

# Integrating ORCID into SobekCM at the University of Florida

Valrie Minson (PI), Mark Sullivan (CO-PI), Laurie Taylor (CO-PI)

## Integration Plan and System Information

The George A. Smathers Libraries at the University of Florida (the Libraries) respectfully request \$19,995 in Sloan Foundation support to integrate ORCID identifiers (iDs) into the Open Source SobekCM repository software (<https://code.google.com/p/sobekcm/>), developed at the University of Florida (UF) and now used by institutions around the world.

The University of Florida (UF) was established in 1853 and is a comprehensive, research university and among the nation's most academically diverse public institutions. UF is one of only 17 public, land-grant institutions in the Association of American Universities<sup>1</sup> and hosts 16 colleges and more than 200 centers, bureaus, and institutes. UF faculty numbers more than 4,000. The library includes 96 faculty members who provide research services to the colleges and departments.

### About SobekCM

The SobekCM Open Source software (<http://ufdc.ufl.edu/sobekcm>) was created in 2006 by software developer Mark Sullivan and powers digital collections and libraries which present and maintain scholarly, library, archival, and museum materials. SobekCM was initially developed to meet the needs of the State University Libraries. Today, SobekCM supports all UF Digital Collections (UFDC), including the UF Institutional Repository at UF (IR@UF), and more than 500 digital libraries and collections owned by partnering institutions.

The UFDC includes 112,017 titles which contain 342,670 total items and over eight million pages of material in the digital collections. Collectively these materials feature the work of more than 5,000 scholars linked to thousands of materials and, since January 2013, have been viewed more than 40,750,941 times. The IR@UF, a collection containing the intellectual output of UF researchers, includes more than 23,000 titles with more than 1,967,000 pages of material. Since January 2013, these items have been viewed more than 590,000 times.<sup>2</sup>

In addition to detailed collection usage, SobekCM supports a robust suite of functionality tailored to meet a broad audience of users including librarians, curators, and scholars. Features include:

- support for all file formats,
- different types of user levels and permissions,
- online submission capabilities,

---

<sup>1</sup> <http://www.ufl.edu/about-uf/>

<sup>2</sup> <http://ufdc.ufl.edu/statistics/usage/ufir>

- metadata editing (Dublin Core, MODS, MARC, DarwinCore, GML, METS, and more),
- integrated tracking and reporting for workflow management,
- unique serial display and browsing,
- quarterly usage statistics delivered via automated emails to registered contributors,
- Google Map based geographic search and display, and
- support for multiple user-interface languages, including French and Spanish.

UF is committed to supporting and enhancing SobekCM as a versatile, scalable platform utilized by a wide range of institutions. The project team seeks to leverage the existing SobekCM architecture to improve common authorship needs and improve data exchange across institutions.

### **Integrating ORCID iDs**

The SobekCM software currently enables users and contributors with the capacity to edit basic information about submitted material, including primary author(s) and other contributors. UF-affiliated authors are assigned standard IDs which are provided to all faculty, personnel and students. SobekCM provides authenticated access to this information through UF's Shibboleth<sup>3</sup> and to other institutions where integration with Shibboleth is available. SobekCM also supports new registered users from both within UF and from affiliated scholars approved for submitting materials to the IR@UF.

The project team plans to integrate a workflow into SobekCM's submission process. This will extend functionality by prompting all existing and new UF contributors to create ORCID iDs as well as to populate their ORCID profile with publications, grants, and other research output. SobekCM supports multiple methods for automated discovery, search, and harvest of data including OAI-PMH, XML returns for all searches and browses with pointers, and the existing JSON API.<sup>3</sup> Additionally, SobekCM utilizes microservices, enabling ease of future extension and integration. By integrating ORCID iDs with SobekCM, ORCID resources and data also will be available through SobekCM, for direct use by the faculty in general, or for consumption into additional university resources.

## **Outreach**

To support campus-wide adoption, UF Librarians will conduct outreach, provide training, and support promotion and availability of ORCID iDs to selected UF and external contributor groups. For example, UF Librarians representing all disciplines at UF will hold workshops to help scholars simultaneously create new SobekCM accounts and ORCID iDs. Librarians also will conduct surveys and usability assessments for the new interfaces to assess this process, gather feedback and enhance the system. Plans for these activities are detailed in the "Rationale" section.

