

Gaps in affiliation indexing in Scopus and PubMed

Cynthia M. Schmidt, MD, MLS; Roxanne Cox, MLS; Alissa V. Fial, MA, MLIS; Teresa L. Hartman, MLS; Martha L. Magee, MSA, MLS

See end of article for authors' affiliations.

DOI: <http://dx.doi.org/10.3163/1536-5050.104.2.008>

Objective: The authors sought to determine whether unexpected gaps existed in Scopus's author affiliation indexing of publications written by the University of Nebraska Medical Center or Nebraska Medicine (UNMC/NM) authors during 2014.

Methods: First, we compared Scopus *affiliation identifier* search results to PubMed *affiliation keyword* search results. Then, we searched Scopus using affiliation keywords (UNMC, etc.) and compared the results to PubMed affiliation keyword and Scopus affiliation identifier searches.

Results: We found that Scopus's records for approximately 7% of UNMC/NM authors' publications lacked appropriate UNMC/NM author affiliation identifiers, and many journals' publishers were supplying incomplete author affiliation information to PubMed.

Conclusions: Institutions relying on Scopus to track their impact should determine whether Scopus's affiliation identifiers will, in fact, identify all articles published by their authors and investigators.

Keywords: "Databases; Bibliographic"[Mesh]; "Organizational Affiliation"[Mesh]; "Efficiency, Organizational"[Mesh]; "Research"[Mesh]; "Writing"[Mesh]

The University of Nebraska Medical Center (UNMC) licenses Scopus, in part, because Scopus indexes author affiliations with affiliation identifiers that promise to make identification of UNMC's publications easy [1, 2]. Scopus's affiliation identifiers are part of a controlled, indexing vocabulary in which each affiliation is worded in only one way. Searchers can easily use the "affiliation identifier" browser to find and select all the affiliation identifiers representing the various colleges and institutes at their universities. All the documents associated with the identifiers can then be retrieved with a single click. When searching with identifiers for their institutions, searchers can be assured that the retrieved publications actually have authors from their institutions.

Anyone who has attempted to use keywords to search for an institution's publications will appreciate the simplicity of an "affiliation identifier" search. Keyword searches depend on the search engine's identification of terms in the affiliation

information that authors and publishers have, respectively, included in the article and submitted to the literature database (PubMed, Scopus, Embase, etc.). Because an institution's authors record their affiliations in a myriad of forms, searchers may be forced to search with a string of relatively nonspecific keywords—city names, state names, and so on—and then check each record to determine whether it does, in fact, represent the publication of an author affiliated with the searcher's institution. This is a time-consuming and tedious task.

The authors began to use both Scopus and PubMed searches to populate the library's "Who's Writing" at the UNMC page [3] and hoped to dispense with the PubMed searches. Discrepancies between the PubMed and Scopus search results led us to suspect gaps in Scopus's affiliation indexing. We searched the literature for any published studies of the quality of Scopus's affiliation indexing but found none.

To identify gaps in Scopus's affiliation identifier indexing of UNMC and Nebraska Medicine (UNMC/NM) affiliated articles, we compared results of a Scopus *affiliation identifier* search to the results of



Supplemental Appendix A, Appendix B, and Table 2 are available with the online version of this journal.

