



ENHANCED INSTITUTIONAL REPOSITORY INDEXING, LINKING AND DISPLAY: FACILITATING USER ACCESS AND PUBLICLY FUNDED RESEARCH COMPLIANCE

Judith Russell - Dean of University Libraries, University of Florida
Todd Digby - Chair, Library IT, University of Florida

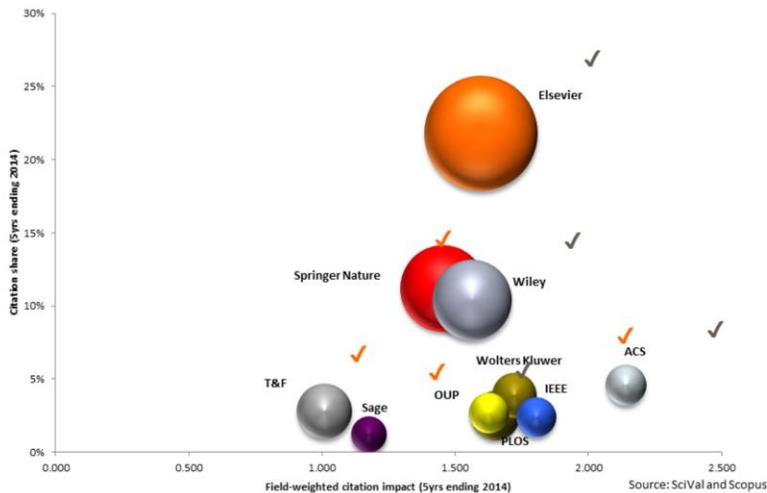
UF George A. Smathers
Libraries
UNIVERSITY of FLORIDA

Todd and I are going to present two related pilot projects, one with Elsevier and the other with Chorus, that demonstrate one way academic libraries and publishers can work together for interoperability between information systems in ways that can generate benefits for both. Before beginning these projects, I had discussions over several years with my Provost, Vice President for Research, and members of the faculty, including the Faculty Research Council, about the need to document and provide improved access to faculty research -- without placing additional burdens on our already very busy faculty. With the development of the White House Office of Science and Technology Policy (OSTP) mandates for deposit of the results of federally-funded research, these discussions began to include the word COMPLIANCE! All of these discussions eventually ended with the plea - "Why can't the library solve this problem -- without placing additional burdens on our already very busy faculty." UF authors publish approximately 8,000 scholarly journal articles a year. Unfortunately, our faculty do not have a culture of deposit of their article manuscripts in the Institutional Repository (IR@UF), although many do rely on it for deposit of other content. Obviously, it would be very expensive and time consuming for library staff to track down all of those manuscripts – and responding to our demands for those manuscripts would definitely place (can you say it with me?) additional burdens on our already very busy faculty.



When I first saw this slide, the word compliance was, as it is now, in the center, but there was one less figure on the slide. I went up to Howard Ratnor after his presentation about CHORUS and told him that the person on my campus who was losing sleep over compliance wasn't on his slide: the Vice President of Research. I acknowledged the need to focus the initial development of CHORUS on publishers and funding agencies, but pointed out that this was a three legged stool and there were three figures from academic institutions (the librarian, the VP research and the researcher) who were not participating in the design of the CHORUS system. I suggested that when they were ready, UF and other academic institutions should participate in the development of CHORUS to ensure that it met our needs as well as those of publishers and funders. A few months ago, Howard contacted me and said they were ready to expand the partnership and asked if UF would participate. I quickly agreed. This offer came about in part because of the pilot project we had already begun with Elsevier, which CHORUS members had been observing closely.

Portfolio Analysis: UF Articles by Publisher 2014



Bubble size represents number of UF-authored articles published in 2014

✓ Pilot participant and CHORUS member publisher
 ✓ CHORUS member publisher (as of January 2016)

3

The most frequent question I get asked about these projects is this one: Why Elsevier? Why Chorus? I think this slide answers that question. 8 of the 10 publishers most selected by UF authors are CHORUS members. Elsevier has the largest volume of any publisher, followed by Springer/Nature and Wiley. Automated solutions for identification and access to articles by UF authors from multiple publishers reduces the burden on our academic faculty, the library faculty and staff, and by sharing data on compliance, on the staff of the office of compliance of the VP research.

The Smathers Libraries are engaged in two projects that demonstrate collaboration between academic libraries and publishers for interoperability between information systems that benefit for both.

- **Elsevier** Pilot Project
- **CHORUS** Pilot Project

As Judy has explained, we are engaged in two related pilot projects, one with Elsevier and the other with Chorus, that demonstrate one way academic libraries and publishers can work together for interoperability between information systems in ways that can generate benefits for both.