---

<sup>3</sup> For more on UF identity management, and future updates which would all be integrated with SobekCM, see: <http://identity.it.ufl.edu/process/technologies/shibboleth/>

## Stakeholders

Integrating ORCID iDs into SobekCM is of interest to many UF stakeholders and external adopters. Scholars from each of the 16 colleges, and their respective scholarly output, are represented in the institutional repository [IR@UF] along with other SobekCM partner repositories. Initially, groups selected for outreach and promotion during the grant period will include UF's Institute of Food and Agricultural Sciences (IFAS) with more than 900 faculty in 16 academic departments, including the School of Forest Resources and Conservation and the School of Natural Resources and Environment, 12 Research & Education Centers, and 68 County Extension locations throughout Florida.

Other campus stakeholders and future beneficiaries include the Division of Sponsored Programs (DSP) for grant related materials, Office of Institutional Planning and Research for providing analytic services, and units actively using enterprise systems such as VIVO. The project team also will provide outreach to campus entities in utilizing ORCID iDs for further data integration.

Beyond benefiting UF stakeholders, the team will integrate ORCID iDs to support the Digital Library of the Caribbean (dLOC; <http://www.dloc.com/>). dLOC is a cooperative digital library of Caribbean cultural, historical, and research materials held in archives, libraries and private collections in the Caribbean and circum-Caribbean. The Libraries serve as technical leader for the 37 partner member institutions within the dLOC. UF will collaborate with scholars affiliated with dLOC from partnering institutions to provide training and support.

In addition, this integration will provide for more robust linkages between the different researchers and contributors. Consequently, this will provide a richer searching and discovery interface to our roughly six million visitors annually.

## Rationale

Author disambiguation is a foundational challenge for the global research community. Associating publications, grants, and data sets to appropriate authors or creators is universally problematic given name duplication between authors, as well as name variants that exist for a single author. ORCID is positioned to be part of the solution by providing persistent digital identifiers to distinguish one researcher from another. However, without integration of ORCID identifiers in primary data systems, the connection between researcher and research output is not accomplished.

Integration of ORCID iDs with SobekCM will benefit UF stakeholders and partners by:

- improving author disambiguation challenges
- enhancing annual faculty publication reporting processes, specifically for the Institute of Food and Agricultural Sciences (IFAS), which is a partnership between UF, Florida A&M University, and Florida Cooperative Extension Service (including federal, state, and county governments)
- supporting researcher profiling and assessment systems (e.g., VIVO, Academic Analytics)
- positioning UF for additional progress in author disambiguation as ORCID iDs are nationally or internationally adopted

## Use cases:

**ORCID creation for authors:** A chemistry professor wants to contribute an open access article to the IR@UF and does not have an ORCID iD. The researcher understands the role of an identifier in demonstrating his research impact and logs into the IR@UF and the interface allows him/her to easily create an ORCID iD.

**ORCID creation for co-authors:** An author submits a journal article to the IR@UF, listing four co-authors. Based on the citation information, SobekCM with the ORCID API finds possible ORCID iD matches. IR@UF prompts the author to verify the coauthors based on the matches found through ORCID iDs.

**Associate publications, data sets, presentations, poster sessions, audio, video, and other scholarly and creative works:** An agricultural science researcher, when listed as a PI or lead author, submits authored papers, training videos, data sets, conference posters, reports, grant proposals, and other scholarly works to the IR@UF. The researcher would like to locate all materials in the IR@UF where he/she was a collaborator, and to display the collaborative and single author works together. The researcher logs in, and locates and connects a full list of materials with his/her ORCID identifier.

**Support institutional data efforts:** A Data Curation Coordinator for a campus research center wants to include ORCID identifiers in a locally-managed profiling system and is already using the SobekCM API for UF authored publication data. The coordinator updates the query for the SobekCM API to include available ORCID iDs.

**Leverage ORCID identifiers for supporting external data sources, and disambiguation:** An Analyst in the Office of Institutional Planning and Research is developing a new reporting methodology to enhance faculty publication tracking and reporting on all authored publications already in the IR@UF as well as publications held in other publisher database. The Analyst uses the ORCID iD to disambiguate publications as needed.

**Reduce author disambiguation issues:** A librarian would like to locate all the research on a professor who has published under several names (e.g., John Smith, J. John Smith, J. Smith, Jorge J. Smith, etc.) and locates the author and ORCID iD in the IR@UF and then uses the ORCID iD to find additional research materials by this author.

**Create unit publication lists, for reporting (and more):** The Dean of Research for the Institute of Food and Agricultural Sciences hopes to create a list of all faculty papers within his 28 departments for publication reporting. His/her Reporting Coordinator uses the list of ORCID iDs to retrieve the list of publications.