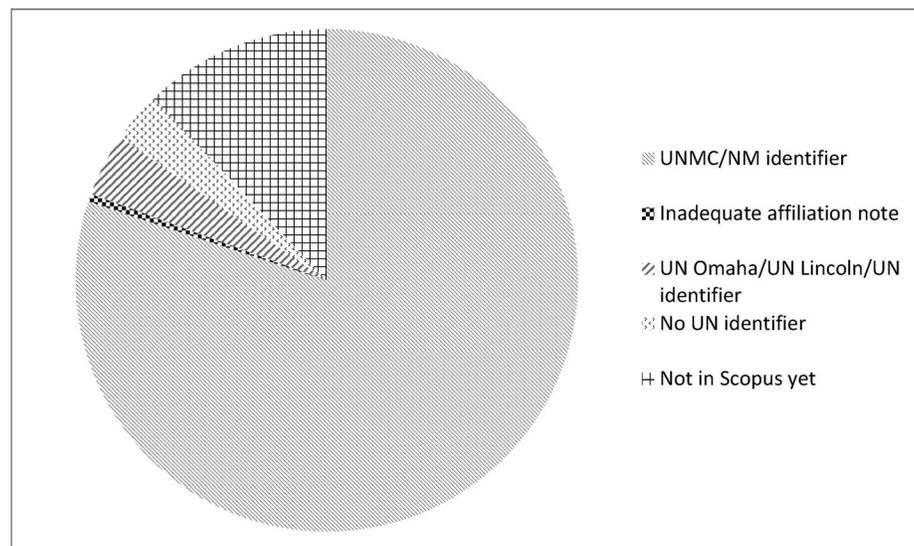


Figure 1

University of Nebraska Medical Center (UNMC)/Nebraska Medicine (NM)-authored articles published during 2014 according to their representation and/or affiliation designation in Scopus

PubMed and Scopus *affiliation keyword* searches. We chose to study 2014 publications for UNMC/NM authors. Studying this time period seemed reasonable as PubMed records began including all author affiliation information submitted by the publisher in October 2013 [4].

METHODS

Phase I: Scopus affiliation identifier search results compared to PubMed affiliation keyword search results

On January 28, 2015, Scopus was searched using all Scopus's UNMC/NM-specific affiliation indexing terms (affiliation identifiers), and PubMed was searched using UNMC-, Nebraska Medicine-, Nebraska-, and Omaha-associated affiliation keywords. The search strategies used are available in online only Appendix A. The search results were sorted into the following groups:

1. articles represented in both groups of search results
2. articles represented in the PubMed affiliation keyword search results and not in Scopus affiliation identifier search results
3. articles represented in the Scopus affiliation identifier search results but not among the results of the PubMed affiliation keyword search

Efforts were made to determine why articles represented in groups 2 and 3 had not been retrieved

by the Scopus and PubMed searches, respectively. During this process, records for epub-ahead-of-print and in-press articles were removed from consideration.

Phase II: Scopus affiliation keyword search compared to PubMed affiliation keyword search and Scopus affiliation identifier searches

After it became obvious that neither the PubMed affiliation keyword search nor the Scopus affiliation identifier search would provide a complete list of UNMC/NM-authors' publications, an additional Scopus search using UNMC-, Nebraska-, and Omaha-associated affiliation keywords was run. UNMC-authored articles not retrieved by the previous searches were studied. Efforts were made to determine why records for these articles had not been retrieved by the previous searches.

RESULTS

During phase I, a total of 1,206 articles published by UNMC/NM authors during 2014 were identified: 733 articles were identified by both the Scopus affiliation identifier and the PubMed affiliation keyword searches; 253 UNMC-authored articles were identified by the Scopus affiliation identifier search only; and 219 by the PubMed affiliation keyword search only (Figure 1). The reasons that the records found by only

	Retrieved by the PubMed affiliation keyword search	Retrieved by the Scopus affiliation identifier search
Corresponding records exist in Scopus, but Scopus records do not contain a UNMC-specific affiliation identifier	70	NA
Corresponding records exist in PubMed, but PubMed records do not include a UNMC/Nebraska-specific affiliation keyword	NA	193*
Corresponding record expected to be present in Scopus but is not	10†	NA
Corresponding record not expected to exist in the other database	139‡	60§
Total records	219	253

* 184 of the PubMed records corresponding to these 193 Scopus records have no author affiliation or only 1 author affiliation listed. This indicates that many journal publishers are not sending full author affiliation information to PubMed.
 † These 10 PubMed records were fully indexed MEDLINE records. Scopus claims to contain records for all articles indexed in MEDLINE.
 ‡ These 139 PubMed records were either PREMEDLINE records or PubMed records for items deposited in PubMed Central that will never be indexed by MEDLINE.
 § These 60 Scopus records included 52 records for articles in journals not indexed by MEDLINE, 5 records for item types not indexed by MEDLINE, and 3 records for articles in journal issues that had not yet been added to PubMed.

