

5-2015

Linking Old Librarianship to New: Aligning 5-Steps of The Innovator's DNA in Creating Thematic Discovery Systems for the Everglades

L. Bryan Cooper

Florida International University, lbcooper@fiu.edu

Margarita Perez Martinez

Florida International University, perema@fiu.edu

Follow this and additional works at: <http://digitalcommons.fiu.edu/glworks>

 Part of the [Agricultural Science Commons](#), [Aquaculture and Fisheries Commons](#), [Archival Science Commons](#), [Biochemistry Commons](#), [Biodiversity Commons](#), [Botany Commons](#), [Cataloging and Metadata Commons](#), [Collection Development and Management Commons](#), [Environmental Policy Commons](#), [Environmental Studies Commons](#), [Marine Biology Commons](#), [Online and Distance Education Commons](#), [Ornithology Commons](#), [Other Biochemistry, Biophysics, and Structural Biology Commons](#), [Plant Biology Commons](#), [Public Policy Commons](#), [Recreation, Parks and Tourism Administration Commons](#), [Scholarly Communication Commons](#), [Systems Biology Commons](#), [Technology and Innovation Commons](#), [Terrestrial and Aquatic Ecology Commons](#), and the [Zoology Commons](#)

Recommended Citation

Cooper, L. Bryan and Perez Martinez, Margarita, "Linking Old Librarianship to New: Aligning 5-Steps of The Innovator's DNA in Creating Thematic Discovery Systems for the Everglades" (2015). *Works of the FIU Libraries*. Paper 24.
<http://digitalcommons.fiu.edu/glworks/24>

This work is brought to you for free and open access by the FIU Libraries at FIU Digital Commons. It has been accepted for inclusion in Works of the FIU Libraries by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

L. Bryan Cooper lbcooper@fiu.edu - Margarita Perez perema@fiu.edu

Introduction / Thesis

Librarianship in the pre-digital era relied heavily on creation and publication of subject-based bibliographies. Librarians were regularly promoted based on such practice. Today, comparable bibliographic and curatorial work has value if applied to metadata selection for new databases and search systems – as national data trends reveal insatiable demand for access to the exponential growth of information.

LibQual's Martha Kyrillidou writes: “No library can ever have sufficient information content that would come close to satisfying this appetite.” And “...our discovery tools are not quite maximizing the value libraries can deliver. There is a lot of room for improvement in this area!”

At Florida International University Libraries, teams tested and implemented: **1)** Curation of metadata for “discovery” in an open-source next-gen catalog system (eXtensible Catalog); **2)** Curation of metadata and associated pdfs and media through Archive-it and **3)** Selection of domains and directories for targeted “thin-metadata” searching using a Google API for Content Management Systems.

The philosophy and practice of FIU's experimentation parallels five concepts found within the Innovator's DNA (2011). **Associating, Questioning, Observing, Networking and Experimenting** help frame how librarians can use new tools to add value to information by collecting metadata and born-digital material in ways comparable to past practice.

I-DNA Concepts / Examples

ASSOCIATING (Analogies – Metaphors -- Relationships)

- Pre-electronic bibliographies and collection development ~ OCLC WorldCat grows as biblio-collection building tool ~ Growth of OCLC and non-MARC schemas linking to resources renders metadata as source of curation for digital material (vendor, open access, archival, state, local, scientific and other agencies/organizations).
- Manual transcription of cards in catalogs ~ OCLC institutional sharing of records ~ Growth of utilities (e.g. MARC-edit) for download, cross-walk and upload of diverse metadata with URLs ~ “Hubs” harmonize and consolidate open metadata for discovery; e.g. Digital Public Library of America (DPLA).
- Typewritten bibliographies from card/print union catalogs ~ “Word processed” bibliographies harvested from online catalogs via Endnote, Procite, etc. ~ Bulk metadata selection/syndication increasingly possible to/from world scale systems, with thematic collocation via z39.50; OAI-PMH, etc.
- Early value added bibliographic organization in cooperatives for individual scholars (LC Union Catalogs, OCLC) ~ later large-scale metadata aggregation (e.g. DPLA; Trove), mined by public ~ Targeted Digital Humanities data extraction w/Hathi-Google big corpus book data, British Library records, etc. ~ Business “big data” aggregation and resultant mining for corporate research/decision making.
- Collection development and catalog growth ~ Digital library growth with curation of print to digital ~ Larger curated aggregates (DPLA, Trove, Hathi, Internet Archive) ~ “Reverse flow” curated smaller thematic “citizen” collections (DPLA “exhibits;” Trove “lists;” Hathi “collections;” Internet Archive).

