

Martin Fenner

Alan Danskin  
Amanda Hill  
Daniel Needham

# ISQ

## INFORMATION STANDARDS QUARTERLY

SUMMER 2011 | VOL 23 | ISSUE 3 | ISSN 1041-0031

TOPIC: ORGANIZATION AND PEOPLE IDENTIFIERS

Jody DeRidder

Janifer Gatenby  
Andrew MacEwan

Geoff Bilder

Louise Timko

INTERNATIONAL  
STANDARD NAME  
IDENTIFIER (ISNI)

THE OPEN RESEARCHER &  
CONTRIBUTOR ID (ORCID)

THE NAMES PROJECT

STANDARD ADDRESS  
NUMBER (SAN)

INSTITUTIONAL IDENTIFIERS  
(<sup>1</sup>) AND ISNI



**I**f I like a novel of one author, I look to see what else he or she has written. The same is true if I find somebody who writes well on topics that interest me, so I would like to find all he or she has written in any form—whether book, article, or blog. The same applies to all creators—composers, performers, film directors, researchers, artists, producers, publishers, and others.

The importance of the creator in determining selection is clearly important to me; titles give me an idea about subject whereas creators give me an idea about quality and pertinence. Yet existing online databases including search engines, online book shops, and to a lesser extent library catalogs, are poor in grouping by creator. In most Internet sites, when you click on “more by this author, composer, etc.” typically a keyword search is launched that returns both imprecise and incomplete results. The user cannot be sure that what is proposed is, in fact, created by the same author. Further searches are often necessary for identity identification.

Identification of creators and other contributors is also critical to societies who administer rights information and royalty payments. These societies in their day-to-day work to channel the incoming royalty payments, often find themselves in a position of searching for information to accurately identify the rights holders. As the market is shifting from physical towards digital

# A **new** system for name identification

JANIFER GATENBY AND ANDREW MACEWAN

and from purchased copy to licensed access, the accurate and unambiguous identification of the parties involved is even more important, not only to rights management systems but also to trade organizations, publishers, aggregators, distributors, and retailers. The current efforts of libraries, rights management societies, and trade organizations to disambiguate creators are arduous, and their efforts are duplicated.

Identifiers serve as short hand for the metadata that differentiate one identity from another. The idea of a simple unique identifier that everyone in the world could use to identify the same person, corporate body, or similar entity has been proposed many times during the last 30 years. Finally, an international standard (ISO 27729) for an International Standard Name Identifier (ISNI) has been approved to respond to this need. ISNI has been designed as a bridge between existing proprietary right holder identification systems, such as the Interested Party Identifier, and resource discovery tools, such as the Virtual International Authority File (VIAF). By sharing a common identifier that is global in scope, data within and across databases can be accurately linked, thus providing the infrastructure for significantly improved name searching. Moreover, by also sharing the metadata resources that

further describe the unique identification, ISNI participants are cooperating to achieve high quality data and at the same time realize processing efficiencies. Rights management societies have an additional constraint in that the data provided to them is in the most part bound by legal confidentiality clauses and is linked to sensitive payment information. ISNI, for them, enables linking into other databases via a neutral identifier that is one step removed and enables other public domain data, in particular VIAF information, to be used as the publicly facing metadata.

The International Confederation of Societies of Authors and Composers (CISAC), the International Federation of Reproduction Rights Organisations (IFRRO), the International Performers' Database Association (IPDA), ProQuest, OCLC, and the Conference of European National Librarians (represented by Bibliothèque Nationale de France and the British Library) met over the period of three years, firstly as participants on the working group that developed the ISNI standard and subsequently for founding the ISNI International Agency (ISNI-IA), which will administer the ISNI assignments and registration. The ISNI-IA, an unprecedented cross-domain alliance, was incorporated in

CONTINUED »

London on December 22, 2010. The consortium members will bring together data from more than 300 rights management societies and 26,000 libraries worldwide for the initial ISNI implementation.

