# Against the Grain

April 2009

# Standards Column -- Working to Solve the Problems of Name Authority -- The International Standard Name Identifier and Other Projects

Todd Carpenter
*NISO*, tcarpenter@niso.org

Follow this and additional works at: https://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

## Recommended Citation

and departmental liaisons discuss how best to spend the allocations, with humanities and social science programs more interested in books, and programs in sciences, engineering, management and economics more interested in databases.

**AUC** has also established with **YBP** an automatic shipment program for Middle East Studies (MES) books. Because MES is very important to **AUC** and orders are submitted from many different departments, piles of publisher catalogs, promotional fliers, **Amazon** print-outs, and so on have always poured into

Acquisitions year-round from all over the university. This led to an enormous duplication of work as liaisons and Acquisitions received the same orders over and over again. Even with all of this duplicated effort, we did not always have a sufficiently comprehensive MES collection.

Our solution was to create an electronic slip Notification profile for MES in **GOBI**, then to chart the weekly results over the first few months. We then switched to

the automatic shipment option and adjusted the non-subject parameters as we analyzed the results in our electronic slips. The books for MES now arrive automatically, pre-processed, and ready to shelve, saving a huge amount of effort and assuring a more comprehensive collection.

These steps are modest enough, admittedly, but in the Middle East there are only a few institutions employing such methods, and most of them are American universities like **AUC**. 🌲

---

# Standards Column — Working to Solve the Problems of Name Authority — The International Standard Name Identifier and Other Projects

by **Todd Carpenter** (Managing Director, NISO, One North Charles Street, Suite 1905, Baltimore, MD 21201; Phone: 301-654-2512; Fax: 410-685-5278)
<tcarpenter@niso.org> *www.niso.org*

Identifying the author of a text has been critical to efficient operation of a library for as long as there have been libraries. As the number of authors have grown over time, so too has the issue of keeping straight all of the authors and their works. Being able to explicitly and authoritatively identify all authors has been the holy grail of authority files since the advent of computer managed cataloging in the 1970s. The expansion of authority files from book authors to include authors of journal articles and other publications has compounded the problems, since book authors are only a relatively small component of the broader authoring community. In addition, the problems have continued to expand over time as non-English and increasingly non-western-language content grows in availability and interest within scholarship.

There are many issues with name identification. Within bibliographic records there is a core group of biographical information that helps to identify a single individual. These are most often, sex, dates of birth and death, nationality, occupation, and language. This complexity is compounded by corporate, fictional or non-human entities that are listed as authors.

There are many known problems with disambiguating content creators. A catalog record might have the same name but contain different biographical information about their life, which would lead to false negative connections between records. Another scenario is when two different people share the same name or share similar biographical information, which creates a false positive

connection between different people. Although less common, are situations where there are different persona but in fact are the same person, for example when people use pen names or change their names. This is most difficult in situations when the creator is actively interested in masking their identity for some reason. Finally in cleaning up systems, there are issues of ambiguity in matching, because of translations, combinations of missing or inaccurate data, or incomplete information.

Despite all of these challenges, many organizations and initiatives have over the years begun developing their own identification registries. Some of the earliest article repositories, such as **arXive** and the **Research Papers in Economics (RePEc)** included author identification services. Commercially-run systems such as those supplied by **Elsevier (Scopus Author Identifier)** and **ThomsonReuters (ResearcherID)** also support name identification and associate services. **ProQuest** provides a different approach with **Scholar Universe**, that is compiled from lists of faculty members, their publicly available resumes and then associates the individuals with their published works. Each of these services provide "**Web 2.0**" interaction so that authors can correct or adjust their information.

In 2007, the **JISC Repositories and Preservation Program**, the **British Library** and **Mimas**, a data centre at the **University of Manchester**, began investigating: "…the potential for the development of a Name Authority Service and factual authority for digital repositories, to support cataloguing,

metadata creation and resource discovery in the repository environment (**JISC**, 2007)." The **Names Project**, as it is called, is developing a prototype for a name and factual authority service for use by UK repositories of research outputs.

The **International Federation of Library Association's** *Functional Requirements for Authority Data: a Conceptual Model (***FRAD***)* outlines a series of entity descriptions compiled from *Functional Requirements for Bibliographic Records* (**FRBR**); and *Guidelines for Authority Records and References* (**GARR**).

