Project Descriptions/Background Information

The Management of Scholarly Identity

A CNI Workshop

Sheraton Inner Harbor, Potomac Room
Baltimore, MD
April 4, 2012

1. arXiv (ORCID)
Simeon Warner

I have been involved with arXiv (http://arxiv.org) since 1999. For many years we have wanted to link author information across arXiv and the services of our partners ADS and INSPIRE (formerly SPIRES) to give more complete linking across high-energy physics and astrophysics. In 2009 we introduced opt-in public author identifiers (http://arxiv.org/help/author_identifiers) at arXiv that support public profiles and a "my articles" widget that users can embed in their web pages. There has been steady uptake. We also experimented with Facebook integration based on author identifiers.

I am a member of the board of directors of ORCID Inc. (http://orcid.org), which aims to create an author identifier registry service used throughout the scholarly domain. A phase one self-claim service will go online this summer.

[Amy Brand, Assistant Provost for Faculty Appointments, Office of the Provost, Harvard University, who was unable to attend the workshop, provided the following description of ORCID]
ORCID is an acronym that stands for Open Researcher and Contributor ID. ORCID, Inc. ORCID was formed in 2010 and is operated exclusively for the charitable purpose of enhancing research attribution, navigation, and discovery. ORCID will accomplish this through the creation of a global registry of unique, persistent, public identifiers for individual researchers.

As the body of existing scholarly research continues to grow and globalize, it is becoming increasingly difficult to reliably identify the author(s) of specific research contributions. The confusion over author identity arises from a variety of sources including: (i) the existence of common names like John Smith or Mary Jones, (ii) name changes upon marriage and divorce, effectively splitting a researcher's output record in two, (iii) widely varying style rules among journals; and (iv) the explosion of research output in countries that do not use the roman alphabet.

Attribution ambiguity thwarts the ability of researchers, institutions, scholarly publishers, and funders to track authorship, collaboration, and the relationship between investment and impact, and to navigate the research literature more generally. ORCID will resolve the author ambiguity problem by creating a numerical identification system for researchers, with an open online

registry linking authors to their scholarly works. A version of the database will be available to the entire scholarly community and the public free of charge.

ORCID is uniquely positioned to accomplish this task. Generally, the barrier facing the creation of such an identity system has been obtaining a critical mass of profiles that will make the system useful. Reluctance to use existing identity systems has come from a variety of factors: lack of trust in the platform and organization running the identity system, identity system parochialism, the amount of work involved in researchers setting-up and maintaining a profile, and the lack of tangible benefits from registering a profile. ORCID will overcome these factors by:

- Being a transparently governed non-profit organization that has gained the trust of the research and university community and whose data and source code is available under recognized open licenses;
- Having a board and advisory committees with a representative sample of stakeholders (Under ORCID's by-laws, the majority of board members will always be from nonprofit organizations);
- Creating an identity system that transcends discipline, geographic, national and institutional boundaries;
- Encouraging institutions to seed ORCID profiles on behalf of their researchers and (given the researcher's permission) to manage those profiles;
- Encouraging third parties (including institutions, funding agencies and publishers) to integrate ORCID profiles into their manuscript/grant management and research assessment systems in order to make researcher use of those systems easier and more efficient; and
- Creating an open and transparent linking mechanism between ORCID and other current author ID schemes.

2. Bowker (ISNI)
Beat Barblan

Bowker and the ISNI Standard

Bowker /ProQuest is one of the Founding Members and the very first Registration Agency of the International Standard Name Identifier (ISNI), a new ISO-certified identifier for the identification of "public identities" associated with intellectual works. The initial ISNI database was launched last fall with data primarily from VIAF and currently has roughly 1 million assigned ISNIs and several million records ready for processing and assignments.

Bowker and ProQuest are contributing name data from Books In Print, ProQuest Dissertations, and Scholar Universe. We are also working with societies such as the American Musicological Society and Authors' Guild to register names for their members.

Beat Barblan is on the ISNI Board and runs Bowker's Identifier Services business unit, which includes the ISBN, ISTC and DOI agencies. Bowker just became the first ISNI Registration Agency (press release shortly). The timing is especially great since the ISNI Board and ORCID representatives met just last week to explore/discuss options for close collaboration between the two organizations. Also, after going through all the necessary (and very long) processes

required, the International Organization for Standardization will finally publish the ISNI standard on March 15th.

3. California Digital Library
Patricia Cruse

I am participating in CNI's Management of Scholarly Identity Workshop as the Director of the University of California Curation Center (UC3) at the California Digital Library. UC3 has developed a range of services that support data intensive scholarship – from a data management planning tool, to an identifier service, to a repository for managing, sharing, providing access to, and finally, to publishing data. Underpinning these services is the recognition that researchers must have the means to control, track, protect, cite, share, and get credit for their research. Key in these services is the need to clearly identify every person (creator, editor, reviewer, aggregator, etc.) involved in these activities. Services that support people identification must be scalable and able to respond as people move throughout their careers. The following are some of the ways we are approaching this challenge.