QUESTIONING:

- Digital growth leads to a proliferation of metadata schemas and taxonomies, based on community requirements. To what extent can/should diverse metadata sets, increasingly diversified in collection silos, be re-purposed, harmonized and centralized for greater end-user discovery efficiencies?
- How does/can this work align with **Ranganathan's Laws of Library Science (1931)?**
- How and to what extent do librarians seize upon software developments to apply traditional decision making to curate relevant and useful yet dispersed thematic metadata and digital information, with the end-user in mind?

OBSERVING: & NETWORKING

- Observing:** Vis-à-vis Internet:
 - Libraries implement both commercial and open-source discovery systems and can play role in managing metadata sets to enhance content and access.
 - Growing plethora of cross-walk and bibliographic utilities can support metadata collection, migration and syndication.
 - Various large and smaller scale aggregates allow sub-collections to be built (Hathi; DPLA; OMEKA, Archive-it, etc.)
 - Google allows thematic search through API (free/charge).
 - Commercial cross-walking software – Mapforce, etc. are growing to meet cross-walking needs to make metadata more fluid.
 - Etc.
- Networking:** Bridging gaps between social/professional networks can help libraries innovate to end-user need. **I-DNA concept:** Professionals who network outside their rank in the administrative chain and outside their immediate industry or profession develop needed innovative ideas/solutions because of the broader networks.

FIU reaches out to diverse partners to assess metadata and data issues and needs.



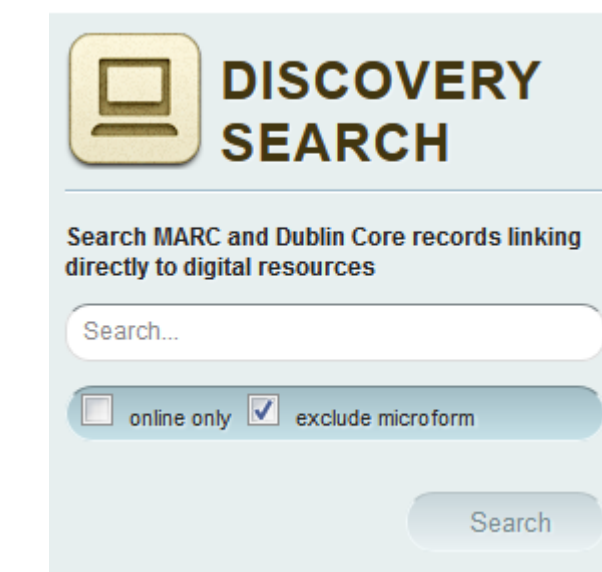
- Initial organizational meetings; surveys; outreach: Local scientists; South Florida Water Management District (SFWMD); United States Geological Service (USGS); South Florida Natural Resources Center (SFNRC); Everglades National Park Museum; Long-Term Ecological Restoration Network (LTER).
- Outreach expands to Hathi; DPLA (re: JSON-LD); Miami-Dade Public Library (Romer MARC records).
- MARCEdit software developer.
- eXtensible Catalog open source project manager
- MASmedios XC software consultants.

Question 1: FIU and partners identified, through survey, issues related to data and metadata access and use. *To what extent can FIU assist with the following identified needs, and how efficiently and effectively?*

- Aggregate and make available partner and other metadata** to increase access and save time of the searcher (partner metadata is sometimes hidden behind systems operating behind a firewall; no holdings are set with OCLC, etc.)
- Promote greater awareness** of and access to select partner collections in silos; also material hidden in CMS.
- Link government reports and original related datasets**, ideally in one system.
- Support diverse Everglades end-users** – scientific, educational (upper and lower), public policy, historical, and environmental advocacy communities as top targets/users.
- Improve public access;** support public information offices and staff.
- Provide assistance to make items more discoverable** from each partner, including through digitization; metadata sharing.
- Consolidate Local, State and Federal documents and scientific reports**, including historical photographs; metadata for datasets, etc.

Question 2: To what extent do outcomes reflect **Ranganathan's Laws** (updated), including: 1st Law *Resources are for use* – balancing modern mechanisms for housing, preservation and improving access; 2nd and 3rd “Long Tail” related Laws -- *Every user his/her record or document; and every record or document its user*; the 4th Law -- *save the time of the user;* and the 5th Law of *organic adaptation* (to change in the technological era).