In identifying a person, **FRAD** outlines the following characteristics of a person: Dates, titles, designations, gender, place of birth/death, country, place of residence, affiliation, address, language, field, profession.

There are also issues with the relation to a name compared with a person, again from **FRAD**:

- real name
- pseudonym
- secular name
- name in religion
- official name
- earlier name
- later name
- alternative linguistic form
- other variant name

Within the question of rationalizing a system of classifying and uniquely identifying a person, there could be a variety of names that

a person might have used in their lifetime, be they pen names or other pseudonyms, changes in legal names via marriage or otherwise, or variant names some might use inconsistently (**Rob** or **Bob** versus **Robert**). These problems are compounded with non-Latin written names, which are romanized so that they can be entered into English cataloguing records, such as the variety of romanization of **Peter Illiach Tcharikovsky**. These identities can all be linked within and between different identities, such as **Paul McCartney**: his participation as a member of the **Beatles**/the **Wings**/solo career, his writing, his music publication, his **Apple Music** holdings, his movie career etc.

According to the **IFLA** Website:  *http://www.ifla.org/VII/d4/wg-franar.htm*, the draft *Functional Requirements for Authority Data*, which was released last year, has been approved by the Standing Committees of the **IFLA** Cataloguing Section and the **IFLA** Classification and Indexing Section in March 2009. The final text is being prepared for publication as part of the **IFLA** Series on Bibliographic Control.

The **Virtual International Authority File** (*http://viaf.org/*) is a joint project of the **Library of Congress** (**LC**), the **Deutsche Nationalbibliothek** (**DNB**), the **Bibliothèque nationale de France** (**BnF**), and **OCLC**.  The project's goal is to combine name authority information from these various institutions, cross match the data, and link the library authority file information into a common system.  As such, this compiled database will be a test case for the creation of a centralized name identifier authority file, which could comprise a large segment of what will eventually become the core **ISNI** registry.

At present, the **VIAF** project contains approximately 9.2 million personal name authority records.  Clustering this information narrows the number of authority files to 7.7 million persons or personas.  This data is derived from approximately 40 million bibliographic records.  As the pilot develops additional records and data will be drawn from the current participants as well as from additional national libraries.

Much of the work described here ties to ongoing work at **ISO** on the **International Standard Name Identifier** (also know as Draft International Standard 27729), which is one of the activities of **ISO** technical committee 46, subcommittee 9 (TC 46/SC 9) on identification and description.  **NISO** is the Secretariat for TC 46 / SC 9 and coordinates this work.  The **ISNI** is a method for uniquely identifying the public identities of authors and contributors to media content such as books, music, movies, television programs, and serial-publication articles.  Once published and operational, the **ISNI** will provide a tool for disambiguating names that might otherwise be confused, and will link the data about names that is collected and used in all sectors of the media industries.

Each of these projects is inter-related in some way, and while they address certain specific issues in their own community.  Last spring the **Coalition for Networked Information** (**CNI**) hosted a seminar in Washington, D.C. entitled "Authors, Identity Management and the Scholarly Communication System," which explored these issues.  As part of that meeting, more than a dozen identity projects were identified as being underway at a various stage.  It is highly unlikely that most, or even many of these disparate projects will be consolidated into a single "identifier to rule them all." However, there is a critical need to at least establish common understanding of the various projects and possibly coordinating to establish common forms and interactions among the projects being developed.

### Additional Resources

**International Standard Name Identifier**: *http://www.isni.org/.*

**JISC**, **NAMES Project** *http://names.mimas.ac.uk/.*

*The International Standard Archival Authority Record for Corporate Bodies, Persons and Families* (*ICA*) and the *Essential Data Elements for Internationally Shared Resource Authority Records* (*IFLA*): A Comparison and Report *http://www.icacds.org.uk/eng/ifla.pdf.*

**Amanda Hill**, "What's in a Name?  Prototyping a Name Authority Service for UK Repositories" *http://ie-repository.jisc.ac.uk/154/1/Names_ISKO2008_paper.pdf.*

**The International Federation of Library Association's Functional Requirements for Authority Data: a Conceptual Model** (**FRAD**) *http://www.ifla.org/VII/d4/FRANAR-ConceptualModel-2ndReview.pdf.*

Presentation by **Joan Lippincott** about the **CNI Meeting** on Author Identities (final report is not yet available) *http://www.delos.info/files/pdf/DLFoundations2008/9_LippincottDLFoundations08.pdf.*  🐦