

Networking Names

Karen Smith-Yoshimura

**Program Officer
OCLC Research**

In collaboration with:

Grace Agnew, Rutgers University
Laura Akerman, Emory University
Genevieve Clavel, Swiss National Library
Joan Cobb, Getty Research Institute
Michele Crump, University of Florida
Thom Hickey, OCLC
Amanda Hill, University of Manchester, Names Project
Deborah Kempe, Frick Collection and Frick Art Reference Library
Ralph LeVan, OCLC
Amy Lucker, New York University
John MacColl, OCLC
Dennis Meissner, Minnesota Historical Society
Suzanne Pilsk, Smithsonian Institution
Michael Rush, Yale University
Jon Shaw, University of Pennsylvania
Laura Smart, California Institute of Technology
Daniel Starr, Metropolitan Museum of Art
Robert Wolven, Columbia University



A publication of OCLC Research

Networking Names

Karen Smith-Yoshimura, for OCLC Research

© 2009 OCLC Online Computer Library Center, Inc.

All rights reserved

April 2009

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 1-55653-412-4 (978-1-55653-412-6)

OCLC (WorldCat): 319639019

Please direct correspondence to:

Karen Smith-Yoshimura

Program Officer

smithyok@oclc.org

Suggested citation:

Smith-Yoshimura, Karen. 2009. Networking Names. Report produced by OCLC Research.

Published online at: <http://www.oclc.org/programs/reports/2009-05.pdf>.

Contents

Introduction	4
Problem space.....	4
Objectives of a Cooperative Identities Hub	5
Characteristics of a Cooperative Identities Hub.....	5
Entities to be covered	5
Benefits.....	6
Requirements	6
Functions	7
Attributes and data elements	8
Cooperative Identities Hub target audiences	9
Use case scenarios.....	10
Use Case Scenarios 1 – Academic libraries and scholars (<i>Laura Akerman, Michele Crump, Amy Lucker</i>).....	10
Scenario 1-A: Academic researcher.....	11
Scenario 1-B: University department.....	12
Scenario 1-C: Library cataloger	12
Scenario 1-D: Student.....	14
Scenario 1-E: Scholar	14
Use Case Scenarios 2 – Archivists and archival users (<i>Dennis Meissner, Michael Rush</i>).....	16
Scenario 2-A: Archivist.....	17
Scenario 2-B: Archivist	17
Scenario 2-C: Archival researcher	18
Scenario 2-D: Archival researcher	19
Use Case Scenarios 3 – Institutional repositories (<i>Amanda Hill, John MacColl, Suzanne Pilsk, Jon Shaw</i>)	20
Scenario 3-A: Researcher.....	20
Scenario 3-B: Institutional repository manager	21
Scenario 3-C: Institutional repository manager	22
Scenario 3-D: Institutional repository manager	23
Scenario 3-E: Research information office manager	24
Notes.....	25

Table

Table 1. Cooperative Identities Hub target audiences	11
--	----

Introduction

The fifteen members of the RLG Partners Networking Names Advisory Group¹ collaborated on articulating the problem space that the research community needs to address and identify components of a “Cooperative Identities Hub” that would have the most impact across different target audiences. The group developed use case scenarios that provide the context in which different communities would benefit from aggregating information about persons and organizations, corporate and government bodies, and families, and making it available on a network level. This report summarizes the group’s recommendations on the functions and attributes needed to support the use case scenarios.

Problem space

- Information sufficient to identify and distinguish people and organizations with similar names is widely dispersed and these names are represented in multiple languages and scripts.
- The preferred form of a name depends on context. There is no one form that is used across all communities and languages.
- Identification and distinction require contextual information and cross-references that are often lacking.
- Information about creators or personal/corporate subjects of works is needed across communities—libraries, archives, museums, digital library production, institutional repositories, publishers, etc.—as well as by users in the Web networked environment.
- National authority files provide scant information to distinguish entities with same or similar names.
- Privacy or data protection issues in some environments restrict access to information.
- Information about entities resides in many different environments and cannot be used easily outside those environments.
- Institutions need to manage names across databases, systems, units and domains.
- Authority work is costly and limited to a small group of trained contributors.
- Expertise from many user communities is untapped.

