals (see table) from [*al-Maktaba al-shāmila*](http://www.shamela.ws/) into an open, standardised file format (XML) following the [Text Encoding Initiative (TEI)](https://www.tei-c.org)’s guidelines[[22]](#footnote-22) and link each page to digital facsimiles from various sources, namely [EAP](http://eap.bl.uk/), [HathiTrust](http://catalog.hathitrust.org/), and [*Arshīf al-majallāt al-adabiyya wa-l-thaqāfiyya al-ʿarabiyya*](http://archive.alsharekh.org/). We also model the periodicals with light structural mark-up for articles, sections, mastheads and bylines and other bibliographic information.[[23]](#footnote-23) Bibliographic metadata on every article and in common standardised formats such as BibTeX and MODS (Metadata Object Description Schema) is then automatically generated from the modelled source.[[24]](#footnote-24) Thus, we provide the advantages of truly digital editions and a means to verify the text layer and our mark-up against facsimiles of the original artefact. Linking page breaks to facsimiles, although trivial, is extremely labour-intensive because they seemingly did not matter enough to the anonymous transcribers of *al-Maktaba al-shāmila* to be consistently recorded. Each of the c.8000 page breaks in [*al-Muqtabas*](https://github.com/tillgrallert/digital-muqtabas) and [*al-Ḥaqāʾiq*](https://github.com/OpenArabicPE/journal_al-haqaiq), therefore, needed to be manually marked up by volunteers.[[25]](#footnote-25) My gratitude goes to Dimitar Dragnev, Talha Güzel, Dilan Hatun, Hans Magne Jaatun, Xaver Kretzschmar, Daniel Lloyd, Klara Mayer, Tobias Sick, Manzi Tanna-Händel and Layla Youssef, who have contributed their time to this task.

OpenArabicPE’s corpus of periodical editions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Title | Place | DOI | Release | Volumes | Issues |
| [al-Ḥaqāʾiq](https://www.github.com/openarabicpe/digital-haqaiq) | Damascus | [10.5281/zenodo.1232016](https://doi.org/10.5281/zenodo.1232016) |  | 3 | 35 |
| [al-Ḥasnāʾ](https://www.github.com/openarabicpe/journal_al-hasna) | Beirut | [10.5281/zenodo.3556246](https://doi.org/10.5281/zenodo.3556246) |  | 1 | 12 |
| [al-Manār](https://www.github.com/openarabicpe/journal_al-manar) | Cairo |  |  | 20 | 387 |
| [al-Muqtabas](https://www.github.com/tillgrallert/digital-muqtabas) | Cairo, Damascus | [10.5281/zenodo.597319](https://doi.org/10.5281/zenodo.597319) |  | 9 | 96 |
| [al-Ustādh](https://www.github.com/openarabicpe/journal_al-ustadh) | Cairo | [10.5281/zenodo.3581028](https://doi.org/10.5281/zenodo.3581028) |  | 1 | 42 |
| [al-Zuhūr](https://www.github.com/openarabicpe/journal_al-zuhur) | Cairo | [10.5281/zenodo.3580606](https://doi.org/10.5281/zenodo.3580606) |  | 4 | 39 |
| [Lughat al-ʿArab](https://www.github.com/openarabicpe/journal_lughat-al-arab) | Baghdad | [10.5281/zenodo.3514384](https://doi.org/10.5281/zenodo.3514384) |  | 3 | 34 |
| **total** |  |  |  | 41 | 645 |

## The presentation layer

TEI XML is certainly not *simple* and comes with a steep learning curve. Bidirectional XML with left-to-right tags and right-to-left text is not particularly *accessible* either, even if specialised XML-editing software would better support it. However, TEI is both *sustainable* and *credible* as an underlying format and the direction of a script is only relevant if rendered on a two-dimensional surface for human readers and editors.

