**Registering Researchers in Authority Files**

**Karen Smith-Yoshimura Daniel Hook**

OCLC Research Symplectic Limited

**Micah Altman Wolfram Horstmann**

Massachusetts Institute of Technology University of Oxford

**Michael Conlon Andrew MacEwan**

University of Florida British Library

**Ana Lupa Cristan Philip Schreur**

Library of Congress Stanford University

**Laura Dawson Laura Smart**

Bowker California Institute of Technology

**Joanne Dunham Melanie Wacker**

University of Leicester Columbia University

**Thom Hickey Saskia Woutersen**

OCLC Research University of Amsterdam

We are issuing this draft report for community comment. We plan to publish the final report in June 2014.

Please direct correspondence and feedback to:

Karen Smith-Yoshimura

Program Officer

karen\_smith-yoshmura@oclc.org

**Registering Researchers in Authority Files**

The pressure on higher education to demonstrate value has led to institutions’ drive to improve their reputation and ranking. Ranking tables rely in part on how often the works produced by the institutions’ scholars are cited in professional and academic journals. This presents institutions with a challenge: how can they accurately measure and reflect the entire scholarly output of all their researchers? The same information about a specific researcher may be represented in multiple databases, and only a subset interoperates with each other.

Scholarly output impacts the reputation and ranking of the institution. Three global university rankings of particular interest to research libraries—[Times Higher Education](http://www.timeshighereducation.co.uk/world-university-rankings/), [Academic Ranking of World Universities](http://www.shanghairanking.com/) and [QS Top Universities](http://www.topuniversities.com/)—all use citations as a factor in determining rankings, skewing the results towards universities focusing on the sciences rather than the humanities. Notes QS: “*Citations…are the best understood and most widely accepted measure of research strength.”[[1]](#endnote-1)* This focus on citations also puts added weight to authors of *journal articles,* which are not usually represented in national authority files. This absence exacerbates the issue of unclear or mis-attribution. The rise of bibliometrics and its extension, altmetrics—the attempt to measure the *impact* of a work including mentions in social media and news media—strengthens the need to uniquely identify researchers and correctly associate them with their scholarly output.

Over the past year the OCLC Research Registering Researchers in Authority Files Task Group has been considering how to make it easier for researchers and institutions to more accurately measure their scholarly output. A number of approaches to providing authoritative researcher identifiers have emerged, but they tend to be limited by discipline, affiliation, or publisher. The Task Group developed use cases and functional requirements for researcher identifier management systems then compared the functional requirements against a sample of currently available systems to identify gaps, challenges and opportunities. How can we best utilize the various types of researcher identifiers to more accurately identify researchers and their work for better research outcomes?

Registering researchers in some type of authority file or identifier system has become more compelling as both institutions and researchers recognize the need to compile scholarly output. Library name authority files manage identities in library catalogs to enable users to easily and quickly find all the known works—individual monographic publications such as books, musical scores, sound recordings, etc.— associated with a given person. They are created to control the names of authors or creators of works and names that are subjects of works, such as biographies, as they are recorded in library catalogs. Librarians control names by recording as many names by which an identity is known in an authority record together with other differentiating information. Traditionally a library authority record contains sufficient information to uniquely identify the name in a library’s catalog, so will be more likely to contain differentiating information such as dates of birth , known affiliations, etc. in records for common names than in records for unusual names. The goal is to ensure that an author search on the catalog will retrieve all and only those works associated with that author. However, since it is common for libraries to share a common authority file, such as the international Library of Congress/[Name Authority Cooperative Program (NACO)](http://www.loc.gov/aba/pcc/naco/index.html) Name Authority File, an authority record may have to uniquely identify a name in the context of hundreds of libraries’ catalogues representing millions of works and millions of authors.