EXPERIMENTING:



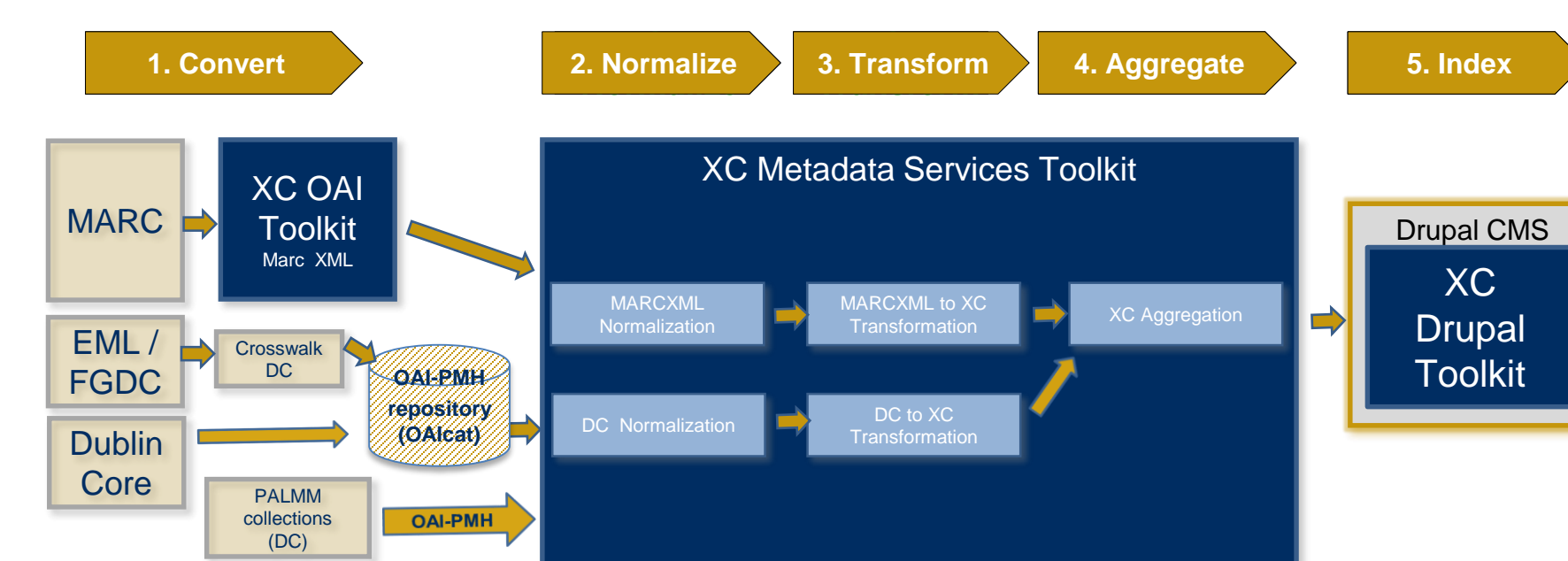
Goal: Harmonize MARC, DC, FGDC, EML schemas into open source XML schema (eXtensible), using Metadata Services Toolkit (MST); MarCEdit; and XSLT cross-walks. Test time commitment and document eXtensible Catalog ability to serve as ongoing Discovery interface.

- Partially “hidden” or un-syndicated metadata identified for aggregation (ENP, MDPL Romer Collection; GIS Center; FCE-LTER). Also, GPO/MarCive records with URLs from the ILS.
- To help *“save the time of the user”* a total of 4,493 records from seven unique silos, plus one cross-digital Florida library collection (PALMM), are aggregated in one XML database with SOLR faceted search.
- Metadata pointing to scientific data sets is collocated alongside historical material that could help scientists establish base-line temporal data.

