```xml
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
<fileDesc>
<title>Scholarly Digital Editions of Right-to-Left Script Texts</title>
<respStmt>
<convenor xml:id="eAM">Adam Mestyan</convenor>
<convenor xml:id="eSCS">Sabine Schmidtke</convenor>
<convenor xml:id="eMMT">Maria Mercedes Tuya </convenor>
</respStmt>
<respStmt>
<resp>Date</resp>
<date>22 March 2024</date>
</respStmt>
<respStmt>
<resp>Place</resp>
<place>Institute for Advanced Study</place>
<place>Princeton, NJ</place>
</respStmt>
</fileDesc>
</teiHeader>
</TEI><?xml version="1.0" encoding="UTF-8"?>
```

Scholarly Digital Editions of Right-to-Left Script Texts

Workshop
22 March 2024
White-Levy Room, Institute for Advanced Study
Princeton, NJ

Convenors: Adam Mestyan (IAS and Duke University),
Sabine Schmidtke (School of Historical Studies, IAS), and
Maria Mercedes Tuya (Digital Scholarship@IAS)
Coordinator: Uta Nitschke-Joseph (IAS)

8:30 am - Breakfast, introductions
8:50-9:00 am - Welcome

Panel One, Chair: **Amy Singer** [9:00 am ET (2 pm CET)]

9:00-9:20 am - **Elena Pierazzo** (Université de Tours), "Digital Philology: from the past to the future"

9:20-9:40 am - **Till Grallert** (Humboldt-Universität zu Berlin), "Open Arabic Periodical Editions: a framework for bootstrapped scholarly editions outside the global north"

9:40-10:00 am - **Walter Young** (McGill University), "Moving Towards the Digital Commentary Complex"

10:00-10:20 am - Discussion
10:20-10:35 am - Break

Panel Two, Chair: **Maria Tuya [**10:35 am ET (3:35 CET)]
10:35-10:55 am - **Jan Thiele** (Consejo Superior de Investigaciones

Científicas) and **Sabine Schmidtke** (IAS), "Arabic Classical Text Editor and TEI XML"

10:55-11:15 am - **Daniel Stoekl** (Centre national de la recherche scientifique) and **Hayim Lapin** (University of Maryland), "eRabbinica: New Ways for Born Digital Editions of (Classical Rabbinic) Literature"

11:15-11:35 am - **Abdurrahman Atcıl** (Sabancı University), "Creating a Digital Thematic Research Collection: Sources of Ottoman Legal History (SOLEH)"

11:35-11:55 pm - Discussion

12:05-1:30 pm - Lunch break

Panel Three, Chair: **Sabine Schmidtke** [1:30 pm ET (6:30 pm CET)]

1:30-1:50 pm - **Adam Mestyan** (IAS and Duke University), "Digital Editions of *Sharīʿa* Court Records – Examples from Damascus, 1918"

1:50-2:10 pm - **David Vishanoff** (University of Oklahoma), "Are Current Trends Leading to a Dead End? Two Case Studies on the Limits and Potential of Digital Methods for Arabic-Script Scholarly Editions"

2:10-2:30 pm - **Theodore S. Beers** (Berlin University Alliance) and **Mahmoud Kozae** (Freie Universität Berlin), "LLM-Enhanced Software Development for Overcoming the Multi-Decade Digital Disruption of (Arabic) Textual Scholarship"

2:30-2:50 pm - Discussion

2:50-3:05 pm - Break

Panel Four, Chair: **Adam Mestyan [**3:05 pm ET (8:05 pm CET)]

3:05-3:25 pm - **Osama Eshera** (University of Maryland), "The Dynamic Critical Apparatus in Dabīrān's Scholarly Editing

Environment: Examples from Critical Editions of Two Works by Avicenna"

3:25-3:45 pm - **Hugh Cayless** (Duke University), "Tools and Workflows for Digital Editions in Arabic and Other RTL Languages"

3:45- 4:05 pm - Discussion

4:05-4:30 pm - General Discussion

4:30 pm - End of workshop

6 pm - Dinner

Abstracts

Elena Pierazzo, *Digital Philology: from the past to the future*

More than thirty years of digital editing have borne fruit: nowadays we have methods, standards, theories, and editions. Yet we are far from having exhausted all the possibilities offered by the digital sphere, and new developments are still happening at an ever-faster rate with Artificial Intelligence bound to change the world as we know it. Research possibilities are still before us and are even more open, and many digital editorial projects aim to study new texts, or old texts with new methods, or to develop new methods to study and display all of the above. The talk will concentrate on digital-driven changes on the heuristics of textual scholarship; the change in the perception of scholarly work, and the elaboration of new editorial models. In fact, these editions challenge the rigid separation between edition of sources and critical editions, as well as challenging editors to reinvent themselves as encoders, data managers and cultural facilitators.