Objectives of a Cooperative Identities Hub

The Hub's primary objective is to enable all users to uniquely identify entities either stored in the Hub or extracted from other data resources. The Hub itself is to be open to searches and updates by other software applications that can retrieve and supply information uniquely identifying entities. The Hub will also allow people to add information about the entities reported and enable contributing agencies to retrieve Hub records corresponding to their own records. People can also add new entities not yet represented.

Characteristics of a Cooperative Identities Hub

1. Provide framework for concatenating and merging authoritative information about entities now hidden within library, archival, museum and other contexts.
2. Serve as a gateway to all forms of names authorized or used in other contexts without preferring one form of name over another.
3. Use a social networking model to broaden the view of "authority work" beyond NACO contributors. This includes authors who can add links to their own works (or delete links mistakenly attributed to them).
4. Provide a switch for users (machine applications) to extract relevant information for re-use in their own contexts and to contribute new information.
5. Create a federated trust environment to authenticate and authorize contributors (including software applications) that can add, modify, combine or delete entity information.

Entities to be covered

- Persons and personas (including pseudonyms), organizations, corporate and government bodies and families.

Benefits

1. Increase metadata creation efficiency and make better use of existing metadata.
2. Make it easier to identify works by or about the same entity regardless of language or discipline.
3. Enable users to determine preferred form of name within their own contexts.
4. Enable contributing agencies to augment their own name data resources.
5. Expose information about personal and corporate bodies beyond their original contexts and bring them into the “network flow.”

Requirements

1. Define a core set of elements and attributes that different user communities can extend as needed suited to their own contexts.
2. Assign a unique identifier for each entity.
3. Track sources of each entity, all its alternate forms of names, source-specific identifiers, and any additional information contributed.
4. Isolate attributes by data source so that users of the data can make their own determination about its value and reliability.
5. Support and define relationships between and among entities over time and space.
6. Fully support multilingual, multiscript data elements using Unicode.
7. Develop a data structure format for harvesting data from different sources.
8. Develop an editorial policy that balances currency and completeness with the need for authority and consistency.
9. Establish guidelines for sustainability of the Hub.

Functions

Search: Search, retrieve, print or download records in response to a query for an entity. Support searches by both individuals and software applications. Cluster together different entities with the same or similar names. Provide actionable links to other resources and pass queries to other databases.

Edit: Authenticated users will be able to:

- Add information to existing entities, including adding links to other resources for the same entity.
- Merge or group together two or more existing records that describe the same entity, retaining all source identifiers.
- Split an existing entity into two different entities.
- Delete or flag for deletion an entity.

Batch update: Contributing agencies will be able to send and retrieve batches of records (additions, updates, deletions). Offer interactive push mechanisms (Atom publishing, SRU record update) and pull mechanisms (OAI-OMH, FTP) to update the Hub from contributors' databases.

Add: Authenticated users will be able to create a new entity not yet represented containing the core set of elements and attributes to uniquely identify it from all others.

Discussion: Support social discussions that clarify or debate on proposed (or actual) changes or additions to entities.

Revision or change history: Track all modifications to each entity to provide an audit trail.

Rollback: The Hub will retain earlier versions of entities and will be able to reverse changes.

Attributes and data elements

All elements will be able to be qualified by dates, language, source and other attributes that could help in verification. Source of each data element in an entity record will be required.

1. Life dates (normalized to a standardized form, including BCE dates). Consider using the Extended Date/Time Format from the Library of Congress (available online at <http://www.loc.gov/standards/datetime/>).
2. At least one form of name. All names will be normalized to the same format, with pointers to the original form of the name. Encode what is the last name, first name and other names. Information that indicates which form of name is used in which context. Other, variant forms should include:
 - a. The preferred form used in a given source (e.g., national authority files).
 - b. Form(s) used in citations (by Abstracting and Indexing services, publishers)
 - c. Dates when form was used, if applicable.
 - d. Form of transliteration, if applicable.
3. Gender, if available.
4. Life events. Each event will include associated dates, if known, with options (single date, start and end dates, date range) and attributes (exact, approximate, inferred).
 - a. Place(s) of origin of the entity (including place of birth and place of death, if available). Geographic encoding, if available.
 - b. Place(s) of entity's output (if different from place of origin). Geographic encoding, if available.
 - c. Institutional affiliations, if applicable and known. Link to the institutional entity, which may have various forms, if represented in the Hub.
 - d. Knowledge domains/subject areas/field(s) of expertise for person or profession, titles, functions, activities or occupation. Implies some standardized list, including multilingual variations (chemistry = Chemie = chimie = kagaku).

