



OpenITI AOCP presents:

# Digital Publication of Right-to-Left Script Corpora

Date:

June 29 - 30, 2023

Presented by:



Location:

The University of Maryland, College Park

Funding provided by:



## Welcome Workshop Participants!

The Open Islamicate Texts Initiative Arabic OCR Catalyst Project (OpenITI AOCP) team would like to thank you for agreeing to participate in this experts workshop and share your expertise and experience with all in attendance.

Before we delve into more specifics on the workshop itself, let's go over the logistical details.

John Mullan—the incredible digital specialist of OpenITI AOCP—has already sent you your flight and accommodation information in separate emails (**if for some reason you did not receive this email please let us know as soon as possible**). Some of this information is repeated below. Be sure to review everything below in detail because there is additional information provided here that is not included in those emails.

## Workshop Background

This workshop is a part of the Open Islamicate Texts Initiative Arabic OCR Catalyst Project Phase II (OpenITI AOCP II)—a three-year project generously funded by [The Mellon Foundation](#).

OpenITI AOCP II is led by Osama Eshera (Roshan Institute for Persian Studies, UMD), [Matthew Thomas Miller](#) (Roshan Institute for Persian Studies, UMD), [Taylor Berg-Kirkpatrick](#) (University of California, San Diego), [Sarah Bowen Savant](#) (Aga Khan University), [David Smith](#) (Northeastern University), and [Raffaele Viglianti](#) (Maryland Institute for Technology in the Humanities, UMD).

This workshop will bring together scholars working with digital corpora who are addressing questions including: digital text collation, annotation and encoding, and digital publication of texts. Participants will be invited to share and work collaboratively on approaches and solutions to the digital publication of text corpora in general, but with a focus on challenges pertaining to text in right-to-left scripts. The workshop has two core goals: 1. To discuss the mechanics of encoding texts and the process of creating

a digital edition (with a particular focus on multi-text editions). 2. To discuss mechanisms for the dissemination of digitally encoded texts, in ways which will allow other scholars to engage with the complexity of the textual tradition while taking account of long-term sustainability. The workshop as a whole will focus on these issues as they relate to right-to-left scripts and the challenges that these scripts present (especially where many existing solutions and software cater to left-to-right use cases). As part of the workshop, we will examine and (where possible) experiment with existing software solutions and evaluate their potential for our use cases.

## ***Transportation***

Please use taxis, rideshare (Uber, Lyft, etc.), or light rail to travel between the airport and the hotel. You can submit these receipts to John for reimbursement. An original receipt is required. If for any reason obtaining a taxi or rideshare is difficult for you, please let us know as soon as possible and we will schedule transportation for you. DC traffic can be quite bad, depending on the time your flight arrives. *If you are departing Dulles International Airport (IAD) for The Hotel via taxi or rideshare, especially near rush hour times (6:30am-9:30am and 4:00-6:30pm), please prepare for a lengthy car ride (sometimes as long as two hours) by using the restroom and grabbing something to eat in the airport before departing.*

Guests arriving at Ronald Reagan Washington National Airport (DCA) or Washington Dulles International Airport (IAD) have the option of using the D.C. metro, which can be faster than hiring a car at rush hour. *For guests arriving at DCA*, follow signs to the airport metro station, then take the yellow or green line towards Greenbelt. Exit the metro at College Park - U of MD station. *For guests arriving at IAD*, follow signs to the airport metro station, then take the silver line towards Largo and exit at L'Enfant Plaza station. From L'Enfant Plaza, take the green or yellow line towards Greenbelt station and exit at College Park - U of MD station. The Hotel at the University of Maryland is a short ride away on rideshare or bus from the College Park station, or a fifteen minute walk.

# Digital Publication of Right-to-Left Script Corpora



## ***Accommodations***

You will all be staying at [The Hotel](#) at University of Maryland (UMD). The Hotel is Located across from the university's entrance at 7777 Baltimore Ave, College Park, MD 2074. You may check in after 4:00 p.m. and must check out before 12:00 p.m.

