

The Digital South Caucasus Collection: An International Digital Library Project

Jasmine Smith, New York University, New York City, NY, United States

Abstract

The Library of the Institute for the Study of the Ancient World (ISAW) at New York University aims to provide access to scholarship on the ancient world for the widest audience possible, both domestically and internationally, via our digital library collections. Our latest project, the Digital South Caucasus Collection (DSCC), is an international collaboration between ISAW and archaeological institutions in the south Caucasus that seeks to preserve and provide open access scholarship on the ancient south Caucasus. This poster describes some of the initial challenges and approaches to building a digital library project during a global pandemic. The project is very much a work in progress, but has already provided various insights into collaborating with international institutions on digital library collections.

Introduction

The Institute for the Study of the Ancient World (ISAW) at New York University (NYU) supports doctoral and postdoctoral education and research in the ancient world. ISAW particularly encourages interdisciplinary work within ancient studies and the implementation of digital methodologies. The ISAW Library, a branch of the NYU Division of Libraries, supports not only the ISAW community, but also the wider community of scholars of the ancient world through its commitment to providing access to research materials.

To support these communities, the ISAW Library has created digital collections of scholarship in order to provide access to research materials that are difficult to find. These collections are often created in partnership with ISAW Faculty and Research Associates. For example, our **Digital Central Asian Archaeology (DCAA)** collection was a collaboration between the ISAW library, ISAW faculty, and various partners from other institutions to create a digital library of scholarship on the history and archaeology of Central Asia. The collection consists of archaeological reports and surveys, many of which were published in limited runs in Uzbekistan and not widely held, and so in need of digital preservation. This material was digitized by the SilkRoDe Digital Library Project, but due to several factors, the project was unable to build the digital infrastructure necessary to serve the material to the public. The ISAW Library thus agreed to preserve some of this material in NYU's institutional repository and make it searchable and discoverable through an open-source content management system.

Our current initiative, the **Digital South Caucasus Collection (DSCC)**, is a work-in-progress and began as a result of the need expressed by an ISAW Research Associate. The DSCC aims to

provide access to digitized copies of scholarship on the ancient South Caucasus (i.e., the region now defined by modern Georgia, Armenia, and Azerbaijan), while also preserving this scholarship in an NYU repository. The main motivations for this project are to serve the research interests of our scholarly community and to provide access to scholarship on the ancient world to the widest possible audience. Indeed, this project aims to serve not only members of our direct community, but also an entire area of research within ancient studies. This poster provides a brief introduction to the DSCC project and describes some of our initial challenges and approaches to building a digital library project during a global pandemic, particularly one reliant on international collaboration.

Background

The DSCC project began in 2018 at the instance of Dr. Karen Rubinson, an ISAW research associate who is an established scholar of the ancient south Caucasus and a leader in scholarly organizations dedicated to this field. The topic of a digital library first occurred at meetings of these organizations; Rubinson explained that scholars discussed the struggles of finding research materials, leading to colleagues often asking each other for PDFs of scholarly publications. The COVID-19 pandemic created similar situations with scholars creating groups on social media platforms such as Facebook to share and ask colleagues for pdfs, due to lack of library access.

Given the success of the ISAW Library's recent development of the DCAA, Rubinson suggested trying to build a similar digital collection of scholarship on the ancient south Caucasus in collaboration with the ISAW Library. The digital collection is intended primarily to solve the issue of scholars' inability to access easily the research materials they need. Rather than contacting other scholars to find PDFs, publications in what would ultimately become the DSCC would be discoverable through multilingual metadata and made available in a downloadable format. Furthermore, the project would create international collaborations and cooperation between NYU and academic institutions in the south Caucasus. An online collection of this material is of particular importance as some library catalogs of institutions in the south Caucasus that hold much of this material are not available online and do not provide sufficient metadata needed for discoverability. The DSCC aims not only to provide downloadable formats for all material in the collection, but also to provide metadata in original languages when possible.

Due to staffing changes in 2019, the DSCC project encountered a slower start than originally anticipated. Challenges with collection development, technical development, and the closures of libraries during the COVID-19 pandemic in early 2020 also contributed to a slower pace for the project. However, towards the end of 2020 the

project made several advances, particularly in our international and interdepartmental collaborations that will support the sustainability of this forthcoming digital library.