- e. Associated entities: name, role and what the relationship is—e.g., founder, successor, predecessor, pseudonyms, personae, family relationships, mentor, arranger, translator, editor, illustrator, co-author, publisher or printer).
 - f. Contextual notes that can help uniquely identify the entity.
5. Associated entities: name, role and what the relationship is for those not represented within a life event.
 6. At least some works associated with the entity, with actionable links to other sources for more. (Meant to help distinguish persons, not provide comprehensive output.) Include the relationship with the work (e.g., editor, author, arranger).
 7. Language(s) used in communications by the entity.
 8. Short biographical history.
 9. Unique identifiers from each source used to populate/enrich the entity record.
 10. Date entity record was created and each date modified.
 11. Links to contact information.

Cooperative Identities Hub target audiences

Target groups would be those that would use the Hub for three main functions:

1. Managing their own information resources (M)
2. Contributing new information (editing existing entities or adding new ones) (C)
3. Information seekers who search the Hub for information and may also discuss or comment on existing entities (S)

Table 1. Cooperative Identities Hub target audiences

Target Audience	Manage	Contribute	Search/Comment
Abstracting & Indexing services (also data source)	M	C	S
Aggregators (also data source)	M	C	
Archivists	M	C	S
Authority control vendors	M	C	S
Authors		C	S
Authors' and publishers' royalty collection agencies	M	C	S
Librarians	M	C	S
Citation seekers			S
Copyright registries	M	C	S
Data curators	M	C	S
Data marketers			S
Fact checkers and journalists		C	S
Funding agencies	M	C	S
Genealogists		C	S
Institutional repositories	M	C	S
Museum curators/custodians	M	C	S
Museum dealers/auction houses	M	C	S
Publishers	M	C	S
Researchers		C	S
Scholarly societies	M	C	S
Search engines	M		S

Use case scenarios

These scenarios were selected to give the strongest illustrations of important features and benefits of the Hub from among many other scenarios and contexts generated from advisory group discussions.

Use Case Scenarios 1 – Academic libraries and scholars

Laura Akerman, Michele Crump, Amy Lucker

Community: Faculty, staff, students and others involved in academic research, and librarians and other information service workers who support research activities and provide collections and access to them through discovery systems.

Scenario 1-A: Academic researcher

Context: Faculty, staff, students and other academic researchers browsing in the library’s “next generation” search interface. The researcher locates a specific name and wants to know more information about the entity. The researcher clicks on the “More information” button located next to the name in the bibliographic record. Clicking on the button passes the name to WorldCat Identities and retrieves information from the Hub about the specific name the researcher located.

Actions:

- The researcher enters a name into the library’s search interface and retrieves a bibliographic record.
- The researcher wants to know more information about the name on the record and is prompted to do so by clicking on the button next to the name (“More information”).
- Clicking on the “More information” button retrieves from the Hub biographical information and a list of or links to creative works associated with the name.
- The researcher can click on the individual creative work/title, which prompts a search for the work within the library discovery system.
- If the work is available the researcher discovers the location of the item. If the work is not owned by the library the researcher is directed to an Amazon-like resource from which the item might be purchased (print or electronic) and/or to WorldCat (or another union or consortium catalog) to see if the item is at another library from which the researcher may request it via Interlibrary Loan.

Outcomes: The researcher successfully verifies the name and receives additional information about the creative works produced by the entity. If the Hub is updated with accurate biographical information and links to the creative works produced by the entity, researchers will derive maximum benefit. If the Hub lacks these elements, researchers may not find the Hub useful.

Impact: Allowing access to the Identities Hub within the search results of the unified discovery interface will streamline the researcher’s information gathering process. Ease of access to information will be very attractive to researchers. Such a facility would require an API for mashups.

Scenario 1-B: University department

Context: A University Department secretary wants a comprehensive list of beloved and prolific Professor J. Greybeard’s publications for an obituary. The secretary knows that another J. Greybeard exists and that his work is pornographic in nature. The secretary wants to honor the life work of the late professor so he or she must verify the professor’s biography and obtain an accurate list of his scholarly work.

