

Response to PCC Policy Committee on the
September 8, 2017
Supplementary Charge to the
PCC TG on Identity Management in NACO

October 18, 2017

Members of the PCC Task Group on Identity Management in NACO¹

Introduction

As early as the [2010 strategic plan](#), the Program for Cooperative Cataloging has been stating its willingness to collaborate with other parties. “PCC welcomes interaction—both influencing and being influenced by other metadata communities.” This vision is increasingly being realized in the area of authority data and the management of identities.

What does identity management look like in a linked data environment?

Summary

- Libraries will use more data sources than the LC NAF
- Libraries will create data in places other than the LC NAF
- Libraries will need to accept data quality outside the control of library standards
 - Granularity of this data may vary
- Data delivery systems will need to utilize identifiers rather than unique strings for entity matching in data quality control. This will make contextual information about the entity very important to accurately identify entities
- Changes in one platform will initiate changes in other platforms that reference an entity’s identifier

Participating in a linked data environment involves reusing data from sources beyond library identity standards, rather than replicating “external” data within our existing data sources (LC NAF). By using existing identifiers from other sources, libraries are acknowledging that this data is of sufficient quality. It may not be about getting more entities into LC NAF but rather enabling use of other types of identifiers in BIBCO/CONSER records, to allow for forms of names from external data sources. When the same entity is represented in both LC NAF and an external source, links between the records will be desirable. All of this will have a trickle-down effect that will impact the implementation of the ILS and the end-user experience.

¹ Amber Billey, Michelle Durocher (PoCo liaison), Paul Frank, Jean Godby, Stephen Hearn, Violeta Ilik, Jennifer Liss, Andrew MacEwan, John Riemer (chair), Erin Stalberg, Diane Vizine-Goetz.

We are already creating data in other domains, e.g. the PCC ISNI Pilot.² Perhaps we will have a placeholder in our file saying we have done work elsewhere or a federated search of identity sources will support cross-domain awareness of identity data. This will need to be determined.

Content standards, best practices, and schemas will vary across our partners--indeed our own library data is largely not RDA-compliant; therefore, the PCC should pursue a system that honors a degree of data autonomy. We need systems that disseminate changes to other systems. Frequency of data dissemination may be a challenge.

We need to rely primarily on identifiers versus text strings. We need more information than the shorthand of a unique text string (institutional affiliation, discipline a scholar works in). We won't see domain-controlled strings, but rather something looser. Our need includes the technological infrastructure to go along with supporting this, e.g. an aggregation of label, attribute, and relationship data linked to the identifier.

A larger labor pool to create and manage identities will be needed. Each distinct entity needs to be covered by an identifier. Most institutions need identifiers created for local researchers and organizational units and, as such, they also need efficient ways to get this work done. Most institutions have assigned people IDs in their personnel system, but the data attached to that ID may not be shareable. Researchers also need a public ID which may be managed personally or institutionally.

What role will PCC play in the emerging identity management environment?

Summary

- PCC will lead **advocacy and outreach**
- PCC will provide **education and training**
- PCC will lead **infrastructure development--policies, systems of communication, technology, recommended data sources**

Advocacy and outreach. The changes to the concepts of identity management described in Section 1 of this report will require significant leadership effort from PCC. The value of expanding the universe of respected authorities for cataloging beyond the traditional MARC21 and LC NAF sources and includes these benefits:

- Better integration of library metadata with other identity registries

² For a summary of the background and goals for the pilot, please see: <http://ow.ly/NCpA30g1PTP>

- Better leveraging of identity management efforts in other domains
- Contributing to the coordination of cross-domain identity management
- Developing effective cooperative relationships with identity management agencies for the purpose of achieving the above three aims

These goals will require PCC to extend itself in new ways, as modeled in part by the relationship being explored between PCC and the ISNI (International Standard Name Identifier) organization. The goals will also require advocacy directed both at the library community and at the larger organizations and communities which support them.