Traditional library practice is to select a “preferred name form” that is used in the library’s catalog. This preferred name form may differ from one community and language to another. For example, “Confucius” is used in Anglo-American communities and “孔子” in Chinese, Japanese and Korean communities. The potential to link between different authority files which may have very different preferred forms for the same name has been demonstrated by the [Virtual International Authority File (VIAF)](http://viaf.org). It uses the differentiating data contained in library authority records and the associated works they are linked to in library catalog records to cluster the authority records together and then assigns that cluster a unique identifier. Authority files and the catalogs with which they are associated are a rich resource of curated data that can support data linking and semantic web applications.

A registration file and an authority file may serve two distinct functions. A registration file strives to create a unique identifier for a given entity. An authority file, while doing the same, may impose additional constraints such as formulating the text string associated with the entity according to specific rule sets (e.g., [Resource Description and Access](http://www.rda-jsc.org/rda.html) or RDA), including variant forms of name, identifying a preferred form of name, adding data about the entity, or linking multiple identities for the same entity (such as pseudonyms).

## Executive summary and recommendations

The functional requirements for registering researchers and our associated recommendations depend on who the stakeholder is. From eighteen use case scenarios[[2]](#endnote-2) we identified seven stakeholders: researcher, funder, university administrator, journalist, librarian, identity management system, aggregator (including publishers). An individual stakeholder may assume more than one role (e.g., a librarian and an identity management system may both need to disambiguate names.) Their needs are summarized in Table 1.

*Table 1: Stakeholders and needs*

|  |  |
| --- | --- |
| Researcher | Disseminate research |
| Compile all publications and other scholarly output  |
| Find collaborators |
| Ensure network presence correct |
| Funder | Track funded research outputs |
| University administrator | Collate intellectual output of their researchers |
| Journalist | Retrieve all output of a specific researcher or track a given discipline |
| Librarian | Uniquely identify each author  |
| Identity management system | Associate metadata, output to researcher |
| Disambiguate names |
| Link researcher's multiple identifiers |
| Disseminate identifiers |
| Aggregator (includes publishers)  | Associate metadata, output to researcher |
| Collate intellectual output of each researcher |
| Disambiguate names |
| Link researcher's multiple identifiers |
| Track history of researcher's affiliations |
| Track & communicate updates |

The criteria for selecting which of the various identifiers to use will depend on the stakeholder. Whatever identifier systems attract the “critical mass” representing the peers you wish most to track are the one(s) you should consider using yourself. Key recommendations for each stakeholder follow.

**Researcher:**

* Obtain a persistent identifier before submitting any output. Ask your librarian or university administrator if you are unsure which identifiers are most suitable or don’t know how to get one.
* Disseminate your persistent identifiers on all external communications—faculty profiles, email signature, professional networks, LinkedIn, or anywhere you communicate with your peers.
* Include the ISNI of your organization(s) and funders in the research output that you submit. Search [isni.org](http://isni.org/search). If your organization does not have an ISNI, it can request one through an ISNI Registration Agency.[[3]](#endnote-3)
* Report errors on your metadata (affiliations, attributions, etc.) or if you’re represented in the same system more than once to your librarian or university administrator if you can’t correct the error yourself**.**

**Librarian/University administrator:**

* Assign persistent identifiers to authors if they don’t already have one. This includes electronic dissertations in institutional repositories, papers or datasets uploaded to research websites, and articles to journal aggregators.
* Integrate researchers’ external identifiers within library applications and services as appropriate.
* Tout the benefits and reasons for researchers to register, use and diffuse their identifiers.
* Find out from your identity management system or aggregator provider how to report errors.
* Provide guidance and training materials on why using persistent identifiers is important, good practices on where to include them and how to report errors.

**Funder:**

* Insist that all researchers who receive grants have and use a persistent identifier.