Presenting facsimiles and text side-by-side for the validation of the latter with the help of the former is core to OpenArabicPE’s claim of credibility. Therefore, and in order to make the editions accessible to readers with as little overhead as possible, we heavily modified the [TEI Boilerplate](http://dcl.slis.indiana.edu/teibp/) for the use with Arabic texts. Our adaptation of the TEI Boilerplate also provides access to a table of contents, limited browsing to neighbouring issues, and bibliographic metadata on the article level.[[26]](#footnote-26) By rendering XML files locally in a web browser on the reader’s computer with XSLT and CSS, we remove the need for backend servers and internet connections. We added a simple system to provide multiple localisations and most parts of the interface have been translated to Arabic. Editions can be downloaded, distributed through USB keys, and run locally. This also removes the need to produce reading surrogates whenever the main TEI XML was changed.

Linking to digital facsimiles already available on the internet is commonly legal and since hosting and serving large files is expensive, it is the preferred option from our project’s point of view. If we were to produce our own facsimiles through scanning, I would recommend uploading them to the [Internet Archive](https://archive.org/), which currently develops IIIF support.[[27]](#footnote-27) From the readers’ perspective, linked online facsimiles require stable, fast and affordable internet connections. Downloading them once and serving them locally would be the preferred option but will most likely violate user agreements, licences and copyright depending on vendors and jurisdictions (It is, nevertheless, trivial since the links are an inherent part of the TEI XML files’ <facsimile> node).

Distributing digital corpora of hundreds of periodical issues without a backend severely restricts the ability to search and browse across issues. The (academic) reference manager [Zotero](https://www.zotero.org) is a staple of the open-source community, which, among many other things, allows sharing an unlimited amount of bibliographic references for free. I, therefore, opted to host all bibliographic metadata in a constantly updated public [Zotero Group](https://www.zotero.org/groups/OpenArabicPE) instead of building a static website with pre-computed indexes, tables of contents etc. Each reference in Zotero links to the relevant section in the periodical editions through a stable URL. Readers can, therefore, use either Zotero’s web interface or standalone client as a port of entry to all of OpenArabicPE’s editions. Full-text search across all issues and editions is provided by means of automatically updated sitemaps and Google’s [Programmable Search Engines](https://programmablesearchengine.google.com/) and Search Console.

## Facilitating re-use through licenses

Within the academic framework we are operating in, we depend on explicit licences to facilitate the use and re-use of our data and tools.[[28]](#footnote-28) We assume that the content of periodicals published across the Eastern Mediterranean before 1920 is in the public domain even under the most restrictive definitions (i.e. US copyright law). The enormous amount of human labour required for digitising cultural artefacts, on the other hand, is commonly actively hidden—within academic historiography, for instance, using digital remediations is still frowned upon as cheating (everybody does it but only few talk about it).[[29]](#footnote-29) Retaining copyright of our own editorial contributions in the form of [Creative Commons Attribution-ShareAlike 4.0 International](http://creativecommons.org/licenses/by-sa/4.0/) is at least a reminder that all contributors need to be transparently credited for their work. We share our tools under a [MIT licence](https://opensource.org/licenses/MIT) whenever possible.

## The infrastructure

Editions and tools are maintained using a combination of .git and [GitHub](github) for distributed version control. Each edit is transparently credited to a contributor with an email address and a timestamp. Github also provides free hosting of the TEI Boilerplate-based webview for our editions through [GitHub Pages](https://pages.github.com/). Github, of course, is a commercial platform and nowadays owned by Microsoft. They might go out of business at any point in the future but they are a core staple of the open-source movement. The platform itself is built on open-source software, protecting us against a lock-in of our data, and its services are free-of-cost to the user.

While platforms for shared version control, such as GitHub, are great for collaborative editing and hosting data, they cannot replace an archive or long-term repository. We use [Zenodo](https://www.zenodo.org) for this purpose. Zenodo is an open and free-to-use, EU-funded research data repository originally developed by the European Organization for Nuclear Research (CERN). Zenodo registers (versioned) [Digital Object Identifiers (DOI)](https://doi.org) and integrates with various Open Science infrastructures such as scholarly aggregators of metadata or [Open Researcher and Contributor ID (ORCID)](https://orcid.org/) for unambiguously identifying contributors. Zenodo also integrates with GitHub by automatically archiving each release of data and tools and recording all contributors as found in the .git repository.