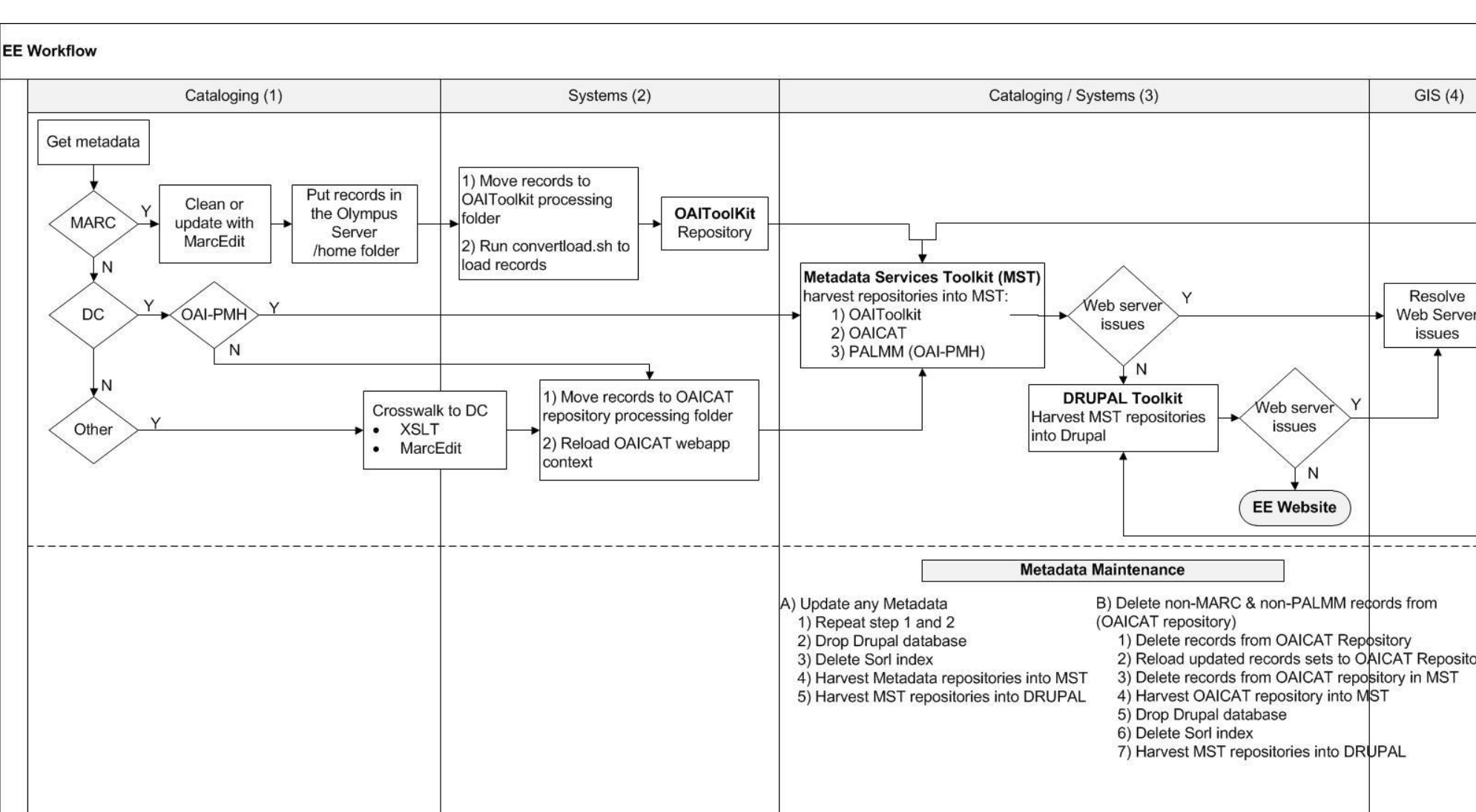
Step 1: Acquiring Metadata

Partner Institution	Num. Rec.	Metadata Schema
Everglades National Park	5	MARC
HathiTrust	85	MARC
SUS Libraries (ILS)	402	MARC
Miami Dade Public Library	164	MARC
FIU GIS Center	12	EML
Florida Coastal Everglades Long Term Ecological Research (FCE LTER)	125	EML
South Florida Information Access (SOFIA)	250	FGDC
PALMM Collections	3,450	Dublin Core
Total	4,493	XC