Making data intensive research citable is crucial in enabling researchers to control and get credit for their scholarly output. UC3 is an active member of the DataCite Consortium which is creating global citation services for data making data easier to find, access, reuse, and repurpose. DataCite and ORCID are proposing (negotiating with funder) to build an interoperability network which will make it possible to 1) directly reference data 2) track use and reuse of data 3) create links between data, subsets, articles, rights statements, and 4) create links to all people involved in the data object's life-cycle.

UC3 delivers DataCite services via EZID, our identity service. EZID's infrastructure is designed to scale to include the identification services for people and organizations. Readiness for ORCID and for describing persons and organizations is part of the reason EZID assigns ARKs (archival resource keys - indentifiers) to EZID customers. A component of the EZID identity service is the NAAN Registry (Name Assigning Authority Numbers), which assigns a unique identifier to organizations. UC3 is an active partner in the NSF funded DataONE initiative. DataONE is a consumer of the EZID service and we expect as DataONE's user community grows internationally there will be an increasing need to leverage identity services for people. DataONE's user community includes faculty, researchers, government agencies, NGOs, societies, etc. so an identity service must meet the needs of this diverse community.

Along with an infrastructure needed to support identity services we also must develop organizational, social, and cultural practices that support identity services for people. I co-chair the DataONE Sustainability and Governance Working Group and we are actively examining data governance issues that support data attribution, citation, reuse, and repurposing. In December DataONE organized a data governance workshop to work through some of these issues. We expect that our findings will be widely shared with the community and put into practice via DataONE.

4. Cornell University (VIVO)
Dean Krafft

VIVO is an open source semantic web application originally developed and implemented at Cornell. When installed and populated with researcher interests, activities, and accomplishments, it enables the discovery of research and scholarship across disciplines at that institution and beyond. VIVO supports browsing and a search function which returns faceted results for rapid retrieval of desired information. Content in any local VIVO installation may be maintained manually, or it may be brought into VIVO in automated ways from local systems of record, such as HR, grants, course, and faculty activity databases, or from database providers such as publication aggregators and funding agencies.

The rich, semantically structured data in VIVO support and facilitate research discovery. Examples of applications that consume these rich data include: visualizations; enhanced multi-site search through VIVO Search; applications such as VIVO Searchlight, a browser bookmarklet which uses text content of any webpage to search for relevant VIVO profiles; and the Inter-Institutional Collaboration Explorer, an application which allows visualization of collaborative institutional partners, among others.

The VIVO Ontology is a linked open data standard that supports the interchange of information about researchers, their publications, their activities and interests, and their organizational and professional context. The Clinical and Translational Science Award (CTSA) Consortium Steering Committee recently encouraged institution-wide adoption and implementation of research networking tools which provide linked open data using the VIVO Ontology, and several other researcher profiling systems are already providing RDF data using the VIVO Ontology.

Reference: Dean B. Krafft, Nicholas A. Cappadona, Brian Caruso, Jon Corson-Rikert, Medha Devare, Brian J. Lowe, and VIVO Collaboration, (2010) *VIVO: Enabling National Networking of Scientists*. In: *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, April 26-27th, 2010, Raleigh, NC: US. http://journal.webscience.org/316/

Dean B. Krafft is the Chief Technology Strategist and Director of IT for the Cornell University Library. His interests include research data management and curation, digital archiving and preservation, and the use of semantic web technologies to support the discovery of and access to research, researchers, and scholarly information resources. He is the Principal Investigator for the Cornell University subcontract on the NIH-funded VIVO researcher profiling project (http://vivoweb.org).

5. DuraSpace
Bradley McLean

DuraSpace supports the communities and manages the code for the DSpace turnkey institutional repository and Fedora digital repository framework open source software efforts, and develops and operates the DuraCloud preservation infrastructure and services project. Both of the repository systems either implicitly create (or consume references to) scholarly identities as part of their intrinsic access control, ingest workflow, attributes of repository items, and display of repository items. DSpace is currently in use at over 1,200 institutions, and Fedora components are used at more than 200.

DuraSpace's interest is in supporting a reasonably bounded set of identity standards and integrations with systems recognizing scholarly identity across its repository and preservation platforms.

Bradley McLean is the Chief Technology Officer of DuraSpace.

6. Elsevier
Chris Shillum

In my role at Elsevier, I am responsible for the Scopus Author Profiling system, which has attempted to automatically disambiguate authorship among the 46 million bibliographic records contained in Scopus, a large abstracting and citation database. The resulting author profiles are made available in Scopus itself, are available for license as custom data, and are also being used as a key data source in SciVal, Elsevier's suite of tools and services offered to research institutes and funding agencies to assist with research planning and performance analysis.

I am also a member of the board of ORCID (www.orcid.org), a community-wide, non-profit initiative aiming to solve the author/contributor name ambiguity problem in scholarly communications by creating a central registry of unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID and other current author ID schemes.

7. Internet2/InCommon
Kenneth Klingenstein

Internet2 has been active in the development of federated identity, through InCommon and Shibboleth, and in the development of Internet-scale access control, through schema such as eduPerson and software tools such as Grouper. We have also been deeply engaged with international R&E federations and US Gov activities in this area.