Till Grallert, *Open Arabic Periodical Editions: a framework for bootstrapped scholarly editions outside the global north*

I will present my project Open Arabic Periodical Editions

(OpenArabicPE, 2015–) as a framework for bootstrapped scholarly editions outside the global north that addresses the severe impact of the multidimensional digital divides for RTL scripts and digitised cultural artefacts of societies from the global south. Without any external funding and thus relying on volunteer labour and free-to-use software and infrastructures, OpenArabicPE has published digital editions of six periodicals published in Baghdad, Beirut, Cairo, and Damascus between 1892 and 1918 with a total of 41 volumes, 645 issues and more than six million words.

The project is influenced by ideas of minimal computing and pirate care. The guiding principles for every part of the tool chain and workflow are accessibility, simplicity, sustainability and credibility. The project originated from a simple proposition: unite manually transcribed digital texts from shadow libraries, such as al-Maktaba al-Shamela, with digital facsimiles from academic scanning efforts in order to validate the former through the latter and to, ultimately, produce the necessary ground truth for modern character recognition algorithms. All texts are marked up in XML following the guidelines of the Text Encoding Initiative (TEI) and each issue is modelled as a single file with relatively light structural mark-up for sections, articles, headers, and bylines, as well as page breaks linking the text to the facsimiles. A set of authority files is maintained for the disambiguation of named entities. Finally, a web-view based on the TEI Boilerplate provides a local GUI for reading the text and the facsimiles side by side.

The presentation will discuss our workflows and infrastructures with a view to their wider applicability.

Walter Young, *Moving Towards the Digital Commentary Complex*

The online, digital, critical edition-translation of Shams al-Dīn al-Samarqandī's ʿAyn al-Naẓar, completed with the guidance, TEI infrastructure, and interface design of Frederik Elwert (Ruhr-Universität Bochum), was only the partial realization of a larger vision: a first, but by no means complete, exploration of what might be accomplished between XML editor and GitLab, within (and sometimes expanding) the parameters set by the Text Encoding Initiative. The more complete vision and eventual goal is the electronic seating of an edited grundtext at the base of its entire (edited) commentary

complex, the whole hyperlinked in such a way that any sharḥ, ḥāshiya, or taʿlīqa, along with its attendant apparatus, may be consulted with a click, in a series of pull-down menus and collapsible windows. In this talk I will first review what I think we accomplished in the ʿAyn al-Naẓar project and its limits, then I will put forward a series of practical ideals—in effect, an edition-user's wish-list—using as example the commentary complex around al-Samarqandī's Risāla fī Ādāb al-Baḥth, with the aim of (hopefully) informing future online, digital-edition interface designs.

Jan Thiele and **Sabine Schmidtke**, *Arabic Classical Text Editor (CTE) and TEI XML*

In our presentation, we will share our experience with producing TEI-XML digital editions with the Classical Text Editor Software (CTE). We will present two examples: (1) The first is a conversion of a print edition (a Muʿtazili text based on two fragments preserved in the Firkovitch Collection, RNL St Petersburg), and we will talk about the process of transforming a human-eye-readable text into a machine-readable text; (2) the second example discusses issues with a new edition of Ibn Kammūna's Tanqiḥ, and particularly the question how a digital edition may solve the problem of adequately representing the complexity of a textual transmission in several recensions.

Daniel Stoekl and **Hayim Lapin**, *eRabbinica: New Ways for Born Digital Editions of (Classical Rabbinic) Literature*

eRabbinica is a shorthand name for a collection of overlapping projects to produce born-digital multi-witness editions of classical rabbinic texts. The projects share an evolving pipeline of semi-automated tasks beginning with transcription and culminating in TEI/XML editions, produced for publication in TEI Publisher. Additional steps involve text-to-image alignments, text alignments, lexical and morphological analyses, and the construction of one or more critical apparatus. The developed TEI Publisher format allows the user to simultaneously work with multiple views of the same edition, or to toggle between them.

Abdurrahman Atcıl, *Creating a Digital Thematic Research Collection: Sources of Ottoman Legal History (SOLEH)*

Many Ottoman legal sources originally written in the Arabic script have been digitized, transcribed, transliterated, and translated into European languages, with the results today dispersed across a range of different digitally accessible platforms. While this makes it possible for scholars to carry out rudimentary word searches and apply basic metadata filters to aid them in their research, the dispersed nature of this material and the limited functionality available on the various platforms that house it continue to pose an obstacle.

The OTTOLEGAL project team is working to address these issues through its new platform, SOLEH—a digital thematic research collection focused on Ottoman legal history. We curate a selection of *kānūnnāme*s, *fermān*s, fatwas, *risāle*s, and *ahidnāme*s, providing original texts alongside transliterations, English translations, and annotations, which include identifiers for actors, resources, places, contexts, and metadata. By developing SOLEH as a "generous interface," we offer users the ability to explore our documents via customized illustrations of diverse legal contexts, utilize advanced search functions to locate desired content, pursue data to their original textual sources, and more, all within a digitally curated environment.