Step 2: eXtensible Catalog Metadata Management



Step 3: EE Workflow



Step 4 : EE Time Study

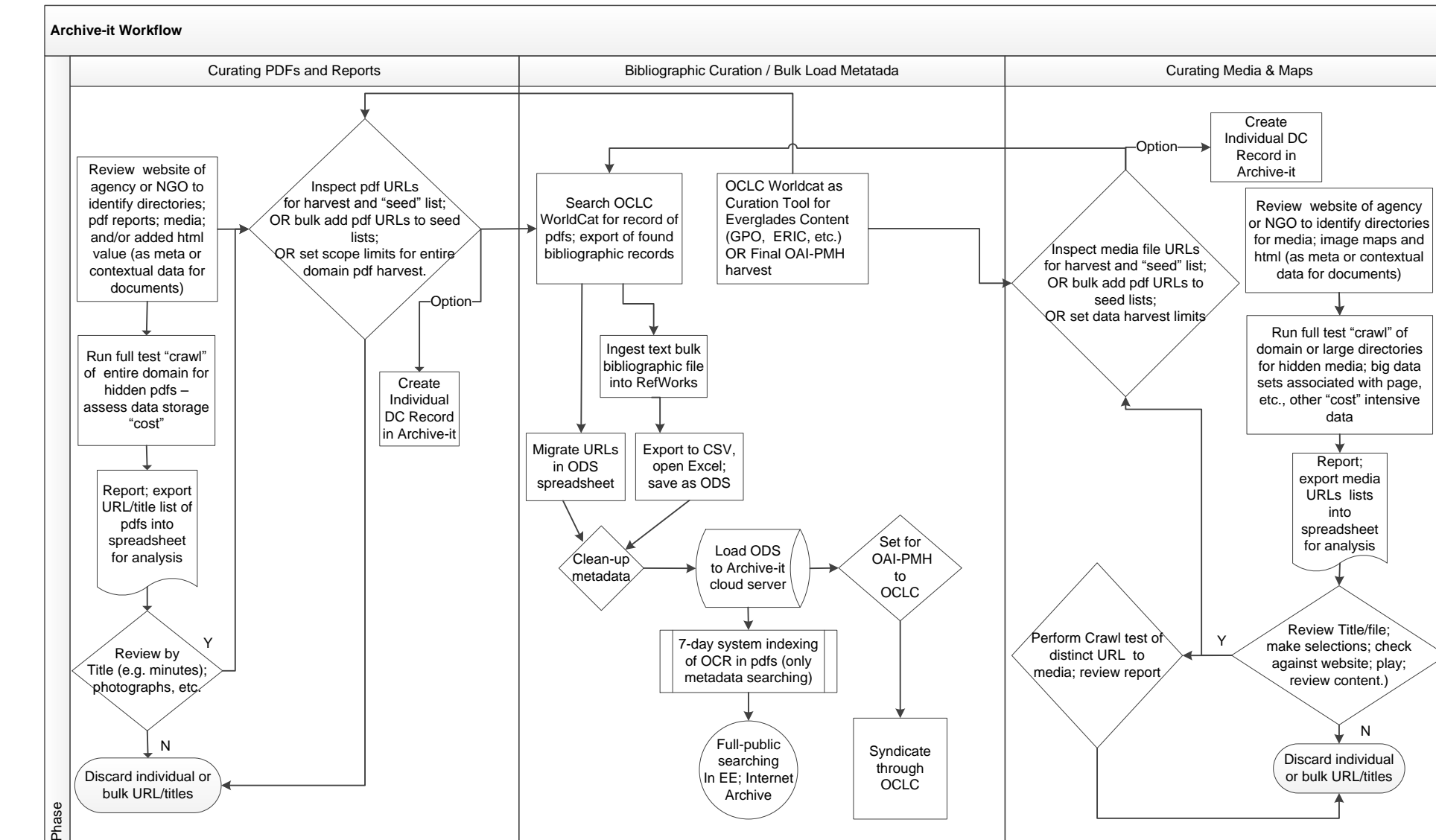
Time Management Metadata maintenance	Resource	Hours
Metadata		
Acquired Metadata	Cataloging	4
Metadata Cleanup	Cataloging	8
Crosswalk to DC using MarCEdit	Cataloging	4
Crosswalk to DC using XSLT	Systems	2
Metadata Services Toolkit (MST)		
Delete and reload OAI Toolkit server	Cataloging/Systems	.5
Delete and reload OAI-PMH server	Cataloging/Systems	.5
Harvest OAI Toolkit, OAI Cat, PALMM repositories into MST	Cataloging	3
Troubleshooting	Cataloging/Systems	4
Drupal Toolkit:		
Drop Drupal database	Systems	.5
Delete SOLR index	Systems	.5
Harvest MST repositories into Drupal Toolkit	Cataloging	1
Total		28