### ISNI Architecture

ISNI cannot be administered in the same way as other resource-oriented identifiers such as the ISBN, where publishers are pre-allocated ranges of identifiers which they progressively apply to their new publications. Creators are not bound to any one publisher or distribution outlet or to any one domain and they publish and collaborate across national borders. Singers are often also composers and may write autobiographies or other books; as an example, Paul McCartney has written poems and a book for children. Thus, allocation and administration of ISNI needs to be controlled centrally in order to avoid, as much as possible, duplicate identifiers for any one creator.

Yet, a fully centralized system would be unmanageable on a global scale. The task of collecting the data, ensuring its quality and completeness, and then disambiguating it will be better done with a global network of participating organizations. It is also important that the assigned identifiers be diffused as widely as possible among databases and indexes accessible over the Internet to facilitate the correct linking and exchange of data. After careful deliberation, the system eventually settled on was a centralized database for requesting and referencing the ISNI identifiers with an international network of registration agencies, responsible for

collecting high quality metadata and sharing in the tasks of disambiguation and diffusion. In addition to the registration agencies, some organizations will make their systems and databases available to the central system for verification and reference, as illustrated in Figure 1.

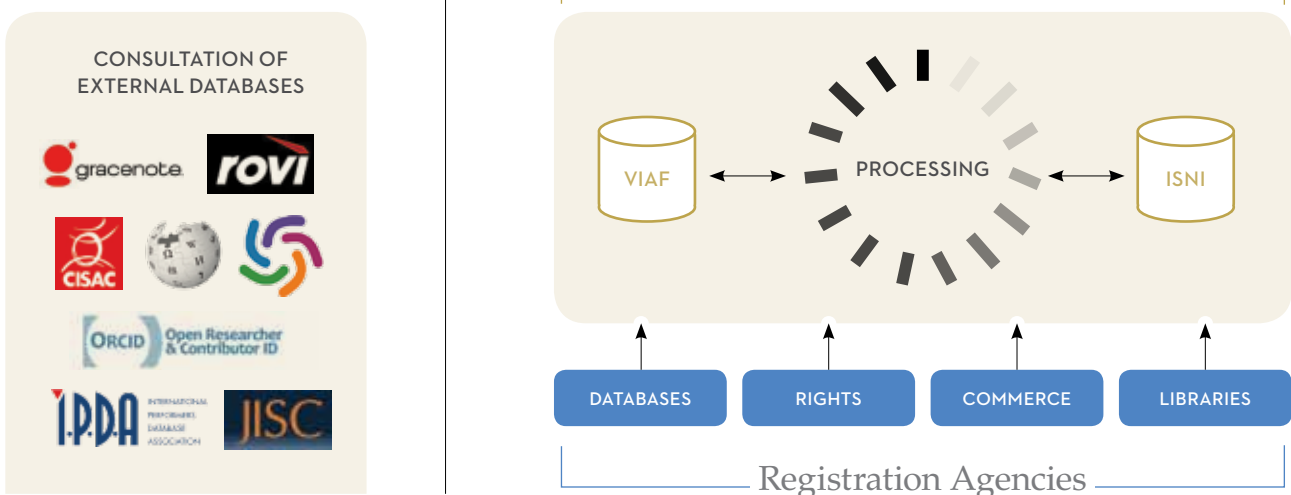
### The Initial Database

In order for ISNI to reach its potential, industry-wide use is a key factor for encouraging adoption by research communities and internet databases generally. The ISNI-IA will ready the ISNI for industry adoption by first creating the initial ISNI database, which will allocate ISNI identifiers by processing data from the databases of the consortium's founding members and affiliate organizations. The ISNI assignment system will be launched in the third quarter of 2011 with an initial database of assigned ISNIs. The base cross-domain file for ISNI is VIAF, the Virtual International Authority file, created over the last six years by combining the authority files of 19 major sources, mostly national libraries, and including the NACO/LC file, an international cooperative library name authority file representing over 300 libraries worldwide. VIAF currently contains 14 million names.

### Name Disambiguation

Confidence and quality are being emphasized in the creation of the ISNI database. The matching techniques of VIAF have been adapted and employed in the ISNI system. Data files from the founding members of the ISNI-IA are progressively being matched against VIAF. ISNIs are being allocated where

Figure 1: The ISNI architecture





Will Smith, born 25 September 1968 is an American actor, film producer, and pop rapper who performed under the pseudonym *The Fresh Prince*

Source: Wikipedia: [http://en.wikipedia.org/wiki/Will\\_smith](http://en.wikipedia.org/wiki/Will_smith)

OR



Will Smith (born 1971) who wrote “How to be cool”

Author: [Will Smith](#)  
 Publisher: London : John Murray, 2007.  
 Edition/Format: Book : English  
 Summary: What is cool? How do you become cool?