**Identity management system/Aggregator:**

* Design your system so that the provenance or source (organization or agency) of each data element is tagged. This information is important for your system users to assess the “trustworthiness” of the information displayed, especially when you have similar information from multiple sources. You’ll need this information to pass on error reports or corrections.
* Establish maintenance mechanisms to:
	+ Correct information about a researcher.
	+ Merge identities representing the same person.
	+ Split entities representing different researchers.
* Establish protocols to communicate changes and corrections to the original source
* Create framework to identify privacy and rights issues. Be willing to share information for matching information between different systems even if the information is not displayed such as birth dates.
* Support batch searching and updating. Enable organizations to export thousands of names at a time to obtain identifiers.
* Address interoperability of standards for both formats and data elements.
* Include the identifiers used in other systems.
* Link researcher identifiers to the institutions or agencies they are affiliated with.

## Challenges

Uniquely identifying each researcher so that each individual can be associated with his or her scholarly output faces these challenges:

* A scholar may be published under many forms of names. Abbreviated given names used in journal articles are generally absent in national authority files and authors that publish only in journal articles may not be represented in authority files at all. If a scholar’s work is translated, the transliteration of the scholar’s name in non-Latin scripts— such as Arabic, Chinese, Cyrillic, Hebrew, Japanese *katakana*, Korean *hangul,* or any of the Indic scripts— makes it difficult to rely on text string matching to determine if two authors represent the same person or not.
* Multiple people can share the same name, requiring additional attributes or metadata to distinguish them such as discipline or research topics, institutional affiliations, or links to publications. This is particularly true for Chinese names, where 87% of the population in China shares just 100 family names (compared to the United States where 90% of the population uses 151,671 family names). 270 million Chinese have the family name of Li, Wang, or Zhang—and that’s not counting all the overseas Chinese.[[4]](#endnote-4)
* Some researchers already have multiple profiles or identifiers, which may not be linked. A researcher may have profiles or identifiers in systems such as Academia, Google Scholar, ISNI (International Standard Name Identifier), Mendeley, Microsoft Academic, ORCID (Open Researcher and Contributor ID), ReseachGate, Scopus, VIAF (Virtual International Authority File), and VIVO as well as be represented in the institution’s CRIS (Current Research Information System). The scholar’s web presence may thus be fragmented. Sometimes scholars deliberately maintain distinct identities (e.g., publishing in different subject areas, writing under pseudonyms, etc.). Privacy control is an additional layer of complexity to consider when developing mechanisms for associating identifiers.
* Information related to a researcher or the researcher’s scholarly output that is updated in one system may not be reflected in other systems that include the researcher’s work. The current researcher ID information flow represents a complex ecosystem, as illustrated by Dr. Micah Altman’s diagram (see *Figure 1.)*
* Interoperability of standards among different identifier systems for both formats and data elements is a huge challenge.

## Possibly emerging trends

This field is changing so quickly that it is hard to tell whether a couple of examples represent isolated occurrences or indications of an emerging trend. The task group has identified the following as “possibly emerging trends”.

