

SobekCM's Community Ecosystems & Socio-Technical Practices



Presented by Mark V. Sullivan
June 10th, 2014






Introduction to SobekCM

Is an open source digital library system

Originally built at and maintained by UF Libraries
and

Co-developed in partnership with a variety of
institutions, collaborative projects, and other
community ecosystems



SobekCM : File Formats and Folder Structures

- SobekCM is open to all file formats, resource types
- Versions of a single object can be stored together (e.g., oral history audio, video, transcript, & published copy)
- SobekCM utilizes “pair-tree” folder structure, similar to California Digital Library holding metadata and other digital resource files

SobekCM : Metadata Support

- Supports the following metadata types
 - **METS**
 - MODS
 - MARC21/MARC XML
 - Dublin Core
 - EAD
 - GML/KML
 - TEI
 - VRACore
 - DarwinCore
 - Custom schemas

SobekCM : Technological Stack

Built upon a technology stack including:

- Windows IIS web server
- C# (as well as Javascript)
- Microsoft SQL Server
- Solr

- IIPimage for JPEG2000 views
- ImageMagick
- Ghostscript
- etc..



SobekCM Community Ecosystems

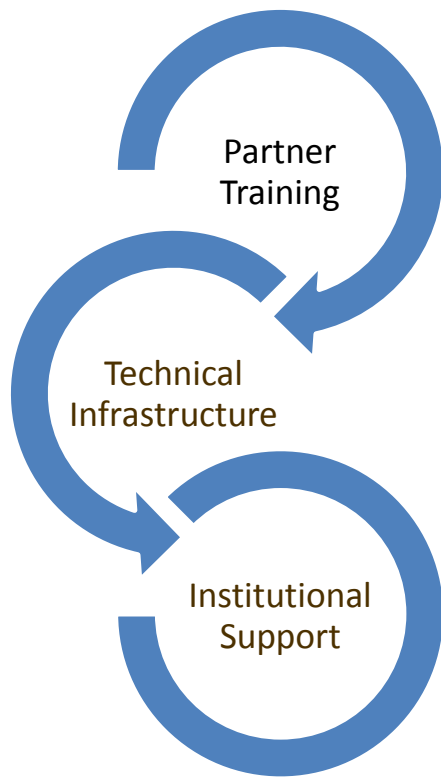
1. Digital Library of the Caribbean
2. University of Florida Research and Computing
3. GLAMS (Galleries, Libraries, Archives, Museums)

(Each community supports one Universal Community)



Digital Library of the Caribbean

dLOC's diverse partners serve an international community of scholars, students, and citizens by working together to preserve and to provide enhanced electronic access to cultural, historical, legal, governmental, and research materials.