UF/Elsevier Project - Goals and Discoveries

Phase I Goals:

- Metadata and links for ~31,000 articles by UF authors in IR@UF from 1949 - 2016
- Indexing based on metadata, including abstracts

Phase I Discoveries:

- More gold OA publishing than anticipated (601 articles by 1,443 unique UF authors between 2009 and early 2016)

Phase II Goals:

- Streaming final articles or accepted manuscripts to IR@UF
- Indexing based on metadata and full-text
- Learning from usability testing

5

We divided the Elsevier project into two distinct phases.

Phase I was focused on locating and harvesting metadata on articles that were written by University of Florida authors.

The most difficult problem was identifying UF authors. The words University and Florida appear in the names of a number of institutions! We have long supported ORCID IDs, but we now also strongly support institutional identifiers.

The most surprising discovery was the high number of Gold Open Access articles published by UF authors in Elsevier journals. We are doing further analysis of open access publishing by these authors and expect to survey them to learn more about their motivation and source of funds for the article processing fees. We hope some of them will become champions for a campus-wide open access policy.

8 to 10 of these authors published over 20 OA articles with Elsevier in these 7 years; one published 30.

Phase II is focused on streaming final articles or accepted manuscripts to IR@UF to provide access for users with licenses, indexing based on both metadata and full-text, and learning from usability testing.

UF/Elsevier Project - Phase I: Implement the Elsevier API Infrastructure to Deliver Enhanced Repository Services

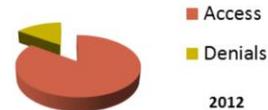
- Increase comprehensiveness of coverage of Elsevier-published content by UF authors through the **IR@UF**.
- Provide subscribers with access to the best available (published) version through the **IR@UF**.
- Integrate published articles with other **IR@UF** content.
- Allow users to return to their IR search results when reading final articles from **ScienceDirect**.

6

During Phase I we implemented the Elsevier API Infrastructure. The primary goal was to increase the comprehensiveness of coverage of Elsevier-published content by UF authors through the IR@UF and to provide subscribers at UF and other institutions with access to the best available (published) version. Because of the volume, it was necessary to automate the process through the use of the APIs for locating the UF author articles and then retrieving the metadata to integrate these published articles with other IR@UF content.

Implementation Using the IR@UF on the Sobek^{CM} Platform

- Coverage: Over 30,000 articles by UF authors from 1949 forward.
 - Metadata, abstracts, and (in Phase II) full text for indexing in the IR@UF.
 - Links to ScienceDirect for access.
- Includes both open access and subscription articles.
- Full access for users with subscriptions; alternative access for other users (Phase II).
 - 95% of ScienceDirect usage is by users with subscriptions.



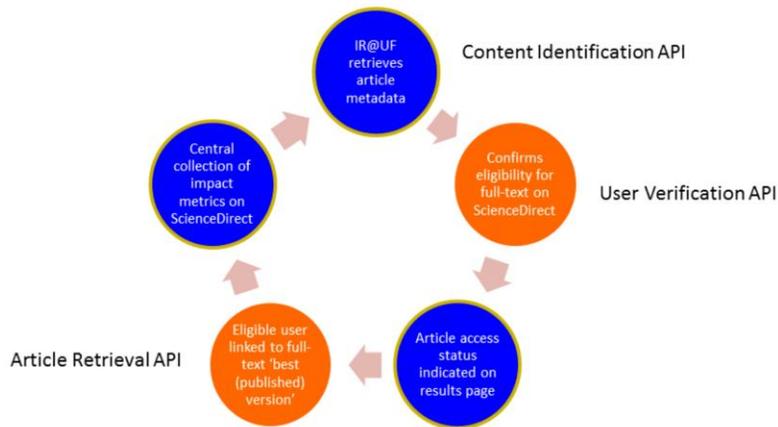
7

The Institutional Repository at the University of Florida is built upon the Sobek^{CM} platform. This open source platform was developed at the University of Florida and is also used at a limited number of other institutions. Given the custom nature of this platform, we had dedicated staff involved in implementing the API's that were provided by Elsevier free of charge.

During this phase we were able to locate over 30,000 articles by UF authors, from 1949 forward, and provide links back to ScienceDirect for access. This included both open access and subscription articles. Access to these articles was then controlled by the User Verification API, which determined if the user was located at an institution with a subscription.

The historical number of denials for non-subscribed users is relatively small, and is frequently corrected by login to a VPN server. Nevertheless, we worked with Elsevier to identify a solution, being implemented in Phase II, to provide the ability for non-subscribers to view the final manuscript.

Elsevier APIs for Institutional Repositories



Adapted from Spears, IFLA 2016

8

As I mentioned earlier, to integrate the Elsevier article infrastructure with our Sobek^{CM} - based Institutional Repository, we are making use of three Elsevier-provided APIs. These are the Content Identification API, the User Verification API and the Article Retrieval API. This diagram shows the stages of the project workflow where these APIs are used.

