



OPEN ACCESS

Necessary but not sufficient: unique author identifiers

Andrew Marc Harrison,¹ Anthony Mark Harrison²

¹Medical Scientist Training Program, Mayo Clinic, Rochester, Minnesota, USA

²Health Psychology Section, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

Correspondence to

Dr Andrew M Harrison, Medical Scientist Training Program, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA; Harrison.Andrew@mayo.edu

Accepted 5 September 2016

Published Online First

23 September 2016

For better or worse, English is the predominant language used by the international scientific and medical communities to disseminate knowledge. The 26 characters of the Latin alphabet are also arranged in names: non-unique patterns. At the time of the origins of modern biomedical research, names may have been relatively unique, at least within the biomedical research community. However, this is no longer the case.¹ We now possess the capacity to visualise atoms using atomic force microscopy. We also possess the capacity to launch telescopes into space to peer into distant galaxies. However, biomedical researchers do not possess the capacity to automatically distinguish between two researchers who happen to share the same, or similar, names. One decade after the publication of articles on this subject in *PLOS Medicine* and *PLOS Blogs*,^{2–4} the embarrassment of this realisation is eclipsed perhaps only by the continued need to plea for a solution to this ‘intractable’ problem.

Before the National Institutes of Health (NIH) of the USA and its National Library of Medicine (NLM) launched the modern PubMed system, the math, physics and computer science community solved this problem with the creation of arXiv in the early 1990s. Like modern digital object identifiers (DOIs) for unique electronic documents, this largely self-curated system linked non-unique, ‘clickable’ author names with unique author identifiers. Although arXiv and self-curation are not without flaw, this problem has plagued the biomedical research community since at least the inception of arXiv over two decades ago. As a dearth of electronic archival technology is not the problem,⁵ what continues to drive this problem?

When the biomedical research community was relatively small (approximately one to three authors per publication), the first–last/corresponding author paradigm sufficed. At least as recently as the 1970s,

biomedical researchers could still publish dozens of pages meticulously describing how something seemingly as trivial as ‘dirt’ on electron microscopy slides was actually a seminal scientific discovery.⁶ With the modern pressure of word limits, it cannot be known how much insight into this process of discovery of new knowledge is now lost to the need for concision. International collaborations with thousands of physicists now relegate authorship to alphabetical appendices.⁷ In the case of one of the first genomics publications with >1000 authors,⁸ the archaic first–last/corresponding author paradigm was maintained.

By the 1950s, it was ‘too much to expect a research worker to spend an inordinate amount of time searching for the bibliographic descendants of antecedent papers’, which led to the creation of an impact factor.⁹ Initially used in part by libraries to select the best journals to purchase, the use of the term impact factor in this context is different from its modern use by the Science Citation Index (Thomson Reuters). By the 2000s, the need for an index to quantify individual researcher productivity led one physicist to create the h-index.¹⁰ However, when the Royal Society of Chemistry attempted to determine the most impactful chemist by h-index, this task was deemed almost intractable due to the amalgamation of researchers with the name Tanaka K.¹¹ This use of the Western-driven (surname/family name|given/first name|middle initial) system is particularly problematic for Asian biomedical researchers in general: Japan, China and especially Korea, where only a few surnames predominate and middle names often do not exist.

The NIH recently announced a novel Relative Citation Ratio to better measure the true impact of scientific articles.¹² However, the NIH/NLM National Center for Biotechnology Information



CrossMark

To cite: Harrison AM, Harrison AM. *BMJ Innov* 2016;**2**:141–143.



(NCBI) SciENcv system, which allows biomedical researchers to link unique 'My NCBI Bibliographies' with NIH Biosketches, as well as automatically pull US federal grant information from the NIH Electronic Research Administration system ('eRA Commons'), is still not fully linked with the PubMed Advanced Search Builder. Related to the launch of the NLM 'computed author display' in 2012, these systems include 'unique' author search functionality algorithms.

This subject is not new.^{13 14} However, the solution to this problem requires innovation and leadership.¹⁵ Many unique author identifier systems already exist: ORCID, Google Scholar, Mendeley, Scopus, ResearcherID, ResearchGate, etc. Some are open access. Others are proprietary. Some are based largely on self-curation, but all contain some automated component. Several are even linked together. However, every biomedical researcher cannot create and maintain dozens of 'unique' identifiers. The time has come for 'DOIs for authors'. Beyond peer-reviewed publications, a universal unique author identifier system would allow researchers to better track and document the totality of their true scientific productivity: textbooks, textbook chapters, teaching, computer coding, Wikipedia editing and more. The implications of such a system are self-evident,¹⁶ including everything from academic advancement to research funding and plagiarism.