## ***Workshop Location***

The workshop will take place in the McKeldin Special Events Room (Room #6137 @7649 S Library Ln, College Park, MD 20742). Osama Eshera will meet you in The Hotel lobby each morning at 8:40 to walk with you from The Hotel to the workshop venue, but in case you need to navigate the campus on your own, [here is a link](#) to a campus map. We can also arrange transportation for any participants who would prefer not to walk for any reason.

## ***Food***

All meals during the workshop will be provided. We will also be hosting a (optional) pre-workshop dinner and drinks on Wednesday, June 28th at 7:30 p.m. at [The Hall CP](#) (located behind The Hotel) for those who arrive early enough (and aren't too jetlagged to attend). For breakfast each morning you will receive a \$12 voucher that can be used at [Bagels N Grinds](#), which is located off of the lobby of The Hotel. We will have some pastries, fruit, and coffee in the meeting room for snacking during the morning hours of the workshop, but if you want a sizable breakfast please eat breakfast before we depart The Hotel for the workshop each morning. Lastly, if you will be missing the group dinner on the second day of the workshop due to an early flight time, please feel free to request a boxed meal to go, or you may purchase dinner at the airport and we will reimburse you for it.

## ***Wireless Internet***

Guests have two options for wireless internet access while on UMD's campus. If your home institution uses eduroam, simply connect to the internet as you normally would. If your home institution does not use eduroam, you will need to connect to UMD's guest wifi. For more information on how to use UMD's guest wifi, please visit

[bit.ly/openiti\\_wifi\\_info](http://bit.ly/openiti_wifi_info). Please note that guest accounts must be renewed every 24 hours.

## ***Reimbursement***

We will reimburse you for travel to and from the airport as well as for any meals you purchase while traveling. U.S. citizens and permanent residents will need to fill out a [W-9 form](#), while international travelers will need to fill out an [X-9 form](#) and a [W-8 BEN form](#). **We will need receipts of all purchases for which you would like to receive reimbursement. All materials for reimbursement must be submitted to [jmullan@umd.edu](mailto:jmullan@umd.edu) by August 4, 2023.**

## ***Workshop Participants***

**A. Sean Pue**, Associate Professor of Hindi Language & South Asian Literature and Culture, Michigan State University  
*Project Affiliation: Jugaad CoLab*

**Bronson Brown-deVost**, Postdoctoral Fellow, Theologische Fakultät, Georg-August-Universität Göttingen  
*Project Affiliation: Qumran Digital*

**Chander Shekhar**, Professor of Persian, University of Delhi

**Daniel Stökl Ben Ezra**, Directeur d'Études, École Pratique des Hautes Études (EPHE), Université Paris Sciences et Lettres; Co-director, eScripta  
*Project Affiliation: eScripta*

**David Smith**, Associate Professor of Computer Science, Northeastern University  
*Project Affiliation: OpenITI*

**David Vishanoff**, Associate Professor of Religious Studies, University of Oklahoma

# Digital Publication of Right-to-Left Script Corpora



**Elena Pierazzo**, Professor of Digital Humanities, Centre d'Études Supérieures de la Renaissance, University of Tours; Former Chair, Text Encoding Initiative (TEI) & TEI Manuscripts Special Interest Group

**Farrukh Shahzad**, Chief Librarian, Forman Christian College University

**Fatemeh Savadi**, Postdoctoral Researcher, McGill University

*Project Affiliation: Dabīrān*

**Gregory Crane**, Professor of Classical Studies & Computer Science, Tufts University; Editor-in-Chief, Perseus Project

*Project Affiliation: Perseus Digital Library*

**Hugh Cayless**, Senior Digital Humanities Developer, Duke Collaboratory for Classics Computing, Duke University

**Intisar Rabb**, Professor of History & Law, Harvard Law School; Director, Program in Islamic law, Harvard University

*Project Affiliation: SHARIAsource*

**Janelle Jenstad**, Professor of English, University of Victoria; General Editor & Director, The Map of Early Modern London (MoEML)

*Project Affiliation: MoEML*

**Jonathan Parkes Allen**, Mellon Humanities Postdoctoral Fellow, Roshan Institute for Persian Studies, University of Maryland, College Park