* Acknowledgement that the need for persistent identifiers for researchers has become widespread. We are seeing increased use of both ISNI and ORCID identifiers to disambiguate names. More broadly, Wikipedia, search engines such as Google and the open web community have been investing efforts into disambiguating names.
* Registration files are being used more than authority files to identify researchers.
* Universities are assigning identifiers to researchers. We have noted five different approaches:
	+ Assigning ORCIDs to authors when submitting electronic dissertations in institutional repositories. (Harvard)
	+ Automatically generate preliminary authority records from publisher files (Harvard pilot).
	+ Assigning ISNI identifiers to all university researchers (LaTrobe)
	+ Assigning local identifiers to researchers (Stanford’s Community Access Profile)
	+ Using UUIDs (Universally Unique Identifiers) to map to other identifiers like ORCID (Oxford)
* The increasing number of open data or public access mandates, with the call that publicly-funded research be accessible to all, will also increase the demand for researchers to have—and use—persistent identifiers.
* National programs have emerged to register all their researchers, such as the Dutch Digital Author Identification (DAI) system and the Lattes Platform in Brazil.
* More researchers will have multiple identifiers in multiple systems.
* Researcher websites are asking participants to have ORCIDs. Academic open source environments have started to integrate researchers’ identifiers into their platforms, such as the ORCID Adoption and Integration (A&I) Program.[[5]](#endnote-5)
* Recognition that there is no *one* central authority file is growing. Recently the Program for Cooperative Cataloging is considering changes to include references to other systems other than LC/NACO.
* Publishers have started to mark up their websites in [schema.org](http://schema.org/), allowing more linking between library and non-library domains.
* Interoperability between systems is increasing:
	+ ISNIs may be automatically added to LC/NACO authority records.
	+ ISNI and VIAF have established interoperability procedures.
	+ ORCID and ISNI are coordinating their services. ORCID now includes organization identifiers to be cross-walked with their ISNI organization identifiers.[[6]](#endnote-6) The ORCID and ISNI boards recently signed a Memorandum of Understanding defining forms of interoperation, investigating synergies and differences between their systems, and how to share or link identifiers. ORCID has released a beta lookup system to search and retrieve ISNI identifiers while inside ORCID.[[7]](#endnote-7)
	+ More Current Research Information Systems (CRIS) are integrating with VIVO.
* More metadata formats represent researchers and their output, including: MARC, XML, MODS, MADS, ZThes, TEI, Common European Research Information Format (CERIF), VIVO RDF. Most if not all CRIS systems in the United Kingdom are CERIF-compliant. More organizations and universities are joining the Consortia Advancing Standards in Research Administration Information (CASRAI) to jointly develop a common data dictionary and advance best practices for data exchange and reuse.
* More machine processing is surfacing the need for more standardization of both formats and data elements to represent the metadata about researchers.

## Methodology

The OCLC Research Registering Researchers in Authority Files Task Group comprises specialists from The Netherlands, United Kingdom and the United States with different perspectives: researcher, librarian, publisher, ORCID or ISNI Board member, LC/NACO contributors, Program for Cooperative Cataloging, VIAF, VIVO, CRIS system (Symplectic) and a national researcher system. The group communicated through conference calls, email and shared documents and ideas using a web-based project management and collaboration tool, Basecamp.

The group developed eighteen use case scenarios around the seven types of actors: researcher, funder, university administrator, journalist, librarian, identity management system, aggregator (including publishers). For this work, we defined researcher as “anyone who produces or is in the process of producing scholarly output.” By “registering” we mean assigning a persistent identifier—a unique reference to a single individual that does not change over time or between systems—usually with some attributes such as discipline, institutional affiliation, or titles of the individual’s publications.

The group derived a set of functional requirements from the use case scenarios. We augmented a list of 100 research networking/identifier management systems compiled by Dr. Michael Conlon[[8]](#endnote-8) and selected 20 systems to profile based on two criteria:

1. The system had to have significant takeup or “mind-share” by researchers.
2. The system had to represent researchers with a persistent, unique and publicly accessible URI.

We wanted to end up with a representative sample of different researcher information/identifier systems. For some system types we profiled only one system to represent a category. We mapped each profile to the functional requirements and identified gaps (functional requirements not met) and overlaps. Dr. Michah Altman diagramed the relationships among the systems. We solicited external feedback on our use case scenarios, system profiles, functional requirements, and the researcher ID workflow diagram as “work in progress” in July 2013.

From this research, we identified changes in the field, possibly emerging trends, opportunities, and recommendations targeted for each stakeholder.