We have a broad set of interests in a broadly scoped definition of scholarly identity. These include:

- Attribute management for collaboration – creating bundles of attributes (e.g. the Research and Scholarship Bundle) and including key attributes (such as the ORCID identifier) in eduPerson
- Cyberinfrastructure – bridging federated logon with national computational resources such as Cilogon – www.cilogon.org, and addressing Science Agency data set access controls
- Collaboration platforms – providing open-source platforms of integrated (identity, access control, provisioning) applications and addressing the unique needs of scholarly processes such as ad hoc enrollment services
- Integration around the scholarly record – working with LTI, VIVO, on campus infrastructure use, trusted citations, ScienCV, etc.
- Leverage, leverage, leverage – there is considerable potential for powerful leverage – using the ORCID identifier for account linking, using InCommon metadata and business processes to support scholarly identity, aligning business models, etc.

Ken Klingenstein is Senior Director for Internet2 Middleware and Security.

8. Iowa State University
Greg Davis

Iowa State University (ISU) is in the process of going live with Phase 1 of our institutional repository project. ISU is using the Digital Commons platform from BePress. We have recently staffed the position of Digital Repository Coordinator for our Library (start date April 16, 2012). The first collection to be loaded into our repository will be all ETDs born digital at ISU since 2006. The ISU Library has a cataloging department who manages the authority control process related to our library management system, but they do not have enough experience yet with our repository to understand authority control issues as they relate to our repository. The CNI workshop is timely for the ISU Library, as we expect the meeting to contain information about authority control issues related to institutional repositories. As our repository is in the very beginning stages of development, I expect we will be able to serve as a test bed for any new and/or promising best practice to emerge from the CNI workshop.

9. Johns Hopkins University
Jing Wang

We are in the state of identity crisis, especially in the realm of semantic web. The crisis is not due to the lack of identifier, but because of too many of them. Uniform Resource Identifier (URI) is critical in the semantic web layered architecture. When adopting semantic web technology to facilitate the scholar discovery and knowledge sharing, the identification of the scholar is a more complex issue than the identification of their publications. Technically, what does URI and URI resolution mean as far as scholarly identity is concerned? From the policy and privacy point of view, what constitutes scholarly identity? We are interested in an identity ecosystem framework that provides overarching standards and technology for effective management of scholarly identity across organizations. We are interested in how ORCID and existing institutional identity management system fit in this framework.

I have been working with VIVO in the past year at Johns Hopkins University. VIVO is a semantic web platform that enables the discovery of research and scholarship through the linked data of researcher interests, activities, and accomplishments. The question we encountered is what the URI of our faculty in VIVO shall resolve to. On the one hand, faculty want control over their profile as they do with their personal web page and CV. On the other hand, each faculty member has a unique identifier within the university, which is linked to a rich set of authoritative data about his/her scholar activities and accomplishments. The problem is the disconnection between the faculty controlled web profile and the authoritative data associated with their unique identifier. In addition, privacy and confidentiality are major concerns as we attempt to manage scholar profile through the linked data. We envision an identity ecosystem framework will enable effective management of scholar identity not only within the university but also across organizations.

10. JSTOR
Heidi Helminiak

The challenges and opportunities faced by ITHAKA in the area of scholarly identity are consistent with the rest of the scholarly publishing community – it is in our interest to provide accurate author information on a growing number of diverse content types, traditionally journals through our JSTOR service, but increasingly to books, primary source materials, and other scholarly content.

We are in full support and have representation to the ORCID initiative, but are interested in keeping up with and helping shape the conversation around all aspects of scholarly identity. In 2011 JSTOR launched our Current Scholarship Program, and in 2012 we are adding books to the JSTOR service. This, in addition to our more than 55 million pages of journal archival content, primary source, pamphlet collections and other monographs, means we face the growing likelihood that authors in one content set have also authored in others. Management and disambiguation of scholarly identities is important to JSTOR and would allow us to provide semantic enrichment of data, as well as expand services we provide to authors, researchers, students, etc. Such data could allow us to deliver enhanced search results or provide a dashboard to authors where they understand the impact of their published work.

Heidi Helminiak is a Product Manager at JSTOR. Her primary area of responsibility is developing features and services for end users, including researchers, faculty and authors.

11. Library of Congress
Ann Della Porta, Kevin Ford

Ann Della Porta is the chief of the Integrated Library System Program Office at LC, where she is responsible for the management and operation of the Library's integrated library management system, electronic resource management system, openURL resolver, Encoded Archival Description (EAD) finding aid system and LC's handle server. In her previous position at LC she served as coordinator for cooperative cataloging projects and the secretariat for the Program for Cooperative Cataloging (PCC). She managed the NACO, SACO and BIBCO programs. Under her direction NACO participation tripled from approximately 100 NACO institutions to over 300 and several new international partners joined the program. In her current position she works with other LC units and the NACO partners to implement major changes in authority data exchange in support of that program. Ms. Della Porta is currently working on a demonstration project to implement Shibboleth at LC.

