

Author Names Project: sharing metadata between communities

Brief Project Description

Understanding that the future of authority control (managing Author Names in library catalogs) will be structured as linked open data, the Author Names Project will target existing metadata streams external to the library and develop a workflow tool that successfully captures author-supplied metadata of interest to publishers and libraries and integrates it within library authority control systems. The project will build a web-based application that meets the needs of libraries and publishers so that incentives for collaboration are high. As a step toward transforming how we work with partner communities, this project will move us toward working in a linked data environment where Author identifiers will uniquely identify content creators in Harvard library catalogs. This project seeks to derive benefits now and in the future by improving efficiency of our existing workflows and allocation of staff resources, while also forging relationships that will serve us in developing next-generation linked data systems between communities.

Summary Goals

The goals of the project are to collaborate with a community external to the library (publishers) to accomplish the following shared goals:

1. Develop a proof-of-concept system that leverages existing Author Names metadata for use in library systems, specifically attempting to funnel data toward creation of Name Authority Records and eliminate redundant work performed within libraries and partner communities.
2. Create connections between and participate in initiatives that will serve to generate unique identifiers for content creators, so that librarians are involved in prototyping systems to move this work into a linked open data environment and away from text-based authority control records.

Project Steps

1. Perform outreach to external and internal communities that create Author Names¹ metadata of interest to libraries. This includes publishers, university archives, peer institutions participating in the [ECIP](#) program, and communities beginning to develop author registries for assigning author/ creator unique identifiers ([ORCID](#), [ISNI](#), [UK Names Project](#), etc.). Learn how elements of their Author metadata function within their business model so that any developments from this project can be a win-win for libraries and for the project partner. [This step is completed for purposes of identifying the best use case for this project but will proceed informally during system development to ensure that what is built will have the greatest potential for expansion and interoperability with other systems of interest.]
2. Identify use cases, such as small- and medium-sized publishers, where author metadata is entered into informal systems that could benefit from a structured user interface, such as a web-based application with a form interface, that would export data for use in their local desktop systems via simple tabular text output and would also capture and export any desired output for libraries for use in establishing author names in the national Name Authority file.
3. Develop web-based application to capture Author-generated metadata for publisher and library systems, incorporating rich content that will serve to disambiguate names and establish unique author identities.

What problem will it solve for the Harvard Library?

¹ using the umbrella term “Author” very broadly to mean: writers, scholars, researchers, content-creators of all kinds

To establish new Author names in the national Name Authority file during the cataloging process, librarians currently spend time researching and/or rekeying information that already exists in publisher systems. This is a laborious process that could be eliminated if author metadata were made to flow into library systems *early and from the source* (i.e. author-created metadata from publishers.) This project will help eliminate redundant effort within libraries, repurpose existing metadata in such a way that it is openly shared between communities, and develop richer connections with a community (publishers) that is such a vital part of the flow of scholarly communication and creative content of interest to universities.

How does it fit with existing activities within and outside of Harvard?

Aiding in discovery of an author's works and differentiating them from other authors' works are central functions of a library catalog. Understanding that publishers and other entities use Author metadata for their internal and business needs, this project will be aimed at developing best practices and/or shared tools for reducing duplicated effort and to provide rich discovery metadata for library users more quickly.

The project will provide professional growth opportunities for team members as the project will expand awareness of metadata practices from communities outside of the library. This will undoubtedly increase potential for creative solutions to metadata exchange in other spheres.

The goals of this project have intersections with Harvard's Interoperability Initiative. The project will be conducted in such a way as to adhere to the principles of open-source, openly shared and collaborative development to maximize the benefits of the tool that is built.

How is it different from other similar initiatives?

There are a number of exciting initiatives happening in parallel (with some dotted line relationships between them) focused on developing robust Unique Identifiers for authors, creators and researchers within different community contexts (ORCID, ISNI, and the UK Names Project, as mentioned above), but their efforts do not directly tie into library Authority control work. I discussed the project with and incorporated input from two publishers, two ORCID board members, an ECIP program participating institution, and OCLC Research; they all expressed support and interest. I have not been able to identify a similar project where the work on the Author Names project would be duplicated effort.

Outreach to Potential Partners

Numerous very fruitful conversations and/or email exchanges were conducted in preparation for submitting this proposal to identify the most viable directions for the project. We conducted outreach to the colleagues at Harvard, MIT, Stanford, OCLC Research, MIT Press, and Tupelo Press. We plan additional outreach to Harvard University Press, Harvard University Archives and the Office of Scholarly Communication/DASH. Interest and support for this project have been expressed by Scott Wicks, Head of Information & Technical Services (ITS) and other ITS managers.

Project Team Members

- **Michelle Durocher**, Head of Metadata Management and Interim Head of Metadata Creation, will be Project Manager. Developing innovations to leverage existing metadata is central to the goals of ITS.

Critical resource people to the project from the library community are Harvard's representatives to the national [Program for Cooperative Cataloging](#): Mary Jane Cuneo (NACO rep), Steven Riel, (CONSER rep), Bruce Trumble (SACO rep) and Isabel Quintana, NACO cataloger. They will perform the key role of mapping the data

export to the RDA NACO authority record and validating results. Their expertise will inform our use of all author-supplied metadata from the project that we can take advantage of via new RDA publically-displayable fields that will benefit researchers.

Establishing formal project partners from publishing communities will be the first step in the project so as to base the tool's specifications on a rich cross-section of publishers' needs.

Appendix – Technical information

I performed outreach to a library web developer, Enrique Diaz, to help me spec the elements needed for the technical component of the Author Names project. We identified the following aspects that would likely be needed.

The Author Names tool would be an open source web-based application that includes temporary database storage to capture data until data is exported to other applications. Other features would include:

1. Tool can support “user sessions” to ensure that only authorized submissions are allowed.
2. Web form interface to capture Author input into various fields, some required and some optional. The fields would be determined by mirroring questions asked on the Author Questionnaires that publishers currently distribute to authors, usually as Word docs or PDFs. Fields needed for libraries would be among the required elements.
3. A Harvard administrative function to grant access to each participating publisher.
4. Once a publisher-side Administrator is established by Harvard, each publisher would require an interface for administrative and access management for their employees:
 - Each publisher would need the option to customize a “data profile” so that the fields they choose to include in their text export maps cleanly to the fields within their internal desktop database system (FileMaker seems to be the commonly used application.) Additionally helpful, though not essential, would be the ability for the publisher to determine which optional fields to include or not include in the publisher-specific instance of the form that is presented to their Author clients; this would create the most flexible tool that help attract wide adoption by many publishers.
 - In addition to the above technical customization, the publisher's Administrator role should be able to customize the author experience in some way by applying a style sheet or at minimum uploading a company logo so that the Author form has a look that is compatible with the publisher's stationery or website.
 - Internally, each publisher would need to set up User accounts for employees who may generate session-specific form links for Authors and download data from the tool.
5. Data exported from the tool would go in 2 streams: one for publishers and one for the library system:
 - Publishers would need an export file for each Author submission in order to import data to their internal system.
 - The export file for the Library would include selected fields formatted into a MARC RDA authority record. We would review and upload to the national authority file manually during the project pilot; automated upload to OCLC and VIAF is the end game.