

Hossein Dehdarirad, James Thomas

EPPI Centre, Social Research Institute,
University College London, London, UK.
h.dehdarirad@ucl.ac.uk

Investigating the trends in the abstract and reference sections of OpenAlex records

Background

- OpenAlex: A comprehensive free database with over 250 million scholarly records.
- Abstract and citation information is useful in (automated) evidence surveillance workflows.
- Limitation: Metadata associated with these records can be unstable.

Aim

What proportion of OpenAlex records contain abstracts and references, and how do these proportions change over time?

Conclusion

- Improved metadata completeness over time.
- Changes in abstract length and reference numbers may hinder the reproducibility of research conducted on OpenAlex.
- If records improve over time, then acquiring the newest records for living review updates risks incorrectly excluding them if they are lacking key data.
- Major publishers' requests for the removal of non-open-access abstracts from OpenAlex may result in a decrease in data availability for next analyses.

Methods

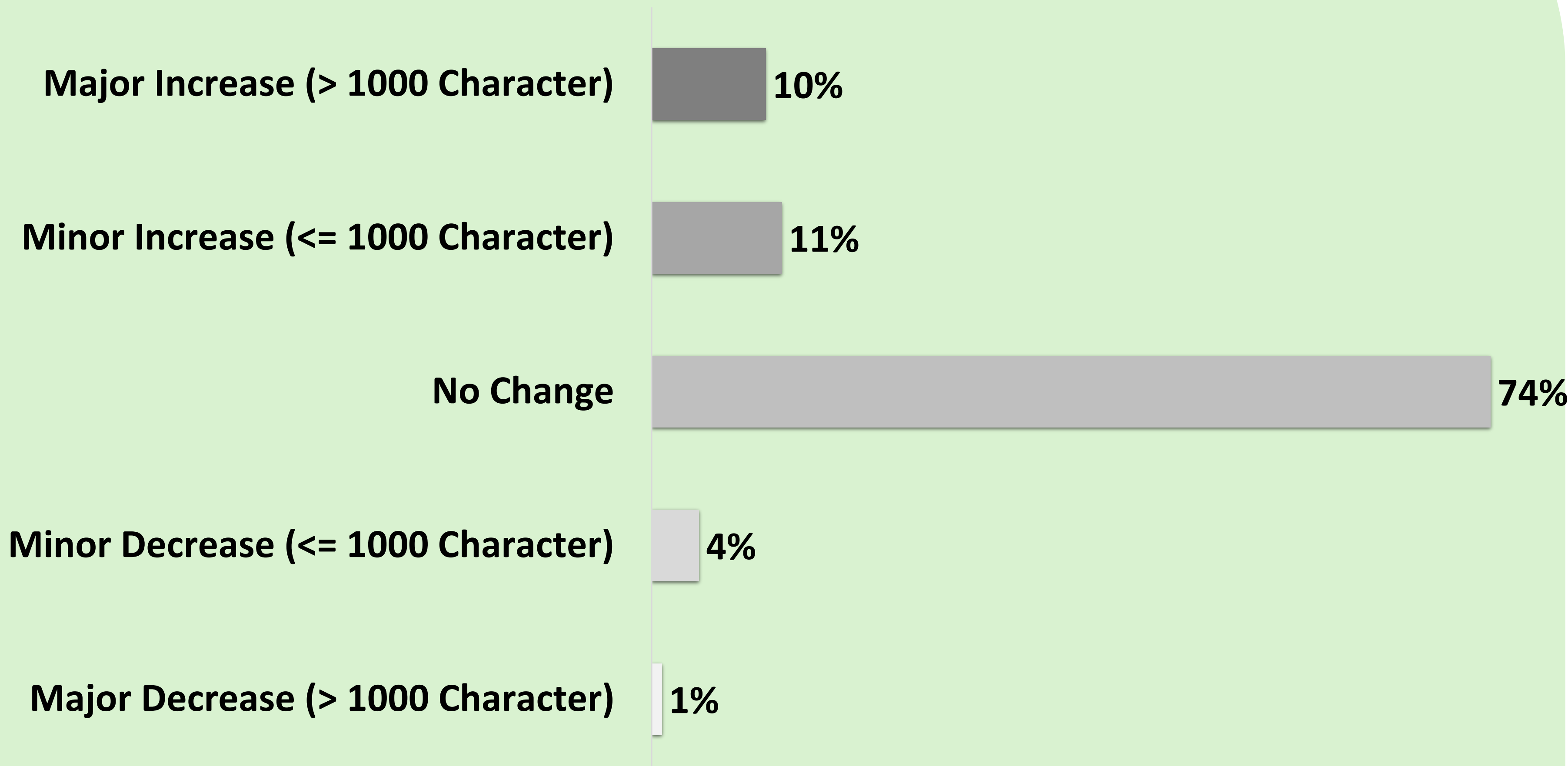
Data: 495,607 records added to OpenAlex in June 2023 (original set)