Kevin Ford works in the Network Development and MARC Standards Office (NDMSO) at the Library of Congress where he administers LC's Linked Data service, ID.LOC.GOV. The ID service publishes the entire LC Name Authority File (8 million bibliographic identities) as linked data. Kevin also participated in the development of MADS/RDF, which provides a more granular way to describe library authority data, such as personal or corporate names and their relationships to earlier- or later-established forms for those names. More recently, Kevin became the primary technical contact for LC's NACO (Name Authority Co-Operative) partners and provides technical support for the NACO nodes (LC, OCLC, British Library, US National Library of Medicine, and SkyRiver).

12. Microsoft Research
Greg Tananbaum

I am keenly interested in understanding how to support and leverage emerging standards to develop Microsoft Academic Search (MAS) author profiles. MAS has, to date, generated nearly 20 million author profiles from the data we index. These profiles contain helpful information like a list of papers this researcher has written, his or her G and H indexes, links to and visualizations of co-authors, and so forth. Disambiguating these records, facilitating corrections, and updating data are key priorities for Microsoft Research. Additionally, MAS author profiles are open - they can be accessed and reused for noncommercial purposes. This means that MAS has the potential to play a key role in propagating any solutions that emerge from discussions such as this.

Greg Tananbaum is a special advisor to Microsoft Research on the Microsoft Academic Search project. In this capacity, he engages with the publishing and library communities to help maximize the scope and usage of the service. Greg has worked on a variety of projects with Microsoft Research in the past five years. Additionaly, he serves as a consultant to publishers, libraries, universities, and information providers as owner of ScholarNext (www.scholarnext.com). Other ScholarNext clients include SPARC, the American Heart Association, and Annual Reviews. He has been President of The Berkeley Electronic Press, as well as Director of Product Marketing for EndNote. Greg writes a regular column in Against the Grain covering emerging developments in the field of scholarly communication. He has been as an invited speaker at dozens of conferences, including the American Library Association, the Society for Scholarly Publishing, the Association of Professional and Learned Society Publishers, and Online Information UK. He holds a Master's Degree from the London School of Economics and a B.A. from Yale University.

13. National Information Standards Organization (NISO)
Todd Carpenter

The National Information Standards Organization (NISO) is the only ANSI-accredited standards development organization in the space between publishers, libraries, and software providers, each of whom play a critical role in the management of scholarly identity, either from aggregation, an identification, or a description perspective. NISO provides a neutral forum where standards, best practices and industry research are created for the media and research distribution community.

NISO has been instrumental in the creation, promotion and adoption of a variety of identification, metadata and interoperability standards. Well known standards such as Dublin Core, DOI, the MARC exchange format, SUSHI, and SERU have been standardized within NISO. Recently, NISO published a best practice on improving the single-sign-on experience for patrons of library resources (ESPRESSO) and has completed work on the Institutional ID (I2).

Internationally, NISO represents US interests within ISO on a variety of issues related to information distribution and documentation. Also, as the international secretariat for the ISO technical subcommittee on identification and description and identification, NISO has been intimately involved with the establishment and maintenance of international trade and

metatadata standards, such as ISBN, ISSN, ISAN, ISRC, and the ISTC. Specifically related to identity management, ISO recently published the International Standard Name Identifier (ISNI).

14. NCBI/NLM/NIH
Bart Trawick

Over the past 7 years, I have been responsible for developing online tools to support NIH-funded scientists and the NIH Public Access Policy. The NIH Public Access Policy requires NIH-funded scientists to archive their peer-reviewed scholarly works in PubMed Central, the National Library of Medicine's (NLM) free digital repository of biomedical and life sciences journal literature. The NLM also provides a free bibliographic tool for users of PubMed called My Bibliography. In early 2010, My Bibliography was expanded to include tools to assist scientists with NIH Public Access Policy compliance. Now scientists can use My Bibliography to make publication/grant associations, easily determine Public Access compliance for their body of work, share their bibliography with colleagues, and populate reporting documents with citations from their bibliography. My Bibliography helps to network the NIH with its funded scientists and scholarly works, and future developments for the system include possibly leveraging it as a mechanism to support the disambiguation of authors with similar names in PubMed.

During the past year I have also become involved with planning for a proposed, federal-wide profile management system called SciENCV (Science Experts Network Curriculum Vitae). SciENCV will contain the data needed for CV's, federal forms, and expertise identification as well as for many other potential uses. The system will offer access and API's for the open development of applications. It will also provide a capability for agencies, institutions, etc. to obtain data for pre-populating and validating the many reports and forms that are currently handled manually. The goals of the system are to reduce the administrative burden for researchers and institutions and to produce data that is needed to document the results of science investments.

15. National Institutes of Health
Debbie Bucci

Debbie Bucci was the Integration Services Center Program Lead at Center for Information Technology (CIT), National Institutes of Health (NIH), U.S. Department of Health & Human Services (HHS), in that role she served as the principal program manager for both the NIH Federated Identity Service and the enterprise Service Oriented Architecture (SOA) efforts. In 2008, along with her other colleagues, she received the 2008 NIH Director's Award in recognition of her work with NIH Federated Authentication, 2009 InformationWeek 500 top 20 Innovators awards and the 2011 Kantara IDDY (Identity Deployment of the Year) Award for NIH iTrust.

