

# SobekCM's Community Ecosystems & Socio-Technical Practices

**Proposal Submission:** Open Repositories 2014 (June 9-13, Helsinki, Finland)

**Presenters:** Mark V. Sullivan ([marsull@uflib.ufl.edu](mailto:marsull@uflib.ufl.edu)) and Laurie N. Taylor ([laurien@ufl.edu](mailto:laurien@ufl.edu))

**Abstract:** Academic research and cultural heritage institutions in the Caribbean—spanning over one million square miles of water, with hundreds of islands, many countries, and a multiplicity of languages and cultures—have long-standing collaborative relationships for preservation, access, and supporting scholarly activities. The real-world complexities of the Caribbean ecosystem led to collaborative creation of the Digital Library of the Caribbean (dLOC, [www.dloc.com](http://www.dloc.com)). The variety, complexity, and diversity of needs across dLOC's many partners directly supported development of the SobekCM Open Source Software (<http://sobekrepository.org>). SobekCM developed in collaboration with dLOC's institutional, scholarly, publishing, and other communities. For data needs, SobekCM is growing across new communities. Activities include collaboration with UF (dLOC Technical Host Partner) Research Computing and the Sunshine State Education & Research Computing Alliance (SSERCA) to integrate SobekCM as a front-end for high-performance and big data computing systems. This presentation will cover SobekCM's socio-technical (people, policies, procedures) practices for supporting and connecting ecosystems with user-focused development with user communities from galleries, libraries, archives, museums, universities, scholarly groups, publishers, and others, all contributing to the system development, use, and integrated connections with other systems and activities. This includes SobekCM's integration with rich data support through connection with research computing systems and communities.

## Introduction

Academic research and cultural heritage institutions in the Caribbean—spanning over one million square miles of water, with hundreds of islands, many countries, and a multiplicity of languages and cultures—have longstanding collaborative relationships for preservation, access, and supporting scholarly research and teaching. The real-world complexities of the Caribbean ecosystem led to collaborative creation of the Digital Library of the Caribbean (dLOC, [www.dloc.com](http://www.dloc.com)). The variety, complexity, and diversity of needs across dLOC's many partners directly supported the development of the SobekCM Open Source Software (<http://sobekrepository.org>).

Being parallel to and intertwined with the development of dLOC, the SobekCM software relies on user-focused development. SobekCM has developed in collaboration with dLOC's institutional, scholarly, publishing, and other communities, as well as in collaboration with other groups and systems. Currently, the dLOC and SobekCM communities are expanding with new technologies and communities to meet data needs. This includes recent work to build capacity using SobekCM as a front-end to high-performance computing systems available at the

University of Florida (UF is the dLOC Technical Host Partner) and with the Sunshine State Education & Research Computing Alliance (SSERCA) community.

This presentation will focus on socio-technical (people, policies, procedures) for supporting and connecting repository ecosystems with the example of the SobekCM Open Source repository software and its user-focused development with user communities from GLAMs (galleries, libraries, archives, museums), universities, scholarly groups, publishers, and others all contributing to the system development, use, and integrated connections with other systems and activities. The presentation will cover SobekCM's current integration with rich data support through connection with Research Computing systems and communities.

### **SobekCM's Community Ecosystems: the Digital Library of the Caribbean (dLOC)**

The Digital Library of the Caribbean (dLOC) is a collaborative digital library with 37 international partner institutions, affiliated partners, publishers, scholars, and other community members who contribute materials and expertise to the shared repository using the SobekCM Open Source software. dLOC partners contribute to and rely on the SobekCM software. dLOC's strong voice in the technical development flows from dLOC's inclusive and equitable partner structure. dLOC is a collaborative community with institutional partners who defined the shared governance model to support partner needs as well as the shared needs of many related and diverse communities, including scholarly and publisher communities. Created both as a repository and a community for building capacity, dLOC will celebrate its tenth anniversary in 2014, and its already major achievements and ongoing activities with high levels of growth and development in the community, shared collaborative capacity, and technologies.

dLOC's success is founded in large part to its connection to and development from existing communities, being born from the work of a group within the Association of Caribbean University, Research and Institutional Libraries (ACURIL) and being born from longstanding collaboration across the Caribbean for preservation and access with institutions in many countries and in Florida. dLOC and SobekCM's mutually informing and supporting development represents the normal process for SobekCM's practices that connect and activate collaboration and community with human and technical systems.

### **SobekCM's Community Ecosystems: UF Research Computing and Data**

The University of Florida (UF) has been a key connector for this work as a technical partner in the Caribbean for many years and technology transitions, including the relatively recent transition from microfilm to digital and with dLOC's founding. UF's contributions to SobekCM have been in collaboration with and informed by many partnerships. For instance, as the largest information network in the State of Florida, the UF Libraries have longstanding histories in providing technological infrastructure for collaborative, geographically diverse projects including dLOC, the Florida Digital Newspaper Library, cooperatives in smaller areas such as creating satellite digitization offices, and collaborative geographically distributed topical projects

such as the Government Documents Digital Library for Federal Depository Libraries in Florida, Puerto Rico, and the US Virgin Islands. The uniqueness of these projects and the contributions from their communities provided detailed functional and non-functional requirements for SobekCM. Supports demanded of these projects include strictly technical features as well as ensured support for scalability, sustainability, interoperability, individual branding, ease of complete online submission, and socio-technical supports for collaboration with scholars, curators, and other expert contributors.

In addition to inter-institutional collaboration, SobekCM is part of rich intra-institutional collaboration at UF, with the most recent work on integrating data repository systems. As the oldest university in Florida, UF has long traditions of advanced research and strong research entities within it, as with UF Research Computing, and strong ties to larger collaboratives as with the Sunshine State Education & Research Computing Alliance (SSERCA). SSERCA is comprised of Florida's public and private universities and is designed to bring together geographically distributed institutions and their resources (high-end computing facilities, massive data storage systems, specialized research instruments, and high-speed networks). SSERCA, like dLOC, is designed to leverage the connected resources and thereby enable a collective impact that far exceeds the sum of individual parts. SobekCM's initial creation and ongoing development show the benefits of socio-technical activities for activating, bridging, and integrating repository systems, both human and technical.

### **SobekCM's Socio-Technical Development Practices**

As SobekCM passes its eighth year of development and enters its third year as open source software, our methods for exploring new ways to keep the system current and relevant are based on community-sourcing with direct programmer and developer interaction with members of the community, keeping the system focused on community, local needs while ensuring wide, global access supporting millions of user views each month. Development has increasingly emphasized ease of customization, simplicity of set up, and increasing the extensibility to encourage adoption. This work has empowered the community with increasing independence and self-reliance as well as increasing integration into other systems and practices, as several of the institutions for which the university originally hosted materials have adopted, customized, and integrated the system locally.

In order to serve researchers and communities, SobekCM increasingly adds and integrates innovative features as demonstrated by the fully on-line creation of materials, the new geographic module (including geo-placement of map overlays and points of interest), and structural metadata module. UF is adding dataset support, a local authority system to enable linked data, and ability to easily link materials into the Common Core Standards taxonomy. Other community members are expanding support for oral histories (both display and creation), "touring" functionalities using mapped materials, integration with ArcGIS for specific high-end GIS related needs, further enhanced project tracking for curation, and more. These new features

are made possible by other community members continuing to actively engage with a wide variety of contributions, from technical through testing and informed feedback on local and larger community needs. With user and community engagement at such a high level developers benefit from abundant and direct insight into community needs and from the opportunities made possible by SobekCM's socio-technical practices which engage across community ecosystems.