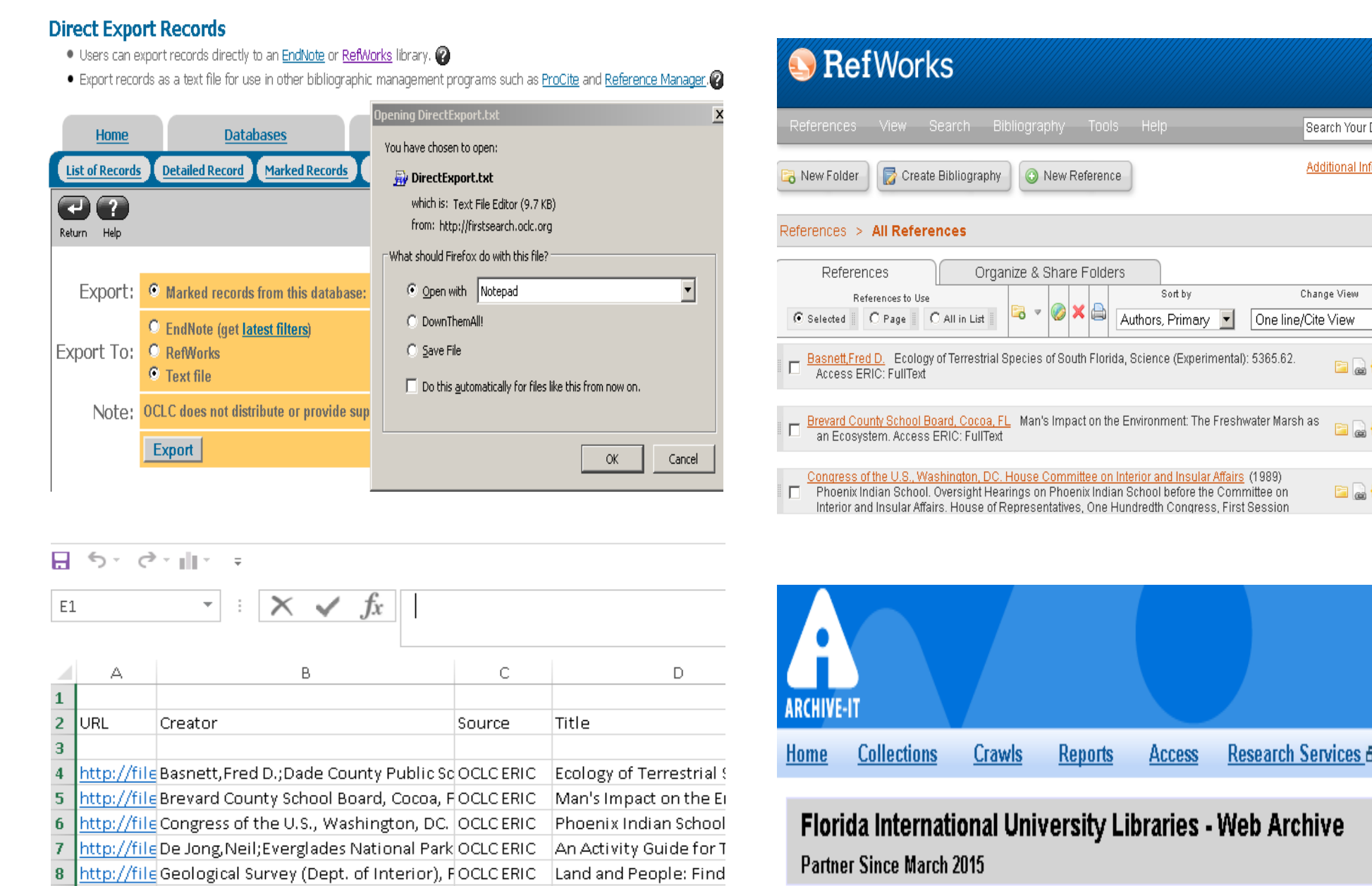
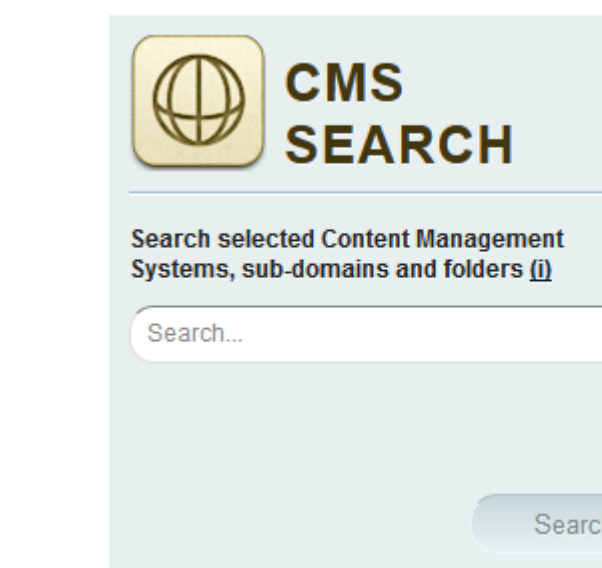
Goal: Archive-it test curation and bulk ingest of government and non-government material in pdf and other formats; OCR indexing and search; develop bulk metadata work flow using OCLC export; RefWorks, CSV file ingest from ODS (Open Document Spreadsheets).