Debbie's current role is the IT Architect for the SciENCV project. The SciENCV mission is to create a shared, voluntary profile aggregation system for all individuals who receive or are associated with research investments from federal agencies, in order to:

Reduce administrative burden for researchers and government in federal grant submission and reporting requirements
Enable discovery about researcher expertise, employment, education, and professional accomplishments
Allow researchers to describe their contributions in their own language.

16. Northwestern University
Gary Strawn

Northwestern University has just implemented the SciVal Experts software paired with VIVO, to begin deploying a research networking platform across its faculty and schools. These platforms make heavy use of names as the primary identifier and need to be able to link with accuracy to a variety of other data sources - indexing databases, aggregators of scholarly content, library bibliographic records, institutional repositories - to yield profiles for analysis of citations and subject expertise. The Northwestern University (NU) Library, partnering is a key partner in this pilot, seeks to ensure that the implementation will be seamless, flexible, and interoperable.

The Library has been active for a long time in what has until recently been called authority control; and in the past several decades has been particularly busy with the automation of as many aspects of that work as possible. Important milestones are the generation of MARC authority records from bibliographic information starting in 1994, the loading and updating of the LCSH, LC/NACO, MeSH, ULAN and AAT authority files in the local system (with associated error detection, correction, and reporting), and the mapping of corresponding headings between LCSH and MeSH, and LC/NACO and ULAN. One important aspect of the authority loader involves the attempt by the program to identify and handle different personal authors who write under the same name.

We are in the early planning stages of a migration to a new type of library services platform, which will mean that we need to build on existing metadata relationships while also ensuring future harvesting both to ingest from other sources and to transfer into repositories and preservation systems. Attending this workshop will help us combine our traditional expertise in bibliographic formats with these newer applications that span a much broader set of standards.

17. OCLC
Thomas Hickey

Thomas Hickey is Chief Scientist at OCLC, and for the last several years he has been in charge of the implementation of the Virtual International Authority File (VIAF). VIAF is a merge of library authority records identifying people, corporations, places, works and expressions. Files from 20 countries are combined into 20 million record clusters. VIAF is moving from a research project to an OCLC supported service this year.

Dr. Hickey is also involved in some other names projects. He has served as an advisor to the UK Names project, is on the board of ORCID (Open Researcher ID), and VIAF is sharing data with ISNI (International Standard Name Identifier), which OCLC runs. OCLC also runs the digital author identification system in the Netherlands that creates identifiers for Dutch

researchers and incorporates the names into the authority system of the National Library of the Netherlands.

Beyond names, he is interested in information retrieval systems, general bibliographic processing and parallel programming. VIAF was implemented using a home grown version of map-reduce, which we are currently porting to the Hadoop framework.

18. RefWorks/ProQuest
Marie Linvill

At ProQuest, researcher and faculty profiles have been a core focus of our content and service offerings and strategy for more than a decade. Our involvement with researcher profiles began with the Community of Science (COS) Expertise database, a user-generated profile database launched 1989 with funding from the Johns Hopkins University as a means for institutions to capture information on faculty data on publications, credentials, and employment history with the aim of promoting corporate consulting. In addition, under a separate business, built in the 2002-2004 timeframe and acquired in 2006, we developed the Scholar Universe database, an editorially-created profile database, now comprising roughly 3 million faculty and researcher profiles (www.scholaruniverse.com), and underpinning our Author Resolver service and API. We learned first-hand the benefits and pitfalls of both approaches to creating and maintaining faculty profiles. These two profiles systems were then merged into a best of breed offering, which is currently incorporated into COS Pivot (Pivot.COS.com), a web-based funding and collaboration workflow tool designed specifically to meet the needs of a university research office.

In this arena, ProQuest has been an early and enthusiastic participant and supporter of efforts such as VIVO, ORCID and ISNI. In addition to this focus on profile data and listening to the needs of a university's Research Management directors, we also benefit from the valuable input from the librarian and information professional community we receive via the diverse family of ProQuest brands, which include Serials Solutions, eBrary, Bowker, and Dialog, to name a few. ProQuest is consistently ranked in the top 100 of the InformationWeek 500, an annual listing of the nation's most innovative users of business technology. For ProQuest, that technology is devoted to connected people with vetted, reliable information, and in the specific case of researcher profiles, this is for the purpose of providing funding opportunity and collaborator recommendations.

Serving as Director of Content Development, a senior product management leader for COS Pivot, and as ProQuest's designated representative of ProQuest to this workshop, business unit of ProQuest, Marie Linvill has been involved with faculty profile data and the associated challenges and opportunities since 2002. Marie is currently spearheading ProQuest's initiative around data management tools which will increase the fluidity of the profile content that is at the heart of ProQuest's information offerings.

19. Thomson
Tiffani Pillifant, Ellen Rotenberg, Berenika Webster

BACKGROUND

For more than 50 years, Thomson Reuters has been the leading provider of citation data and research indicators to the global scholarly community. Each day millions of researchers from thousands of organizations visit Thomson Reuters' destinations, such the Web of KnowledgeSM, InCitesTM and Research In ViewTM to analyze, understand, and make decisions. We recognize that increasingly our customers also value the ability to integrate high-quality Thomson Reuters' data and indicators into their research management infrastructure. These activities include building and maintaining institutional and national repositories, developing researcher profiles, and using data for research or evaluation all have a common requirement — the accurate identification of researchers and proper association with their scholarly activities and outputs.