Investigation Time: November 2024 (updated set)

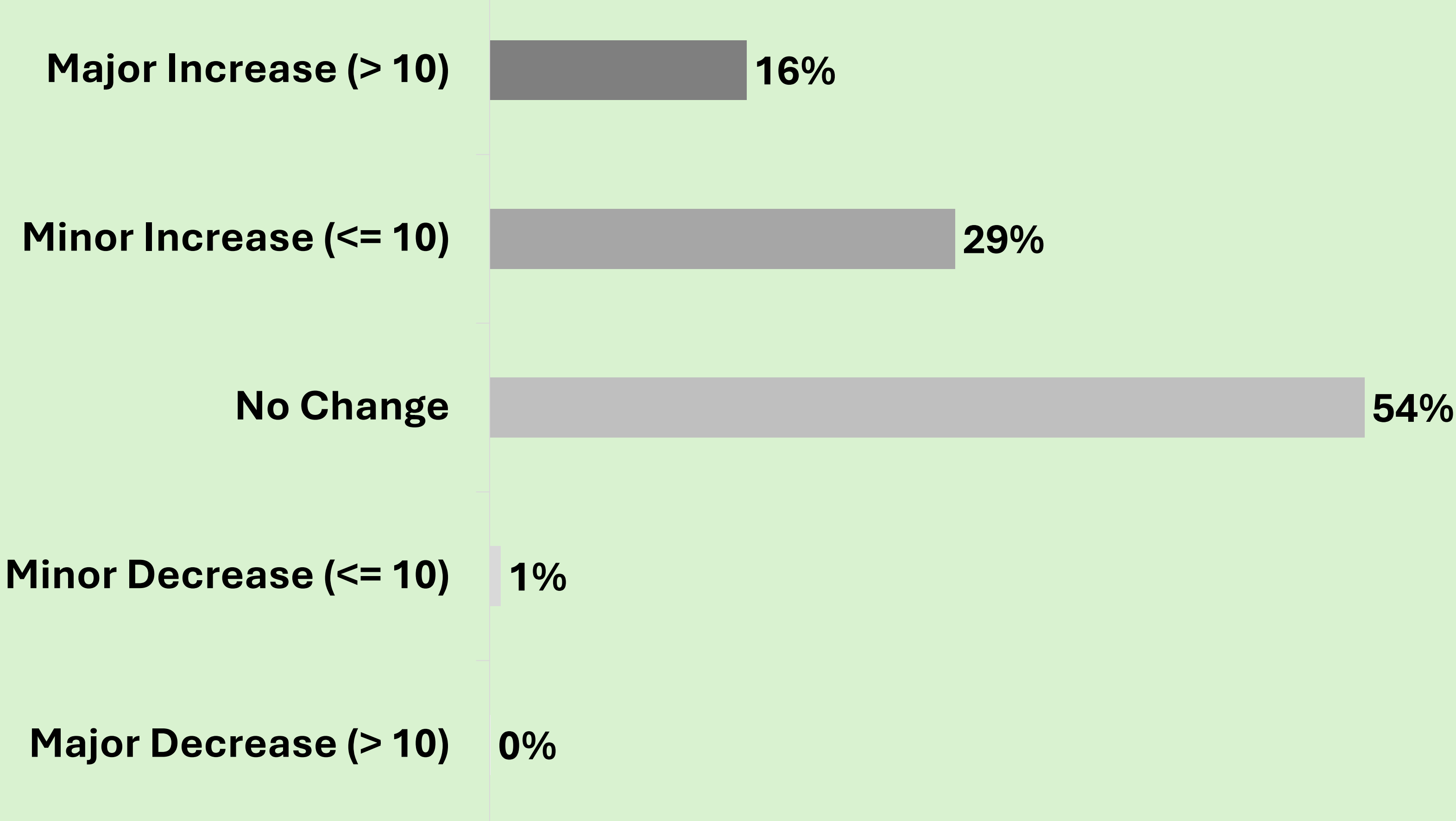
Tool: Python scripting with OpenAlex API