Actions:

- The secretary enters name J. Greybeard into the search portal of the Hub.
- The search results offer two authority records for the name J. Greybeard.
- The secretary can differentiate between the entities by examining the biographical details and noted creative works entered under each name.
- The secretary selects the correct Professor Greybeard record—a composite from several sources that include additional information through links to citations in WorldCat, Wikipedia, Google Scholar and other online sources.

Outcomes: The scenario succeeds if each entry for the name J. Greybeard contains sufficient information about the person so that it is apparent to the researcher that these are two very different individuals. The Hub provides depth to each entry when it supplies links to other sources that further confirm the entity while offering detailed information.

The scenario fails if the entries in the Hub are not accurate and confuse the researcher as he or she attempts to distinguish entities with the same name.

Impact: The Hub benefits the casual researcher as well as the scholarly researcher by providing a central source for verifying names. From this central source, the researcher—through active links—may be directed to even more information located in other “reliable” sources.

Scenario 1-C: Library cataloger

Context: Library catalogers creating authorities for NACO or another cooperative can draw data from Hub records. A library cataloger (or metadata creator), while cataloging an item, may find no established record in the cooperative shared authority file used for this catalog (e.g. NACO Authority File). Perhaps no works are found under that name in the shared bibliographic file (e.g. WorldCat). As a cooperative participant, the cataloger needs to create a new authority record but has very little information to supply.

Stakeholders include cataloging librarians and staff responsible for name authority management. Users include other catalogers and end-users of library discovery systems.

A prerequisite for the enhanced workflow as described is availability of Hub link, search and capture function interfaces to external software such as the Connexion client. Other name authority cooperatives might not technically be able to use these features, but catalogers could still benefit from consulting the Hub and drawing data from records there.

Actions:

- A cataloger, after determining that no NACO record exists for the creator of a work being cataloged, highlights the name in the bib record and clicks a link to the Hub in Connexion.
- The name is captured and searched in the Hub.
- The cataloger finds a record from the Union List of Artist Names for the same name string, with enough information to make him or her fairly sure the record is for the person in question.
- The cataloger clicks a button to capture some Hub “core elements” derived from the ULAN record, translates them into MARC21 authority fields and inserts them into a NACO authority workflow in Connexion.
- The cataloger reviews, revises and augments the authority record with other information from the work being cataloged, or elsewhere, and contributes it to NACO.

Outcomes: The scenario succeeds if additional information is found in the Hub often enough that searching this resource is a worthwhile part of authority record creation workflow. The scenario fails if the Hub doesn’t have more information on a name very often, or if searching the Hub misses such information a high percentage of the time.

Impact: If enough contributions from different sources are made that do not overlap, so that the Hub is likely to yield information about names that is not found in NACO or other cooperative authority stores, being able to capture and use Hub data could greatly enhance and streamline cooperative authority creation work and lead to richer and more informative records. Incorporating the Hub search and capture into workflow tools such as Connexion could add to the attractiveness of the service.

Scenario 1-D: Student

Context: A pre-dissertation student in geo-informatics wants to find out more about the T. Smith who published an article on alluvial fan processes. Has he published any more articles? There are millions of other T. Smiths; how can he or she find articles only by this one? The student knows from a blurb at the bottom of the article, that at the time it was published, T. Smith taught at The University of California at Santa Barbara.

Actions:

- The student searches the Identities Hub and retrieves many T. Smiths.
- The student limits his or her search by affiliation keywords “santa barbara.”
- This limits results to a single record. The article isn’t listed there, but there are two other articles on closely related areas of geo-informatics. The student deduces that the record describes the person he or she is interested in.
- The Identities Hub has links to Web of Science and other article aggregations and to online journal articles authored by this person.
- The student has learned from the Identities Hub record that the author’s first name is “Tony,” so he or she can search other databases with the full name.

Outcomes: This scenario succeeds only if the Hub also includes reliable institutional affiliations with a standardized way of representing institutions and includes authors of journal articles. The scenario fails if the Hub does not include institutional affiliations to distinguish authors with the same name and authors as cited in articles. The scenario fails if both these attributes are lacking.

Impact: The Hub would benefit students and researchers looking for a specific author that shares the same name as many others. NACO authority records do not include forms of names used in article citations, and the Hub could provide an important bridge to authors represented in both article and monographic literature.