PEER REVIEW

Even prior to publication, proper name recognition and attribution of scholarly works is essential to the integrity of the scholarly publication process. During peer review, authors and reviewers are routinely evaluated not only on the merit of the current manuscript in question, but on their prior publications as well. Scenarios involving Editors submitting manuscripts as authors to their own journal, plagiarism checks, "shotgun" application methods, and lack of expertise in a given subject area are all evaluated as possible barriers to publication – and all rely on the accurate identification of authors, co-authors, and reviewers of a paper. In addition, as publishers move to a more "contributor-centric" view of their relationship with consumers and contributors, they wish to employ new technologies for single sign-on and comprehensive views for users that require users to be unique and properly identified. New industry-wide initiatives such as ORCID offer promise to assist with this problem and ScholarOne will look to incorporate these as they are available in the community.

Tiffany Pillifant
Director, Product Management
ScholarOne

AUTHOR DISAMBIGUATION: CLAIMING SYSTEMS AND IDENTIFIERS

Thomson Reuters understands the importance of author identity and proper attribution of scholarly works. To help alleviate issues of finding the right 'Jill Jones', we have developed tools that combine computational techniques with end-user feedback to provide an accurate view of an author's publications. Using a proprietary algorithm that analyzes relationships between publications, Web of Science records are grouped into distinct author sets. Users can provide input, validation, and adjustment into clusters through ResearcherID, by claiming individual publications and/or author sets; author validated publications display the author's ResearcherID number, as well as a link to their online profile. We will also look to include Industry-wide initiatives (e.g., ORCID) into this process as those projects progress. ResearcherID provides author-level metrics and visualization tools for analysis of co-author and citing articles networks. This unique researcher identification and disambiguation service has been integrated across Thomson Reuters applications (ScholarOne Manuscript Central, Web of Knowledge, EndNote, and Research In View), and is being used by institutions around the globe to aid in publication list management—either by the institution or a proxy at the institution through a suite of APIs and an administrative UI.

Ellen Rotenberg
Senior Manager

PROFILING AND ALLOCATION OF CREDIT FOR SCHOLARLY WORK

The need to manage, validate, and assess researchers' scholarly record is increasingly seen as

the responsibility of both the institution and the individual author. To allow for both efficient and transparent management of bibliographies and records of activity, Thomson Reuters has been developing new approaches and tools to allow authors (faculty) and institutions create and manage their profiles. The aim of our efforts was to remove the burden of data collection from an individual author (faculty) by creating an automated system, Research in View, which can search rich data sources residing within TR, institutional data sources (such as HR, grants, registry and institutional repository systems and EndNote libraries) as well as other web-based resources like library catalogues, ResearcherID, etc., to create and continuously update author (faculty) information.

Author (faculty) profiles are enriched by analytics including citation and other indicators to allow for internal assessment, and to support benchmarking of authors' and institutional activities. These analyses can be carried out for individual authors as well as user-defined groups of authors.

An infrastructure to continuously update the scholarly requires not high level of integration with content sources described above but also integration with internal identity systems such as Shibboleth to accommodate for researcher movement between organizations. As standards and 'national level' initiatives emerge in this area, Thomson Reuters is eager to work with the community to incorporate these into the Research In View infrastructure. Examples include: ORCID (planned), Eurocris' CERIF (currently), CASRAI and VIVO ontology (under development) for compliant data exchange and interoperability.
Dr Berenika M. Webster
Product Manager, Research Analytics

20. Tufts University
Jeff Kosokoff

Jeff Kosokoff is Director of the Edwin Ginn Library and Information Technology at Tufts University's Fletcher School of Law and Diplomacy. Jeff works closely with Fletcher's academic dean and represents Fletcher within the Tufts community on issues of scholarly communication and identity. Like all scholarly entities, Tufts faces business needs that critically involve issues of identity, including tenure and promotion, publicizing faculty activities and expertise, and copyright and author's rights. Jeff has been a member of Tufts' Faculty Information Committee since its founding in 2008; the committee has been working to foster collaboration and awareness about various faculty information and scholarly identity efforts at the University.

One of the core challenges that Tufts University faces is systematically identifying collaborative opportunities between faculty within and beyond Tufts whose common research interests may not be readily apparent. In a Spring 2011 report (http://provost.tufts.edu/wp-content/uploads/White-Paper-on-Collaboration-at-Tufts-May-2011.pdf), Tufts' Provost called for enhancing the University's capacity for collaborative research matchmaking, using, among other things, information resources that catalog faculty expertise, interest, and scholarly output. Tufts' identity management efforts in progress include an implementation of Harvard Catalyst's Profiles as part of our Clinical & Translational Science Institute (CTSI), development of new processes and systems at the Tufts School of Medicine and The Fletcher School to manage and present faculty information, moving towards implementation of a research administration system, updating our identity and access management technologies, and exploration of scalable methods to gather and standardize faculty publication information from disparate sources.