## Problems of bootstrapping

The main problem of bootstrapping is the reliance on external data, tools, and services, which are beyond our control. Every one of these dependencies will eventually break. The only question is when. The main mitigation strategies are open and widely supported standards and a minimal software stack. On the data level, the links to the facsimiles are the most volatile component and we already encountered three major instances of link rot. Two were caused by vendors moving servers and changing protocols: The British Library moved the Endangered Archives Programme to IIIF in 2017 and *Arshīf al-majallāt al-adabiyya wa-l-thaqāfiyya al-ʿarabiyya* moved to a new domain in 2019. In another instance, Hathitrust removed the facsimiles of Princeton’s copy of *al-Haqāʾiq* (vol.1, 1910) from the public domain in 2016.[[30]](#footnote-30)

On the representation layer, TEI Boilerplate was the only viable option to render TEI XML files in the browser without a backend in 2015 and it still works. However, web browsers and the TEI have moved on. All major browsers still support XSLT transformation but with two limitations: First, the web has moved on to Javascript and JSON. XSLT support is stuck with version 1, which was superseded by XSLT 2 in 2007 and XSLT 3 in 2017. If (!) browsers continue to support XSLT, it won’t be further developed. Second, browser vendors tighten the screws on security. Running XSLT transformations on local files is considered attempted cross-site scripting and, consequently, aborted by Google’s Chrome. Firefox will transform local XML files but only with remotely hosted XSLT. At the time of writing, only Apple’s Safari allows for rendering our editions without any internet connection. Third, [CETEIcean](https://github.com/TEIC/CETEIcean) has matured into a Javascript-based alternative to TEI Boilerplate with the support of the TEI consortium.[[31]](#footnote-31) All future developments of OpenArabicPE’s webview will therefore move to CETEIcean.

Hosting is another challenge. I originally decided to directly render XML files as available in the GitHub repositories. This made sure that readers always saw the most current version of the data without any additional steps. [Rawgit](https://github.com/rgrove/rawgit) allowed to directly serve files from any branch of a GitHub repository with the correct content type. But the service was shut down in 2019 because people ran malware exploits directly from their GitHub repositories.[[32]](#footnote-32) OpenArabicPE has therefore moved to GitHub’s own [GitHub Pages](https://pages.github.com/) service, which serves files from the gh-pages branch.

# Conclusion

This paper contrasted the promises of digitisation with the stark reality of digitised cultural artefacts from the Global South against the backdrop of Arabic periodicals from the late Ottoman Eastern Mediterranean as an experience of multiple, compounding layers of *inaccessibility*—depending at the same time on the material artefact itself, remediated through neo-colonial power structures, and on one’s own position within a globalised hegemony of anglophone neoliberal capitalism.

I then introduced our project Open Arabic Periodical Editions as a proposal for and concrete implementation of bootstrapping existing data, tools, and infrastructures into a framework to produce and distribute open scholarly periodical editions under the affordances of the Global South. The tools and infrastructures outlined above are not specifically tailored to Arabic texts or periodicals and can be applied to material in other languages and scripts and from other regions (the language of the presentation layer depends on the language of the editions and defaults back to English). Current developments in open, machine-learning based, and language-agnostic approaches to OCR are aided by OpenArabicPE through re-purposing our editions as ground truth for training models and will eventually remove the dependency on pre-existing digital texts.