## Landscape overview

The 20 systems we profiled fell into 10 categories or types:

* Authority hubs, providing a centralized location of authority records for multiple institutions (6): Digital Author Identifier (DAI) in The Netherlands; Lattes Platform in Brazil; LC/NACO Authority File; Names Project in the United Kingdom; ResearcherID; Virtual International Authority File (VIAF)
* Current Research Information System (CRIS), which stores and manages data about research conducted at an institution and integrates it with data from external sources: Symplectic
* Identifier hubs, providing a centralized registry of identifiers (2): International Standard Name Identifier (ISNI); Open Researcher and Contributor ID (ORCID)
* National research portal, providing access to all research data stored in a nation’s network of respositories: National Academic Research and Collaborations Information System (NARCIS) in The Netherlands
* Online encyclopedia, a compendium of information divided into articles which includes references to the works by scholars: Wikipedia
* Reference management, a system to help scholars organize their research, collaborate with others, and discover the latest research: Mendeley
* Research and collaboration hub, a centralized portal where scholars in a particular discipline can work together: nanoHUB
* Researcher profile systems, networks that facilitate professional networking among scholars: Community of Scholars; Google Scholar; LinkedIn; SciENcv; VIVO
* Subject author identifier system, a registration service to link scholars with the records about the works they have written: AuthorClaim
* Subject repository, a discipline-based centralized repository to facilitate scholarly exchanged in the fields covered: arXiv

Each profile includes: URL for the site; year started; purpose; description; scope; sources; content; size; who it is used by; public functions; restricted functions; which other research network or identifier systems it interoperates with; overlaps with other systems; whether it supports linked data; access methods; metadata schema; licenses; fees; responsible agency; references.[[9]](#endnote-9) Some of the information was not available publicly, and where possible we interviewed the system’s developers to provide as complete a profile as possible.

Comparisons were more meaningful where we had profiled more than one system within a category. For example, the comparison of just the scope and size of authority and identifier hubs illustrates a key difference between those that focus *only* on researchers and those that are very broad, with millions of people represented, where we cannot tell how many researchers are included. The Lattes Platform is the largest of the researcher-only databases profiled; it includes Brazilian researchers in all disciplines plus those outside Brazil who work with them. The total number of researchers worldwide is estimated to be about nine million in 2012[[10]](#endnote-10).

A link to the spreadsheet compiling all the characteristics of all the system profiles is included in the list of supplementary data.

*Table 2: Scope and size of Authority and Identifier Hubs profiled – size as of Dec 2013*

|  |  |  |
| --- | --- | --- |
| Hub | Scope | Size |
| Digital Author Identifier | Researchers in all Dutch CRIS & library catalogs | 66K |
| Lattes Platform |  Brazilian researchers and research institutions | 2M people, 4K inst. |
| ISNI  | Data from libraries, open source resource files, commercial aggregators, rights management organizations. Includes performers, artists, producers, publishers  | 7M total; 720 K researchers |
| LC/NACO Authority File | Persons, organizations, conferences, place names, works  | 9M total; ? researchers |
| ORCID | Individual researchers plus data from CrossRef/Scopus, institutions, publishers | Over 500K |
| ResearcherID | Researchers in any field, in any country | 250K |
| VIAF | Library authority files for persons, organizations, conferences, place names, works  | 26M people; ? researchers |

These authority and identifier hubs overlap. For example, ISNI is a VIAF contributor, and has loaded and matched records from the Dutch DAI. LC/NACO Authority File records occasionally include ISNI or ORCID identifiers and are also contributed to VIAF. There are some key differences in approaches. ISNI consolidates data from multiple databases while ORCID requires researchers to self-register.

Systems characterized as Current Research Information Systems and researcher profile systems similarly include those that focus *only* on researchers, while those like Google Scholar and LinkedIn serve millions of people with an unknown percentage representing researchers.

*Table 3: Scope and size of CRIS and Researcher Profile Systems profiled—size as of 2013*

|  |  |  |
| --- | --- | --- |
| Profile System | Scope | Size |
| CRIS (e.g. Symplectic) | Institutional data sources plus article & citation, reference databases, bibliographies. Author names, identifiers, affiliations, bibliography, grants, professional activities  | > 100K |
| Community of Scholars | Researcher-created profiles plus over 70 article & citation databases | > 3M profiles |
| Google Scholar | Authors of publications in commercial, institutional, & web sources indexed by Google Scholar. | Millions; ? "verified" research profiles |
| LinkedIn | International user base; profiles created by individuals & organizations. | 200M; ? researchers  |
| SciENcv | National Institutes of Health pilot to link researchers to their grants and output. Data from federal and non-federal sources. Embedding ORCIDS | Planned: >100K |
| VIVO | Sources: funding agencies, institutions, scholar self-reports, open source & publisher data. Profiles include all scholarly output & research, teaching, service activities | 1M people, 10M publications |