View of Results with Elsevier Articles from IR@UF

The screenshot displays the IR@UF search results page. The header includes the logo and navigation tabs: HOME, SEARCH OPTIONS, BRIEF VIEW, TABLE VIEW, and TABULAR VIEW. Below the header, there are search filters and a list of results. Two results are highlighted with red circles around the 'Publisher version' link:

- Result 41:** Paleoclimate of Southwestern China for the Past 50,000 yr inferred from Lake Sediment Records. The 'Publisher version' link is circled in red with the text 'You have access'.
- Result 42:** Paleoecology of shallow-marine carbonate environments, middle Eocene of Peninsular Florida. The 'Publisher version' link is circled in red with the text 'Check access'.

9

This is a screen capture of a couple of these Elsevier articles as they appear in the IR@UF. The item with Check Access circled was published in 1990. UF does not own the backfile for that journal. Consequently the Elsevier API identifies (correctly) that a user accessing the article from the UF IP range may not be entitled to access. A user from another institution with access to the backfile would see the message You have access. The API presents results that are specific to the status of the individual user.

View of Results with Elsevier Articles from IR@UF



Publisher version
Open access

Prevention of Atrial Fibrillation by Bucindolol Is Dependent on the Beta1389 Arg/Gly Adrenergic Receptor Polymorphism

Links: ([external resource](#) | [internal citation](#))

Publication Date: 2013-08-31

Creator: O'Connor, Christopher M. (author)
Fiuzat, Mona (author)
Davis, Gordon (author)
Abraham, William T. (author)
Anand, Inder S. (author)
Liggett, Stephen B. (author)
Bristow, Michael R. (author)
Section of Cardiac Electrophysiology, University of Colorado
Denver, Denver, Colorado (host institution)
Aleong, Ryan G. (author)
Sauer, William H. (author)
Murphy, Guinevere A. (author)
Port, J. David (author)

Publisher: American College of Cardiology Foundation. Published by Elsevier Inc.

Format: Pages 338-344

Subjects: Arrhythmia
genetics
Norepinephrine
Heart Failure
beta adrenergic receptors

Source Institution: American College of Cardiology Foundation. Published by Elsevier Inc.

10

UF wanted to highlight Gold Open Access articles by its authors – not just to say You have access, but to identify the articles as full open access articles, available to all users.

As noted earlier, we are currently working with Elsevier on the tagging and displaying manuscripts that are available to users without subscriptions.

UF/Elsevier, Phase II: Refine the API Infrastructure and Expand Enhanced Repository Services:

- Provide an access option for users without a subscription: viewing of the post-embargo accepted manuscripts (2013 forward).
- Offer full text searching through the IR@UF, with links to the published article on ScienceDirect.
- Usability testing by Elsevier and UF.
- Research on open access publishing by UF authors and use of Elsevier metadata for other University purposes, including compliance.

11

Phase I was completed several months ago, and we are now well along in Phase II.

During this phase we are working towards providing an access option for users without a subscription: viewing of the post-embargo accepted manuscripts (2013 forward).

Additionally we are working on enabling the full text searching of articles through the IR@UF. This will still provide results that will link to the published full-text article on ScienceDirect.

Usability testing by Elsevier and UF will continue to be conducted, and we both expect to make modifications based on what we learn.

UF will also do research on open access publishing by UF authors and the use of Elsevier metadata for other University purposes, including compliance.

Benefits of Collaboration Through Linking

Why link to the published articles on ScienceDirect?

- Maximizing research impact for articles by UF authors
- Delivering the best available (published) version on ScienceDirect
- Displaying the search results with data sets and related content in the IR@UF and the published article with links to related content on ScienceDirect
- Assuring the reliability and trustworthiness of content



12

We have identified benefits of collaboration through the linking of these articles on ScienceDirect.

It is our hope that this will maximize the research impact for articles by UF authors by providing additional visibility and access.

It will also deliver the best available (published) version on ScienceDirect, that may include data sets and other related content within the IR@UF.

This approach assures the reliability and trustworthiness of content since any corrections will automatically be provided through the linked content.