1. Alex Gil and Élika Ortega, ‘Global Outlooks in Digital Humanities: Multilingual Practices and Minimal Computing’, in *Doing Digital Humanities: Practice, Training, Research*, ed. Constance Crompton, Richard J Lane, and Ray Siemens (Abingdon: Routledge, 2016), 29 [↑](#footnote-ref-1)
2. E.g. Adīb Muruwwa, *al-Ṣiḥāfa al-ʿArabiyya: nashʾatuhā wa taṭawwuruhā. sijill ḥāfil li-tārīkh fann al-ṣiḥāfa al-ʿarabiyya qadīman wa-ḥadīthan* (Bayrūt: Dār Maktabat al-Ḥayyāt, 1961); Shams al-Dīn al-Rifāʿī, *Tārīkh al-ṣiḥāfa al-Sūriyya*, 2 vols (al-Qāhira: Dār al-Maʿārif bi-Miṣr, 1969); Yūsuf Asʿad Dāghir, *Maṣādir al-dirāsa al-adabiyya*, 2 vols (Ṣaydā: al-Maṭbaʿa al-Mukhliṣiyya, 1950); Yūsuf Aḥmad Dāghir, *Qāmūs al-ṣiḥāfa al-Lubnāniyya 1858-1974* (Bayrūt: al-Maktaba al-Sharqiyya al-Kubrā, 1978); Jūzīf Ilyās, *Taṭawwur al-ṣiḥāfa al-Sūriyya fī miʾat ʿām: 1865-1965*, 2 vols (Bayrūt: Dār al-Niḍāl, 1982); Yūsif Q. Khūrīya, *al-Ṣiḥāfa al-ʿArabiyya fī Filasṭīn 1876-1948* (Bayrūt: Muʾassasat al-Dirāsāt al-Filasṭīniyya, 1976); Ibrāhīm ʿAbduh, *Aʿlām al-ṣiḥāfa al-ʿArabiyya* (al-Qāhira: Maktabat al-Ādāb, 1948), <http://hdl.handle.net/2333.1/v9s4n21b>; Munīr Bakr Tikrītī, *al-Ṣiḥāfa al-ʿIrāqiyyq wa-ittijāhātihā al-siyāsiyya wa-l-ijtimāʿiyya wa-l-thaqāfiyya min 1869-1921* (Baghdād: Maṭbaʿat al-Irshād, 1969), <http://hdl.handle.net/2333.1/gf1vhp1w>; ʿAbd al-Razzāq Ḥasanī, *Tārīkh al-ṣiḥāfa al-ʿIrāqiyya* (Baghdād: Maṭbaʿa al-Zahrāʾ, 1957), <http://hdl.handle.net/2333.1/m63xspj1> [↑](#footnote-ref-2)
3. A map of the global distribution of surviving copies of Muḥammad Kurd ʿAlī’s monthly *al-Muqtabas* (Cairo and Damascus, 1906–1918) is available at <https://tillgrallert.github.io/s0fb41f86/map-holdings_al-muqtabas-vol_1-9.png>. [↑](#footnote-ref-3)
4. Muḥammad Kurd ʿAlī, for instance, gifted library copies of *al-Muqtabas* 2 from 1907/08 to his friend Jūrj Fākhūri in 1914 and to the library of al-Ṣalāḥiyya college in Jerusalem in 1916. The former is now held by the University of Minnesota and a facsimile is made available through [Hathitrust](http://catalog.hathitrust.org/). The later ended up in the al-Aqṣā Mosque’s library in Jerusalem and was digitised by the British Library’s [Endangered Archives Programme](http://eap.bl.uk/). Handwritten note on [*al-Muqtabas* 2(1), p.1](https://tillgrallert.github.io/digital-muqtabas/xml/oclc_4770057679-i_13.TEIP5.xml#pb_1.d2e814). [↑](#footnote-ref-4)
5. Mohamed M El-Hadi, *Union List of Arabic Serials in the United States: The Arabic Serial Holdings of Seventeen Libraries*, Occasional Papers (University of Illinois at Urbana-Champaign. Graduate School of Library and Information Science) ; No. 75. (Urbana: University of Illinois, Graduate School of Library and Information Science, 1965); Derek Hopwood, *Arabic Periodicals in Oxford: A Union List* (Oxford: St. Antony’s College, 1970); Mohammed M. Aman, *Arab Periodicals and Serials: A Subject Bibliography*, Garland Reference Library of Social Science (New York: Garland, 1979); Fred De Jong, ‘Arabic Periodicals Published in Syria Before 1946: The Holdings of Zahiriyya Library in Damascus’, *Bibliotheca Orientalis* 36 (1979): 292–300; Zāhir Iḥdādan, *Bībliyūghrāfiyā al-ṣiḥāfa al-Jazāʾiriyya* (al-Jazāʾir: al-Muʾassasat al-Waṭaniyya li-l-Kitāb, 1984); Yūsuf Quzmā Khūrī, *Mudawwanat al-ṣiḥāfa al-ʿArabiya*, ed. ʿAlī Dhū al-Fiqār Shākir, vol. 1: Miṣr (Bayrūt: Maʿhad al-Inmāʾ al-ʿArabī, 1985); Gerhard Höpp, *Arabische und islamische Periodika in Berlin und Brandenburg 1915 - 1945* (Berlin: Verlag Das Arabische Buch, 1994); Touraj Atabaki and Solmaz Rustămova-Tohidi, *Baku Documents: Union Catalogue of Persian, Azerbaijani, Ottoman Turkish and Arabic Serials and Newspapers in the Libraries of the Republic of Azerbaijan* (London: Tauris Academic Studies, 1995). [↑](#footnote-ref-5)
6. Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy* (Evanston: Northwestern University Press, 2019), <http://www.jstor.org/stable/10.2307/j.ctv7tq4hg>, c.f. Paul Gooding, *Historic Newspapers in the Digital Age: ‘Search All About It’* (London: Routledge, Taylor & Francis Group, 2018), 149–57; Nanna Bonde Thylstrup, *The Politics of Mass Digitization* (Cambridge: The MIT Press, 2018), 79–100. [↑](#footnote-ref-6)
7. I am part of the endeavour to gather and openly share information on all holdings of Arabic periodicals; Philip Sadgrove et al., *Jara’id: A Chronology of Nineteenth Century Periodicals in Arabic (1800-1900). A Research Tool*, ed. Adam Mestyan and Till Grallert, 2012, <https://projectjaraid.github.io>. [↑](#footnote-ref-7)
8. I compiled information on the accessibility of platforms serving digitised Arabic periodicals, which is available at <https://tillgrallert.github.io/s0fb41f86/table_accessibility-of-digitised-collections.html>. [↑](#footnote-ref-8)
9. On the need for Arabic interfaces see Nadirah Mansour and Marwa Gadallah, ‘al-Iḥtiyāj li-l-wājihāt al-iliktrūniyya bi-l-lughat al-ʿArabiyya’, <https://twitter.com/NAMansour26/status/1274361436215574529>; David Jospeh Wrisley and Najla Jarkass, ‘RTL SoftwareLocalization and Digital Humanities: The Case Study of Translating Voyant Tools into Arabic’; David Joseph Wrisley and Najla Jakas, ‘On Translating Voyant Tools into Arabic’, 9 June 2016, <https://djwrisley.com/on-translating-voyant-tools-into-arabic/> [↑](#footnote-ref-9)
10. Till Grallert, ‘Catch Me If You Can! Computational Approaches to Track the Late Ottoman Ideosphere of Authors and Periodicals in the Wasteland of the “Digitised” Arabic Press’, ed. Simone Lässig, *Geschichte Und Gesellschaft*, no. Special Issue ‘Digital History’ (2020). Things have been improving over the years and many catalogues can now display Arabic script. But re-cataloguing is costly and not a priority for vendors. For an example, see EAP’s digitisation of the Ottoman newspaper *Yenī Taṣvīr-i Efkār* from Istanbul, which was catalogued with the faulty Arabic title *Taṣwīr Afkār* and as published in Jerusalem; <https://eap.bl.uk/archive-file/EAP119-1-18-1#?c=0&m=0&s=0&cv=0>. [↑](#footnote-ref-10)
11. Volker Märgner and Haikal El Abed, eds., *Guide to OCR for Arabic Scripts* (London: Springer, 2012), <https://doi.org/10.1007/978-1-4471-4072-6>. is still relevant for current platforms; see the Internet Archive’s claim that Arabic is “currently not OCRable”; e.g. <https://archive.org/details/1_20191109_20191109_1843>. Recent years have seen huge advances with the application of machine-learning algorithms; c.f. Benjamin Kiessling et al., ‘Important New Developments in Arabographic Optical Character Recognition (OCR)’, *Al-ʿUṣūr Al-Wuṣṭā* 25 (2017): 1–13, <https://www.middleeastmedievalists.com/wp-content/uploads/2017/11/UW-25-Savant-et-al.pdf>. They will eventually find their way into HathiTrust etc.; see ‘HathiTrust Research Center Awards Three ACS Projects for 2020’, HathiTrust Digital Library, 2020, <https://www.hathitrust.org/htrc-awards-three-acs-projects>. But without GUI’s (in Arabic!) and due to their reliance on server-side computing, which needs to be paid, these also remain largely inaccessible to the general public. [↑](#footnote-ref-11)