Current Research Information Systems like Symplectic pull information from a variety of abstract & citation databases such as ArXiv, CiNii, DBLP, figshare, Google Books, Mendeley, ORCID, PubMed, RePEc, SciVal, Scopus and Web of Science. Community of Scholars includes content from over 70 ProQuest and CSA proprietary databases and other certain verified publications such as ABI/INFORM, ERIC and PubMed; information is also pulled from the scholar's personal and institutional websites, if it exists. SciENcv is a planned pilot to provide a shared, voluntary researcher profile system for individuals who receive or are associated with research investments from federal agencies.  Embedded ORCID identifiers in the grant application workflow will serve to better track research and understand the impact of the National Institutes of Health’s funding.

The same information about a specific researcher may thus be represented in multiple databases, and only a subset interoperates with each other. Altman’s diagram of the researcher ID information flow illustrates the complexity of the current ecosystem. The ovals on the left show the various stakeholders: libraries, publishers, individual researchers, ISNI registration agencies, ORCID and VIVO members, national research institutions. The dotted lines show the information flow among them, for example, book publishers provide information to libraries. The drums in the centers represent various types of aggregated databases: Virtual International Authority File, ISNI, ORCID, CrossRef, National platforms, VIVO. The lower right represents institutional-based databases such as Harvard Profiles, Stanford’s Community Academic Profiles, an institution’s CRIS or Institutional repository. Note that these operate separately from the libraries’ work on authority records on the top left. The information flows to either controlled information sources on the far right, or uncontrolled information sources on the Internet shown on the top (Google Scholar, LinkedIn, Mendeley). [Micah will provide more info about the diagram, which may be revised.]

A key question is how corrections or updates can be communicated between systems. Researchers are frustrated when they see errors in their profile or works incorrectly assigned to them, or works missing. Even if the information is corrected in the local instance, it often does not get reflected in the aggregated databases or hubs.



*Figure 1: Researcher ID information flow*

## Changes in the field

Since the task group started its work in September 2012, we have seen a number of changes in the field:

* JISC stopped funding The Names Project, encouraging researchers to get ORCIDs instead.
* Elsevier acquired Atira (a CRIS system) and Mendeley.
* Thomson Reuters acquired AVEDAS, the supplier of CONVERIS, another CRIS system.
* Elsevier and University College London launched the “UCL Big Data Institute” (building on Elsevier’s acquisition of Mendeley), a new innovation lab to “**tackle the challenges researchers face as they seek to forecast trends, synthesize information from thousands of research papers, and show the potential societal impact of their research so it will be eligible for funding.”**[[11]](#endnote-11)
* ISNI continues to increase the number of researchers represented, now including identities from 12 research-based sources as of January 2014: American Musicological Society, AuthorClaim, British Library Theses, Digital Author Identifiers (DAI, Dutch), JISC Names (UK), Modern Languages Association, OCLC Theses, ORCID and DataCite Interoperability Network (ODIN), Proquest Theses, RePec, Scholar Universe and Electronic Tables of Contents (ZETOC).
* ISNI officially submitted a “notice of inquiry” to the United States Copyright Office to make ISNI part of the author’s process of obtaining copyright.[[12]](#endnote-12)
* More research sites are using or experimenting with ORCID to identify their researchers, including PubMed, nanoHUB and SciENcv.
* Wikipedia has been adding ORCIDs, ISNIs, and VIAFs via an “authority control template[[13]](#endnote-13)”; VIAFs and ISNIs have been added to Wikidata (the source for all language-Wikipedias).[[14]](#endnote-14)
* The Library of Congress’ Bibliographic Framework Initiative has proposed as its model for the future linking local authority files to external authority sources.
* VIVO or Harvard Profiles (using VIVO format) has become more widespread, with more than 100 institutions in the US plus universities in Australia and China