Scenario 1-E: Scholar

Context: A scholar-generated subject portal has name information to contribute to the Hub. Many scholars create subject- or discipline-related portals or reference resources. Some portals, particularly those with a historical dimension, have a strong focus on individuals and other creative entities, and collect unique information about them and their relationships to other entities, places and works. For example, a subject portal on early American shape-note music could have data with

rich relationships between composers and editors, tunebooks, families and locations. The portals may have a scholar-developed data organization that doesn't follow a standard authority schema.

As stakeholders in the Hub, portal creators want to share their identities data to support both cross-institutional collaboration in their field and scholarship in general. Academic libraries and librarians help with sharing in their roles as information service providers. Other scholars and librarians who search, harvest and/or augment the data outside of the original portal could also be stakeholders and users.

A precondition for sharing is a well-defined schema and format for batch contributing the Hub's "core elements." The schema should allow expression of detailed relationships among named entities and between entities and works. The Hub should support common record transfers by methods such as OAI-PMH or FTP, and selective download/harvesting of the core elements by source collection, among other selections.

Actions:

- A scholar (or group of scholars) is aware of the Hub and wants to contribute data from a subject portal to it. Hub documentation clearly details the format and procedure for contributing Hub core elements, and the scholar makes a request for batch contribution.
- The Hub managers review information about the portal and samples of the source data and approve the inclusion.
- With help of a library metadata specialist, submission records are prepared by mapping and transforming data from the portal's format to the Hub core elements. Shared tools or strategies created by other contributors could be useful here.
- The records are contributed to the Hub and are available to scholars everywhere.
- Curators of a collection at another institution, in a related subject area, learn of the portal's name records in the Hub. Their catalogers download "core elements" records for this collection—perhaps using the Connexion client to transform them into MARC Authority format.
- Catalogers manually enhance many records with information from their own collection, contribute the enhanced records to the NACO authority file and use them in their library's discovery interface.
- The new NACO records are eventually incorporated into the Hub. The Hub display for each entity merges and links to information from both sources.

Outcomes: The scenario succeeds when quality, unique data about identities that has been created by scholars in a non-standard way can be contributed to the Hub that is of value to scholarship and to the work of librarians and metadata specialists, and that can be captured and/or enhanced by others.

The scenario fails when the unique data about identities stays in its “silo” because there’s no pathway for “nonstandard” name data to get added to the Hub, or because mapping and transforming the data for inclusion is too difficult. It also fails if the submission review lets in biased or inaccurate data.

Impact: As scholars with name information become aware that they can contribute to the Hub, they recognize that “publishing” some of their data there adds to the usefulness and visibility of their work. As scholars contribute more, the Hub becomes “the place to go first” for scholars looking for basic information about people and organizations related to information resources.

Access to data from smaller, subject-related data stores expands the available information for catalogers/metadata creators and saves significant time in creation of NACO authorities or other identities records.

Use Case Scenarios 2 – Archivists and archival users

Dennis Meissner, Michael Rush

Community: Three archival stakeholder groups will most frequently use Hub services:

1. Archivists who are creating descriptions of the entities represented in an archival collection,
2. Archivists who are hoping to link to existing description(s) of one of these entities from a description of an archival collection, and
3. Users of archives who are seeking to collocate resources relating to a particular entity or to add a comment to the description of an entity.

EAC-CPF (Encoded Archival Context—Corporate Bodies, Persons and Families) databases will be the primary archival point of interaction with the Hub. EAC-CPF is an XML encoding standard, under development, for encoding and communicating descriptions of the entities who create or are represented in archival collections. The information in those records will include a control number, the identity of that entity, a rich description section (dates, place, legal status, occupation, activities, legal mandate, organization chart/genealogy, narrative, chronologies), relations (links to other CPF entities or to resources themselves) and sources.

Scenario 2-A: Archivist

Context: An archivist or his or her proxy will create narrative histories and/or chronologies, and document identity and resource relationships, in a new EAC-CPF record and will then contribute information about that record to the Hub. The Hub is EAC-CPF-aware and can access APIs for EAC-CPF repositories.

Actions:

- Archivist searches Hub to determine whether a Hub record exists for the entity. The archivist may revise EAC-CPF record based on existing Hub information.
- Archivist submits minimal information about EAC-CPF record to Hub, e.g., ID and persistent network address.
- Hub client queries EAC-CPF repository and harvests additional data fields from relevant record(s) to enhance Hub record and to create links to EAC-CPF resources.

