

AUTHOR IDENTIFIER OVERVIEW

by Martin Fenner

Abstract

Unique identifiers for scholarly authors are still not commonly used, but provide a number of benefits to authors, institutions, publishers, funding organizations and scholarly societies. This report gives an overview about some of the popular author identifier systems, and their characteristics. The report also discusses several important issues that need to be addressed by author identifier systems, namely identity, reputation and trust.

Introduction

We have long assigned unique numbers to genes, species or stars, and have used unique identifiers for scholarly works for more than 10 years, but unique identifiers for authors are still fairly new and not yet in widespread use (1). Unique author identifiers are useful for the following reasons (2-8):

1. Researchers want to find potential collaborators, and want an easier way to get credit for their scholarly activities,
2. Institutions want to collect, showcase and often evaluate the scholarly activities of their faculty,
3. Publishers want to simplify the publishing workflow, including peer review,
4. Funding organizations want to simplify the grant submission workflow and want to track what happened to the research they funded, and
5. Scholarly societies want an easier way to track the achievements of their members.

The reason that unique identifiers for authors are not as commonly used as unique identifiers for scholarly contributions is not that they are not needed, but rather that they are something rather difficult to implement. In this report I want to summarize the status quo and some of the important issues that need to be addressed by an author identifier system. Throughout the text I will use the term author in the broader meaning of a creator of scholarly works, in most instances this term could be replaced by researcher, scholar or contributor.

Status quo

Some popular author identifier systems for scholarly researchers are listed in table 1. While some systems have been around for more than 10 years, several new systems have emerged in the last three years and there clearly is an increased awareness for unique author identifiers (9, 10). The ORCID and PubMed Author ID system have been announced (11), and are expected to become publicly available later this year. With the exception of the few countries with mandatory author identifiers such as Brazil and the Netherlands, and some specific disciplines, author identifiers are still not widely used.

In addition to unique author identifiers for scholarly works, we also see the emergence of identity systems with a much broader scope. The International Standard Name Identifier (ISNI) system will cover all creators of creative works, including artists, musicians. And OpenID has become the de facto standard for identification and authentication of internet users.

The overview of existing systems is not only helpful to describe the status quo, but also to understand the different approaches to author identification that these systems have taken. In the following sections I want to focus on three important aspects: identity, reputation and trust.

Identity

In its simplest form an author identifier system provides an unique identifier to a person. The identifier could be given to everybody who asks for it – as with the OpenID system (<http://openid.net/>) – or could be given to all authors of creative works – as is intended for the International Standard Name Identifier (ISNI) system (<http://www.isni.org/>) – or could be given only to someone actively involved in scholarly work. In the latter case we have to think about the definition for scholarly work, and here two approaches are in use. One option would be to assign the identifier upon graduation with a science degree, and this is what Brazil and the Netherlands are doing. The problem is that this approach might not catch all authors of scholarly works, and this is why some author identifier systems, including AuthorClaim (<http://authorclaim.org/>) and Researcher ID (<http://www.researcherid.com>) are open to registration by everybody. The other option would be to assign an author identifier when someone has created a scholarly work, most commonly this would mean a scientific paper or book chapter. This is the approach taken by the ArXiv Author ID (http://arxiv.org/help/author_identifiers) and the Scopus Author ID systems (http://help.scopus.com/robo/projects/schelp/h_auteursch_intro.htm).

Until now we have talked about unique author identifiers being assigned proactively, most commonly when an author decides to get an identifier. The much more complicated situation is the retrospective assignment of unique identifiers to authors, including authors that are no longer actively doing scholarly work. Scopus Author ID is an example of a service that does name disambiguation, and ORCID (<http://www.orcid.org/>) is also working on name disambiguation.

This retrospective assignment only works if another person – or a computer algorithm – can unambiguously identify a particular person. There are actually two problems to solve: different people might have the same name, a situation particularly prominent in China and Korea (12, 13). And we have to solve the opposite problem where different names all point to the same person. A reason for this could be name changes, e.g. through marriage, or several different spellings of the same name – this is common for names from countries such as China using non-latin alphabets, but also a problem for countries using the latin alphabet, e.g. because of an umlaut in a German name. Name disambiguation is inherently difficult, and the algorithms are at best 95-98% perfect.