Source: Wikipedia: [http://en.wikipedia.org/wiki/Will\\_Smith\\_%28comedian%29](http://en.wikipedia.org/wiki/Will_Smith_%28comedian%29)

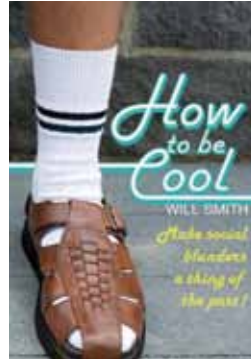


Figure 2:

How can you then tell one Will Smith from another?

there are more than three VIAF sources or two independent sources. Each allocated ISNI is assigned a confidence level in the data itself and another for the level of confidence in the matching. Thus, high confidence is placed on matching data, collected independently where one of the sources has direct contact with the identity, as is the case with a royalty management society.

The amount of data required for distinguishing one identity from another varies by the amount a name has been used and by the number of name variants. Core ISNI metadata consists of:

- ➔ Name of public identity (e.g., surname, forename, prefix, suffix)
- ➔ Name variants (e.g., Bernard Shaw, George Bernard Shaw, G.B. Shaw, and the name in a non-Roman script, but not including pseudonyms that are considered different identities)
- ➔ Creation class or classes
- ➔ Creation role or roles
- ➔ URI to contributing source or sources

Yet this data is frequently not sufficient for disambiguation. Many other data elements are collected that play a role in differentiation such as co-authors, affiliations, and publishers, but the most significant are the titles of works in which the identity has played a role in the creation, and birth and death dates. For institutions, other elements are more important, e.g., geographic location.

Where there is any doubt, multiple records for potentially the same identity are being kept in the database for manual

review and an ISNI is not assigned until disambiguation is finalized. Once an ISNI is made available for widespread use and diffusion, it is hard to change or correct it. If two identifiers have been issued for the same identity, mapping from a deprecated ISNI to a correct one can be communicated and the ISNI reference database can be consulted, which will accept both the correct and deprecated ISNI and point to the correct metadata. It is considerably more difficult to diffuse corrections where one identifier needs to be split into two. The original ISNI needs to be deprecated and then point to two different identifiers, such that human review is necessary to select the correct metadata.

One of the challenges is to unite data from separate data sources where different criteria are favored for disambiguation. VIAF, for example, uses resource titles and, to a lesser extent, dates. How can you then tell one identity from another? For example Will Smith, born 25 September 1968 is an American actor, film producer, and pop rapper who performed under the pseudonym *The Fresh Prince*. But he is not the Will Smith (born 1971) who wrote “How to be cool” (see Figure 2).

The assignment system works best in an online, interactive mode where multiple records can be presented and either the best chosen or proof given that a separate identity is involved. Here the registration agencies (RAGs) will play an essential role in assuring the completeness and quality of the request data before it is submitted and assisting in disambiguation by providing additional information as demanded in the system’s response. Consultation of external sources will play a role. Also important is the need for the creators themselves to be able to provide input to their

CONTINUED »

