

Enhancing Metadata and Improving Discoverability for Digital Collections

**ALCTS CaMMS Cataloging Norms Interest Group
American Library Association Midwinter Meeting
January 26, 2019**

***Dave Van Kleeck
Chair, Cataloging & Discovery Services
University of Florida***

***Chelsea Dinsmore
Chair, Digital Support Services
University of Florida***

PRINT SEND SHARE

UFDC

UNIVERSITY of FLORIDA
DIGITAL COLLECTIONS

Search Collection:

Go



UFDC HOME

ADVANCED SEARCH

TEXT SEARCH

BROWSE PARTNERS

The University of Florida Digital Collections (UFDC) hosts more than 300 outstanding digital collections, containing over 13 million pages, covering over 78 thousand subjects in rare books, manuscripts, [antique maps](#), [children's literature](#), newspapers, [theses and dissertations](#), data sets, photographs, [oral histories](#), and more for [permanent access and preservation](#). Through UFDC, users have free and [Open Access](#) to full unique and rare materials held by the University of Florida and [partner institutions](#).

<http://ufdc.ufl.edu/>

METADATA CHALLENGES

INPUT

First items entered in early 1990s.

Variety of input standards over the years.

Inconsistent curatorial oversight.

MAINTENANCE

Hundreds of collections to maintain.

Finding staffing and other resources.

Limitations of hosting platform.

Theses & Dissertations from the University of Florida

IR @ UF



HOME ▾

ADVANCED SEARCH

TEXT SEARCH

BROWSE BY ▾

VIEW ITEMS ▾



PRINT



SEND



ADD



SHARE

Search Collection:

Go

The **University of Florida Theses & Dissertations Collection** within the *Institutional Repository at the University of Florida (IR@UF)* will eventually include all theses and dissertations from the University of Florida. The collection currently includes:

- Open Access Theses & Dissertations, originally submitted electronically (ETDs)
 - [Data, data sets, and supplementary data related to theses and dissertations](#)
- [Projects in Lieu of Theses](#), submitted electronically but which would otherwise not be available online ([submission help](#))
- Retrospective dissertations, digitized from print-only dissertations ([information on this service for dissertation authors](#))
- [UF Undergraduate Honors Theses \(new in 2013/2014\)](#)

Usage statistics for each ETD are available under the "description," and overall usage statistics are available [here](#).

For dissertations not covered by the [Retrospective Dissertation Scanning Policy](#), thesis and dissertation authors can grant permissions to have their theses and dissertations added to the *University of Florida Theses & Dissertations Collection*. For print-only theses and dissertations, the permissions form and mailing address [are available here](#). Authors can also grant permissions and submit electronic theses and dissertations through the [online self-submittal form here](#).

<http://ufdc.ufl.edu/ufetd>

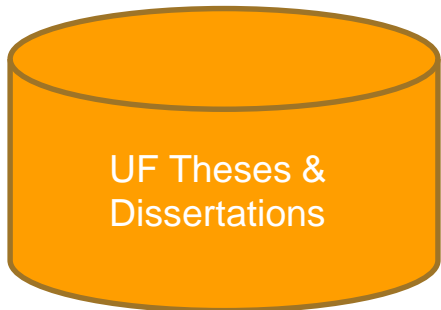
Initial Steps

- Select thesauri to compliment existing LCSH in records.
- Pull data fom UFDC = OCR'd full text + existing metadata.
- Run extracted metadata through MAIstro software to enhance record metadata.

Hypothesis

Enhance subject access to the materials by adding terms to the records from the JSTOR thesaurus, based on the OCR'd full text of each document and selected via machine assisted indexing.

University of Florida Digital Collections



XML Records & Full
Text Exported from
UFDC for Analysis



XML CMS
System



Staff Review Panel



Enhanced Metadata
Added to UFDC
Records



Updated XML
Records
Returned to
UFDC



MARC Records
Exported



Subjects

Subjects / Keywords: alligator, detectability, estimates, habitat, population, sightability, survey
Interdisciplinary Ecology -- Dissertations, Academic -- UF

Genre: Electronic Thesis or Dissertation
bibliography (marcgt)
theses (marcgt)
Interdisciplinary Ecology thesis, M.S.