For the rare biomedical researcher with a truly unique last name, or at least last name and first initial, perhaps this is not a major concern. However, for the Tanaka Ks and Harrison AMs of this world, it is. As long as these researchers continue to publish in differing academic fields, manual curation will continue to struggle in the absence of unique author identifiers. However, we already know that this system is fundamentally problematic.¹¹ Maybe some biomedical researchers will eventually add or invent additional middle names.⁶ (We will not even touch the subject of name changes,¹⁷ which is a complex legal matter in the USA and can be a protracted process of obtaining a 'deed poll' in the UK.) However, when the Tanaka Ks and Harrison AMs of the biomedical research world begin to publish within similar fields,^{18 19} and/or together in collaborative scientific endeavours, what will happen then?

The solution to this problem is for PubMed to shift to an arXiv-like, self-curation system, which requires not only this continued plea but also vision and leadership from the highest levels of the international biomedical research community. The pathway to achieve this solution is not trivial and not unique. One pathway to reach this solution is for PubMed to adopt an existing unique author identifier system, such as ORCID, which is already used by many publishing groups. Another option is for PubMed to create its own unique author identifier system, which already

partially exists in forms such as eRA Commons and SciENcv. No pathway will be free. Although self-curation has worked well for arXiv, a comparatively greater amount of supervised-curation, which is already the case for proprietary systems such as Scopus, may be required for biomedical researchers to mitigate some of the flaws of self-curation. It should also be noted that the worldwide 'PubMed research community' is significantly larger than the worldwide 'arXiv research community', which increases the challenge of implementation of this solution.

Any pathway to this solution should also optimise implementation time, which is already an area of active informatics research. However, the complexity of the relationship between clever biomedical researchers,²⁰ publishing groups and funding organisations continues to increase. Thus, a renewed push for urgency for this change is needed from the increasingly fast-paced communities of science and medicine.

Twitter Follow Anthony Mark Harrison at @antmarkharrison

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- 1 Taha K. Extracting various classes of data from biological text using the concept of existence dependency. *IEEE J Biomed Health Inform* 2015;19:1918–28.
- 2 Falagas ME. Unique author identification number in scientific databases: a suggestion. *PLoS Med* 2006;3:e249.
- 3 Cave R. Unique Author Identification. PLOS Blogs, 2006. <http://blogs.plos.org/plos/2006/11/unique-author-identification/>
- 4 Polychronakos C. Unique author identifier; what are we waiting for? *J Med Genet* 2012;49:75.
- 5 Zhang MW, Yeo LL, Ho RC. Harnessing smartphone technologies for stroke care, rehabilitation and beyond. *BMJ Innov* 2015;1:145–50.
- 6 Carpenter ATC. The recombination nodule story—seeing what you are looking at. *BioEssays* 1994;16:69–74.
- 7 Aad G, Abbott B, Abdallah J, *et al.* Combined measurement of the Higgs Boson Mass in p p collisions at s= 7 and 8 TeV with the ATLAS and CMS experiments. *Phys Rev Lett* 2015;114:191803.
- 8 Leung W, Shaffer CD, Reed LK, *et al.* Drosophila muller f elements maintain a distinct set of genomic properties over 40 million years of evolution. *G3 (Bethesda)* 2015;5:719–40.
- 9 Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas. *Science* 1955;122:108–11.
- 10 Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 2005;102:16569–72.

- 11 Broadwith P. End of the road for h-index rankings. *Chem World* 2013;10:12–13.
- 12 Hutchins BI, Yuan X, Anderson JM, *et al.* Relative Citation Ratio (RCR): a new metric that uses citation rates to measure influence at the article level. *PLoS Biol* 2016;14:e1002541.
- 13 Garfield E. The role of the medical librarian in SDI systems. *Bull Med Libr Assoc* 1969;57:348–51.
- 14 Enserink M. Scientific publishing. Are you ready to become a number? *Science* 2009;323:1662–4.
- 15 Fairall L, Bateman E, Cornick R, *et al.* Innovating to improve primary care in less developed countries: towards a global model. *BMJ Innov* 2015;1:196–203.
- 16 Coats AJ. Ethical authorship and publishing. *Int J Cardiol* 2009;131:149–50.
- 17 Warner ME, Warner NS, Warner LL. Medical marriages: time-sensitive bliss. *Mayo Clin Proc* 2013;88:213–5.
- 18 Harrison AM, McCracken LM, Bogosian A, *et al.* Towards a better understanding of MS pain: a systematic review of potentially modifiable psychosocial factors. *J Psychosom Res* 2015;78:12–24.
- 19 Harrison AM, Heritier F, Childs BG, *et al.* Systematic review of the use of phytochemicals for management of pain in cancer therapy. *Biomed Res Int* 2015;2015:1–8.
- 20 Chapman C, Slade T. Rejection of rejection: a novel approach to overcoming barriers to publication. *BMJ* 2015;351:h6326.