Everything you wish you didn't have to know about metadata matching

Dominika Tkaczyk, Crossref

What is metadata matching?

article

DOI: 10.5555/my-doi

Dept. of Mechanical Eng., The Ohio State University

[5] A. Smith, "A Very Important Paper," The Journal, vol. 3, no. 1, pp. 68-80, 2001
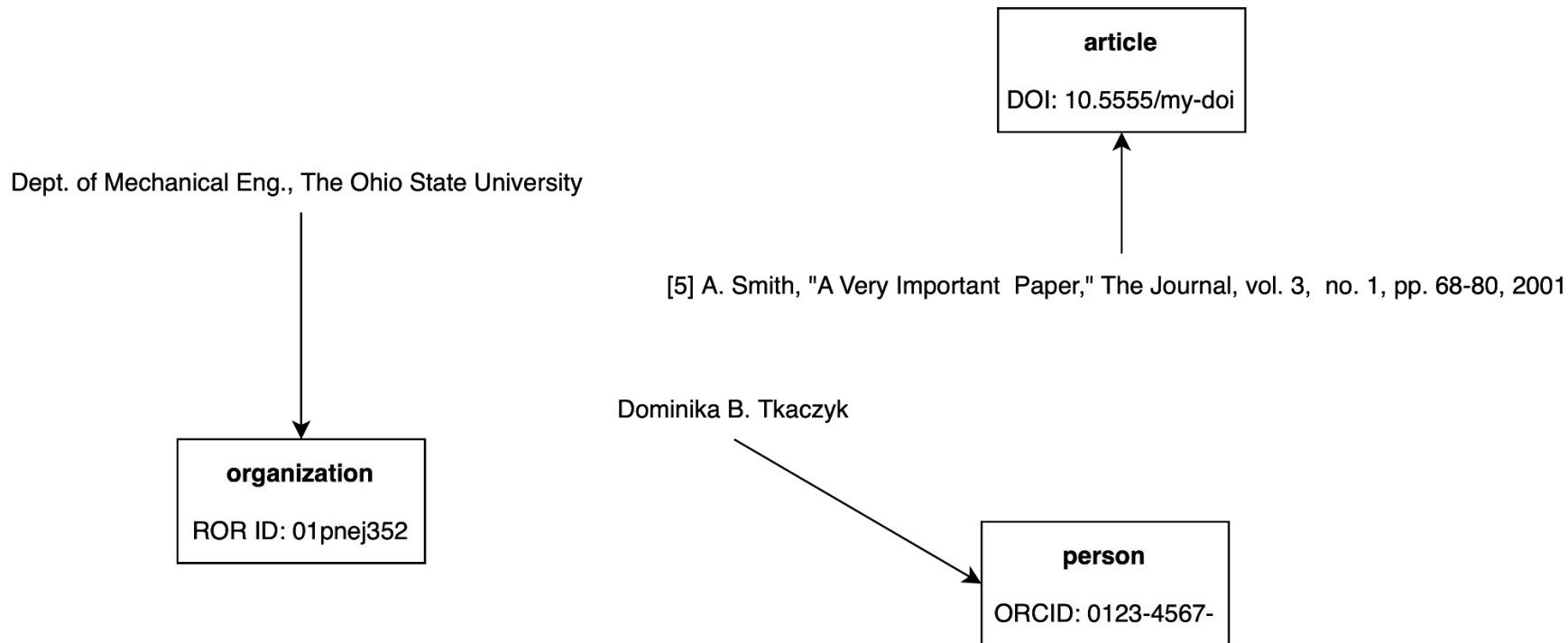
Dominika B. Tkaczyk
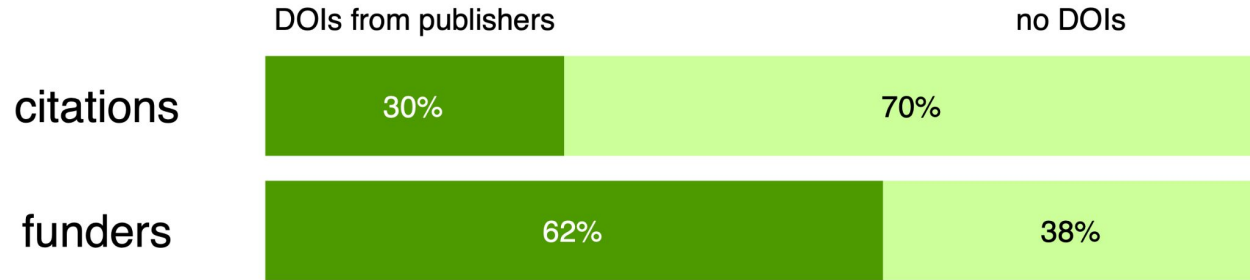
organization

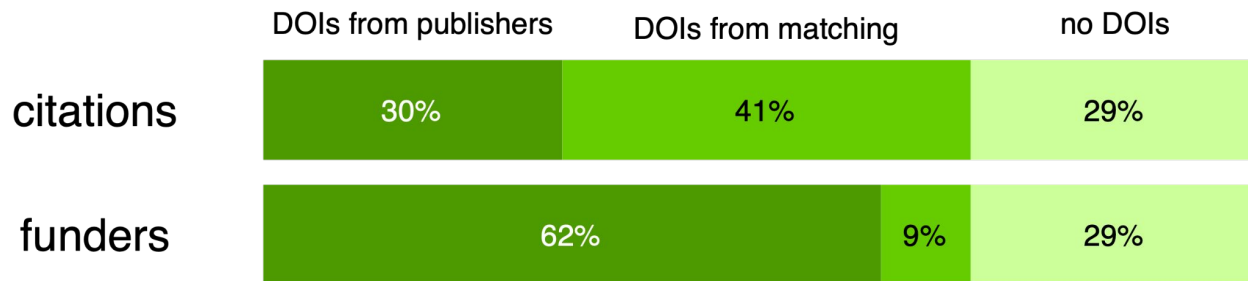ROR ID: 01pnej352

person

ORCID: 0123-4567-

What is metadata matching?

Why do we need matching?

What do we do at Crossref?

	DOIs from publishers	DOIs from matching	no DOIs
citations	30%	41%	29%
funders	62%	9%	29%

- citation matching inserts DOIs for 41% of citations
- funder matching inserts funder DOIs for 9% of funder mentions
- research around grant matching
- plans to address preprint matching
- plans to address affiliation matching
- modular matching API

Questions?

As time goes by, do we need to do less and less matching?

No:

- The fraction of citations deposited with DOIs does not increase over time.
- With new identifiers emerging, there are new types of matching needed.
- Typically we collect mentions of items BEFORE we start giving them identifiers.

Questions?

Isn't a simple title lookup enough to match a citation?

No:

- Sometimes we do not know the title of a citation.
- The title in the citation and the title in the cited article might differ.
- There might be multiple articles with the same title.

Questions?

Can we be 100% sure that all citation links we see in the data are correct?

No:

- If we require the matching to be very confident about a link, it will miss a lot of valid links because of small differences in the metadata.
- If we require the matching to be only somewhat confident about a link, we will introduce a lot of incorrect links.
- Usually we choose something in between, but on a large scale, it will result in a few incorrect citation links.

Thank you for listening!

dtkaczyk@crossref.org

Crossref