Original Subject Terms

From: *Effects of Habitat Type and Structure on Detection Probabilities of American Alligators (Alligator mississippiensis) During Night-Light Counts*,
Cameron Blair Carter (2010),
<http://ufdc.ufl.edu/UFE0041794>

Enhanced Subject Terms

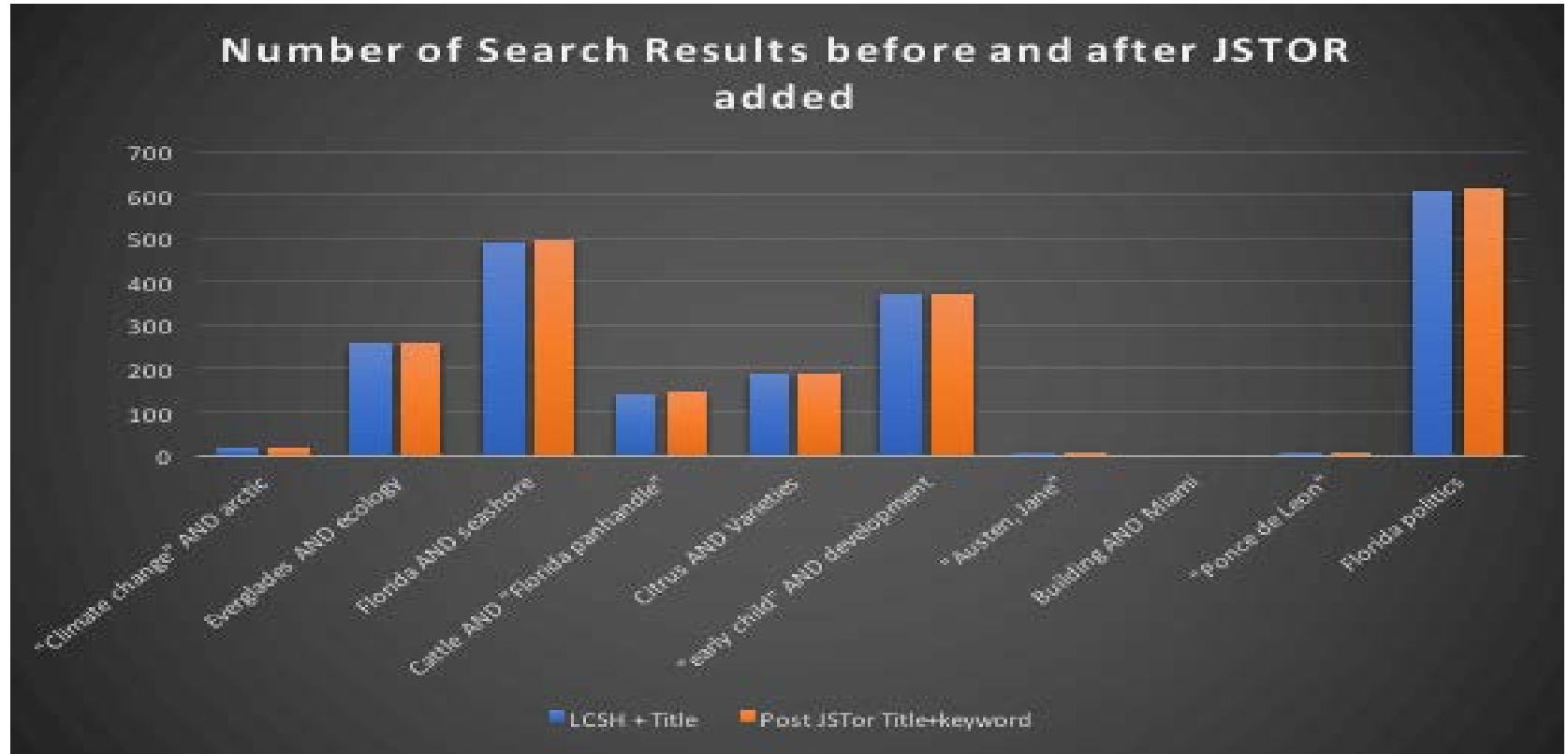
Subjects

Subjects / Keywords: alligator, detectability, estimates, habitat, population, sightability, survey
Interdisciplinary Ecology -- Dissertations, Academic -- UF