Investigated Metadata: Abstract, Reference, Open Access Status, Scholarly Work Type

Changes in Abstract Length Over Time

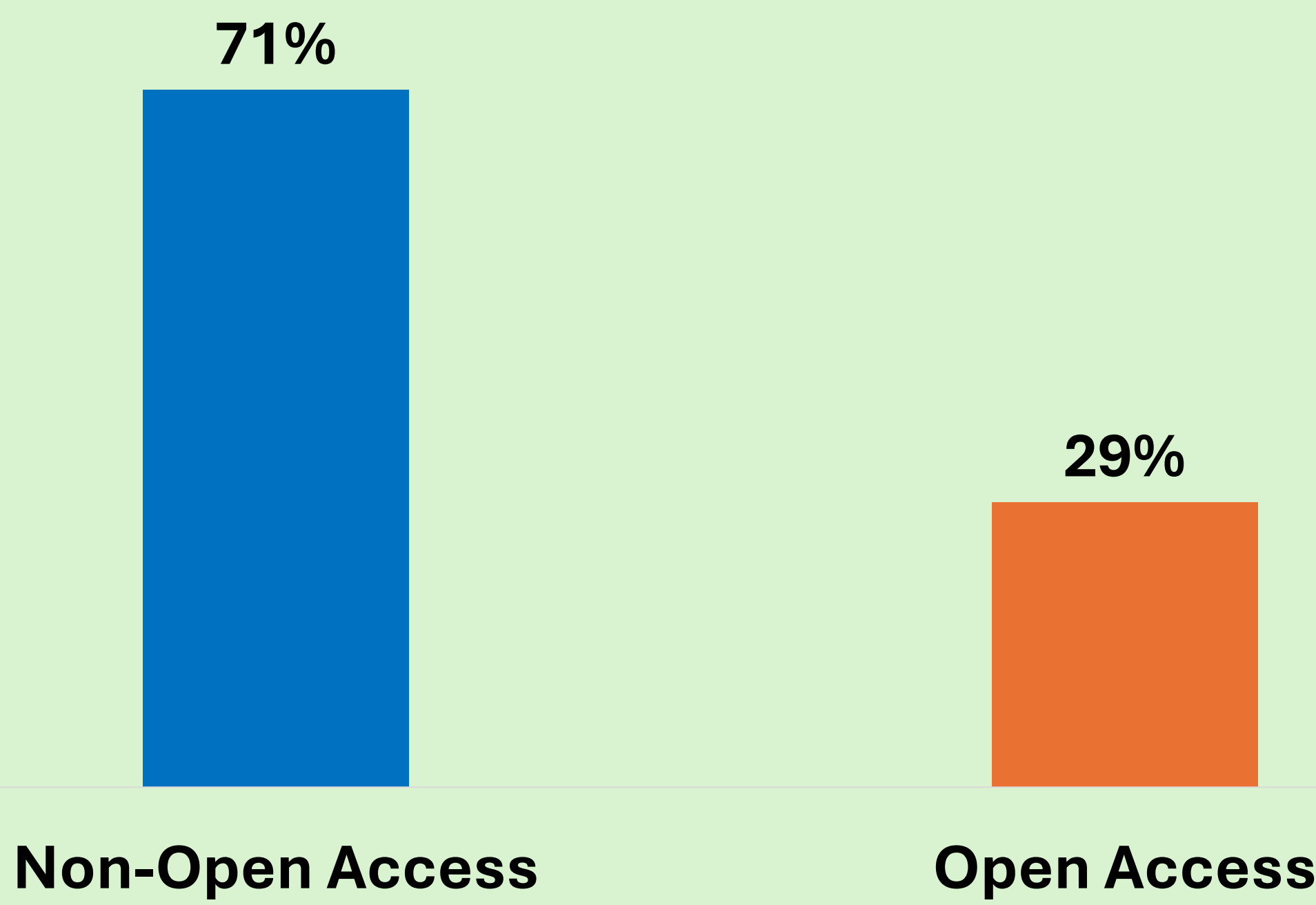


Changes in Reference Number Over Time



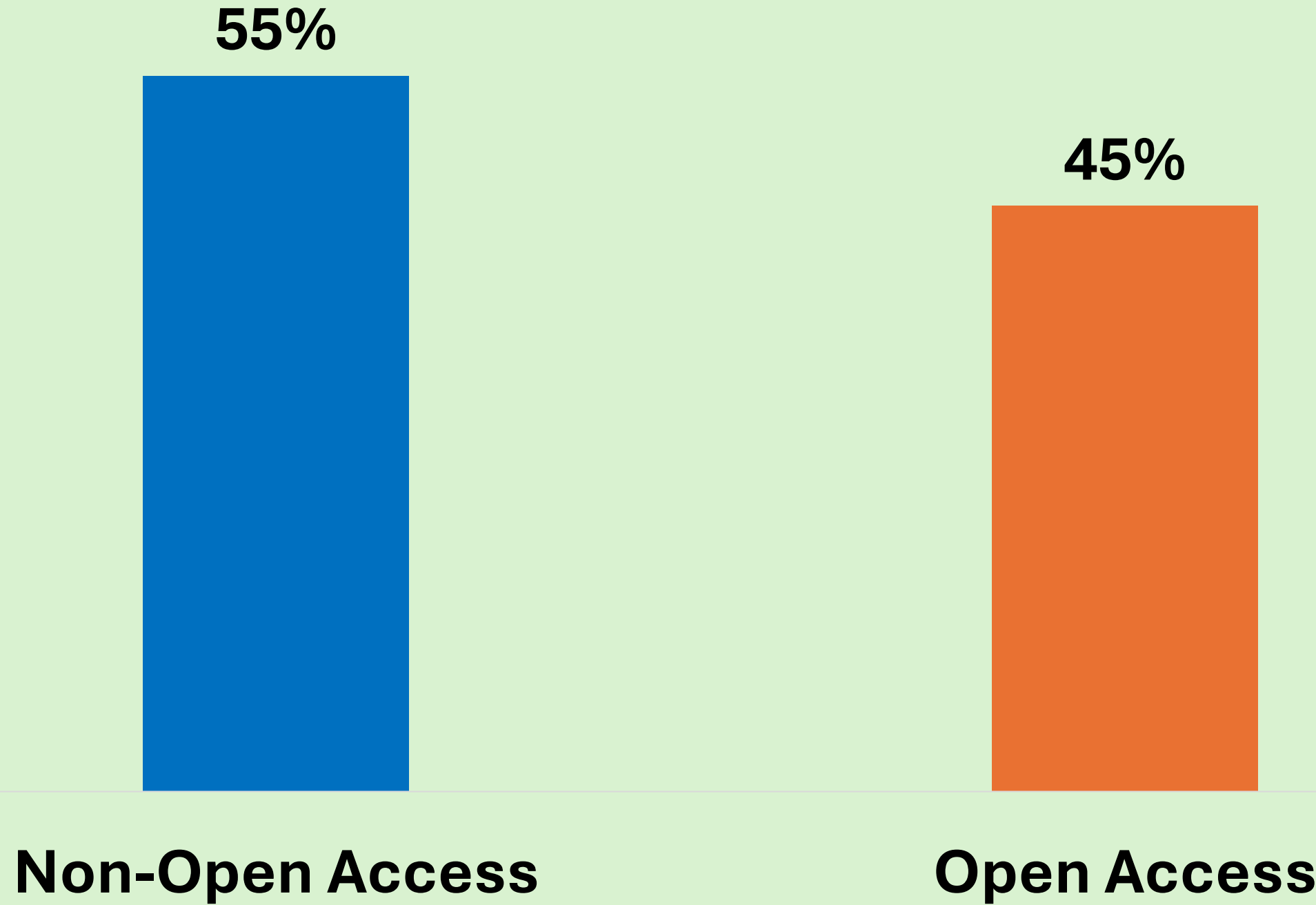
Records with Empty Metadata (November 2024)

Abstract (N=121,190)



- Empty Abstract records reduced 39% to 24%.
- Top three work type:
 - Article (42%)
 - Book chapter (26%)
 - Para texts (8%)

Reference (N=190,545)



- Empty Reference records reduced 57% to 38%.
- Top three work type:
 - Article (43%)
 - Book chapters (20%)
 - Preprint (14%)