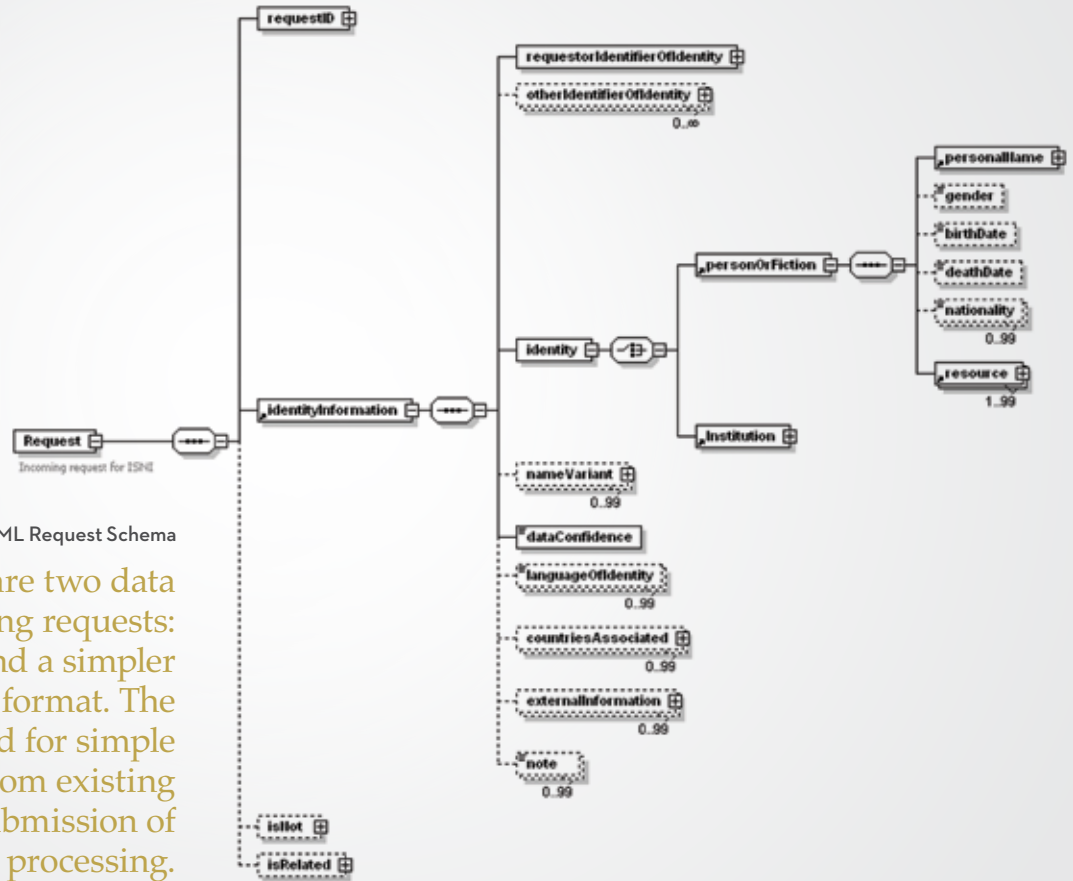


Figure 3: ISNI XML Request Schema

There are two data formats for making requests: an XML schema and a simpler tab delimited format. The latter is envisaged for simple mapping from existing databases and submission of a file for bulk processing.

metadata, correcting it where necessary. It is envisaged that the RAGs will provide such interfaces, such as the planned VIAF interface in the xA (extended authorities) service.

### Business Model and the ISNI System

The initial database is being funded by the founding members of the ISNI-IA. Once the database and system are launched, requests for ISNIs will be placed through RAGs, for which there will be a small charge to cover reasonable costs as per ISO's RAND (reasonable and non-discriminatory) policy. All players are encouraged to distribute allocated ISNIs as widely as possible. There will be a free search service providing basic metadata for an allocated ISNI with name and ISNI search options. The search service will be available via the ISNI-IA webpage and machine to machine using the Search/Retrieval via URL (SRU) protocol; a downloadable search box that can be included in external web pages is also being considered. The name forms and other information made available in ISNI search responses will most often be sourced from VIAF, protecting the confidential nature of data from other contributors.

Requests will come into the system via the RAGs. At least one RAG will make a web-based individual request system, open to all, with a charge in accordance with the RAND principle, mentioned above. There are two data formats for making requests: an XML schema (see Figure 3) and a simpler tab delimited format. The latter is envisaged for simple mapping from existing databases and submission of a file for bulk processing. With bulk processing, 100% assignment cannot be guaranteed. There will always be a residue requiring further analysis. Requests using the XML request schema can be submitted either in bulk or interactively using the Atom Publishing Protocol.