12. *al-Maktaba al-shāmila* was first published on CD-ROM and went online in 2005; ‘ʿan al-shāmila’, al-maktaba al-shāmila, accessed 6 July 2020, <http://www.shamela.ws/index.php/page/about-shamela>. For evaluations of *al-Maktaba al-shāmila* as the basis for scholarly corpus building efforts see Yonatan Belinkov et al., ‘Shamela: A Large-Scale Historical Arabic Corpus’, *arXiv Preprint arXiv:1612.08989*, 2016; Maha Alrabiah, A Al-Salman, and E S Atwell, *The Design and Construction of the 50 Million Words KSUCCA* (The University of Leeds, 2013). [↑](#footnote-ref-12)
13. We found that *al-Maktaba al-shāmila* consequently removed footnotes and terms in scripts other than Arabic. Transcribers normalised the text and occasionally left Arabic notes in the text, from which we can deduce that they were Arabic speakers. [↑](#footnote-ref-13)
14. There is a huge and unmarked gap in *al-Maktaba al-shāmila*’s transcription of *al-Muqtabas* 5(7) between [p.463](https://tillgrallert.github.io/digital-muqtabas/xml/oclc_4770057679-i_54.TEIP5.xml#pb_61.d1e2036) and [p.466](https://tillgrallert.github.io/digital-muqtabas/xml/oclc_4770057679-i_54.TEIP5.xml#pb_64.d1e2045), which is reproduced in [this “fakesimile”](https://archive.alsharekh.org/MagazinePages/Magazine_JPG/AL_moqtabs/Al_moqtabs_1910/Issue_7/605.JPG). [↑](#footnote-ref-14)
15. C.f. Babalola Titilola Aiyegbusi, ‘Decolonizing Digital Humanities: Africa in Perspective’, in *Bodies of Information: Intersectional Feminism and Digital Humanities*, ed. Elizabeth Losh and Jacqueline Wernimont, Debates in the Digital Humanities (Minneapolis: University of Minnesota Press, 2019), 434–46, <https://dhdebates.gc.cuny.edu/read/d02c3ed5-0c55-4de9-88de-5f543fecd130/section/c9862793-ef00-4d6b-a30c-2f3d354e4e94#ch23> [↑](#footnote-ref-15)
16. For a recent conceptualisation of the term see Massimo Ragnedda, ‘Conceptualising the Digital Divide’, in *Mapping the Digital Divide in Africa: A Mediated Analysis*, ed. Bruce Mutsvairo and Massimo Ragnedda (Amsterdam: Amsterdam University Press, 2019), 27–44, <https://doi.org/10.2307/j.ctvh4zj72.6>. Unfortunately, while acknowledging the multiple socio-technological layers, they focus solely on the internet and do not mention the hegemony of English. [↑](#footnote-ref-16)
17. Chadi, ‘UNDP Latest Poverty Assessment Report: 30% of Lebanese Are Poor’, Blog Baladi, 17 February 2018, <https://blogbaladi.com/undp-latest-poverty-assessment-report-30-of-lebanese-are-poor/>; Rosette Fadel, ‘Third of Lebanese Live in Poverty, Experts Say’, *An-Nahar*, 21 September 2018, sec. Economy, <https://en.annahar.com/article/865485-third-of-lebanese-live-in-poverty-experts-say>. with reference to inaccessible UNDP report “Rapid Poverty Assessment in Lebanon” for 2016. [↑](#footnote-ref-17)
18. According to Central Intelligence Agency, ‘Lebanon’, 2020, <https://www.cia.gov/library/publications/the-world-factbook/geos/le.html>. only 15% of inhabitants have access to a landline and less than 1% have a broadband subscription, while 73% have a mobile phone. [↑](#footnote-ref-18)
19. Corruption and waste are not the only culprits to blame. Israel deliberately damaged and destroyed power plants in the 2006 war. [↑](#footnote-ref-19)
20. Gil and Ortega, ‘Global Outlooks in Digital Humanities’, 29 [↑](#footnote-ref-20)