Alligators (JSTOR)
Population estimates (JSTOR)
Aquatic habitats (JSTOR)
Wet prairies (JSTOR)
Vegetation (JSTOR)
Night lights (JSTOR)
Water depth (JSTOR)
Area Surveys (JSTOR)
Florida -- Everglades
Florida -- Kissimmee
Florida -- Franklin County
Florida -- Taylor County

Genre: Electronic Thesis or Dissertation
bibliography (marcgt)
theses (marcgt)
Interdisciplinary Ecology thesis, M.S.

INITIAL TESTING



KEY TAKEAWAYS



Optical Character Recognition (OCR) results are variable.



All collections are special - One size rarely fits all.



MAI is only machine ***assisted*** indexing - You still need experts.



Standardizing fields in SobekCM is essential.



We need to build a Florida-specific taxonomy for prominent individual, family, and corporate body names.



This is an iterative process!

Looking Ahead

CARIBBEAN NEWSPAPER DIGITAL LIBRARY



HOME

ADVANCED SEARCH

NEWSPAPER SEARCH

MAP SEARCH

VIEW ITEMS

SUBCOLLECTIONS



PRINT



SEND



ADD



SHARE

Search Collection:

Go

☐ Include the document text in search

The **Caribbean Newspaper Digital Collection** is within the [Digital Library of the Caribbean \(dLOC\)](#), a cooperative digital library for newspapers resources from the Caribbean and circum-Caribbean. *dLOC* provides access to digitized versions of Caribbean newspapers, gazettes, and other research materials on newsprint currently held in archives, libraries, and private collections. *dLOC* continues to add historical and current newspapers through ongoing digitization and born-digital curation, expanding the geographic, temporal, political and linguistic variety of the newspapers.

Please see the [invitation for dLOC partners](#) or [contact us](#) to inquire about this project.

To grant permissions to add your newspaper, please [contact us](#) or use the [permissions form here](#).



<http://ufdc.ufl.edu/cndl>

Long-Range Plans

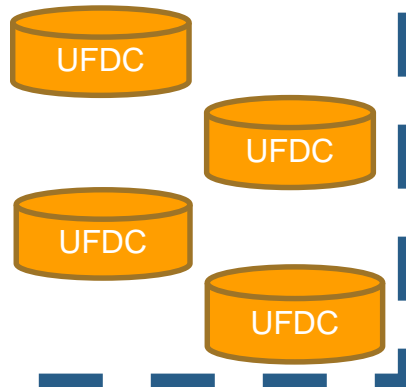
- Apply process to all new collections.
- Retrospectively enhance existing collections.
- Utilize software to clean up and batch correct large numbers of records.
- Develop new workflows to create more efficient use of staff time.

Data Input Panel

File	Get Record	Search Set	
Publication Type	Doc Type	Title	Title Sub Title
BOOKPROC	BK	Phase II Evaluation Findings	The Segmental Concrete C
Author Info		MAI Author Specialty	Corp Author
XXXXXX		Bridge engineering	Civil Engineering Research Foundation, ASCE
		Bridge construction	MAI
		Construction management	

XML CMS System

XML Records Exported to UFDC



Repository of Updated Records

MARC Records Exported from XML



MARC Records Exported to UF Libraries OPAC/Discovery Service

OPAC

Thank You

CONTACT INFORMATION

Dave Van Kleeck
Chair
Cataloging & Discovery Services
George A. Smathers Libraries
University of Florida
352-273-2863
dvankleeck@ufl.edu
orcid.org/0000-0001-5649-321X

Chelsea Dinsmore
Chair
Digital Support Services
George A. Smathers Libraries
University of Florida
352-273-0369
chelseaz@ufl.edu
orcid.org/0000-0002-1349-2572