### Adoption and Relationship with Other Initiatives

ISNIs can be assigned to all entities that create, produce, manage, distribute, or collaborate in creative content including human beings (alive or not), legal entities (such as academic institutions, publishers, and societies), or fictional characters. The scope of ISNI is broad, though ISNI's initial database will only include personal names. Institutions are also in scope and will be included thereafter. NISO's I<sup>2</sup>

(Institutional Identifier) committee has produced a set of metadata for institutions and is recommending adoption of ISNI as the identifier scheme for institutions in the supply chain. (See the I<sup>2</sup> and ISNI article on page 26.)

The European Arrow Project (Accessible Registries of Rights Information and Orphan Works towards Europeana) is a consortium of national libraries, publishers, and collective management organizations that is supported by the European Commission. The consortium favors the use of ISNIs in conjunction with the International Standard Text Code (ISTC) as the fundamental building blocks for rights management administration.

Within the music industry, there is a lot of interest in ISNI, in particular by the record labels wishing to provide ISNIs to accompany their data submissions to multiple exposure and distribution sites. This industry is expected to be an early adopter of ISNI for both performers and composers.

ORCID is an initiative that is seeking to disambiguate researchers and writers of articles in scholarly journals. (See related article on page 10.) Their scope is a subset of the scope of ISNI and there has been communication between the two groups on the possibility of using the same identifier scheme. ISNI's initial database is to include two files from JISC (see Names project article on page 14): names from the Merit project, which provides names of researchers covered in the 2008 Research Assessment Exercise (RAE) data, and Zetoc, a file of authors of theses from ProQuest and ProQuest's Scholar Universe. It would be confusing for the two initiatives, ISNI and ORCID, to be producing identifiers for the same identities in the same time frame. The two systems being developed seem to be complementary, with that of ISNI focusing on name disambiguation and metadata registration and that of ORCID developing a researcher-facing system. Thom Hickey's *Outgoing* blog contained an entry in July 2011, *VIAF and other IDs*, concerning the relation of VIAF, ISNI, and ORCID and a possible pivotal role for VIAF.

## Conclusion

The initial implementation phase of ISNI is concentrating on creating a database of assigned ISNIs of high quality and high certainty, together with a framework enabling name disambiguation that incorporates input from Registration Agencies, verification sources, and the creators themselves. | FE | doi: 10.3789/isqv23n3.2011.02



**JANIFER GATENBY** <janifer.gatenby@oclc.org> is EMEA Program Manager Metadata with OCLC B.V., and a member of both the ISO ISNI Working Group and the NISO Institutional Identifier (I<sup>2</sup>) Working Group.



**ANDREW MACEWAN** <andrew.macewan@bl.uk> is Head of Collection Processing at the British Library and a member of the ISO ISNI Working Group.

## RELEVANT LINKS



**Atom Publishing Protocol, IETF RFC 5023**  
[www.ietf.org/rfc/rfc5023.txt](http://www.ietf.org/rfc/rfc5023.txt)

**European Arrow Project**  
[www.arrow-net.eu/](http://www.arrow-net.eu/)

**Hickey, Thom. VIAF and other IDs. Outgoing [blog], July 2011.**  
[outgoing.typepad.com/outgoing/2011/07/viaf-and-other-ids.html](http://outgoing.typepad.com/outgoing/2011/07/viaf-and-other-ids.html)

**Interested Party Identifier**  
[www.ipisystem.org/](http://www.ipisystem.org/)

**International ISTC Agency**  
[www.istc-international.org/](http://www.istc-international.org/)

**ISNI International Agency**  
[www.isni.org](http://www.isni.org)

**Merit project**  
[www.jisc-collections.ac.uk/Project-Merit/Merit-Search/](http://www.jisc-collections.ac.uk/Project-Merit/Merit-Search/)

**Names project**  
[names.mimas.ac.uk/](http://names.mimas.ac.uk/)

**ORCID**  
[www.orcid.org/](http://www.orcid.org/)

**Scholar Universe (ProQuest)**  
[www.scholaruniverse.com/](http://www.scholaruniverse.com/)

**SRU protocol**  
[www.loc.gov/standards/sru/](http://www.loc.gov/standards/sru/)

**Virtual International Authority File**  
[www.viaf.org](http://www.viaf.org)

**Zetoc**  
[zetoc.mimas.ac.uk/](http://zetoc.mimas.ac.uk/)