

# Making sense of the mathematics literature

Edward Dunne

Executive Editor, *Mathematical Reviews*

American Mathematical Society

*Chorus Forum: How Can We Ensure Research Integrity?*

24 July 2024



# Introduction

The digital age has allowed publishing and research to flourish. The openness of the system has allowed many new participants, such as groups of scientists hoping to provide outlets for their subdisciplines or entrepreneurs seeing opportunity in a new market. Separating the wheat from the chaff is challenging, especially given the high volume.

There two areas of need for integrity of the literature:

- Correctness of the bibliographic record
- Establishing and maintaining standards for the research.

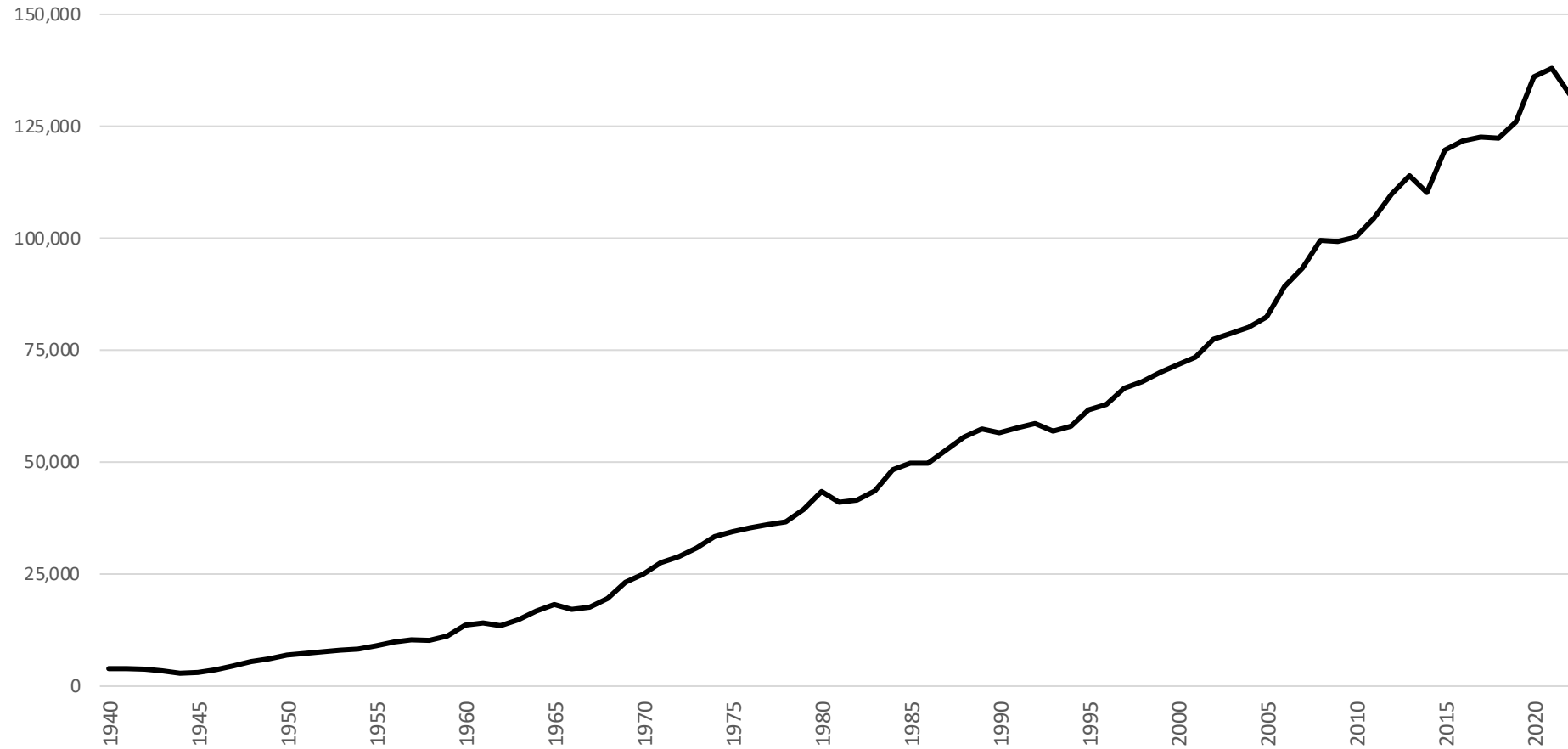
# Introduction

Mathematics is fortunate to have a decades-old indexing and reviewing service, MathSciNet, that can be especially useful for both bibliographic and editorial integrity.