Jeff has a BA in philosophy from the University of California at Santa Cruz as well as Masters degrees in library science and the history and philosophy of science from Indiana University. Jeff has been an active member of the American Library Association, especially in the Collection Development and Management section of ALCTS. Before coming to Tufts, he worked in various collections, technology and reference positions at Connecticut College, DePaul University, Harvard University, and Simmons College. Jeff also worked as a MetaLib and Primo Implementation Librarian at Ex Libris.

21. University of Florida (VIVO)
Michael Conlon

Michael Conlon is Director of Biomedical Informatics, Associate Director of the UF Clinical and Translational Science Institute, and Principal Investigator, VIVO: Enabling National Networking of Scientists, at the University of Florida

Since 1997 Dr. Conlon has worked on issues of identity management and scholarly attribution at the university level, the national and international level. Dr. Conlon has led efforts to create and deploy directory services (LDAP and Active Directory), network id (GatorLink), and UUID (UFID) services at the University of Florida, as well as single sign-on (Shibboleth) services for all major enterprise systems and hundreds of university web sites.

Dr. Conlon has been a frequent participant and presenter at EDUCAUSE and Internet2 workshops regarding identity management and level of assurance.

Since August 2009, as principal investigator of the NIH project VIVO: Enabling National Networking of Scientists, Dr. Conlon has led a team of over 180 investigators and implementers in the development and deployment of VIVO, an open source, semantic web application for research discovery and scholarship. VIVO can be fully integrated with university identity management systems to provide high level of assurance assertions regarding scholarly work.

In May of 2010, VIVO sponsored a workshop in Gainesville Florida on the topic of author disambiguation and identity management for attribution, bringing together representatives from NSF, NIH, OCLC, publishers, NBIC, leading universities and others to address issues of scholarly identity.

The upcoming VIVO conference, August 24-26, 2012 in Miami Florida, will bring together practitioners and stakeholders in the construction of semantic web data and tools for scholarly work. Topics will include scholarly attribution, identity management and the roles of publishers, institutions, government agencies and others.

Dr. Conlon serves on the InCommon working group regarding identity management in the federal agencies, and has participated in planning sessions regarding SciENCV, an inter-agency, next generation, researcher profile system.

Dr. Conlon is a frequent invited speaker on topics of research discovery, identity management, data sharing and the emerging semantic web in support of scholarly work.

22. University of Illinois Urbana-Champaign (BibApp)
Sarah Shreeves

I am the program manager for BibApp, an open source software based here at Illinois that helps institutions create an institutional bibliography and expert finder system (see http://www.bibapp.org/). BibApp is used by a small but steadily growing number of institutions, both here in the US and internationally. The BibApp team has been tracking both the development of ORCID and systems like VIVO, and are committed open standards, linked data, and interoperability. I manage our installation at Illinois (http://connections.ideals.illinois.edu/) and run into many of the challenges in managing scholarly identities including the relatively technical issues of author disambiguation and metadata discrepancies across A&I services to the more social issues like identifying what to include in such identities (heavily dependent on discipline) and researcher concerns about privacy and monitoring (also dependent on discipline).

I also manage IDEALS, the institutional repository at Illinois (see http://www.ideals.illinois.edu/) and so am also interested in how institutional repositories, and systems like BibApp interface with publisher and other efforts to collate publication histories. I am also involved closely in the development of the DMP Tool (https://dmp.cdlib.org/), and am interested in how systems related to data interface with those that tackle more traditional publications.

I think that I can offer the perspective of a developer of a faculty profile management system and an on-the-ground repository manager, as well as someone with experience in developing best practices and standards in national and international settings.

23. University of Minnesota
Stephen Hearn

Like other institutions, the University of Minnesota is experiencing an efflorescence of overlapping identity-related systems, including Elsivier's SciVal Experts, Harvard's Profiles, and a system to track individual tenure-related activities, combined with increasing use of institutional and disciplinary repositories. The University Libraries is already a significant participant in these various initiatives, and sees a developing role for itself as the campus agency responsible for coordinating and managing the identities of scholars and units of the university across these diverse domains.

My own professional background is in library authority control and data management, but I'm finding that traditional library authority files are an inadequate response to the current challenge. It is no longer reasonable to pursue solutions that are wholly under library control. What is needed are concepts and systems which actively engage with multiple sources to aggregate identity-related data into useful and dynamic constellations. Linked data technologies are an important component of these developments, but not in the sense of a particular imposed data format. Rather, it is the concept of open, permeable, interoperating data systems that is crucial to linked data, the idea that each system through the use of well-marked data and reliable, actionable identifiers will be able both to gather data from elsewhere to meet local needs and to provide specified data freely to other systems.

What is really at issue is a renegotiation of who is responsible for identity data management. Traditionally identity management systems have been built to serve the purposes of particular agencies and have been closed to external participation. Increasingly, this is not a viable model for the task we face. These systems need to be open to participation by the entities their identifiers and identity records represent and by their agents. As institutions that are both central and trusted, libraries have an opportunity to play a major role in the way campuses and communities manage the identities, reputations, and relationships of their citizens.