Outcomes: Success occurs with the accurate association of an EAC-CPF record with other records in the Hub and the contribution of important information and relationships not available in other records in the Hub. If the archivist fails to make that association, the Hub record is not enhanced and the EAC-CPF record that is subsequently loaded may duplicate an existing record for the same entity.

Impact: When Hub records successfully associate an EAC-CPF resource with a Hub identity, or can reliably indicate that the identity does not exist in the Hub, then value is added by enhancing information, collocating resources and preventing redundant labor.

Scenario 2-B: Archivist

Context: An archivist intends to create an EAC-CPF record to complement his or her description of a series of archival materials. The entity in question is famous and has deposited additional archival materials in other repositories, as well, so the archivist suspects that an EAC-CPF record may already exist. The archivist first searches the Hub to determine whether an EAC-CPF record exists, or whether the Hub contains other applicable authority data.

Actions:

- Archivist searches Hub to determine whether a Hub record exists for the entity.
- Archivist retrieves a Hub record for the entity which links out to an existing EAC-CPF record in another repository.
- The archivist enhances, or simply links to, the existing EAC-CPF record.
- Archivist submits minimal information about EAC-CPF record, and/or the EAD finding aid, to the Hub, e.g., ID and persistent network address.
- Hub client queries EAC-CPF repository and harvests additional data fields from relevant record(s) to enhance Hub record and to create links to EAC-CPF resources.

Outcomes: Success occurs when the archivist locates a Hub record for the correct entity and is able to locate an existing EAC-CPF record as a result. If successful, the archivist will either harvest or link to that record or, if no EAC-CPF record exists, will use the Hub to locate other appropriate records from which to pre-populate a new EAC-CPF, automatically recording the source in the appropriate EAC-CPF element. Failure occurs if the archivist creates a new, redundant, EAC-CPF record because he or she failed to discover relevant information already in the Hub.

Impact: When Hub records successfully associate an EAC-CPF resource with a Hub identity, as well as with an existing EAC-CPF record if one exists, then value is added by simplifying EAC-CPF creation, collocating resources, preventing redundant labor by archivists in different repositories, and by preventing confusion caused by multiple EAC-CPF records for the same entity or identity.

Scenario 2-C: Archival researcher

Context: An archival researcher desires to identify all significant archival materials related to a particular entity. To optimize both precision and recall, he or she first searches the Hub in the hope of finding a record for the entity that will provide references to archival collections in disparate repositories, and perhaps to related entities, as well.

Actions:

- The researcher searches the Hub and locates a record for that entity or identity.
- The Hub record contains links out to associated resources in various locations.
- The researcher is thereby able to retrieve the full text of archival descriptions and, perhaps, digital resources.

Outcomes: Success occurs when the user accesses metadata on resources by or about a given entity via the Hub. If the researcher fails to find a Hub record, he or she loses the opportunity to both collocate and link to relevant resources.

Impact: When a Hub record successfully brings together disparate resources for researchers, it saves time and uncertainty and adds significant value to research communities. When a Hub record also enables researchers to link directly to resource descriptions and digital objects, time is saved and functionality is leveraged, again adding significant value.

Scenario 2-D: Archival researcher

Context: Archival researchers, in the course of using Hub records or the archival resources to which they link, may notice erroneous or incomplete information about the entities represented. They, and the archivists responsible for describing the entities, benefit from the ability to attach comments to Hub records. Preconditions for success are that records would have to pre-exist in the Hub in order for relationships to be noted, and that the Hub would have to provide a mechanism for feedback to the record maintainer.

Actions:

- The researcher searches the Hub and locates a record for that entity or identity.
- The Hub record contains links out to associated resources in various locations, and the researcher examines their networked descriptions.
- The researcher, a topical expert, notices erroneous dates in the chronology of the entity's life, which appears in a linked EAC-CPF record.
- The researcher attaches a lengthy comment to the Hub record, noting the errors and providing corroborative references.
- The Hub passes a notice about the annotation to the creator of the Hub record.
- The archivist who created the record reviews the comment and makes a change to the EAC-CPF record.

Outcomes: Success can occur in two ways: when later researchers are able to access the annotator's knowledge of the identity, and when archivists are able to accurately update their records based on the annotations. If the user fails to find a record in the Hub, or fails in the act of attaching a comment, then archivists lose a potential opportunity to correct errors or to enhance an EAC-CPF record.