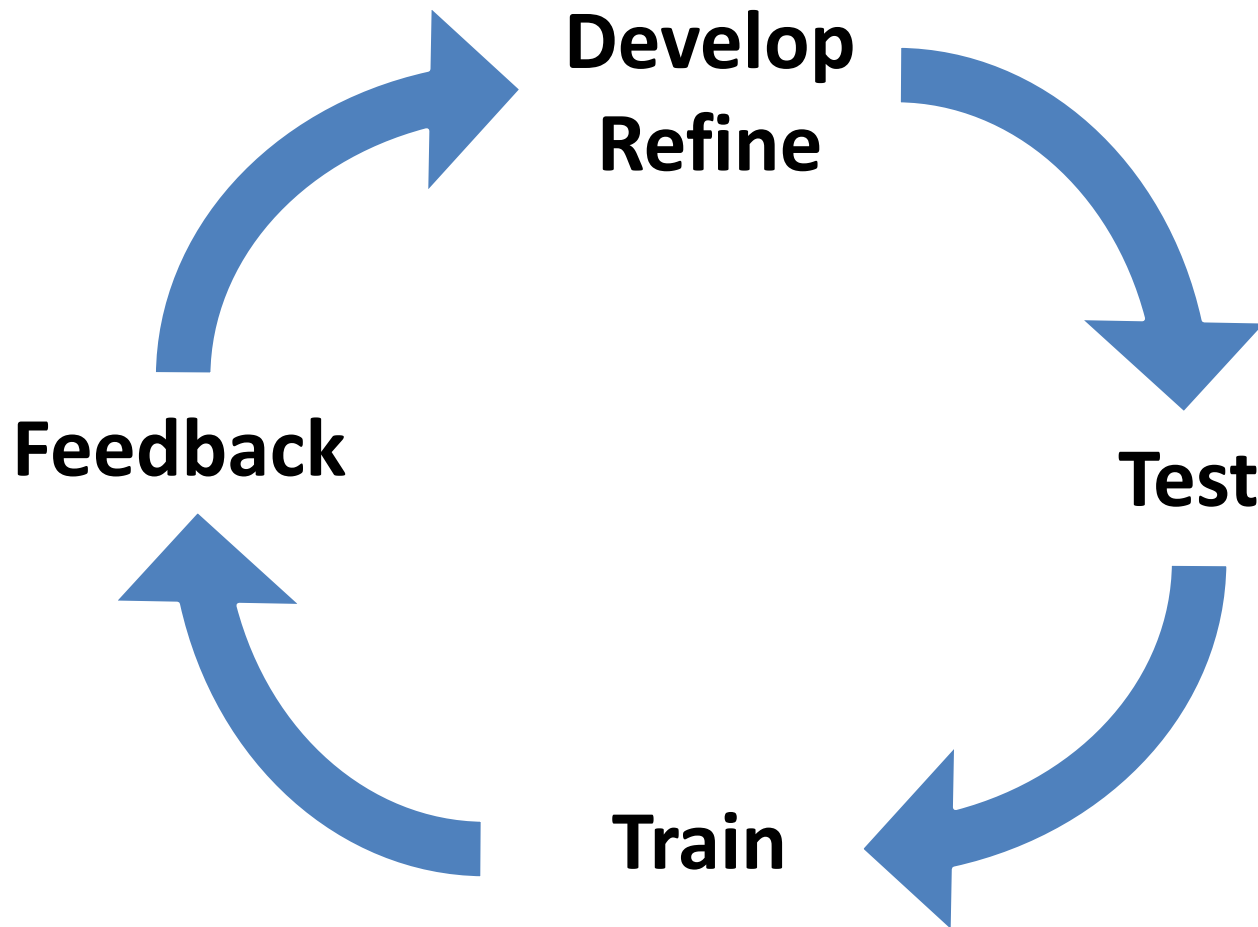


dLOC Quick Facts

- ❑ Content Management System and Long-term Preservation
- ❑ 37 Partners – Caribbean, Europe and US
 - Over 24 million hits since 2006
 - Over 1.8 million pages of open access content
 - 13,000 titles with 80,000 items
- ❑ Training Programs
- ❑ Scholarly Collaborations
- ❑ Educational Outreach
- ❑ Shared Governance
- ❑ New Initiatives



Community Involved Development



dLOC Example 1: Digitization Toolkit

1. As originally developed

- a. Multiple separate windows applications
 - i. Tracking tool
 - ii. Metadata Creation tool
 - iii. Quality Control tool
 - iv. FTP tool
- b. Locally installed MS SQL Express database
- c. Unintuitive and difficult to use
 - i. Difficulty in training
 - ii. Difficulty in local installs

dLOC Example 1: Digitization Toolkit

2. Second Toolkit

- a. Still based on the same pieces as before, but toolkit itself developed completely from scratch again
- b. Single windows form application, all workflow and steps occur in a single window
- c. Using locally installed MS SQL Express database (still)
- d. Problematic for local installs and updates (still)

dLOC Example 1: Digitization Toolkit

3. Completely web-based approach
 - a. Metadata creation and editing templates
 - b. Quality control online
 - c. Multiple file uploads

New web-based templates are now utilized by all projects, such as Institutional Repository and items loaded by internal groups

◀ CANCEL

SAVE ▶

Basic Information

Main Title: ?Other Titles: + ?Resource Type: 1 map : col. ; on 3 sheets 65 x 55 cm. -- nyu -- eng -- 600 ... ?Language: + ?