## Functional requirements by stakeholder

For researchers and university administrators

* Link multiple identifiers a researcher might have to collate output.
* Associate metadata with a researcher’s identifier that resolves to the researcher’s intellectual output.
* Verify that a researcher and his or her scholarly output is correctly represented.
* Register a researcher who does not yet have a persistent identifier.

For librarians

* Create consistent and robust metadata through manual or automatic means.
* Associate metadata for a researcher’s output with the correct identifier.
* Disambiguate similar results.
* Merge entities that represent the same researcher and split entities that represent different researchers.

For funders and university administrators

* Link metadata for a researcher’s output to grant funder’s data.

For aggregators and identity management systems

* Make pre-existing authoritative institutional identifiers in systems (such as institutional Internet Access, eMail, Single-Sign-On, Finance or Human Resources) interoperable with external systems such as ISNI, ORCID or VIAF.
* Link a researcher’s multiple identifiers.
* Determine whether authority record represents a single identity.
* Affiliate a researcher with multiple departments, institutions, disciplines.
* Track history of a researcher’s various affiliations over time.
* Merge entities representing the same researcher & split entities representing different researchers.
* Communicate information to other systems.
* Tag each data element with its source or provenance.
* Support batch searching and updating.
* Support Unicode to record researchers’ names and citations in their own language and writing system.

## Opportunities for next steps

The desire to compile and present to the Internet community all a scholar’s output offers new opportunities. Use of identifiers rather than text strings to represent a researcher could enable a more comprehensive compilation of the researcher’s work. We have already seen examples of ISNI contributing authors as cited in journal articles to VIAF. The widespread use of persistent identifiers could provide the means to link and integrate all scholarly output, including those using non-traditional channels such as datasets, research data, e-prints, presentations, survey results and academic use of social media like blogs.

The changing landscape offers the opportunity to reconsider what the role of authorities is in an ecosystem of registered identifiers. Universities may wish to consider automatically registering researchers and assigning persistent identifiers from the university’s local faculty files. This is an opportunity to shift to more “dynamic” authorities. Rather than wait until you have sufficient information to create a national authority heading, generate a stub with an identifier that someone else can augment or match and merge with other metadata.

Batch-querying identifier databases would allow more linking of a researcher’s multiple identifiers. We believe it’s unlikely that there will ever be one, comprehensive source for all researchers, so the ability to communicate information among systems becomes crucial.

We think there is a huge opportunity for third-party reconciliation or resolution services to provide linking among different identifier systems rather than have each institution attempting to do it on its own.

## Supplementary data

This report summarizes the results of the research conducted by the OCLC Research Registering Researchers in Authority Files Task Group in 2012-2013. Details are in the following supplementary data sets. [Links to be provided in final report; see “[Work in Progress](http://www.oclc.org/research/activities/registering-researchers/progress.html)” page for earlier versions of the first five.]

**Use Case Scenarios:** The eighteen use-case scenarios the task group developed around different stake-holders: researcher, funder, university administrator, journalist, librarian, identity management system, aggregator (including publishers).

**Functional Requirements:** The 43 functional requirements the task group derived from the use-case scenarios.

**Research Networking Systems:** The list of 100 research networking and identifier systems the task group considered, with the ones selected for profiling highlighted. The task group augmented a list originally provided by task group member Dr. Michael Conlon.

**The Researcher ID Information Flow:** Task group member Micah Altman’s diagram of the researcher ID information flow illustrates the complexity of the current ecosystem.

**Research Networking Systems Characteristics Profiles:** The 20 research networking systems the task group characterized.

**Compilation of Research Networking Systems Characteristics Profiles:** The spreadsheet with all 20 research networking systems’ profiles for easy comparison.