Note: “Mathematical Reviews” will often be abbreviated to “MR”.

# Growth of the literature

Math Reviews database: 1940-2021



# Growth of the literature

- With such rapid growth, there is a need for systems or tools to help researchers, librarians, and others to make sense of the literature:
  - To find relevant publications on topics of interest
  - To access the sources
  - To assess the sources

# Bibliographic integrity

As any librarian or archivist will tell you, generating, obtaining, or possessing good materials is only the beginning. You also have to know what it is and how to find it. This is the function of *cataloging* and its close relative, *metadata*.

# Metadata

Essentially every literature database relies heavily on metadata for keeping track of publications. Mostly, metadata are created by publishers. It is how the publishers inform the world of the existence of new issues and new articles. The metadata contains descriptive information about the issues and articles: title, authors, date of publications, pagination, etc.

Sometimes, secondary services, such as databases or libraries, complete, correct, or enhance metadata.

# JATS

Perhaps the most widely used standard for metadata in publishing is JATS: the Journal Article Tag Suite. It provides a common XML format in which publishers and archives can exchange journal content. JATS provides a set of XML elements and attributes for describing the textual and graphical content of journal articles as well as some non-article material such as letters, editorials, and book and product reviews.



# Other (meta)data

- Document identifiers: In the digital age of mathematics and science, perhaps the most important datum is the DOI. It plays the role of the URL for online journals and books. It is often the stand-in identifier for a publication. The DOI is a Persistent Identifier (PID).
- Author identifiers: ORCiDs, MR Author IDs.
- Institution identifiers: Ringgold IDs, ROR IDs, ORCiDs, MR Institution IDs.
- Subject identifiers, such as the MSC from MR and zbMATH.

# Correctness and completeness of metadata

Publishing has rich standards. But what good is a standard if people don't use it or don't use it well?

Problems:

- Incomplete records
- Errors

# Correctness and completeness

Some types of problems we see in metadata:

- Given name / surname combined into a single AuthorName field. This even occurs when separate fields are included in the metadata, with the word “null” or other meaningless information in one of the fields.
- MSCs are often combined with other kinds of keywords, requiring validation and sometimes breaking parsing.
- ISSNs may occur in a separate file from the paper’s metadata file, without a clear link to it.

# Correctness and completeness

Some types of problems we see in metadata (continued):

- The metadata refers to a pre-publication filing despite having been added to an issue.
- The ISSN is wrong. We have at least one journal that generally has the ISSNs for another journal from the same publisher.
- The link to the PDF points to the wrong file or version.
- The DOI is already in use (rare), is incorrect, or is unregistered.

# Correcting and completing: data

Publishers either deliver metadata directly to libraries and databases such as MathSciNet and zbMATH Open, or make it available through websites or APIs. Oftentimes, incomplete delivered metadata can be completed by querying the publisher's website or double-checking with a DOI registry, such as CrossRef.

For publishers who present consistent problems, a process can be adjusted accordingly, such as the case where the ISSN assigned to a particular journal is consistently the ISSN of a different journal.

In many cases, bibliographic information (metadata) must be corrected or completed by an examination of the paper.

# Correctness and completeness

## Some data:

In the period from 7 January 2024 to 3 July 2024, catalogers at Mathematical Reviews processed 61,788 journal articles. The catalogers made 80,976 corrections and additions to bibliographic metadata (not including name verifications; not including subject classes).

## Some specifics:

Institution	42,677
Paging	21,936
Title	13,748
Author	2,280
Editor or author related	225
Misc.	110

# Correcting and completing: authors

Author information is part of metadata. Sometimes it is difficult to know who is who when trying to assign an author identifier.

- Author search in MathSciNet for **Wei Wang** yields 580 matches. Over 250 of them are exactly “Wei Wang”. Others are variations, such as Wei Ping Wang, Wei Wei Wang, and so on.
- **Alexei Myasnikov**: there are two of them affiliated with Stevens Institute of Technology (New Jersey, USA). One will publish as Alexei D. Myasnikov. The other will sometimes publish as Alexei G. Myasnikov. They have 6 joint papers.