Test & Production: Archive-it Collections & Workflow

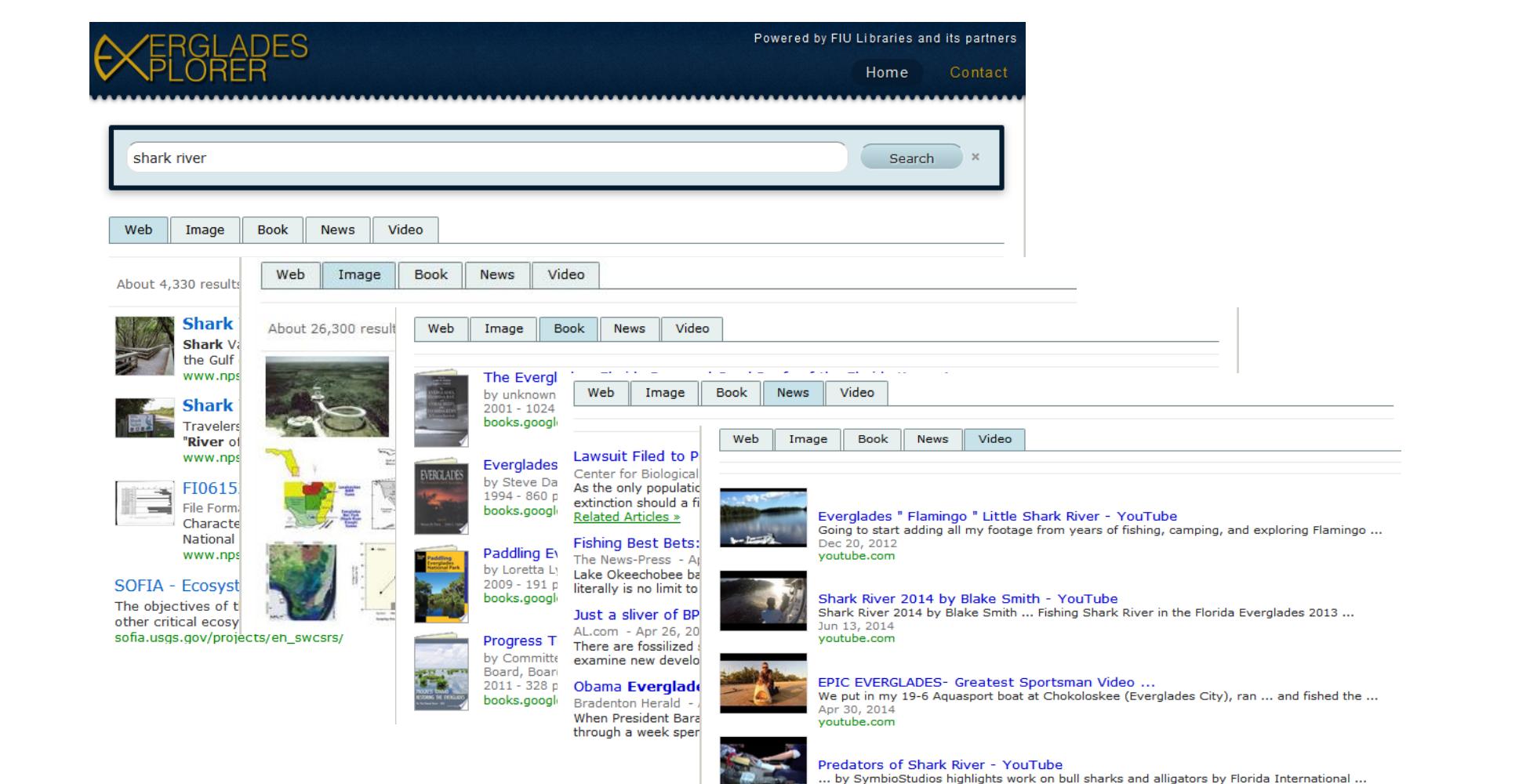
- Everglades Explorer -- **EAPRA (Assorted PDF & Report Archive)**
- Everglades Explorer -- **EMAP (Everglades Map Archive)**
- Everglades Explorer -- **EMEDIA (Everglades Media Archive)**



Utility Ecosystem: WorldCat – RefWorks – Excel/ODS – Archive-it

Goal: Test Google API to determine effectiveness of full domain and directory-level search scoping; apply curatorial principles to effectiveness of Google API to drill down into CMS systems.



