Open Access Reference Citations Service with author-identification

Michael Maune[1], Eberhard R. Hilf[1]

Introduction: $\mathcal{DOARC}$ [2] *Distributed Open Access Reference Citations*, is a new service under development to be served by DINI as part of its emerging OA-Network System[3] and funded by the DFG[4] which aims at creating an interactive reference index for scientific documents. Special emphasis is given to the Open Access (OA) documents posted by the present German OAI-PMH-Institutional Repositories at Universities and large Research Institutions.

One part of it will be a citation-based user interface with tools for authors and readers. The general motivation behind $\mathcal{DOARC}$ is to serve add-on services with regard to citations and specifically exploit and make use of the opportunities that the OA document world offers by its access to the full text documents. This will provide an extra benefit for both, authors and readers and thus boost the way to spread OA and thus in the end add to increase the rate of citations in an OA world.

Specifically, by $\mathcal{DOARC}$ authors will be given a tool, to ensure that they cite correctly, and that their document's references list will be extracted and added to the pool of $\mathcal{DOARC}$ citations. Readers will get a tool by which they can find a document relevant for them by browsing through citations and by a graphical tool which shows the 'content affinity' to other documents in the widely distributed pool of scientific OA-papers.

We will exchange our checked metadata with other citation services and further the know-how for non-commercial citation services. In the interface the user will be able to see references with additional information of high value (enriched metadata). We are integrating the services into a wider European context by joining a new initiative organized by Alma Swan of *Key Perspectives*.

Author Identification for Science: Author identification will be an essential requirement[a], [b] of any citation service. The general specifications differ quite widely from say a passport, see [c], [d]. While for a citizen passport the identification of a person as a legal object is the focus, which leads to all kinds of public controls and enforcements, legislation against forgery etc.. The author-identification in science has a distinct role and thus needs distinct ways of realization.

It is not the legal identification of the author, that concerns the reader, but the information which other publications belong to one author. The reader does not care for whether the individual author has changed his legal name (marriage or not), changed or has different variations of spelling, and transliteration into different languages, whether the author is an indvidual or a group of authors, e.g. *Bourbacki*, which is a synonmym for a group of mathematicians who wrote a whole set of very

---

important textbooks on a new way to formulate the Linear Algebra. They wanted to stay individually anonymous by this.

Recently there have been various services coming up, commercial ones such as ResearcherID by Blackwell, registration induced such as by the reference services *Bibsonomy* and *CiteULike* , and *AuthorClaim* . The latter is unique in that it is open to all, free to use, no obligations, and that it serves just the 'naked' service of author-identification. An author can register to the system, set up a list of known variations of his name and can choose his institution from an instituational database. After completing these steps he than is presented a list of documents whose author he or she may be and which he may claim or reject. So not only the accepted papers of an author are saved, also the rejected ones.

Importantly, the usefulness of *AuthorClaim* rests on the completeness of its document information database, which at present goes already into the 20 Millions. *AuthorClaim* just keeps the minimal necessary information for each paper, which are its title, the names of the authors and a link to a description page.

Inclusion of *AuthorClaim* in $\mathcal{DOARC}$: To build author records, $\mathcal{DOARC}$ will use the *AuthorClaim* [5] system, built by Thomas Krichel and maintained by the Open Library System.

Technically the *AuthorClaim* database is mirrored to $\mathcal{DOARC}$ to allow fast user and author add-on tools. Authors, whose email is allocated by $\mathcal{DOARC}$ are alerted and encouraged to register. Beyond the mirroring of *AuthorClaim* database, $\mathcal{DOARC}$ will also sync its own database with *AuthorClaim.* From every newly harvested repository, which is not already a part of *AuthorClaim* , the required set of metadata is being copied to *AuthorClaim.* By keeping both databases in sync, the best possible service for authors, wanting to claim their authorship, is realized.

References:
a: Hilf, Eberhard R.; Krichel, Thomas: *42nd Street paper: scholarly communications is author-driven rather than reader-driven*; acis.openlib.org/documents/42st.html; 2002

b: Hilf, Eberhard R.; Kappenberg, Bernd; Roosendaal, Hans E.; *Author identification: the benefit of being able to identify researchers uniquely*; The Euroscientist 5; 2008; www.euroscience.org/author-identification,e8115,en.html

c: Severiens, Thomas; *Requirements for author registries*; The Euroscientist 5; 2008; www.euroscience.org/requirements-for-author-registries,28114,en.html

d: The ACIS academic contributor information system consortium: citations in ACIS; 2007; acis.openlib.org/software/doc/citations.html

---

[5] authorclaim.org

[6] Palmer School of Library and Information Science, Long Island University, Faculty of Information Technology, Novosibirsk State University and Open Library Society