Author Names Project – November 2013 Report

With the generous support of the Arcadia Fund, the Harvard Office of Scholarly Communication and the Berkman Center for Internet & Society, the Author Names project team has been working on an ambitious project that aims to have a major impact on how libraries and publishers[1] exchange data in support of discovery of new scholarly and creative content.

The project team has been developing a proof-of-concept system to enable publishers to exchange Author Names/Identity metadata with libraries. This web application, which we are calling OAQ (Online Author Questionnaire), is open source[2] and will utilize APIs wherever feasible to seek interoperability between identity registries, controlled vocabularies and library databases (e.g. ORCID, ISNI, GeoNet Names, VIAF) as a way to create linkages with existing metadata and facilitate author name disambiguation at the start of the publication process.

With the submission of this report, ending our first funding cycle, we would like to request an extension on the project to spend any unspent funds from our initial allocation as well as to request additional development support for use with Berkman to complete application development, testing and refinement.

<u>Background</u>

For those unfamiliar with publishers' business processes, sending Author Questionnaires is a common practice for publishers when bringing a new work to publication, and planning the marketing and distribution for new titles. The information collected with a publisher's Author Questionnaire includes many of the same basic data elements that libraries use when creating Name Authority Records, such as the name under which an author publishes and variant forms of name. In addition, it also includes rich information related to an author's affiliations, educational background, geographical associations, prior works and biographical information. While an Author Questionnaire supports the internal business needs of publishers, it is also a trove of valuable data of keen interest to libraries when establishing author identities, disambiguating between authors, and in linking to external identity registries.

These areas of joint interest are the foundation for the Author Names project and are the compelling rationale for libraries and publishers to develop formal data exchanges in support of resource discovery.

---

[1] The term *publishers* is used here in the broadest sense to mean any organization that is engaged in promoting new scholarly or creative output: university publishers, literary presses, scholarly societies, academic departments and research centers, etc.
[2] The application is being built using Ruby on Rails with a PostgreSQL database.

Project Partners/Supporters

Our current development partners from publishing are Tupelo Press, a small literary press, and Harvard University Press, a large academic/scholarly publisher. A medium-sized academic/scholarly publisher, MIT Press, also provided invaluable information during the proposal phase of the OAQ platform design and development. An additional library involved in the project is MIT Libraries, from whom we have received input and validation for mapping data elements from the publisher questionnaires to RDA/NACO Authority records. Amy Brand, member of the ORCID Board and Harvard Assistant Provost, had expressed support for the project and connected us with a resource person at ORCID to provide technical and conceptual information to the OAQ project team about possible intersections between the two initiatives.

Project Drivers and Goals

A key goal in the OAQ development process has been to create a tool with a low barrier to adoption by publishers and high incentives for publishers to collaborate with the OAQ project team so that the tool can be designed effectively to meet their business needs. OAQ will provide them with a customizable, flexible, secure online platform to perform the routine outreach to authors. While some publishers, like Harvard University Press, are clearly attracted to the project by the goal of improved and early discovery for their authors and their works, other publishers with fewer internal technical resources, like Tupelo Press, are interested in the convenience that an online tool would provide to their authors and the greater technical sophistication that OAQ would provide for their business process; many presses of all sizes are sending Word docs or PDFs via email for their author questionnaires.

Data export from the OAQ application is being designed to flow into two streams: 1) customized publisher data exports and reports, and 2) selected data of interest to libraries that will be mapped to automatically create RDA/NACO Authority records. Once the library-side data export is established, refined and tested for reliability, the OAQ project team would like to extend the functionality to feed directly to the national authority file and VIAF. While the current development track will allow the system to function with current name authority structures and systems, OAQ is being designed with the linked data future in mind.

Development Status of the OAQ Application

OAQ is currently midway in development with a goal to have a beta application available for testing within months if we had funding to continue development. A carefully designed database structure is in place. Publisher-side functionality and features are in active development, with library-side data export streams yet to be developed.

Promotion

As part of performing outreach to the library and publisher communities about the development of OAQ and the underlying project goals, Michelle Durocher presented the project at the Charleston Conference

on November 9, 2013. Her slides and abstract will be shared as part of the conference proceedings. Promotional materials were distributed to attendees so that they will share information with colleagues who may be interested in participating in the ongoing OAQ development or future adoption of the tool. The positive reception at this presentation, with engaged Q&A from attendees, was consistent with the type of reaction that we have been receiving when speaking to publishers and libraries over the course of the project this year.

Future steps -- national initiatives that the Author Names Project aligns well with:

1. The NACO group with the Program for Cooperative Cataloging at the Library of Congress has recently done a call for input from its members to help them move forward in these specific areas that are well-supported by the Author Names project:

   •   Engaging other sectors of the information environment—system developers, service providers, ID registries, cultural heritage institutions, etc.—in exploring the use of URIs and linked data syntax for expressing and managing identity metadata
   •   Taking a lead role in reconfiguring the relationship between library metadata and metadata drawn from other sources and in realigning expectations regarding cooperation and collaboration across sectors in the information community.
   •   Developing tools and techniques outside the LC/NACO Authority File for expressing relationships between identified entities and between relationship categories found in different systems.

I have asked for their formal endorsement, and if possible partnership, to move to the next stage of development with OCLC, VIAF and the data exchange for the national name authority file.

2. Partnership with the Cataloging in Publication program is another ideal fit for the Author Names project.   If the OAQ platform can be developed to a beta-testing phase, I would like to formally request their support and participation in recruiting publishers to test the application with us. From preliminary conversations with library and publisher colleagues, I anticipate enthusiasm from the CIP program.

We look forward to continuing to work on this ambitious project and hope to have a major impact on data exchange between these key communities working together to support resource discovery.

Respectfully submitted,

Michelle Durocher, *project lead*
Christine Fernsebner Eslao
Anita Patel, *developer*
Colin Van Alstine