Impact: Archivists do not have the time to carry out extensive research while describing records or describing entities. When knowledgeable users are able to attach comments to Hub records, significant value is added to archival descriptions.

Use Case Scenarios 3 – Institutional repositories

Amanda Hill, John MacColl, Suzanne Pilsk , Jon Shaw

Community: Repositories of research outputs. These can be institutional databases of journal articles, working papers, conference papers (for example those run by many universities) or they may be discipline-based repositories with similar content (e.g. RepEc, ArXiv).

Scenario 3-A: Researcher

Context: Carol, a researcher, is searching the institutional repository of the University of Anystate and wants to retrieve all materials in the repository that were created by Professor Frederick Jones. Other stakeholders in this scenario include the repository’s managers and the Identities Hub.

Actions:

- Researcher begins to enter the name of the individual. A list of possible matches is presented as the researcher types, allowing Carol to select the person she is interested in from the list.
- Contextual information with the names helps Carol identify the person with a high degree of confidence (area of research, institutional affiliation, etc.).
- Once Carol selects the person, she views a list of the materials within the University’s repository that are associated with Professor Jones.

Outcomes: For this scenario to be successful, there needs to be a unique identifier associated with the creator of the materials and this identifier needs to have been associated with each of the items that Professor Jones has created. It is assumed that the repository will have obtained this identifier from an existing record for Professor Jones in the Identities Hub (dynamically, as part of the first submission of an item associated with him into the repository) or will have created a record in the Hub if no such identity existed and have obtained an identifier as a result of that process. During the submission of any subsequent materials for Professor Jones into the repository, the identifier would be automatically linked to the metadata when his identity as creator is selected.

If the name is not listed, Carol’s mission would have failed. If the name is listed, but the identifier attached to that identity has not been associated with all the relevant materials in the repository,

only some of them will be retrieved, resulting in only partial success. This scenario assumes that all of Professor Jones's materials have been deposited in the repository: if this is not the case, then the scenario will fail.

Impact: The Identities Hub will provide a source of unique identifiers which will help repositories to unambiguously associate an individual with the outputs of his or her research. This avoids problems of imperfect retrieval due to variations of the researcher's name or due to a researcher having a similar name to another individual.

Scenario 3-B: Institutional repository manager

Context: A manager of an institutional repository wants to normalize the names of researchers that he or she already has in his or her repository. He or she wants to use the Identities Hub to establish whether those individuals are represented in the Hub, together with the variations in the form of their names. The manager wants to then associate all the variations for a particular individual within his or her system to a single identity to improve retrieval.

Actions:

- The repository manager sends a batch of names to the Identities Hub and receives a list of potential matches for each of the names. The manager can group or individually approve the matches, adding variants with the indicators of sources of name strings.
- If no match is found or approved by the repository manager, he or she can submit name and indicate source of submission.
- The repository manager can indicate his or her repository's preferred form of name without insisting on anyone else using the same form.
- Future queries can match on submitted IR with names stored in the Hub for consistent recording of matched identities. This will assist institutional repositories with name-form decisions.

Outcomes: The success of the scenario is determined by the repository manager successfully concatenating the names in institutional repositories. Also, variant forms of the names, if any, are added to the entity in the Hub. When adding a new name to the Hub, a heading is established, assigned a unique identifier, and the unique identifier is then added to the researcher's works in the institutional repository. Failure occurs when a manager is unable to disambiguate the names in the Hub. This may lead to the unlikely scenario of the repository manager checking with the researcher in order to determine his or her identity.

Impact: The Hub will allow institutional repositories to determine and establish names, a feature that is not present in most institutional repositories today.

Scenario 3-C: Institutional repository manager

Context: A manager of an institutional repository wants to normalize the top-level institutional corporate names (e.g., ‘University of Edinburgh’) representing affiliated faculty authors in his or her repository. He or she wants to use the Hub to establish the variant forms of name which exist, and to associate all variations for a particular corporate name to a single approved form. This form should be assigned a persistent identifier.

Stakeholders include the repository manager, cataloging librarians and staff responsible for name authority management. Users include other catalogers, faculty author depositors and end-users of library discovery systems.

