

# (Wo)man vs. Machine

Elizabeth Blake

Inera Inc.

The Good, the Bad, and the Ugly in Citations

May 8, 2018

2018 CSE ANNUAL MEETING

# PUBLISHING

IN THE CHANGING CLIMATE OF  
THE RESEARCH ECOSYSTEM

[WWW.COUNCILSCIENCEEDITORS.ORG](http://WWW.COUNCILSCIENCEEDITORS.ORG)



MAY 5-8, 2018  
NEW ORLEANS, LA

# Background on Inera

- ▶ We're in the business of automating certain citation/reference processes, including:
  - ▷ Copyediting
  - ▷ Linking
  - ▷ Validating
  - ▷ Tagging
- ▶ How well these processes can be automated depends on the quality of the input
- ▶ How well the results can be leveraged by publishers depends on the quality of the output

# Quality of Input: Authors

# "Can we count on authors to provide the correct information?"

– Angela Cochran

- ▶ I assumed this was a rhetorical question...
- ▶ For more information, see:
  - ▷ Meyer CA. Reference accuracy: best practices for making the links. J Electron Publ. 2008; 11: <https://doi.org/10.3998/3336451.0011.206> and references therein, especially reference 10
  - ▷ <https://arxiv.org/abs/cond-mat/0401529>
- ▶ Anecdotally, we've learned that since the APA style guide began requiring DOIs (6th edition), the APA has found that ~20% of author-supplied DOIs are incorrect or corrupt in some form or another
- ▶ If you're not editing and validating your references, you're publishing errors

# The Variety of Editorial Styles

- ▶ These types of errors highlight the challenges of:
  - ▷ Relying on authors to provide accurate data
  - ▷ Requiring authors to know and follow multiple editorial styles
- ▶ We've noticed a trend in recent years of publishers being more open to adopting standard styles (e.g., AMA, ICMJE) than they were in the past, which can enhance compliance and accuracy
- ▶ What drives your choice to follow a particular style? Brevity? Completeness? Tradition? Inertia?
- ▶ If you're struggling with citations, it may be worth reviewing your editorial style and asking yourself if it's part of the problem

# Quality of Output: Publishers

# Bad Publisher Behavior

- ▶ As a former copyeditor, I'm not going to call out copyeditors for reference errors that make it to print 😊
- ▶ Instead, I want to point out some choices publishers make that can reduce the quality and discoverability of their citations
- ▶ Oh, the things we've seen...

# If You'd Like Automated Citation Tools to Fail, Consider...

- ▶ Restarting pagination at 1 for every issue in a volume

CDC. General recommendations on immunization: recommendations of the Advisory Committee on Immunization Practices (ACIP) and the American Academy of Family Physicians (AAFP). MMWR 2002;51(No. RR-2):1-35

- ▶ Giving your journal a name that could also be a date, like "October" or "1999"

Levi, P. Cinema by Other Means. October. Winter 2010, No. 131: 51–68

- ▶ Including numerals in your journal name

Benson V, Marano MA. Current estimates from the National Health Interview Survey, 1992. Vital Health Stat 10 1994 Jan;(189):1-269

# If You'd Like Automated Citation Tools to Fail, Consider...

- ▶ Changing your journal name without changing the ISSN
  - ▷ 1977–1992: *International Journal of Obesity* (ISSN 0307-0565)
  - ▷ 1992–2002: *International Journal of Obesity and Related Metabolic Disorders* (ISSN 0307-0565)
  - ▷ 2005–present: *International Journal of Obesity* (ISSN 0307-0565)
- ▶ Not following [NFAIS best practices for assigning article IDs](#) (e-pages)
- ▶ Using an idiosyncratic style for citations to your own publications

# DOI Deposit Errors

Some errors publishers make when depositing metadata to Crossref:

- ▶ Including referees or editors as authors
- ▶ Using the year of deposit rather than the year of publication
- ▶ Leaving out metadata (e.g., issue or pagination)
- ▶ Setting author names or article titles in all caps
- ▶ Using encoding other than Unicode (but declaring the encoding as Unicode) when depositing special characters

# Best Practices for Best Results

- ▶ If you want automated tools to be able to link your citations, there are a few key principles
- ▶ Include enough accurate reference data for a machine to be able find the version of record online (and to disambiguate one reference from another, similar reference)
  - ▷ Numeric info (year, volume, issue, pages) is critical, but text such as titles can help refine the search
  - ▷ Chemistry-style references, which lack article titles, can lead to incorrect links
- ▶ If journal names in references are incorrect or non-standard, links are likely to fail

# The Risk of Leaving Information Out

- ▶ In theory, an automated lookup of a journal reference that includes only the first author, the article title, and the journal name should return a link
- ▶ A lookup of the same reference that includes only the journal name, the volume, the year, and the first page number should also work
- ▶ But if any one of those elements is incorrect in either scenario — the author's name is misspelled, or the first page number is wrong — the link will likely fail or (even worse) point to the wrong record
- ▶ A complete reference often enables tools to link accurately even when there are errors in the reference

# Beyond Linking

Tools that are able to return accurate links can also:

- ▶ Correct bad data in references
  - ▷ Misspelled or missing author names (e.g., expanding out "et al." to include additional authors)
  - ▷ Errors in article titles
  - ▷ Incorrect pagination
- ▶ Flag papers that have been corrected or retracted
- ▶ Flag publications from predatory journals?
  - ▷ No tools that we're aware of yet, but with a reliable list we could create this fairly easily

# Preprints

- ▶ There are emerging editorial styles for citing preprints, but it's in flux
- ▶ Most authors don't know how to cite them... and editors may not have settled on a style either
- ▶ Reference managers (e.g., EndNote) may not support them
- ▶ Preprints are heavily dependent on DOIs, so you'll want to verify that those are correct
- ▶ In general, citations to preprints currently require a lot more editorial oversight than citations to journals

# The Importance of Being Editors

- ▶ Publishing is moving more and more toward machine readability and big data analysis
- ▶ For these initiatives to work, references — and their underlying persistent identifiers like DOIs — need to be accurate and well structured
- ▶ Good tools can help fix many errors in references, but even the best tools can't automate everything
- ▶ Automation **always** requires human oversight, and good editors will **always** be needed to sort out issues that tools can highlight but can't fix

# Thank you!

Elizabeth Blake

Director of Business Development

Inera Inc.

19 Flett Rd.

Belmont, MA 02478 USA

[lblake@inera.com](mailto:lblake@inera.com)

... come say hello at stand 11!