21. Valeria Graziano, Marcell Mars, and Tomislav Medak, ‘The Pirate Care Project’ (Centre for Postdigital Cultures, Coventry University, 2019), <https://pirate.care/>. See also Marcell Mars and Tomislav Medak, ‘System of a Takedown: Control and de-Commodification in the Circuits of Academic Publishing’, in *Archives: In Search of Media*, ed. Andrew Lison et al. (Lüneburg / Minnesota: messon press / University of Minnesota Press, 2019), 47–68, <https://doi.org/10.14619/1501>. [↑](#footnote-ref-21)
22. TEI Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, version 4.0.0, XML (Zenodo, 2020), <https://doi.org/10.5281/zenodo.3413524>. TEI XML is the quasi-standard of textual editing and required by funding bodies and repositories for long-term archiving; cf. , *DFG-Praxisregeln ‘Digitalisierung’* (Bonn: Deutsche Forschungsgemeinschaft, 2016), <http://www.dfg.de/formulare/12_151/12_151_de.pdf>. [↑](#footnote-ref-22)
23. Our TEI customisation for Arabic periodicals is openly available at <https://github.com/OpenArabicPE/OpenArabicPE_ODD>. [↑](#footnote-ref-23)
24. Our code is openly available at <https://github.com/OpenArabicPE/convert_tei-to-bibliographic-data>. It uses <tei:biblStruct> as an intermediary format and can also generate YAML and Zotero RDF. [↑](#footnote-ref-24)
25. In other instances, such as the journals [*Lughat al-ʿArab*](https://github.com/OpenArabicPE/journal_lughat-al-arab) and [*al-Ustādh*](https://github.com/OpenArabicPE/journal_al-ustadh), *al-Maktaba al-shāmila* did provide page breaks that correspond to a printed edition. [↑](#footnote-ref-25)
26. John Walsh et al., *TEI Boilerplate*, version 1.1.0, XSLT, 2016, <https://github.com/TEI-Boilerplate/TEI-Boilerplate>; John Walsh and Grant Leyton Simpson, ‘TEI Boilerplate’, *Journal of Digital Humanities* 2, no. 3 (2013), <http://journalofdigitalhumanities.org/2-3/tei-boilerplate/>. Our version of the TEI Boilerplate is openly available at <https://github.com/OpenArabicPE/tei-boilerplate-arabic-editions>. [↑](#footnote-ref-26)
27. According to example code in the documentation, IIIF support is in development since at least 2015; ‘IIIF Documentation’, Internet Archive, accessed 1 July 2020, <https://iiif.archivelab.org/iiif/documentation>. [↑](#footnote-ref-27)
28. For a great introduction to the conceptual critique of copyrights see Gary Hall, *Pirate Philosophy: For a Digital Posthumanities* (Cambridge: The MIT Press, 2016), 1–9. For a specific application to TEI XML files see Vanessa Hannesschläger, ‘Common Creativity International: CC-Licensing and Other Options for TEI-Based Digital Editions in an International Context’, *Journal of the Text Encoding Initiative*, no. 11 (June 2020), <https://doi.org/10/gg3wts>. [↑](#footnote-ref-28)
29. This is evidenced by the few references to the URLs of works from *al-Maktaba al-shāmila* in academic texts. JSTOR returns a total of 16 hits. [↑](#footnote-ref-29)
30. My inquiries about the legal basis for this decision have not been replied to [↑](#footnote-ref-30)
31. Hugh Cayless and Raffaele Viglianti, *CETEIcean*, version 1.2.1, JavaScript (Text Encoding Initiative Consortium, 2020), <https://github.com/TEIC/CETEIcean>; Hugh Cayless and Raffaele Viglianti, ‘CETEIcean: TEI in the Browser’, in *Proceedings of Balisage: The Markup Conference*, 2018 [↑](#footnote-ref-31)
32. <http://rawgit.com> shut down in 2019. There since emerged a replacement service at <https://rawgit.org/>. [↑](#footnote-ref-32)