

## Authority Control for TAMU Digital Asset Management Ecosystems

### Introduction

Texas A&M Libraries needs to increase the impact of scholarly works published by its community members. In order to increase TAMU scholarship's global impact, OSC provides various tools, services, trainings, and workshops to promote online scholarly identity, open access, research collaboration, and new forms of digital scholarship. Greater access to scholarly works and data, however, is enabled not just by appropriate services, tools, and policies, but also by effective identification mechanisms. Identifiers are important types of metadata that traditionally have been used for entity identification, linking, and referencing in various domains.

#### An identifier is

- A true minimum for accessing, citing, and linking data (Altman & King, 2007)
- A key form of metadata needed for the successful management and use of data stored in institutional repositories (Lynch, 2003)

OAKTrust currently uses NetIDs as its user identifiers. NetIDs support most of the local needs we have within DAMS; however, the boundary of scholarly works continuously expands and becomes blurrier with the use of increasingly diverse media. Also, the rise in multi-disciplinary studies and global collaborative projects requires our DAMS users to employ global identification systems.

### Identifiers for Author Name Disambiguation

Different author identifiers (e.g., ISNI, ORCID, ResearcherID, Scopus Author ID, VIAF, etc.) exist. Three identifiers are selected and reviewed here. The brief descriptions of each identifier help move this discussion forward.

#### *International Standard Name Identifier (ISNI)*

The International Organization for Standardization (ISO) developed ISNI. ISNI identifies public identities across multiple fields of creative activity. People playing a part in creation, production, management, and content distribution chains can be recognized accurately, and the content created from the public identities can be managed effectively. ISNI is allocated to any party that is or was a natural person, a legal person, a fictional character, or a group of such parties, whether or not incorporated. The assignment of ISNI is based on data aggregated from hundreds of bibliographic and rights management databases, including the Virtual International Authority File (VIAF), which is an international collaborative service to aggregate and provide convenient access to the world's major name authority files.

#### *Open Researcher and Contributor ID (ORCID)*

In 2012, the ORCID service was launched by the ORCID community and developed to disambiguate scholars with the same name and make connections between research (e.g., research articles and research data) and researchers. The ORCID community maintains it as a registry service, and it has many participants, such as Elsevier, Taylor & Francis, and CrossRef. The main goals of ORCID are to provide a reliable identifier and to support its communication and authentication. The format of ORCID is compatible with the format of ISNI.

#### *ResearcherID*

ResearcherID was designed by Thomson Reuters in 2008 to solve the ambiguity of authors' names in scholarly communications. Researchers registered with ResearcherID.com are given ResearcherID identifiers. ResearcherID enables researchers to manage their publication lists, check their number of citations, identify future collaborators and avoid author misidentification. Also, ResearcherID information integrates with the data citation index developed by Thomson Reuters, so that researchers can easily discover a publication and its related data from the repository.

### Quality Evaluation

Based on a list of identifier system quality criteria (Lee & Stvilia, 2014) and my conceptual analysis, ISNI, ORCID, and ResearcherID all have similar characteristics in the quality dimensions of uniqueness, persistence, simplicity, opacity, contextuality, compatibility, granularity, and scalability. The major differences are revealed in the following criteria:

- Verifiability: ISNI and ORCID have the same identifier string syntax: 16-digit numbers. The last character in the IDs is a checksum. Otherwise ResearcherID only depends on human check.
- Interoperability: All three identifiers are interoperable with other systems in scholarly communication. ORCID has the highest score and ResearcherID has the lowest score of interoperability.
  - ISNI: [ORCID](#), [VIAF](#), [Wikidata](#)
  - ORCID: [ISNI](#), [ResearcherID](#), [Scopus Author ID](#), [VIVO](#), Publishers ([Elsevier](#), [Taylor & Francis](#), etc.)
  - ResearcherID: [ORCID](#)
- Actionability/Resolution: ORCID and ResearcherID both allow users to locate an object by clicking its identifier string. ISNI only provides a search function in its website ([isni.org](#)).
- Authority/Reputation: ORCID has a better reputation and buy-in among the identification system community.
- Security: ISNI can only be registered through a limited number of ISNI Registration Agencies.

	ISNI	ORCID	ResearcherID
Uniqueness	1	1	1
Persistence	1	1	1
Simplicity			
Opacity	1	1	1
Verifiability	2	2	1
Contextuality			
Compatibility	1	1	1
Interoperability	2	3	1
Actionability		1	1
Granularity			
Reputation	1	2	1
Scalability	1	1	1
Security	2	1	1
<b>Total</b>	<b>12</b>	<b>14</b>	<b>10</b>

The quality evaluation suggests that ORCID is the best option for authority control for DAME.

However, we should also consider the quality of identity metadata provided by each identifier system. Without having quality metadata, a system does not fully support author name disambiguation. The metadata quality of the systems can broadly be divided into two different systems: (1) professionally curated metadata and (2) non-professionally curated metadata. ISNI belongs to the first category, whereas ORCID and ResearcherID belong to the second category. Systems in each category have different pros and cons. ISNIs can only be registered through ISNI Registration Agencies, and informational institutions (e.g., libraries) curate the metadata. On the other hand, ORCIDs and ResearcherIDs are registered by researchers themselves through the system website, and the researchers also curate the metadata. Information professionals have expertise on metadata and information organization, but do not have the researchers' level of domain expertise, and vice versa. Due to this fact, ORCID and ISNI may work complementarily. Each system has a significant risk in maintaining their metadata quality.

### Community Trend and Recommendation

Digital library technologies have somewhat matured and are now slowly improving. Communities are looking for different ways to improve our systems. One of the directions for improvement is in understanding systems' socio-technical context (e.g., tools, instruments, policies, rules, norms, division of labor, and communities) and using current systems efficiently. In this context, system interoperability and a balanced use of metadata are increasingly important. For example, TAMU Libraries' DAMS assessment task force recently submitted an initial report recommending future DAMS to be implemented as DAME, which includes multiple DAMS operating as an ecosystem. Individual systems have their own strengths, and if we use several of them in effective ways with systematic planning, the systems would better serve our library and community. Furthermore, there are many external projects that integrate different systems:

- [Symplectic Elements \(v5.1\) and DSpace](#)
- [VIVO & DSpace](#)
- [VIVO & ORCID](#)
- [FigShare \(data repository\) & Institutional Repositories](#)

Metadata quality control is essential for data curation. To the best of my knowledge, no identification systems provide sufficient quality control for human needs; even professionally curated metadata do not fully meet users' needs. Through OSC's VIVO project, we are experiencing these difficulties. In our process, we use the available technologies or external services, and then curators intervene to control the quality. Lastly, we provide a profile editor so researchers can contribute to their data's management. These three steps may be necessary for library communities based on the resources currently available. ISNIs support the first two steps and ORCIDs support the last step.

### References

Altman, M., & King, G. (2007). A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib Magazine*, 13(3/4). Retrieved from <http://www.dlib.org/dlib/march07/altman/03altman.html>

Lee, D. J., & Stvilia, B. (2014). Developing a data identifier taxonomy. *Cataloging & Classification Quarterly*, 52(3), 303–336.  
doi:10.1080/01639374.2014.880166

Lynch, C. (2003). *Institutional repositories: Essential infrastructure for scholarship in the digital age* (No. No. 226). Association of Research Libraries. Retrieved from <http://www.arl.org/storage/documents/publications/arl-br-226.pdf>