Some of the currently available unique identifier systems are not universal, but limited to a specific discipline (e.g. the ArXiv Author ID to physics, mathematics and related disciplines) or country (e.g. LATTES (<http://lattes.cnpq.br/>) in Brazil or NARCIS (<http://www.narcis.nl/>) in the Netherlands). With this approach we run into problems with interdisciplinary or multinational scholarly works. A good example would be assigning author identifiers to all publications in the multidisciplinary journals Science or Nature. We therefore also need universal identifiers, and Researcher ID, Scopus Author ID, AuthorClaim and ORCID all provide such a service. ORCID is the only service trying to associate the ORCID identifier with other existing author identifiers. This integration is needed so that established specific author identifiers such as LATTES or ArXiv Author ID can be used in parallel with universal identifiers.

Reputation

A unique author identifier in itself has limited value. We have to add meaning to it by associating the identifier with biographic and bibliographic information: where does the author work and has worked in the past, what scholarly works has he created and with whom, what other author identifiers point to the same person, etc. With this information we are building an author profile, and this can be done either by the system issuing the identifier, by the systems that collect scholarly contributions, or by one or more other systems. As there is currently no initiative for a single universal system that holds the scholarly record, profile information for the time being will continue to be distributed and duplicated. All author identifier systems discussed here collect profile information. The profile information is a proxy for the reputation of an author, i.e. the opinion of the scientific community.

While reputation is influenced by many factors, the information that can be collected in an author profile should ideally consist mostly of information collected from other systems using digital identifiers. For scholarly activities we have both discipline-specific identifiers (e.g. pmid PMID for life sciences publications or GIgi for nucleotide sequences) assigned by individual organizations collecting this information and universal digital object identifiers (DOIs) assigned by registration agencies such as CrossRef (<http://www.crossref.org/>) and DataCite (<http://datacite.org/>). Whereas most scholarly publications now have a DOI assigned to them, we are still at the beginning of routinely assigning DOIs to research datasets. We do have universal and unique identifiers for publications and research datasets, but not for the other scholarly activities that could be listed in an author profile, including but not limited to grants, awards, patents, peer review, or teaching. Most unique author identifier profiles are limited in scope to scholarly works, but LATTES, NARCIS, ORCID and PubMed Author ID also look at other scholarly contributions. AuthorClaim, VIAF, Scopus Author ID, LATTES, NARCIS and the Names Project are assigning identifiers to institutions, whereas Researcher ID, ArXiv Author ID and ORCID don't use unique identifiers for institutions.

Not all scholarly activities of an author are public information that can be included in an author profile. Peer review is a good example for an important and valuable scholarly activity where the authors of the reviewed paper or grant do not know the identity of the reviewer. Journals and funding organizations might use unique author identifiers internally to simplify the peer review workflow, but the public author profile will probably at most list the journals and funding organizations for whom the peer review was done.

Related to reputation is provenance, which describes the record of ownership of an object. For a scholarly work provenance not only refers to its authors, but also to the place and time it was published, the other works citing it, etc. When reading a scientific paper or looking at a research dataset, we always do this in the context of its provenance, and this is obviously easier to do with unique author identifiers.

Reputation and provenance in the scholarly context are typically used for knowledge discovery and academic metrics (14). Author profile information collected with the help of unique author identifiers improves knowledge discovery; it becomes much easier to find other scholarly works by the same author or other authors with similar research interests. Academic metrics are increasingly used to make funding and job hiring decisions, and this is done by trying to put the reputation of an academic, department or institution into numbers. Author identifiers simplify academic metrics, but a lot of work

still needs to be done about whether reputation can be put into numbers, how these numbers should be calculated, and whether this is the best approach to forecast the academic productivity of individuals or institutions.

Trust

Identity and reputation are based on trust in the claims made about the author and his scholarly contributions. The individual author has to trust the author identifier system. Most importantly he wants to control the privacy settings of his profile information. Authors also want to know that the author information system is reliable and will be around for a long time to come, and that the information in the system is open, meaning that the data collected by the author identifier system can be freely accessed, exported and reused. Authors also need trust in the organization running the author identifier service, and this has historically been an issue for proprietary systems run by private companies, from Microsoft Passport as single-sign on system for internet users to Thomson Reuters and Elsevier with their Researcher ID and Scopus Author ID services.

Other users of an author identifier system also have to trust the claims made in an author profile. This is not possible in a system that relies on self-claims made by authors – e.g. the AuthorClaim system – but requires verification of these claims. This would typically be institutions for author affiliations, publishers for scholarly publications and data centers for research datasets. Scopus Author ID is an example of a system that primarily relies on external claims. The problem with a system that only uses external claims is that that these claims are much more difficult to do and still will never be 100 % accurate.