Approaches to Collection Development

Our initial approach to the DSCC focused on collection development and the technical development for the collection platform. We approached collection development by executing agreements between NYU and institutions in the south Caucasus in 2018. These agreements give NYU the non-exclusive right to host publications published by these institutions and their predecessor organizations in the DSCC. We currently have agreements with the Georgian National Museum (GNM) and The Institute of Archaeology and Ethnography in the National Academy of Sciences of the Republic of Armenia. We are hoping to conclude a similar agreement with the Institute of Archaeology and Ethnography in Azerbaijan in the future. Once the agreements were signed, we needed a systematic way of gathering publications for the collection.

Our first attempts at gathering publications involved searching various institutions' websites and contacting Rubinson's colleagues at the institutions, asking them to provide lists of their publications or access to any of their digitized material. We quickly realized we would need a more formal partnership with researchers and librarians at these institutions, as such lists were difficult to find or non-existent. We decided to focus on one institution at a time, and on Rubinson's recommendation, we invited a researcher from the Georgian National Museum to join the project team.

Our proposed tasks for collection development included the following: locating bibliographies or lists of scholarship published by the Georgian National Museum, forming a list of these publications that are solely related to the study of the ancient south Caucasus, and using this list to search for digitized copies and print copies that we would then digitize. The website for the DSCC would also include a form for scholars to contribute publications for inclusion into the collection. Bibliographies and catalogs of the publications by the Georgian National Museum and its predecessor institutions were not easy to find online if they were online at all, and there seemed to be no bibliography of their publications from the past 20 years. The researcher at the GNM worked with the GNM's librarians to find a print bibliography of publications by the Archaeological Research Center of the Georgian Academy of Sciences from the 1970s-1990s. The GNM library catalog consisted of a large Excel spreadsheet divided into tabs based on subject area, which they provided for our use. We are currently in the process of scoping this spreadsheet and the print bibliography (which was scanned and sent to the ISAW team) to create a list of publications related to the ancient south Caucasus.

Approaches to Technical Development

We approached the technical development of the collection's user interface platform similarly to the DCAA by creating a prototype using Omeka, an open-source content management system (CMS). Omeka is useful in that it allows users to browse and search the collection, and it allows us to display the metadata and files necessary to make the publications searchable and discoverable. Our prototype consists of approximately 50 items published by the GNM and Institute of Archaeology and

Ethnography in Armenia. We stored PDFs of each item in NYU's instance of Box cloud storage service, while the metadata for each item sits in a google spreadsheet. Our team wrote a series of scripts that package the PDFs and metadata into SAF files. The SAF files are then uploaded to NYU's institutional repository called the Faculty Digital Archive (FDA) using a batch uploader. The FDA returns a CSV file of the data from the SAF files with added FDA handles for each item. The CSV is then uploaded to Omeka using Omeka's CSV Import plugin.

Unfortunately, we have encountered interoperability issues between the FDA and Omeka. For example, we experienced issues with the CSV Import plugin accepting the CSV data from the FDA. The process of batch uploading our data to the FDA and retrieving the URLs also proved to be more complicated than expected. When attempting to create solutions to these issues, they were often erased by Omeka updates. We also attempted to use Omeka S, and found that CSV import works very differently and required manual input for every item uploaded, which was completely unscalable and unsustainable.

Our challenges with the FDA led us to investigate whether other departments at NYU Libraries were having similar experiences. We recently consulted the data services department where our colleagues introduced us to Invenio, an open-source research data management software framework. Invenio would provide our project a preservation and storage environment, replacing the function of the FDA. We are currently in further conversation with the data services team about adopting Invenio, and are also considering Drupal for the collection's front-end user interface.

Impact of the COVID-19 Pandemic

COVID-19 had a significant impact on the timeline of the project, as well as our ability to communicate and collaborate with our international partners. Beyond the signing of the agreements in 2018, we started formal collaboration between the ISAW Library and staff at the Georgian National Museum during the pandemic. This process took the form of video conference calls. Before the pandemic, our international communication primarily consisted of emails.

The pandemic severely affected our timelines due to library closures and collaborators becoming ill with the virus. Differing lock-down periods between the United States and Georgia greatly impacted our collection development efforts in particular. As mentioned previously, we initially attempted to scope online, as a list of the museum's publications proved difficult to find. This led to our need to reach out directly to staff, albeit at the worst possible time. The effect of COVID-19 and the lock-downs caused delays in communication and the retrieval of information. This was clearly not the opportune time to begin international communications and collaboration on a project that needed print materials.