Other Benefits of Collaboration

- Collect information without burden on **UF** faculty publishing in **Elsevier** journals
- Facilitate University oversight of compliance with public access mandates
- Achieve cost savings and efficiencies for the Libraries and **UF** through automation
- Test and refine **Elsevier** APIs to provide smooth scalability of process with future academic collaborators
- Improve understanding of publisher and academic library perspectives and address constraints inherent in these roles

13

Additional Benefits of Collaboration included,

- Collecting information without burden on UF faculty publishing in Elsevier journals
- Facilitating University oversight of compliance with public access mandates
- Achieving cost savings and efficiencies for the Libraries and UF through automation
- Testing and refining Elsevier APIs to provide smooth scalability of process with future academic collaborators
- Improving understanding of publisher and academic library perspectives and address constraints inherent in these roles

Challenges of Collaboration

- Identification of **UF** authors with existing metadata
 - Need reliable identifiers like **Ringgold** and 
 - Funders and publishers should require use of standard identifiers
- Arriving at a common understanding of distinctive approaches and unique roles for content provision by publishers and academic libraries
- Adaptation of unique **IR@UF** platform (SobekCM) with **Elsevier** APIs

14

There were some challenges that were identified through this project, including the identification of UF authors with existing metadata. This process would be greatly improved if we had reliable identifiers like Ringgold and ORCID. It would be very helpful if funders and publishers required use of standard identifiers.

There were also some challenges in arriving at a common understanding of distinctive approaches and unique roles for content provision by publishers and academic libraries.

And finally the adaptation of the unique IR@UF platform (SobekCM) with Elsevier APIs required some adjustments.

Overall, the collaboration was cordial, thoughtful and effective.



Pilot Working Group - Publishers



15

As Judy mentioned earlier 8 of the 10 publishers most selected by UF authors are CHORUS members, which is one of the main reasons we are working with CHORUS in an additional joint project.

7 CHORUS members are already participating in this pilot and we expect more to join.

UF/CHORUS Pilot Project

- Perform usability testing and continuous project assessment
- Extend collaboration to other publishers through
- Extend **Ingest** process to other repositories
- Expand linking of article metadata to data in the **IR@UF** (small sets) or the **UF Research Computing** large-scale data repository (large sets)



16

The UF/CHORUS Pilot Project included extending the collaboration with to other publishers through CHORUS. This will be done by extending the ingest process to content from other publishers, which will expand the linking of article metadata to data in the IR@UF. This may further include working with UF Research Computing on large-scale data repository (large sets). But the primary value will be our ability to assist the Vice President of Research with compliance with funder open access policies.

UF/CHORUS Pilot Project Strategies

Identification

- Identify and collect metadata for articles written by institution staff, faculty and students & published by CHORUS pilot member publishers (October 2015 - January 2017, Scopus / CHORUS)
- Filter and sort by US agency funding (CHORUS)

Compliance

- Develop institutional dashboards, alerting and reporting (CHORUS)
- Collaborate to create mechanisms to track and report on compliance (CHORUS / Institution)

Engagement

- Generate interest and involvement of other stakeholders in the project (e.g., funders, other institutions and CHORUS publishers, preservation services)
- Recommendations to minimize duplication of effort and burdens on researchers
- Explore related metadata, identifier and compliance initiatives and map future interactions

17

The most important goal is facilitating compliance, but additionally the CHORUS project is meant to build upon our work with Elsevier.

Like our Elsevier project, we are working with the CHORUS publishers in identifying articles by UF authors; additionally we are checking the metadata for the funding source; verify deposit in the appropriate funder repository; and then to report to UF.

The report might indicate that Professor Smith published an article based on a DOE grant and it is in the DOE mandated repository, while Professor Jones published an article funded by NSF that is not yet in the mandated repository. A subsequent report might confirm that Professor Jones' article is now in the repository.

The office of compliance would only need to follow up on articles that were not yet deposited.

UF/CHORUS Pilot Project Aspirations

Discovery (Optional Extensions)

- **Indexing:** Harvest metadata and full text from publishers for delivery to participating institutions (Institution/CHORUS)
- **Linking:** Link from portal search results via Crossref DOIs to relevant content on publisher platforms (Institution)
- **Display:** Consider use of an API to allow display of full text within the local search results using content streamed from publisher platforms (Institution / Publisher)

18

UF/CHORUS Pilot Project also includes some farther reaching goals. These include improving indexing by harvesting metadata and full text from publishers for delivery to participating institutions (Institution/CHORUS). Linking from portal search results via Crossref DOIs to relevant content on publisher platforms. Finally, we hope to work with publisher to consider the use of an API to allow display of full text within the local search results using content streamed from publisher platforms (Institution / Publisher).

Contact Us For Additional Information:

Judy Russell
University of Florida
jcrussell@ufl.edu
@JudyRussell

Todd Digby
University of Florida
digby@ufl.edu
@todddigby

Laura I. Spears
University of Florida
laura.spears@ufl.edu
@yenscott

Chelsea Dinsmore
University of Florida
chedins@uflib.ufl.edu
@dnsmore07



Photograph by Catherine Russell

Link to UF / Elsevier search
<http://ufdc.ufl.edu/elsevier>