Stephen Hearn is Metadata Strategist, Technical Services, University Libraries, University of Minnesota

24. University of North Texas
William Moen

VIVO at the University of North Texas

Beginning in Spring 2011, the Texas Center for Digital Knowledge (TxCDK) in the University of North Texas' College of Information started working with VIVO. This was in part to get a better understanding of VIVO as a semantic web technology and to potentially bring the platform to the attention of UNT administration. To further these goals, we have undertaken two VIVO projects:

- UNT's local Research Clusters: In Fall 2011 we began piloting VIVO to explore how it could be used in the context of UNT's research cluster initiative (http://research.unt.edu/clusters/). This is a locally oriented project.
- The iSchools (information schools) movement: In conjunction with Dr. Katy Borner's VIVO implementation team at Indiana University, we are exploring the use of VIVO by the iSchools (see http://ischools.org). At the 2012 iConference, we distributed a proposal to get some iSchools to pilot VIVO. We are implementing an instance of VIVO that will hold information about UNT's College of Information (one of the current 33 iSchools).

Currently, we are working with the official VIVO project partners (a list of those are at http://vivoweb.org/about) in dealing with data issues (determining quality sources of data, acquiring the data, transforming the data, figuring out ingest methods to map source data onto the VIVO ontology, and exploring the potential integration of existing controlled vocabularies (such as Library of Congress Subject Headings) into the effort. Although we are not one of the original seven institutions funded by NIH to develop VIVO, we are an early adopter and are contributing to the VIVO initiative based on our work here.

25. University of Pennsylvania
John Ockerbloom

Technical and social challenges of identity management: The critical role of context

At Penn, as at many other universities, we are keenly interested in collecting and highlighting the scholarly work of our researchers, and their collaborations with scholars elsewhere. To that end, we are building and populating repositories focused not just on our own institution (such as ScholarlyCommons@Penn), but also on worldwide research fields in which our scholars participate (such as the Work and Family Research Network). We are also highly interested in

compiling comprehensive bibliographic records of our scholars, through systems like VIVO and Selected Works. Our success in meeting these goals depends on technical advances, including unifying data through technologies like linked data, authority databases, common author identifiers, and author disambiguation tools. It also crucially depends on the cooperation of our scholars in compiling and maintaining information about themselves and their works.

We have found an understanding of *context* to be critical in meeting both the technical and social challenges of scholarly identity management. For instance, if we aim to link scholarly output with international authority files, we have to deal not only with their focus on books, but also on their subtly different concept of identity. While a scholar usually maintains a single identity in formal scholarship (albeit with changes in affiliation and sometimes name over time), literary identities, the typical subject of library name authorities, are often either shared, or maintained multiply for the same person. Similarly, when we consider identities in campus authentication and email systems (which often have shorter lifespans than personal scholarly identities) or in social networking systems (where pseudonyms and multiple personas are common for scholars as well as laypeople), we also need to understand the differences between the notions of identity in these different contexts if we want to bring them together effectively.

Even in the formal scholarly realm, where there is generally one identity per person, we have found that scholars are more likely to cooperate in building up publication sets if they feel they have control over the context in which their identity and work is displayed. In our institutional repository and in our VIVO pilot, for example, faculty and department heads have wanted the ability to present their work to the public with emphasis of their choosing, highlighting publications they want to show off and downplaying work that does not reflect their current interests and expertise. This conflicts with the desire of administrators and librarians to have a comprehensive record of faculty bibliographies and publications. But when we have been unable to give our clients adequate control over the visibility and organization of information about them in our systems, we have encountered non-cooperation or requests for withdrawal of information, even when the information can be found in other public sources.

Aggregation and selectivity, then, are both of crucial importance in scholarly identity management, yet have a fundamental tension between them, one that can be adressed by taking context into account. Context also helps us understand the subtle but important distinctions in technical and social semantics of identity in different settings. Along with discussing methods and technologies for aggregation and selection, then, we need to discuss the management of context as well. I hope to provide some useful contributions to this discussion based on our observations and experiences at Penn.

26. University of Virginia
Daniel Pitti

The Institute for Advanced Technology in the Humanities (IATH) is working on a project, in collaboration with the National Archives and Records Administration (NARA), to develop a vision and blueprint for establishing a sustainable National Archival Authorities Cooperative (NAAC), both for the archival community and the users of archival resources. We are examining the business, governance, and technological requirements of both developing and sustaining this type of cooperative.

We are involving archivists, manuscript librarians, and scholars, representatives of all federal repositories (NARA, Library of Congress, National Agricultural
Library, National Library of Medicine, Smithsonian Institution, and National Park Service), representatives of funding agencies and foundations, and other stakeholders.

Our work will begin by reviewing the findings of the NEH-funded
Social Networks and Archival Context (SNAC) project, to set the stage for discussion. SNAC has demonstrated its potential to transform scholarly historical research, both by dramatically improving access to resources that document the lives, work, and events surrounding historical persons, organizations, and families, and by providing unprecedented access to the biographical-historical contexts of the people documented in the resources, including the social-professional networks within which the people lived and worked. It is time to begin looking beyond the SNAC project, to develop a sustainable program that will enable the archival community to collaboratively build and maintain a powerful new research tool.

Funding for this initiative is provided by an Institute for Museum and Library Services grant to the University of Virginia.

Daniel Pitti is Associate Director of the Institute for Advanced Technology in the Humanities (IATH), University of Virginia