# Correcting and completing: authors

More examples of author difficulties:

- **Ángel Plastino**: there are two. One sometimes publishes as **Ángelo Plastino**. Both are affiliated with Instituto de Física de La Plata, National University of La Plata, Argentina. They have 85 joint papers.
- Besides many authors to one name problem, a single author can be published under many names. Case in point, **Yuriï Alekseevich Mitropol'ski** is a Ukrainian mathematician with 471 publications published under 41 different name variations.



# Correcting and completing: authors

The majority of authors can be identified programmatically.

Indeed, by matching on a small set of elements, we can match 2/3 of the authors of newly-arriving papers with a unique author in the database.

In many of the remaining 1/3 of cases, there is only a slight variation between the information related to the author of the paper and the information for an author in the database, enough so that the match is “blatant”, if not exact.

For the rest, it takes a human touch. Sometimes, we need to write to an author for clarification.

# Editorial integrity

- Using ISSNs to build profiles of journals provides a way to gather data on the journal.
  - Dominant subjects areas represented in the journal
    - Look for anomalies: *Journal of the Society of Commutative Algebra*<sup>(1)</sup> has more than half its papers in computational fluid dynamics
  - What other journals are citing this journal?
  - What journals are being cited *from* this journal?
    - Can help to identify patterns of *citation stacking*;
    - Can help to identify patterns of *citation cartels*.

(1) Not a real journal

# Editorial integrity and a curated database

There are approximately 1,650 journals currently being indexed in MathSciNet. Historically, we have indexed more than 7,000 journals. Some have stopped publishing, others have changed sufficiently that they count as new journals, and others have been dropped for coverage by Mathematical Reviews.

# Editorial integrity and a curated database

To be included in MR coverage, a journal must meet the criteria of Mathematical Reviews Editorial Statement (for content) and Policy on Indexing Journals (for publishing practices).

<http://www.ams.org/publications/math-reviews/mr-edit>

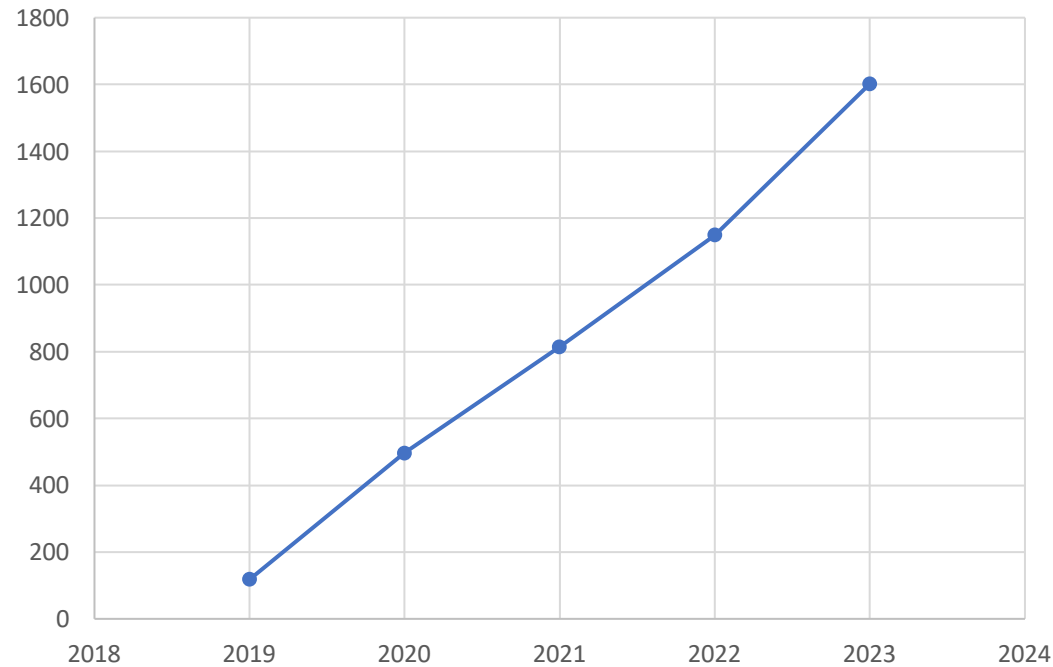
<http://www.ams.org/publications/math-reviews/mrelecjour>

The approval rate for journals proposed for coverage by MR is approximately 17% over the past six years.