Education and training. For a new model of identity management to become effective in our community, a significant effort at education and training will be needed. Librarians need to be made more aware of and familiar with a range of policy and procedural changes:

- Raising awareness about the differences between authority control and identity management, and the environmental factors giving impetus to the changing emphasis
- Laying out the conceptual framework for managing identities in a multi-domain environment
- Integrating selected external identity registries into BIBCO cataloging
- Recording relationships between identity representations in different domains
- Adopting computer coding and metadata encoding conventions to enable cross-domain searching and referencing
- Developing and proliferating new models of participation in identity management beyond NACO conventions

This initiative will have different impacts on different segments of the PCC community. Education and training efforts will be needed both at a general level and in more specialized tasks. Some training may come from external agencies with whom PCC is seeking a stronger working relationship and may represent an opportunity for building mutually beneficial partnerships.

Infrastructure development. All the efforts describe above must be enabled by significant work on infrastructure development--policy frameworks and statements, communication systems and protocols, technical interoperability, and partnership development with identity management agencies. There are network protocols already developed which may provide crucial structural support for interoperability among identity registries and the databases that reference them. PCC should not be attempting to build this technical infrastructure from scratch; rather, the goal should be to leverage existing structures and protocols to ensure better integration between libraries and other networked data providers. The direct challenge which PCC must address will be in developing policies and interorganizational relationships that can guide implementation of these initiatives on a practical path to demonstrable successes.

What will the future look like?

In this section we discuss what is needed to develop an environment or a system that will foster real time data exchange. One system Violeta Ilik compared it to in [her blog post](#) is the one used in the airline industry (GDS). The comparison is not meant to imply that the system needs to be built on similar standards as in the airline industry. It is referenced simply due to one specific feature -- exchange and verification of data in real time. A proposed name of the system is Global Distribution System (GDS) for authors information exchange [see Figure 1].

As described in another [blog post](#) by Violeta Ilik, which received feedback from experts in the area, "this system would be comprised of hubs where all stakeholders would engage in exchange/verification of information about authors. It would be a decentralized system that joins together various software instances so that everyone would be able to see the activities in all of the hubs."