Name Authorities

Creator:
 + ?Donor: ?

Publication Details

Publisher: Name:
Location(s): + ?Manufacturer: Name:
Location(s): + ?Publication Date: ?Copyright Year: ?Edition: ?

◀ CANCEL

SAVE ▶

Other Titles:

EDIT MAIN TITLE ? X

Title Type:

Non Sort: Language:

Title:

Sub Title:

Statement of Responsibility:

Part Numbers:

Part Names: Authority:

CLOSE

Theory of the Haitian meringue

RESOURCE

EDIT

SETTINGS

VIEW

HELP



A0001



Page: Front Cover 1

D: ☒ Cover

Name: Front Cover

A0002

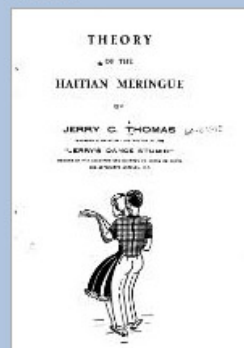


Page: Front Cover 2

D: ☐ Cover

Name: Front Cover

A0003



Page: Page 1

D: ☒ Title

Name: Title Page

A0004



Page: Page 2

D: ☐ Title

Name: Title Page

A0005



Page: Page 3

D: ☒ Foreword

A0006



Page: Page 4

D: ☒ Chapter

A0007



Page: Page 5

D: ☒ Chapter

A0008



Page: Page 6

D: ☐ Chapter

Comments:

COMPLETE

CANCEL

FILE

CONTROLS

ACTIONS

HELP

Find a Location 

Google

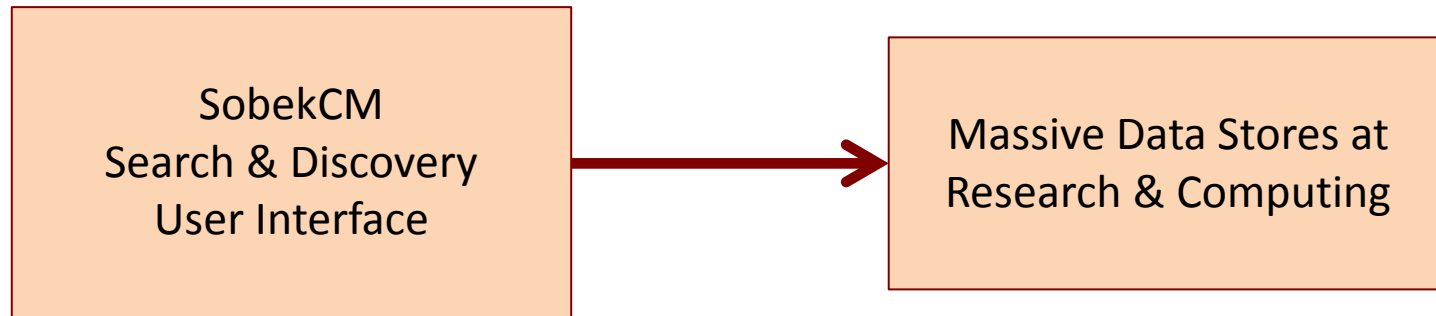
dLOC Example 2: Curator Tools

Managing the collection online:

- Edit (multi-lingual) home pages
- Change images (banners, buttons)
- Add child pages or collections
- Control the look, feel, and behavior of your online collection

University of Florida Research and Computing

New ecosystem of data repository, working with UF Research and Computing supporting research data generated by the university.