# Growth of the literature

## some special contributors

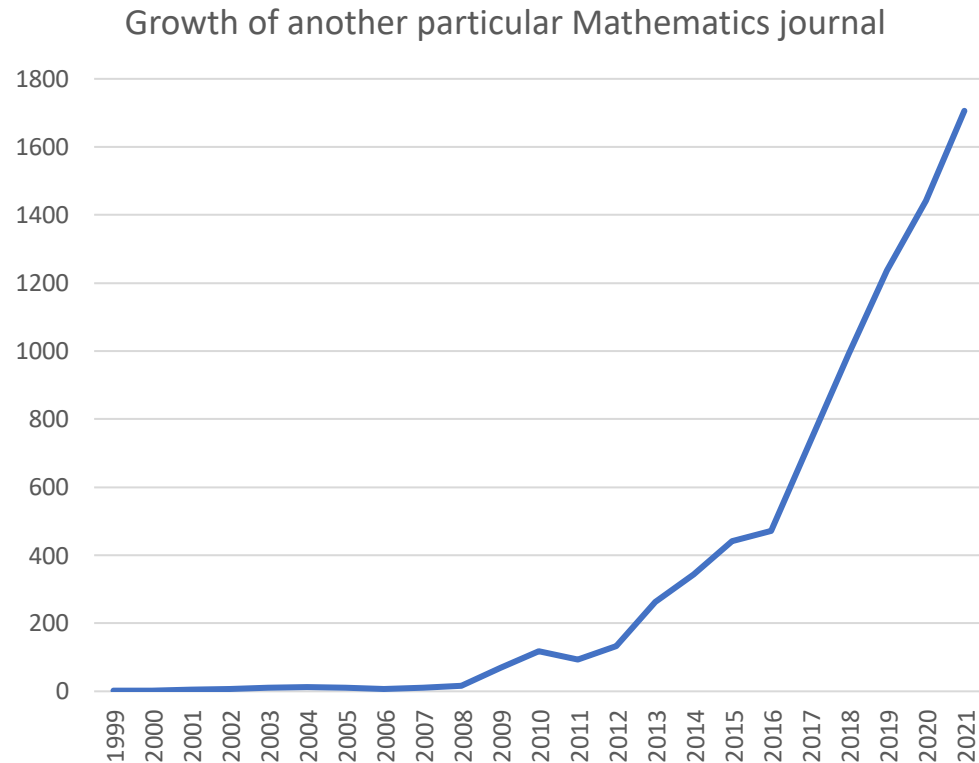
Growth of a particular Mathematics journal



- From 2019 to 2023, the journal grew almost linearly from 118 articles per year to 1602 articles per year.