Table 1
 Categorization of the University of Nebraska Medical Center (UNMC)–authored articles identified by either the PubMed affiliation keyword search or the Scopus affiliation identifier but not identified by both searches

one of the searches were not found by the other search are described in Table 1.

Seventy of the articles identified by the PubMed affiliation keyword search had corresponding Scopus records that were not indexed with UNMC-specific identifiers. A Scopus affiliation keyword search identified another twenty-six records coauthored by UNMC-affiliated authors that lacked UNMC-specific affiliation identifiers in Scopus. The affiliation identifiers that had been applied to the

ninety-six UNMC-authored articles that were not identified by Scopus’s UNMC-specific affiliation identifiers were studied. About half had been indexed with affiliation identifiers specific to other institutions in the University of Nebraska system, although only one of these ninety-six articles actually had an author affiliated with one of the other University of Nebraska institutions. Six of the ninety-six articles had been indexed with a generic University of Nebraska affiliation identifier, and the

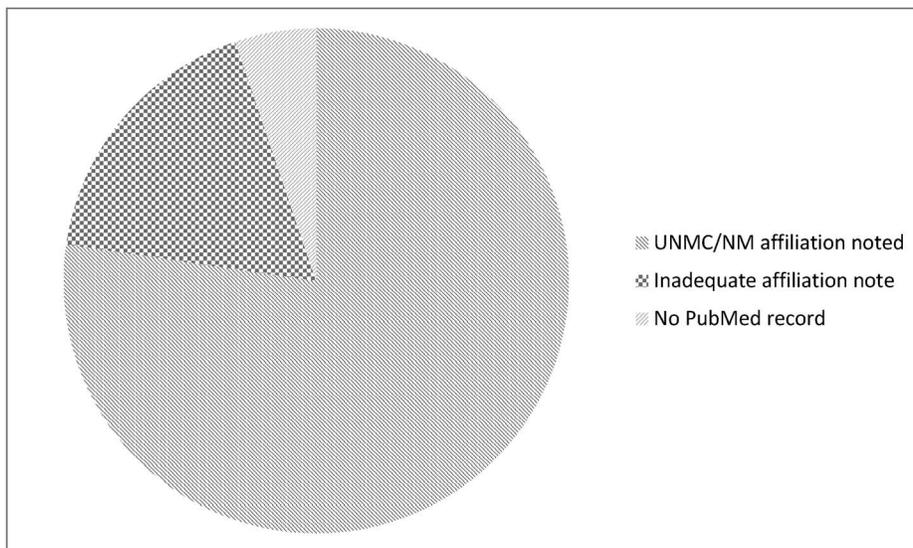


Figure 2
 UNMC/Nebraska Medicine–authored articles published during 2014 according to their representation and/or affiliation notes in PubMed

remaining records had been indexed with none of the University of Nebraska–related identifiers.

The affiliation notes in the ninety-six Scopus records that lacked UNMC/NM-specific affiliation identifiers were then studied. Six of the records lacked UNMC-specific affiliation notes in Scopus. Of the ninety records with adequate affiliation notes, fifty-six contained clear “University of Nebraska” and “Medical Center” or “Medical Centre” affiliation notes (Table 2, online only). Another nine contained a “Nebraska Medical Center” affiliation note. An additional twenty-three contained affiliation notes that one would guess could be easily mapped to UNMC/NM-specific affiliation identifiers by Scopus.

DISCUSSION

Unfortunately, our study shows that about 7% of UNMC/NM’s publications fall through the cracks in Scopus’s affiliation-identifier indexing. We have sent our study findings to Scopus and hope that improved mapping of UNMC-affiliation notes to UNMC-specific affiliation identifiers will result.