*Project Affiliation: OpenITI*

**Joseph Hilleary**, Masters Candidate, Tufts University

**Lorenz Nigst**, Research Associate, Corpus Management, KITAB Project, Aga Khan University Institute for the Study of Muslim Civilizations

*Project Affiliation: KITAB; OpenITI*

**Magdalena Turska**, e-editiones.org

# Digital Publication of Right-to-Left Script Corpora



**Mathew Barber**, Post-Doctoral Research Fellow, KITAB Project, Aga Khan University Institute for the Study of Muslim Civilizations

*Project Affiliation: KITAB, OpenITI*

**Matthew Thomas Miller**, Assistant Professor of Persian Literature and Digital Humanities, Roshan Institute for Persian Studies, University of Maryland, College Park; Director, Roshan Initiative in Persian Digital Humanities

*Project Affiliation: OpenITI*

**Maxim Romanov**, Junior Research Group Leader, The Evolution of Islamic Societies (c.600-1600 CE): Algorithmic Analysis into Social History, Universität Hamburg

*Project Affiliation: OpenITI*

**Mehdy Sedaghat Payam**, Doctoral Candidate, Roshan Institute for Persian Studies, University of Maryland, College Park

**Muhammad Taimoor Shahid Khan**, Mellon Humanities Postdoctoral Fellow, Roshan Institute for Persian Studies, University of Maryland, College Park

*Project Affiliation: OpenITI*

**Nick Laiacona**, President, Performant Software Solutions LLC

**Osama Eshera**, Assistant Research Professor, Roshan Institute for Persian Studies, University of Maryland, College Park; Assistant Director, OpenITI AOCP Phase II

*Project Affiliation: OpenITI, Dabīrān*

**Peter Stokes**, Directeur d'Études, École Pratique des Hautes Études (EPHE), Université Paris Sciences et Lettres; Co-rector, eScripta

*Project Affiliation: eScripta*

**Peter Verkinderen**, Assistant Professor, Centre for Digital Humanities, Aga Khan University Institute for the Study of Muslim Civilizations; Postdoctoral Research Fellow, KITAB Project

*Project Affiliation: KITAB*

# Digital Publication of Right-to-Left Script Corpora



**Raffaele Viglianti**, Senior Research Software Developer, Maryland Institute for Technology in the Humanities, University of Maryland

*Project Affiliation: OpenITI*

**Sajjad Nikfahm Khubravan**, Postdoctoral Researcher, University of New Hampshire

*Project Affiliation: Dabīrān*

**Sarah Bowen Savant**, Professor of History, Aga Khan University Institute for the Study of Muslim Civilizations; Principal Investigator, KITAB Project

*Project Affiliation: KITAB*

**Theodore Beers**, Postdoctoral Research Fellow, Seminar für Semitistik und Arabistik, Freie Universität Berlin

*Project Affiliation: Kalīla and Dimna – AnonymClassic*

**Till Grallert**, Research Associate, Universitätsbibliothek, Humboldt-Universität zu Berlin; Marie Skłodowska-Curie Postdoctoral Fellow, Universität Hamburg

**Wayne Graham**, Chief Information Officer & Director of Informatics, Cultural Networks, and Knowledge Systems, Council on Library and Information Resources

**Wolfgang Meier**, e-editiones.org

## *Schedule (subject to change)*

See <https://openiti.org/2023workshop/> for a full up-to-date program with abstracts.

### **Pre-workshop Event (Optional)**

**Wednesday, June 28th, 2023**

7:30 p.m. - 10:00 p.m.: Welcome dinner and drinks ([The Hall CP](#) located behind The Hotel) for those who arrive early enough (and aren't too jetlagged to attend)

# Digital Publication of Right-to-Left Script Corpora



## **Workshop Day #1: Thursday, June 29, 2023**

*7:30 a.m. - 8:40 a.m.*: Breakfast at *Bagels 'N Grinds* located at The Hotel (not a group event—please just grab breakfast at your leisure)