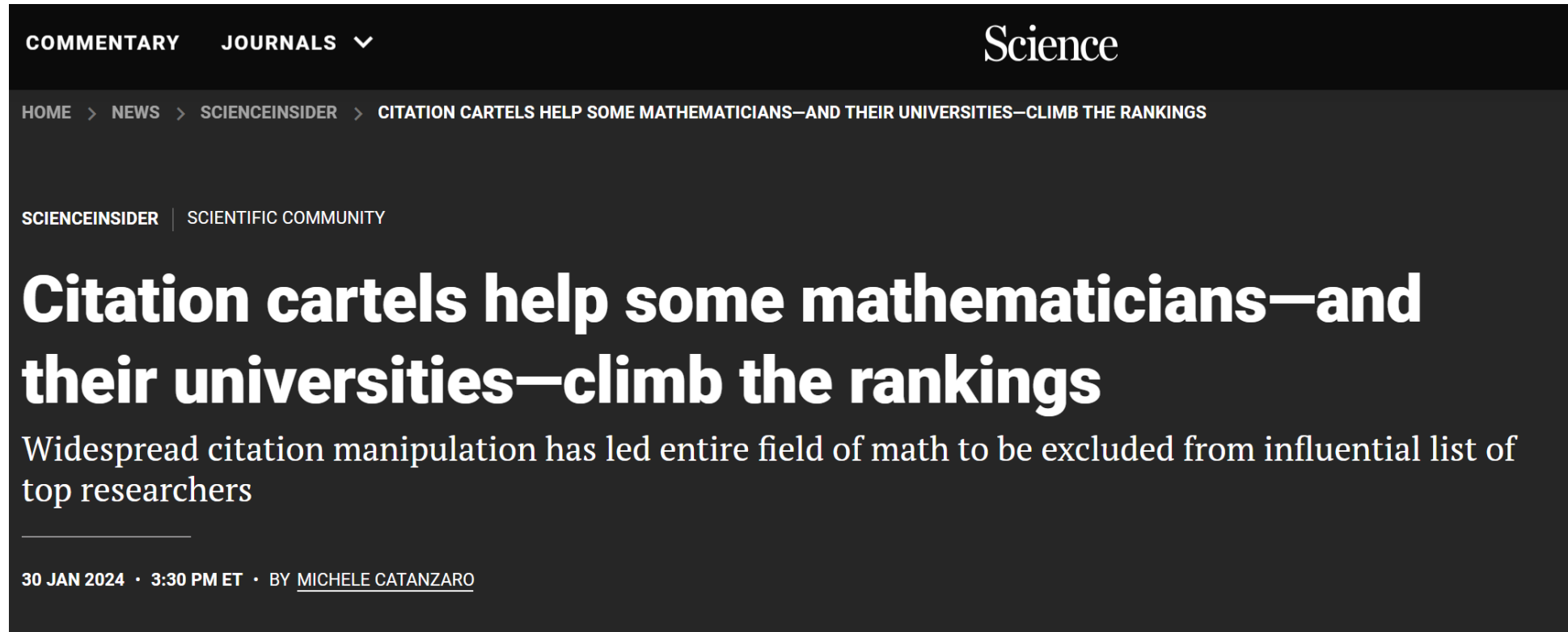
# Growth of the literature

## some special contributors



- From 2007 to 2021, the journal had annualized growth of 33.6% per year.
- In ten years, the publication rate of the journal jumped by an order of magnitude, from ~100 articles per year in 2010-2012 to 1706 articles in 2021.

# Citation cartels



The image is a screenshot of a Science magazine article page. At the top, there is a dark navigation bar with the word "Science" in white serif font on the right. On the left, there are menu items: "COMMENTARY" and "JOURNALS" with a downward arrow. Below this is a breadcrumb trail: "HOME > NEWS > SCIENCEINSIDER > CITATION CARTELS HELP SOME MATHEMATICIANS—AND THEIR UNIVERSITIES—CLIMB THE RANKINGS". Underneath the breadcrumb, it says "SCIENCEINSIDER | SCIENTIFIC COMMUNITY". The main headline is in large, bold, white sans-serif font: "Citation cartels help some mathematicians—and their universities—climb the rankings". Below the headline is a sub-headline in a smaller white font: "Widespread citation manipulation has led entire field of math to be excluded from influential list of top researchers". At the bottom left, there is a line of text: "30 JAN 2024 • 3:30 PM ET • BY MICHELE CATANZARO".

COMMENTARY JOURNALS ▼ Science

HOME > NEWS > SCIENCEINSIDER > CITATION CARTELS HELP SOME MATHEMATICIANS—AND THEIR UNIVERSITIES—CLIMB THE RANKINGS

SCIENCEINSIDER | SCIENTIFIC COMMUNITY

## Citation cartels help some mathematicians—and their universities—climb the rankings

Widespread citation manipulation has led entire field of math to be excluded from influential list of top researchers

30 JAN 2024 • 3:30 PM ET • BY MICHELE CATANZARO

<https://www.science.org/content/article/citation-cartels-help-some-mathematicians-and-their-universities-climb-rankings>

# Citation cartels

Mathematical Reviews was already aware of these types of problems

- Looking at the Mathematics Literature (*Notices of the AMS*, Feb 2019): <https://dx.doi.org/10.1090/noti1799>;
- Don't Count on It (*Notices of the AMS*, Jan 2021): <https://dx.doi.org/10.1090/noti2201>.



# Citation cartels

MathSciNet includes reference lists from selected journals, the “Reference List Journals”. These journals pass through a second round of editorial vetting, and represent about 1/3 of the journals indexed overall.

**Note:** most of the journals being used in cartels, as identified by Domingo Docampo are not reference list journals, hence not contributing to citation counts in MathSciNet.

# Conclusion

The digital age of science has allowed for the production and dissemination of much more research than ever before by removing much of the friction that occurred in the past.

A mixture of tech and human tools can be used to ensure the bibliographic and editorial integrity of the endeavor.

# Acknowledgments

Thank you to CHORUS for providing the opportunity to discuss ways the AMS is working to ensure the integrity of the research literature in mathematics.