Not all our study findings concern flaws in Scopus. Two hundred fifteen articles that in a perfect world would have been identified by the PubMed affiliation keyword search were not retrieved by that search (Figure 2). The publishers of the 215 articles (in 123 journals) had provided “as supplied by publisher” records to PubMed that contained either no author affiliation information, only one author’s affiliation information, or in a few cases, affiliations for several authors. These journals (Appendix B, online only) should be encouraged to take advantage of the change in National Library of Medicine policy and help all their authors’ institutions receive the recognition they deserve.

REFERENCES

1. Scopus: content coverage guide [Internet]. Amsterdam, The Netherlands: Elsevier; 2014 [cited 14 Dec 2015]. <https://www.elsevier.com/_data/assets/pdf_file/0007/69451/sc_content-coverage-guide_july-2014.pdf>.
2. Scopus: features [Internet]. Amsterdam, The Netherlands: Elsevier; 2015 [cited 7 Feb 2015]. <<http://www.elsevier.com/online-tools/scopus/features>>.
3. McGoogan Library. UNMC this week [Internet]. The Library; 2015 [cited 2 Feb 2015]. <<http://www.refworks.com/refworks2/?site=037501144987200000%2fRWWEB104644411%2fUNMCthisweek>>.
4. National Library of Medicine. Changes coming to author affiliations. NLM Tech Bull. 2013(394):b4. <http://www.nlm.nih.gov/pubs/techbull/so13/brief/so13_author_affiliations.html>.

AUTHORS’ AFFILIATIONS



Cynthia M. Schmidt, MD, MLS (ORCID ID: orcid.org/0000-0003-1657-2671), cmschmidt@unmc.edu; **Roxanne Cox, MLS** (ORCID ID: orcid.org/0000-0003-4487-6015) rcox@unmc.edu; **Alissa V. Fial, MA, MLIS** (ORCID ID: [0000-0001-6927-434X](https://orcid.org/0000-0001-6927-434X)) alissa.fial@unmc.edu; **Teresa L. Hartman, MLS** (ORCID ID: orcid.org/0000-0003-1621-6167) thartman@unmc.edu; **Marty L. Magee, MSA, MLS** (ORCID ID: orcid.org/0000-0002-2483-9500) mmagee@unmc.edu; Education and Research Services, McGoogan Library of Medicine, University of Nebraska Medical Center, 986705 Nebraska Medical Center, Omaha, NE 68198-6705

Received March 2015; accepted December 2015

AFTER SUBMISSION INTERACTION

A reviewer asked the authors two questions. Here are the author's responses. –Editor.

If ORCID identification was universal, would that improve searching for specific affiliation? (ORCID is a free digital identifier unique to individual authors)

We do not believe it would have any effect on the ease of affiliation searching. In the event of universal ORCID ID adoption, a librarian working for an institution that has only a few hundred authors or investigators (authors or investigators who rarely move) might, instead of searching for publications naming the institution, search for publications carrying the ORCID IDs of any of the institution's several hundred authors. However, a librarian at a large institution with a continuously changing cadre of thousands of authors would find searching for all the institution's ORCID IDs impractical.

With respect to the articles not identified by the PubMed keyword search, could not some "fault" lie with the authors who incorrectly identify their affiliation? For example, one might use a VA affiliation for work that was done at a VA hospital.

Authors who list alternate affiliations are certainly a problem for those conducting affiliation searches. However, it is unlikely that articles carrying such alternate affiliations would be identified by our study. We only knew about the 215 PubMed records with inadequate author affiliation information, because the corresponding Scopus records contained author affiliation information (either an affiliation identifier or keyword) that could be retrieved by our searches. We only analyzed data for those publications containing a UN- or UNMC-specific affiliation word. The information in the Scopus records came, presumably, from the full-text articles. It seemed unlikely that a check of the full text would reveal that the full-text articles lacked author and contributor affiliations identified by the Scopus records. However, in an effort to respond to the reviewer's comment directly, we checked the full-text versions of a sampling of 55 of the 215 articles with inadequate author affiliation information in PubMed. Each of these 55 full-text articles contained affiliation information for all authors and contributors (including the UNMC affiliated author). In each case, the University of Nebraska Medical Center affiliation was clear.