

Metadata for the metaconsortium

Authors frequently ask the journal to help them get properly listed in PubMed, which we do. However, representing detailed international contributor attribution is no job for a national public library's index.

Every new journal is launched with the ambition 'to be indexed by PubMed'. This index, provided by the US National Library of Medicine (NLM), is used by many researchers to navigate tens of millions of biomedical publications dating back to 1948. So, we commonly receive queries from authors about how to format their manuscripts so that authors are correctly credited. The short answer is that authors must be listed in the paper's byline (the author list under the title) or identified elsewhere in the paper as authors to appear as authors in PubMed. PubMed curators can also identify and index consortium collaborators identified as such within the paper. If contributors are listed as consortium members, but not identified as authors, they will appear in PubMed only as collaborators.

So far, so good, but authors' queries do not stop there; increasingly, we are trying to help with questions such as, "How can we make sure our equal authorship is represented in PubMed?" and "Why doesn't the email address for the corresponding author always appear in PubMed?"

These are frustrating questions because in addition to publishing all the affiliations and contribution roles the authors have provided, we have already provided solutions to many of the problems. You can see this for yourself just a couple of days after online publication of your paper, when the XML tags ('metadata') that we supply automatically appear in PubMed, bearing the equal authorship statements and corresponding author contact details. Unfortunately, once the monthly issue arrives, it is as if ancient scribes had unsheathed their quill pens as the publisher-supplied metadata are systematically scratched out, with consequent loss of statements of equal authorship and accurate affiliations.

Some of the undesirable consequences of this current workflow are that a corresponding author must also be the first published author for PubMed to record his or her email address for correspondence. Even less desirably, if a consortium is the first listed author, the sole listed institutional affiliation of the paper will be that of the second author.

Authors need to be able to provide documentation of their roles in their consortium papers to funders for grants and to committees for career advancement. So that journals can provide this informa-

tion consistently, unique author-controlled identifiers will need to be universally adopted and linked in publisher metadata (Editorial, *Nat Genet.* 41, 1; 2009). The author affiliations and author contribution statements will need eventually to be merged using an agreed tagging protocol with just enough conventions to avoid logical inconsistencies (for example—with apologies to George Orwell's *Animal Farm*—'all authors are equal, but these authors are more equal than the others'). The most highly developed consortium author statements already provide publishers with enough information to do this, by precisely specifying the independent unitary roles for the publisher to tag with metadata. We currently do this with XML, but the traditional print-oriented reader can envisage a familiar output format: separate footnote numbers, each specifying a single affiliation or role, to be used in combinations, but with no one footnote having more than one affiliation or role.

As we said in the January Editorial, we believe that reputation is too important to be controlled by any organization, even one as central to everyday research as NLM's invaluable indexing service. Instead, we favor agreed conventions that allow authors and third-party indices to offer distributed solutions for different applications. All of these solutions would be fed by the metadata provided by authors with their word processors and by publishers with their tagging schemas.

Our recommendation is that PubMed should leave the publisher-supplied metadata as it is supplied if it has ambitions to provide the more detailed author affiliations that authors frequently ask us about. Authors, please think whether you would welcome the wider adoption of existing technical conventions that allow universal and distributed appreciation of your growing reputation, or whether you would rather continue to muddle along trying to extract a reputation from the slowly-evolving customs of a national central library. Whatever solution we end up with, we should keep clear the distinctions between research data collection, analysis and writing. It is generally accepted by editors and readers alike that authorship should be limited to those who actually wrote the manuscript and take responsibility for its accuracy and integrity. We will do everything within our power to help consortium authors to keep this distinction clear. ■