

ORCID – Omeka S Integration

Abstract

This grant aimed toward and integration between Open Researcher and Contributor ID (ORCID) and Omeka S. Omeka S is the next generation of the widely-used Omeka web publishing platform. Omeka S puts special emphasis on the needs of small- to medium-sized institutions and integration with other systems and Linked Open Data (LOD). ORCID provides a mechanism for reliably identifying authors and providing metadata about them via unique identifiers that are shareable on the web. ORCID data, however, is currently overwhelmingly tilted toward researchers in the sciences. This integration will encourage humanists to register an identifier with ORCID, fostering new connections between humanists' research. Omeka S would thus augment ORCID's goal of "enabl[ing] transparent and trustworthy connections between researchers, their contributions, and affiliations" within the humanities, while expanding the utility of Omeka S for users and data aggregators.

Omeka S, the Roy Rosenzweig Center for History and New Media's new version of our popular Omeka <<http://omeka.org>> web publishing platform, is designed specifically with the principles of Linked Open Data in mind. Developed for medium and larger GLAMs, Omeka S uses JavaScript Object Notation-Linked Data (JSON-LD) as its native data format, which makes it possible to enmesh Omeka S in the LOD world. Every Omeka S Resource (item, item set, media) has a URI, and these URIs can be related to internal or external URIs.

ORCID has, since 2012, provided a mechanism to reliably identify authors and provide metadata about them via unique identifiers. Their vision is "is a world where all who participate in research, scholarship, and innovation are uniquely identified and connected to their contributions across disciplines, borders, and time." Currently, over three million ORCID iDs have been registered, with connections to over twenty million associated works <<https://orcid.org/statistics>>.

Researchers create their unique ORCID iD either directly through their site or via another system that uses their API . The researchers use their iD as they publish and engage with others to connect them to institutional affiliations, grants, education, and publications. A key aspect of ORCID's philosophy is to keep ownership and permissions for their data in the hands of the researcher. Authorized data is available on ORCID's site, through connectors to other systems, and through community-produced applications.

Goals

The original grant contained the following goals.

Phase I: Import data from public ORCID iDs

- Create an Omeka S module that will create an item populated with data retrieved from an ORCID ID. ORCID IDs will be authenticated via ORCID's API.

Phase II: Dynamic display and update from public ORCID iDs

- Create an Omeka S module that will retrieve and display data from an ORCID iD in real time. It will also optionally update Omeka S records associated with an ORCID iD.

Phase III ORCID Member Integration

- Create a module that will allow ORCID members to find likely ORCID IDs that match content creators, and prompt them to authenticate or create ORCID IDs.

Progress

Due to budgetary and financial needs at RRCHNM, a director-level decision was made to compress the year-long work plan to have it end on July 31. Obviously, this spurred difficult decisions about what could be completed in that time frame. Phase III was eliminated, and Phases I and II can only be considered partially complete. However, that does not mean that important lessons were not learned, nor that (partially) functioning code is not available. The code is available at <https://github.com/omeka-s-modules/OrcidConnector>. It can serve as a place for others to pick up the work by forking the repository, submitting pull requests, or simply taking the code as examples of some of the needed tasks for integrating ORCID into other systems.

The two essential features for phases I and II were OAuth integration and the display of Linked Open Data supplied by ORCID.

OAuth Integration

ORCID's guidelines are quite specific about each individual researcher establishing their own authorization to connect their ORCID data to another system. This rightly ensures validity of data across systems. Thus, the first step was to create an authentication mechanism within Omeka S with a round-trip between an individual Omeka S installation and ORCID, allowing the user to authorize a data transfer that copies basic ORCID data into Omeka S. Since Omeka S is fairly agnostic about what an "Item", its fundamental record, represents, upon completing the authorization process via ORCID, a new Item of class FOAF (Friend Of A Friend) Person is created in Omeka S. That item can then be related to other items – presumably works and publication to which they have contributed – via common metadata properties such as Dublin Core Creator or Contributor.

One weakness of the approach of using an item to represent a researcher is that the default display label for an item is the Dublin Core Title of the item. It seemed inappropriate to copy the FOAF name data into a Dublin Core Title field, though. Thus, currently the default display label for a researcher appears as "[Untitled]". A user could add their name in the Dublin Core Title field manually, or a theme or module could likely rearrange the data. Addressing this is a known feature request more generally in Omeka S. (See "Display and Customization" in Further Work below). A possible solution is to incorporate a solution into Omeka S's Resource Template system, which allows customization for Omeka S resources of a particular type. Thus, a late-stage addition was to automatically create a resource template for researchers with their data connected to ORCID.

Using Linked Open Data

ORCID's use of Linked Open Data – and their API for accessing it – forced a number of early decisions, each with its own consequences for later design needs. ORCID's API allow data about a researcher to be accessed via different response formats, and early on different response formats did not necessarily return the same data. An API update during the time of the grant work, however, made a true and complete RDF (Resource Description Framework) response available. Since the data available is sure to be updated as researchers update their data within ORCID, it would be a losing battle to directly import and update the data directly into Omeka S. Thus, when a researcher's page is loaded, a request is made to ORCID to retrieve and display the most current data.

This approach has the advantage of displaying the most current Linked Open Data available. However, it also bears the weakness of having a general system for displaying arbitrary and changing RDF graphs. Thus, what is currently displayed is only a limited, hard-coded, set of properties. Because Omeka S has a dependency on the EasyRDF PHP library, the retrieved data is parsed there, then selected out for display. (See "Display and Customization" in Further Work below.)

Further work

Display and Customization

A display label based on FOAF name data should be provided for researcher item resources. This might be available via a resource template setting built into Omeka S, or via a future module customization.

The full range of possible RDF data retrieved from a request on an ORCID iD has not yet been explored. The current hard-coded options are fairly minimal. Expanding to everything available in the response would be the logical next step. A site-specific customization would be even better, and could be achieved via either a configuration screen or with small modules that connect to OrcidConnector's internal configuration settings, with changes to make that possible.

User Experience

Currently, after the OAuth confirmation is complete, the researcher's item resource is created, but it is unclear how to access it. A directly link should be provided – or, better, a redirect – to the resource.