The best trust exists in systems that use claims by both authors and external sources. This is most easily done when the author identifier is used at the time a paper, grant or dataset is submitted, and much more difficult when done retrospectively. Self-claims and external claims not only require a unique author identifier, but also a mechanism for authentication (confirm that this is really author x) and authorization (allow journal y to add publication z to author profile y, but not change the other publications). Authentication and authorization are not a core function of author identifier systems, and can also be provided by standard protocols such as OpenID and OAuth.

Conclusions

Unique identifiers for scholarly authors benefit all involved stakeholders, but are currently not common practice. A number of recent initiatives are addressing this problem and we can expect to see major progress in this area in 2011. Author identification is a complex problem and involves a large number of stakeholders who sometimes have opposing views on some of the issues that need to be addressed. Building an author identifier system is therefore not just about technical challenges, it also requires decisions about openness, privacy, collaboration, business models and other critical issues.

Disclaimer

The author is a member of the ORCID Board of Directors. The views expressed here are his personal opinion.

References

1. Aerts R. Digital identifiers work for articles, so why not for authors? *Nature*. 2008;453:979.
2. Falagas ME. Unique author identification number in scientific databases: a suggestion. *PLoS medicine*. 2006;3:e249.
3. Bourne PE, Fink JL. I am not a scientist, I am a number. *PLoS computational biology*. 2008;4:e1000247.
4. Cals JW, Kotz D. Researcher identification: the right needle in the haystack. *The Lancet*. 2008;371:2152-3.
5. Wolinsky H. What's in a name? *EMBO reports*. 2008;9:1171-4.
6. Enserink M. Scientific publishing. Are you ready to become a number? *Science*. 2009;323:1662-4.
7. Habibzadeh F, Yadollahie M. The problem of "Who". *The International Information & Library Review*. 2009;41:61-2.
8. Thorisson GA. Accreditation and attribution in data sharing. *Nature Biotechnology*. 2009;27:984-5.
9. Credit where credit is due. *Nature*. 2009;462:825.
10. Center GPK. Researcher Identification Primer. 2009; Available from: <http://www.gen2phen.org/researcher-identification-primer>.
11. Fenner M. ORCID or how to build a unique identifier for scientists in 10 easy steps. *Gobbledygook* 2010. Available from: http://blogs.plos.org/mfenner/2010/01/03/orcid_or_how_to_build_a_unique_identifier_for_scientists_in_10_easy_steps/
12. Qiu J. Scientific publishing: identity crisis. *Nature*. 2008;451:766-7.
13. Warner S. *Author Identifiers in Scholarly Repositories*. 2010.
14. Lane J. Let's make science metrics more scientific. *Nature*. 2010;464:488-9.

Name	Organization	Kind	Characteristics	Disciplines	Countries	Year started	Link
AuthorClaim	Open Library Society	Nonprofit	Integrates with databases for institutions (ARIW) and publications (3lib.org). Started as RePEc Author Service, extended as AuthorClaim in 2008.	All, currently mostly economics	All	1999	http://authorclaim.org
LATTES	National Council for Scientific and Technological Development (CNPq)	Government	Part of several databases covering many scholarly activities. Mandatory for all Brazilian researchers since 2002.	All	Brazil	1999	http://lattes.cnpq.br/
VIAF	Online Computer Library Center (OCLC) and 15 national libraries	Nonprofit	Integrates name authority records from several national libraries. Also contains other creators of creative content.	All	Several	2003	http://viaf.org/
NARCIS	Royal Netherlands Academy of Arts and Sciences (KNAW)	Government	Part of a database for publications, datasets and research projects	All	Netherlands	2004	http://www.narcis.nl
ArXiv Author ID	Cornell University Library	Academic	Part of e-print archive (ArXiv)	Physics, mathematics, computer science and related disciplines	All	2005	http://www.arxiv.org
Scopus Author ID	Elsevier	Commercial	Integrates with bibliographic database (Scopus)	All	All	2006	http://www.scopus.com
Names Project	Mimas, British Library	Academic	Identifiers for researchers and institutions.	All	United Kingdom	2007	http://names.mimas.ac.uk
Researcher ID	Thomson Reuters	Commercial	Integrates with bibliographic database (Web of Science)	All	All	2008	http://www.researcherid.com
ORCID	ORCID	Nonprofit	Integrates with bibliographic database (CrossRef) and other author identifier systems.	All	All	2009	http://www.orcid.org
PubMed Author ID	National Library of Medicine (NLM)	Government	Part of several biomedical databases for publications and datasets (NCBI)	Life sciences	All	2010	http://www.pubmed.gov

Table 1. Some popular author identifier systems.