**Impact assessment:** A PhD candidate in the Classics department is working to develop his/her professional portfolio. The candidate submits a poster presented at a recent conference to the IR@UF and creates an ORCID iD in the process. Later, he/she receives usage statistics reports from the IR@UF, and uses these along with his ORCID iD profile page to show his research impact to prospective employers.

**Library outreach:** A lab assistant submits a data set to the IR@UF and is prompted to input an ORCID iD. The assistant reads through the information on “Why an ORCID identifier, and other ways to increase your impact.” He/she creates an iD and contacts a librarian to give a workshop on methods for assessing research impact.

## Work Plan

The project team recognizes the need for full socio-technical support (people, policies, procedures, and technologies) to achieve successful integration. Thus, the work plan includes elements of technological development, outreach, training, and other types of support. The steps outlined in the work plan below are not listed chronologically.

- 1) Create and post position descriptions to hire a programmer and an outreach/support service provider
- 2) Develop programming to integrate the ORCID API into SobekCM to directly query and post to the ORCID API for:
  - a. Creating new ORCID iDs, and
  - b. Searching by name or ORCID iD for authors and co-authors
- 3) Utilize the ORCID API to:
  - a. Create ORCID iDs for all existing UF authors who have registered in SobekCM
  - b. Create an interface for new registering faculty to obtain ORCID iDs
- 4) Conduct usability assessments on the new interface
- 5) Extend SobekCM interface with ORCID iDs to support disambiguation of authors and co-authors:
  - a. Create ability to enter the ORCID iD for co-authors, directly in online system
  - b. Develop user prompts querying the user to verify when co-author matches are found
- 6) Promote and perform outreach on availability and benefits of ORCID iDs:
  - a. Develop support materials for outreach, including scripts for librarians
  - b. Provide training to librarians
  - c. Target outreach and promotion to UF faculty, postdoctoral researchers, clinicians, graduate students, and other affiliated scholars in selected groups
  - d. Target outreach and promotion to Caribbean scholars, both at UF and other institutions, who are approved for submission to dLOC
  - e. Share all outreach materials as Open Access
- 7) Share ORCID and Digital Object Identifiers/PURLs:
  - a. Share ORCID and Digital Object Identifiers/PURLs through SobekCM, making the data available for automatic discovery, searching, and harvesting in an open manner

## Timeline

	Deliverables	Timeline
1	<p>Create and post job notices for two student employees: programmer and outreach service provider</p> <p>Initiate internal outreach by training librarians to liaise and support UF scholars</p>	Month 1
2	<p>Hire student employees and begin training</p> <p>Develop basic interaction with the ORCID API, allowing creation of ORCID iDs and searching by name, ORCID, etc.</p> <p>Finalize web pages for outreach and training</p>	Month 2
3	<p>Integrate programming and ORCID into SobekCM</p> <p>Create the interface for faculty to obtain ORCID iDs when registering with SobekCM</p> <p>Release source code as available</p> <p>Test new programming interfaces for usability</p> <p>Develop training and promotional materials for librarians conducting outreach</p> <p>Demonstrate prototype during a webinar</p>	Months 3-6
6	<p>Create ORCID iDs for UF faculty currently registered in SobekCM who have given permission</p> <p>Conduct broad outreach and promotion for UF and dLOC stakeholders</p> <p>Communicate with researchers about the value of the new iDs</p>	Month 7
7	<p>Host workshops to create ORCID iDs using the <i>IR@UF</i> for each of the UF researcher groups: IFAS, Health Science Center, College of Engineering, Center for Undergraduate Research, and others to be confirmed</p> <p>Provide training and support to dLOC partner institutions</p> <p>Refine integration of ORCID API based on feedback</p> <p>Ensure all outreach, promotion, and technical information is documented and available in the <i>IR@UF</i> for others to use</p> <p>Serve as reference for organizations interested in ORCID integration</p>	Months 8-12

## Budget

Funding for ORCID membership fees has been approved by the Dean of University Libraries, Judith C. Russell and therefore is not reflected in the budget request detailed below.

<b>Expense Categories</b>	<b>Details</b>	<b>Total (including fringe)</b>
Student Programmer	\$15/hour for 20 hours/week for 50 weeks	\$15,690
Student Outreach Service Provider	\$12/hour for 343 hours	\$4,305
Total request		\$19,995

## Bibliography

Additional resources listed and linked in the bibliography for this proposal:

<http://ufdc.ufl.edu/AA00016480/00001/pdf>

# Workflow