*8:40 a.m. - 9:00 a.m.*: Walk with Osama Eshera to McKeldin library

*9:00 a.m. - 9:45 a.m.*: Welcome and introductory remarks (Matthew Miller, Raff Viglianti)

*9:45 a.m. - 10:45 a.m.*: Opening talks

- Digital Editing as Publication, Craft, and Research (Elena Pierazzo)
- Perseus, NLP and Opening up the Human Record (Gregory Crane and Joseph Hilleary)

*10:45 a.m. - 11:00 a.m.*: Break

*11:00 a.m. - 12:30 p.m.*: Curation

- Digital Cairo: Exploring Urban News (Hugh Cayless)
- OpenITI mARkdown: Balancing between a Standard and a Research Method
- Open Arabic Periodical Editions (OpenArabicPE): A Case Study in Digital Editing outside the Global North

*12:30 p.m. - 2:00 p.m.*: Lunch

*2:00 p.m. - 3:30 p.m.*: Annotation

- TEI Publisher and Right-to-Left Texts (Wolfgang Meier and Magdalena Turska)
- Challenges and Potentials in Multi-Script Text-Image Annotation (Peter Stokes)
- Annotating and Publishing Right-to-Left Script Corpora using FairCopy, ArchivEngine, and EditionCrafter (Nick Laiacona)

*3:30 p.m. - 3:45 p.m.*: Break

*3:45 p.m. - 5:15 p.m.*: Dissemination

- KITAB: Corpus, Text Reuse and the Multi-Text Editions Future (Peter Verkinderen, Mathew Barber, Lorenz Nigst, Sarah Savant)
- Tools for Comparative Reading in the AnonymClassic Project (Theodore Beers)

# Digital Publication of Right-to-Left Script Corpora



- “The End is Where We Start From:” The Endings Principle (Janelle Jenstad)

5:15 p.m. - 6:15 p.m.: Break

6:15 p.m.: Leave The Hotel on shuttle for dinner

6:30 p.m.: Dinner ([Pennyroyal Station](#))

## **Workshop Day #2: Friday, June 30, 2023**

7:30 a.m. - 8:40 a.m.: Breakfast at *Bagels 'N Grinds* located at The Hotel (not a group event—please just grab breakfast at your leisure)

8:40 a.m. - 9:00 a.m.: Walk with Osama Eshera to McKeldin library

9:00 a.m. - 9:15 a.m.: Day overview and planning (Raff Viglianti, Mathew Barber)

9:15 a.m. - 10:45 a.m.: Session 1: Strategies for Right-to-Left Encoding (Post-OCR)

- oXygen XML editor for RTL scripts (Hugh Cayless)
- mARKdown and NotePad++ (Lorenz Nigst, Peter Verkinderen)
- FairCopy (Nick Laiacona)
- Free and Open Source Software for XML (Raff Viglianti)

10:45 a.m. - 11:00 a.m.: Break

11:00 a.m. - 12:30 p.m.: Session 2: Enriching Editions: Collation, Additional Markup, Annotation

- AnonymClassic approaches to collation (Theo Beers)
- Performant Software solutions: ArchivEngine and EditionCrafter (Nick Laiacona)
- Collation algorithms: CollateX, Passim (David Smith)
- TEI Publisher annotation module (Magdalena Turska)

12:30 p.m. - 2:00 p.m.: Lunch

2:00 p.m. - 3:30 p.m.: Session 3: Publication Systems

- Scaife and KITAB's Digital Arabic Reader (Greg Crane, Mathew Barber)
- Making and Knowing Reader (Nick Laiacona)

OpenITI AOCP presents:

# Digital Publication of Right-to-Left Script Corpora



- TEI Publisher (Daniel Stoekl, Madgalena Turska)
- Static solutions: CETEIcean (Raff Viglianti, Hugh Cayless, Brown-deVost)

*3:30 p.m. - 3:45 p.m.*: Break

*3:45 p.m. - 5:15 p.m.*: Session 4: Open Discussion and Closing Remarks

*5:15 p.m. - 6:30 p.m.*: Break

*6:30 p.m.*: Dinner ([GrillmarX College Park](#))