Actions:

- The repository manager sends the institutional corporate name to the Hub and receives a list of potential matches for each name.
- The manager can group the matches, adding variants as required.
- The repository manager can indicate his or her repository’s preferred or current form of name.
- As the name is ‘controlled,’ the Hub presents future searchers with a drop-down list of likely matches, showing both current and former names, preferably with dates.
- Future deposits made to the repository can match on submitted name with the names stored in the Hub for consistent recording of corporate name identities.

Outcomes: The scenario succeeds if users depositing in the IR are offered an authorized form of institutional corporate name from which to choose their affiliation. It succeeds also if searches of the Hub against a particular corporate name identify all former names of a corporate entity.

Impact: The unambiguous identification of institutional names is increasingly important in metrically-based assessment of research profile, performance and impact. Persistent naming will help institutions optimize their impact in machine-created indexes, and derivative rankings lists and journalistic and professional interpretations.

Scenario 3-D: Institutional repository manager

Context: A manager of an institutional repository wants to normalize the fully articulated institutional corporate names (e.g. ‘University of Edinburgh: College of Science & Engineering: School of Molecular Biology’) for faculty authors in his or her repository. The manager wants to use the Hub to establish the variant forms of name which exist, and to associate all variations for a particular corporate name to a single approved form.

Stakeholders include the repository manager, cataloging librarians and staff responsible for name authority management. Users include other catalogers, faculty author depositors, and end-users of library discovery systems.

Actions:

- The repository manager sends institutional sub-unit corporate names to the Hub and receives a list of potential matches for each sub-unit name.
- The manager can group the matches, adding variants as required.
- The repository manager can indicate his or her repository’s preferred or current form of sub-unit name.
- As corporate names are so ‘controlled’, the Hub presents future searchers with a drop-down list of likely matches.
- Future deposits made to the repository can match on submitted name with the names stored in the Hub for consistent recording of corporate sub-unit name identities.

Outcomes: The scenario succeeds if users depositing in the institutional repository are offered authorized forms of institutional corporate sub-unit name from which to choose their affiliation. It succeeds also if searches of the Hub against particular corporate names identify all variants, including former names of a corporate sub-unit name, preferably with dates. The scenario fails if the clustering is not made on the basis of the whole hierarchical name string.

Impact: The identification of institutional sub-unit names assists discrete parts of a university—departments, institutes, research centers, etc.—in being credited for the impact of their research. It would help in internal forms of research assessment. Persistent naming will help institutional sub-units optimize their impact in machine-created indexes, and derivative rankings lists and journalistic and professional interpretations.

Scenario 3-E: Research information office manager

Context: A manager of a research information office wants to create faculty description pages using the Hub.

Stakeholders include the research information office manager, the repository manager and faculty. Users include other research administrators and fellow researchers, journalists or members of the public seeking information on faculty members or seeking sources of expertise in an institution.

Actions:

- The research information manager uses the Hub to populate a template customized for his or her university, using an API.
- The research information manager can specify the data elements to draw from the Hub. These could include authorized personal and corporate sub-unit names, publication lists (with links) and dates of publishing activity associated with different universities in the course of an academic career (derived by the Hub from publication records).

Outcomes: The scenario succeeds if the research information manager can download a range of useful information from the Hub which can then be edited by faculty themselves. The scenario fails if too few faculty within a particular institution have identities records with at least a reasonable amount of this information.

Impact: The capacity for institutions to derive faculty description pages from a network-level source could afford a powerful efficiency gain. The ability of the system to develop and adjust these description pages dynamically over time would add to the power of the resource. Impact would be substantially increased if edits made in the local system by faculty and research information managers could be used to update the Hub records. Hub records might also benefit from data which would originate in the local system, e.g., areas of expertise.

Notes

¹ RLG Partners Networking Names Advisory Group members: Grace Agnew (Rutgers University), Laura Akerman (Emory University), Genevieve Clavel (Swiss National Library), Joan Cobb (Getty Research Institute), Michele Crump (University of Florida), Amanda Hill (University of Manchester, UK Names Project), Deborah Kempe (Frick Art Reference Library), Amy Lucker (New York University), Dennis Meissner (Minnesota Historical Society), Suzanne Pilsk (Smithsonian Institution), Michael Rush (Yale University), Jon Shaw (University of Pennsylvania), Laura Smart (California Institute of Technology), Daniel Starr (Metropolitan Museum of Art), Robert Wolven (Columbia University). Staffed by Thom Hickey, Ralph LeVan and Karen Smith-Yoshimura of OCLC Research.