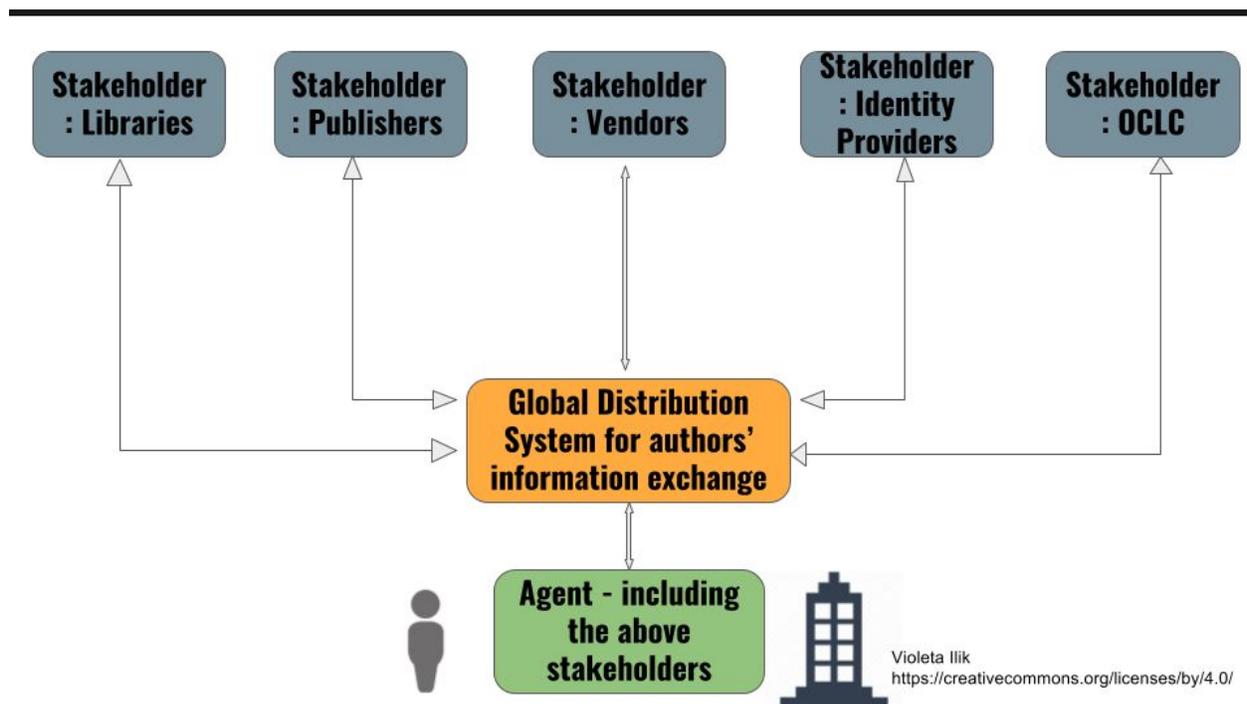


Figure 1: GDS for authors' information exchange

As further described “The proposed solution to the present challenge is a shared information pipeline where all of these stakeholders/agents will be able to share and exchange data about authors. Publishers, vendors, OCLC, libraries, and other stakeholders would have access to and contribute with their own information. This would enable real time data exchange.”

In the same [blog post](#) more specifics are given to the solution of how will the information travel and how will other institutions/organizations/application consume that information. WebSub, a HTTP-based publish/subscribe protocol is recommended for further exploration as the most optimal solution to the problem. The blog post provides details of how will WebSub “provide an environment where each party could posts its evolving version of a description on a channel to which all parties subscribe. Each party would be able to gather the information they need from that channel. In this environment there is no central/correct/unique version of the data -- instead there are many versions that are informed by work being done in different institutions/applications that manage and use identity information. This is a real time information channel fed by and consumed by institutions and applications that manage and use identity information. ResourceSync Change Notifications can be used to create/update/delete links when information about a new or updated description is sent via the URI of the description. The nature of the change (create, update, or delete) and the associated URI are sent through Change Notification Channels as Change Notifications. These notifications “are sent to inform Destinations about resource change events, specifically, when a Source's resource that is subject to synchronization is created, updated, or deleted.” Further details are described in the ResourceSync Framework Specification (ANSI/NISO Z39.99-2017). ResourceSync Change Notification is based on WebSub and software for it already exists.”

In conclusion of the blog post, Ilik states the benefits of this solution “all organizations that manage identity information (LoC, OCLC, ORCID, DBpedia, WikiData, libraries, museums, archives, library system vendors) should have a clear interest in deploying an information sharing pipeline. The most important motivator for all of these organizations to agree on an information sharing pipeline is that all of them would need to work with only one API which would be based on the WebSub protocol. The benefits for all the organizations mentioned above are clear.”

Acknowledgment for this section: Violeta Ilik would like to acknowledge [Sarven Capadisli](#), [Herbert Van de Sompel](#), and [Lukas Koster](#) for their constructive feedback on this idea of creating a real time information channel for which standards and software already exist.

Are there any other thoughts or advice your group wishes to have considered as PoCo develops the new strategic document?

We all will face some key questions about identity management as we exit the current world and move into a linked data environment.

What types of descriptions will libraries want to continue to capture?

What constitutes a useful description?

What data does it contain?

What data does it link to?

What workflow(s) will be needed?

What quality standards will be needed, and what will be tenable? (In other words, what criteria are required of the data sources we might take advantage of, for any of our intended uses?)

Can/should we support data exchange in a way that lets most data harvested from partners remain as-is?

What are the new best practices going to be?

We will guess at best practices, but we will likely end up discovering what they are/should be.

It appears there are many leadership opportunities and potential roles for the PCC in the above.

A spirit of risk taking and experimentation will be needed to get us to where we are going.

Task Group members have a strong sense that, in welcoming a multipolar, multi-file world that involves collaboration with others, we will come out farther ahead of where we are now overall. We will increase the number of participants in identity management work, the resulting amount of identifiers and other metadata, and, in that, we will improve the discovery experience for our users.

+ + + + +

Supplementary Charge to PCC TG on Identity Management in NACO
September 8, 2017

Dear John,

As you may know, the PCC Policy Committee is preparing to draw up a new strategic directions document on November 1, just prior to the Committee's annual meeting. The PCC leadership would like to request that this Task Group reflect on its work to date, other recent developments* and then provide its big-picture, forward-looking thinking on these questions:

- What does identity management look like in a linked data environment?
- What role will PCC play in that environment?

Are there any other thoughts or advice your group wishes to have considered as PoCo develops the new strategic document?

We would appreciate having your group's thoughts in writing by Wednesday, October 18. Please contact Matthew Beacom (matthew.beacom@yale.edu) if you have questions.

Thank you in advance for your support as PCC attempts to build a strong leadership for the future.

Matthew Beacom
PCC Chair