**Compilation of Research Networking Systems Mappings to Functional Requirements:** Each of the 20 research networking systems’ characteristics are mapped to the functional requirements in this spreadsheet.

**References**

1. [https://web.archive.org/web/20140301191458/http://www.iu.qs.com/university-rankings/rankings-indicators/methodology-citations-per-faculty/](https://web.archive.org/web/20140301191458/http%3A//www.iu.qs.com/university-rankings/rankings-indicators/methodology-citations-per-faculty/) [↑](#endnote-ref-1)
2. Link to finalized version of our use case scenarios Word document compilation [↑](#endnote-ref-2)
3. ISNI Registration Agencies are listed at: <http://isni.org/do-you-have-an-isni> [↑](#endnote-ref-3)
4. [https://web.archive.org/web/20140111223125/http://blog.foreignpolicy.com/posts/2013/04/26/why\_do\_so\_many\_chinese\_people\_share\_the\_same\_name](https://web.archive.org/web/20140111223125/http%3A//blog.foreignpolicy.com/posts/2013/04/26/why_do_so_many_chinese_people_share_the_same_name) [↑](#endnote-ref-4)
5. [https://web.archive.org/web/20140112012510/https://orcid.org/content/adoption-and-integration-program](https://web.archive.org/web/20140112012510/https%3A//orcid.org/content/adoption-and-integration-program) [↑](#endnote-ref-5)
6. [https://web.archive.org/web/20140112013110/http://support.orcid.org/knowledgebase/articles/276884](https://web.archive.org/web/20140112013110/http%3A//support.orcid.org/knowledgebase/articles/276884) [↑](#endnote-ref-6)
7. “ISNI and ORCID sign Memo of Understanding”, January 2014. <http://www.isni.org/content/isni-and-orcid-sign-memo-understanding> [↑](#endnote-ref-7)
8. Link to the research networking systems supplementary data set. [↑](#endnote-ref-8)
9. Link to the compilation of the research networking systems profiles supplementary data set. [↑](#endnote-ref-9)
10. The nine million researchers worldwide figure is extrapolated from the World Bank’s 2013 figure of 1,271 researchers (full-time equivalents) per million inhabitants ([https://web.archive.org/web/20140225192253/http://wdi.worldbank.org/table/5.13](https://web.archive.org/web/20140225192253/http%3A//wdi.worldbank.org/table/5.13)) and a 2012 world population of 7.046 billion ([https://web.archive.org/web/20140225193314/http://wdi.worldbank.org/table/2.1](https://web.archive.org/web/20140225193314/http%3A//wdi.worldbank.org/table/2.1)). The World Bank’s definition of researcher: “ Professionals engaged in the conception or creation of new knowledge, products, processes, methods, or systems and in the management of the projects concerned. Postgraduate PhD students (ISCED97 level 6) engaged in R&D are included.” [↑](#endnote-ref-10)
11. [https://web.archive.org/web/20140112043805/http://www.elsevier.com/connect/university-college-london-and-elsevier-launch-ucl-big-data-institute](https://web.archive.org/web/20140112043805/http%3A//www.elsevier.com/connect/university-college-london-and-elsevier-launch-ucl-big-data-institute) [↑](#endnote-ref-11)
12. <http://www.isni.org/filedepot_download/58/336>

 [↑](#endnote-ref-12)
13. The authority control template used in English-language Wikipedia at [https://en.wikipedia.org/wiki/Template:Authority\_control](https://en.wikipedia.org/wiki/Template%3AAuthority_control) has been adapted by other Wikipedias but not all. [↑](#endnote-ref-13)
14. See for example the “Statements” on the Wikidata page for Noam Chomsky at [https://web.archive.org/web/20140112045216/http://www.wikidata.org/wiki/Q9049](https://web.archive.org/web/20140112045216/http%3A//www.wikidata.org/wiki/Q9049) [↑](#endnote-ref-14)