Target CMS Search across select agency and partner sites for html, images, books, news, video, etc., using Google API

- Florida Coastal Everglades (FCE) LTERNET - <http://fce.lternet.edu/>
- South Florida Water Management District (SFWMD) Home Page - <http://www.sfwmd.gov/portal/page/portal/sfwmdmain/home%20page>
- SFWMD America's Everglades project site - <http://www.sfwmd.gov/portal/page/portal/xweb%20protecting%20and%20restoring/americas%20everglades/>
- Library of Congress Reclaiming the Everglades - <http://memory.loc.gov/ammem/collections/everglades/>
- Everglades Digital Library - <http://everglades.fiu.edu>
- National Park Service - <http://www.nps.gov/ever/index.htm>
- Everglades Foundation - <http://www.evergladesfoundation.org/>
- South Florida Information Access (SOFIA) <http://sofia.usgs.gov/>
- Odyssey Earth - <http://www.odysseyearth.com/>
- Eyes on the Rise - <http://www.eyesontherise.org/>
- Florida Keys National Marine Sanctuary - <http://www.floridakeys.noaa.gov>
- Southeast Environmental Research Center (SERC) Water Quality Monitoring Network - <http://serc.fiu.edu/wqmnetwork/>
- The Everglades Foundation Channel - <https://www.youtube.com/user/EvergladesFdn>
- Everglades NPS's channel - <https://www.youtube.com/user/EvergladesNPS>

Results:

- At launch, **Everglades Explorer (EE)** provides access to over 10,000 aggregated unique Everglades related records linking directly to digital resources.
- First meta-harvests into **Everglades Explorer** netted 4,493 SOLR indexed discovery records/digital resources; 6,492 pdfs harvested and OCR key word indexed; and targeted CMS search narrowed to 226,000 “Everglades” kw results, in contrast with 16,400,000 mixed commercial and non-academic results through traditional Google search.
- Curatorial and collection development philosophies begin to be applied to metadata, with growing cross-walking skills (for non-MARC schemas into XML) supporting enhanced discovery by end-users, through EE portal and future possible syndication to State discovery system, Digital Public Library of America, etc.
- Outreach and partnering with federally funded Florida Coastal Everglades LTER brought new digital peer reviewed publications to the FIU Institutional Repository.
- Project partnership with SFWMD lead to FIU Institutional Repository and Everglades Digital Library receiving legacy scanned documents, and metadata for EE.
- MarCEdit EML-MARC crosswalk joint venture started with external developer.
- Staff knowledge and manual data transfer ability grows, strengthening capacity for providing quality discovery.

Conclusions / Questions

- Examples of local material being pulled or lost by agencies and regional CMS's supports need for Archive-it.
- Google API not as efficient at drilling down and locating specific publications as is eXtensible or Archive-it; but API is useful for broader searching across more localized networks.
- Metadata cross-walking and syndication from multiple silos to one provides elevated value from legacy records centralized for end-user, following *Ranganathan's 4th Law – Save time of the user.*
- With ascendant discovery systems, and plans for DPLA Hubs to funnel metadata from states, there is growing importance to efficient metadata quality management, clean up and syndication of legacy and non-MARC metadata, paralleling *Ranganathan's 3rd and 4th Laws.*
- eXtensible Catalog appears to have scaling limits based on MARC emphasis, and need to overwrite entire production server when updating/adding non-MARC metadata from partners using DC load.
- Numerous other Everglades record sets are scattered in silos across institutions and government agencies, leading to potential to expand metadata ingest and significantly grow EE for the Everglades community.
- Archive-it's ability to target and broadly harvest pdfs has great potential to save time and labor, by pulling from scattered internet documents and metadata; its OCR indexing provides fast effective production and improved public access in advance of any additional original or bulk metadata.
- Further studies can compare time to curate legacy or multi-schema metadata, with traditional cost to process physical or purchased items.

Acknowledgements

- The project would not be done without our co-workers' contributions.
- Elaine Dong – Cataloging
 - George Fray – Library Systems
 - Jennifer Fu – Geographic Information Systems (GIS)
 - Daniel Gann – GIS
 - Julian Gottlieb – GIS
 - Wing Ho – Library Systems
 - Daniel Kirsch – GIS
 - Jill Krefft – Institutional Repository
 - Dan McGilliguddy – GIS
 - Jamie Rogers – Digital Collections Center
 - JingYing (Crystal) Zhou – Library Systems

Reference

Dyer, J., Gregersen, H. B., & Christensen, C. M. (2011). *The innovator's DNA: Mastering the five skills of disruptive innovators*. Boston, Mass: Harvard Business Press.