In this presentation, I will discuss the stages of SOLEH's development, including source selection, metadata development, and the construction of our data model for content analysis, as well as our generous interface design and analysis and visualization tools. Furthermore, I will address concerns regarding data interoperability, preservation, and the challenges associated with collating, analyzing, and presenting documents in a digital thematic research collection.

Theodore S. Beers (Berlin University Alliance) and **Mahmoud Kozae** (Freie Universität Berlin), *LLM-Enhanced Software Development for Overcoming the Multi-Decade Digital Disruption of (Arabic) Textual Scholarship*

The Kalīla and Dimna project embarked on developing specialized software to support its editorial activities, prompted by the inadequacy of existing software solutions to meet the project's specific

requirements. This gap highlights a persistent challenge within digital textual scholarship, which has struggled to fully adapt since the advent of computing technology. While the development of specialized software has proven effective for the Kalīla and Dimna project, it should not be universally recommended for all editing endeavors, particularly those facing circumstances similar to those encountered by the Kalila and Dimna project between 2018 and 2022. However, the landscape of software development has undergone significant transformation, notably with the advent of more efficient Large Language Models (LLMs), suggesting that software development could increasingly present a viable solution across a wider range of projects.

Adam Mestyan, *Digital Editions of Sharīʿa Court Records – Examples from Damascus, 1918*

This presentation describes potential schemas for the digital mark-up and publication of twentieth-century Arabic *sharīʿa* court records. I take three types of *sharīʿa* court registers (probate, alimony, and general registers) from October 1918 in Damascus and present possible TEI XML schemes for their digital publication.

David Vishanoff, *Are Current Trends Leading to a Dead End? Two Case Studies on the Limits and Potential of Digital Methods for Arabic-Script Scholarly Editions*

The accessibility and functionality of traditional scholarly editions have been significantly enhanced using digital methods that are now being applied to Arabic-script texts. This progress, however, has resulted in the proliferation of idiosyncratic and short-lived data silos useful only for a small audience and a single purpose. Moreover, even the vitally important shared TEI encoding standards developed to make editions more interoperable appear in fact to be hindering the application of emerging digital methods that have the potential to transform the production and scholarly analysis of texts. This presentation will use two case studies, an online CommentPress edition of al-Juwaynī's Kitāb al-Waraqāt fī uṣūl al-fiqh and a database of annotations on a forthcoming edition of the Islamic Zabūr Dāwūd, to illustrate both the limits of digitally enhanced traditional editions and the potential payoffs of a new direction for digital encoding, publication, and

annotation that permits stable external references to spans of text down to the character level.

Osama Eshera, *The Dynamic Critical Apparatus in Dabīrān's Scholarly Editing Environment: Examples from Critical Editions of Two Works by Avicenna*

This presentation concerns the kinds of complex textual variants that are often encountered in the manuscript traditions of medieval Islamic philosophical works, such as authorial revisions, scribal redactions, and gloss-incorporations. Such phenomena pose significant challenges for determining the stemmatological relationships between manuscripts and, ultimately, for establishing texts in critical edition. This presentation demonstrates how the digital tools and philological-historical methods developed in the framework of the Dabīrān project (www.dabiran.org) help scholars deal with such challenges. Drawing on concrete examples from my forthcoming critical editions of two interlinked works on metaphysics by Avicenna (d. 1037), al-Mabdaʾ wa-l-maʿād and the treatise on Ilāhiyyāt of the philosophical summa Kitāb al-Najāt, I shall outline how we approach the problems of encoding, collating, editing, and representing texts with complicated critical variants. Finally, I shall highlight how Dabīrān's Scholarly Editing Environment is being redeveloped and expanded in order to be interoperable with eScriptorium and other software within the OpenITI ecosystem.

Hugh Cayless, *Tools and Workflows for Digital Editions in Arabic and Other RTL Languages*

The Text Encoding Initiative provides very useful guidance and affordances for digital editions. It is a text-based format, using XML, making it well suited for version tracking systems like Git. TEI allows for the packaging of document and project metadata along with the text and for annotating that text with links to authority lists, gazetteers, and prosopographies. For the DigiCai project, we are working with articles from the newspaper al-Waqaʾiʿ al-Misriyya that concern urban development. I will discuss how we manage the difficulties of right-to-left XML, and how we use GitHub Actions to manage our digital workflows.

Resources

OpenArabicPE

DigiCai project

eScriptorium

OpenITI

OTTOLEGAL project

SOLEH

Classical Text Editor Software (CTE)