Data Support (beta)

Gainesville City Directory Dataset

SHARE

GO ►

Showing 1 to 100 of 9,699 entries

First Previous 1 2 3 4 5 Next Last

PersonID	Directory Title	Directory Year	Page Sequence	Source Line	First Name	Last Name	Middle Name	Title	Sex	Racial Affiliation	Spouse	Historic Note	Home Location
6416	Burtz' Gainesville Directory for 1905-1906	1905	21	Abbott, S S (Lucia), retired, 800 E Seminary	S	Abbott	S		m	white	Lucia		800 E Seminary
4797	Burtz' Gainesville Directory for 1905-1906	1905	21	Abrams, Ella J, 508 West Main N	Ella	Abrams	J		f	white			508 West Main N
4800	Burtz' Gainesville Directory for 1905-1906	1905	21	Abrams, J K (May), bookkeeper, 508 West Main N. Bellphone 33	J	Abrams	K		m	white	May		508 West Main N, Bellphone 33
7914	Burtz' Gainesville Directory for 1905-1906	1905	21	Ackley, J E (Sarah), carpenter, Alabama. N G'ville	J	Ackley	E		m	white	Sarah		Alabama, N Gainesville
2641	Burtz' Gainesville Directory for 1905-1906	1905	21	Adamson, J D, printer, 301 West Main S	J	Adamson	D			white			301 West Main S
	Burtz' Gainesville Directory for 1905-1906	1905	21	Adamson, Millie, 301 West Main S	Millie	Adamson			f	white			301 West Main S

Data Support (beta)

Person		
PK	PersonID	<i>integer</i>
	Directory Title	<i>string</i>
	Directory Year	<i>integer</i>
	Page Sequence	<i>integer</i>
	Source Line	<i>string</i>
	First Name	<i>string</i>
	Last Name	<i>string</i>
	Middle Name	<i>string</i>
	Title	<i>string</i>
	Sex	<i>char</i>
	Racial Affiliation	<i>string</i>
	Spouse	<i>string</i>
	Historic Note	<i>string</i>
	Home Location	<i>string</i>
view details view contents		

Corporation		
PK	CorporateID	<i>integer</i>
U	Corporate Name	<i>string</i>
view details view contents		

Person Occupation Link		
← Person.PersonID	FK	PersonID <i>integer</i>
← Corporation.CorporateID	FK	CorporateID <i>integer</i>
		Occupation <i>string</i>
view details view contents		

Data Support (beta)

Gainesville City Directory Dataset

DESCRIPTION ▾

EXPLORE DATA

DATA STRUCTURE

REPORTS

DOWNLOADS

PRINT

SEND

+ ADD

SHARE

Data Structure / Codebook

[← Back](#)

Person Table Details

Table contains information about all of the people which appeared within the city directory. These may also be linked to a corporate body and/or an historic occupation.

Person

PK	PersonID	integer	Unique key for this person within the dataset
	Directory Title	string	Title of the city directory from which this person's data was derived
	Directory Year	integer	Year of the city directory from which this person's data was derived
	Page Sequence	integer	Page sequence within the city directory that this person's data was derived
	Source Line	string	Original source line from which this person's data was derived
	First Name	string	Last or family name of this person
	Last Name	string	First or given name of this person
	Middle Name	string	Middle name of this person
	Title	string	Title associated with this person (i.e., Mr., Reverand, etc..)
	Sex	char	Normalized racial-ethno affiliation associated with this person by the city directory
	Racial Affiliation	string	Name of any spouse (usually just the first name)
	Spouse	string	General note field
	Historic Note	string	Location of the home, usually cross-streets or an street address
	Home Location	string	Home Location

[view contents](#)



Data Support (beta)

Data Creation, Management, and Sharing

1. Create a dataset online from scratch
2. Edit data online
3. Correctly permission and share data utilizing existing user and user group permissions
4. Alter data structure online



Sobek Digital Hosting & Consulting

Finally, the socio-technical evolution of the software has reached the point where there is now a company for providing hosting of repositories as a cloud solution at a fraction of the cost of hosted solutions.

<http://sobekdigital.com>



FOR MORE INFO:

Contacts:

Mark Sullivan (mark.v.sullivan@sobekdigital.com)

Laurie Taylor (laurien@ufl.edu)

SobekCM Discussion Group:

<https://groups.google.com/forum/#!forum/sobekcm-discuss>

More:

<http://sobekrepository.org>

<http://sobekdigital.com>

<http://www.youtube.com/user/sobekcm>