Future Work and Takeaways

The challenges of using Omeka and the FDA led us to two options: search for another CMS and data repository solution, or build a bespoke system that catered to our needs. Though building a simple web interface attached to a database was doable, we preferred to have a solution with a community or institutional support behind it. Building a new system could potentially lead to sustainability

issues if the staff member who built it were to leave their position. We were also dedicated to having the files stored and preserved in an institutional repository.

Evaluating our options for repositories and content management systems will likely be similar to Cai et al's "Evaluation priorities for choosing a discovery tool"[1]. These priorities include architecture framework, depth of integration, metadata quality, user interface, openness, and interoperability with existing systems. Overall evaluation of the usability of our future systems will also need to be evaluated, and a framework for this evaluation may follow similar models such as those discussed by Xie, 2016 [3]. The sustainability of the project is a top priority and will continue to be taken into consideration as we evaluate new approaches for the collection's technical structure. Sustainability of the project can be ensured when it is a priority during the design process and when our technologies are flexible in meeting our institutional and subject discipline needs [2].

The development of partnerships was crucial to our ability to create the collection. These partnerships became increasingly necessary as we discovered the difficulties of collecting, particularly during library closures due to the COVID-19 pandemic. The agreements between NYU and institutions in the south Caucasus were a necessary step in ensuring we could provide content published by these institutions, but the agreements did not ensure how we would make these publications accessible in the DSCC.

Through partnering with a researcher at the GNM we gained a better understanding of the steps we needed to take in order to gather a list of relevant publications. As mentioned previously, the current GNM library catalog consists of an Excel spreadsheet that is not available online. The spreadsheet was made available to us via our partner at the museum and has been useful in scoping material for the DSCC as well as understanding the level of metadata and description work that will be required.

Despite the setbacks of a global pandemic, staffing changes, and changes in approach, the ISAW Library and our partners are continuing to work toward the creation of the DSCC. Now that we have identified ways of collecting relevant scholarship and a new approach to the digital infrastructure, we are currently applying for funding to support the next phase of the project. The next phase includes digitizing the publications on the list we are currently forming, and hiring a developer to build the website. Our current plan calls for the digitization of publications to occur at the Georgian National Museum and to apply for grant funding to support the hiring of students to scan, as well as the acquisition of scanning

equipment. Once the website is developed and we have digitized approximately 70,000 pages of scholarship from Georgia, we will attempt to repeat this workflow with any necessary adjustments in Armenia, and eventually Azerbaijan once we have a contract.

Conclusion

The Library of the Institute for the Study of the Ancient World is dedicated to making scholarship on the ancient world accessible to the widest possible audience. Our current project, The Digital South Caucasus Collection, seeks to curate, digitize, preserve, and provide access to scholarship on the ancient south Caucasus. While this project is a work in progress with its fair share of challenges, it has offered us new insights and approaches to building our digital library collections. These new approaches have involved prioritizing sustainability when dealing with interoperability issues between open-source content management systems and institutional repositories, and adapting to situations that delay collection development such as closed libraries and lack of library catalog access. As we look toward the future of the project, we will continue to prioritize collaboration between our international partners, as well as our interdepartmental colleagues.

References

- [1] L. International Conference on Asia-Pacific Digital, *Digital libraries : for cultural heritage, knowledge dissemination, and future creation : 13th International Conference on Asia-Pacific Digital Libraries, ICADL 2011, Beijing, China, October 24-27, 2011 : proceedings*. Heidelberg ; New York: Springer, 2011.
- [2] TpdL, "Theory and Practice of Digital Libraries Second International Conference, TPDL 2012, Paphos, Cyprus, September 23-27, 2012. Proceedings," 2012.
- [3] I. Xie, "Discover Digital Libraries : Theory and Practice," 2016.

Author Biography

Jasmine Smith is the Assistant Research Scholar for Digital Initiatives at the Institute for the Study of the Ancient World Library. She received her M.S. from the University of Michigan School of Information where she focused on archives and digital curation. Her thesis, 'Motivations and Expectations for Central Database Creation at an Egyptian Archaeological Site', studied a team of archaeologists planning to move from individualized ways of storing data to a central database system.