

Publication Culture and Review Processes in the Data Management Community: An Open Discussion

Sihem Amer-Yahia
CNRS, Univ. Grenoble Alpes
sihem.amer-yahia@cnrs.fr

Stratos Idreos
Harvard University
stratos@seas.harvard.edu

Sourav S Bhowmick
Nanyang Technological University
assourav@ntu.edu.sg

Wolfgang Lehner
TU Dresden
wolfgang.lehner@tu-dresden.de

Xin Luna Dong
Meta
lunadong@fb.com

Divesh Srivastava
AT&T Chief Data Office
divesh@research.att.com

ABSTRACT

The Data Management community has explored many options in recent years to improve our publication culture and review processes, ranging from innovative journal-conference hybrids that decouple publication from presentation, incorporating journal-style reviewing for conference-style papers, requesting code reproducibility and code/data availability, multiple submission deadlines in a year, new categories of papers, informal shepherding processes, guidelines for diversity and inclusion, automated COI check, and so on. This panel seeks to examine our many experiments, comparing them with other CS disciplines, and help determine (i) have our experiments worked? (ii) what has their impact been? and (iii) can we do better?

CCS CONCEPTS

• **Information systems** → **Data management systems**; • **Social and professional topics** → **Professional topics**.

KEYWORDS

Publication Culture, Review Processes, Scientific Progress

ACM Reference Format:

Sihem Amer-Yahia, Sourav S Bhowmick, Xin Luna Dong, Stratos Idreos, Wolfgang Lehner, and Divesh Srivastava. 2022. Publication Culture and Review Processes in the Data Management Community: An Open Discussion. In *Proceedings of the 2022 International Conference on Management of Data (SIGMOD '22)*, June 12–17, 2022, Philadelphia, PA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3514221.3528413>

1 INTRODUCTION

Sir Issac Newton is known to have remarked “If I have seen further it is by standing on the shoulders of giants.” His remark is relevant to all of science, which has made enormous progress over the centuries because new scientific discoveries are built on and make use of previous discoveries, enabling each generation of scientists to see further by standing on the shoulders of those who came before them.

These discoveries are known to us even today because scientists have always published their findings for others to build on.

Different scientific disciplines have evolved their publication cultures over time in response to the needs of their communities, but peer-reviewed journal publications have remained the gold standard in almost all scientific disciplines. A notable exception is Computer Science with its emphasis on high quality peer-reviewed conference publications, in addition to traditional journal publications, to enable rapid dissemination of new ideas and results for others in the community to build on. The Computer Science community is quite proud of the rigorous review processes we have developed over the last few decades to select high quality papers, with some top conferences and journals having acceptance rates only around 15%.

In recent years, Computer Science has seen an explosive growth with many new emerging areas as well as significantly increased participation from people from different countries around the world. To provide a way to handle this explosive growth, several communities within Computer Science have been experimenting with many different ways to improve our publication culture and review processes. The Data Management community, in particular, has explored many options over the years, ranging from innovative journal-conference hybrids that decouple publication from presentation, incorporating journal-style reviewing for conference-style papers, requesting code reproducibility and code/data availability, multiple submission deadlines in a year, new categories of papers, informal shepherding processes, guidelines for diversity and inclusion, automated Conflicts of Interest (COI) detection and management, and so on. With an increased number of options, our processes have also tended to become more complex in recent years and can vary depending on who is in charge of the process.

This panel seeks to examine our many experiments to improve the publication culture and review processes in the data management community, comparing them with other CS disciplines, and help determine (i) have our experiments worked? (ii) what has their impact been? and (iii) can we do better?

2 DISCUSSION TOPICS

The panelists will present their perspectives on a variety of topics including, but not limited to, the following.

Quality of accepted papers : The numbers show that the journal-style review processes have led to an increase in the percentage of papers that are accepted, often from 15% using

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGMOD '22, June 12–17, 2022, Philadelphia, PA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9249-5/22/06.
<https://doi.org/10.1145/3514221.3528413>

traditional review processes to 25%. Has it also led to an improvement in the quality of the accepted papers? What role have recent requests for code reproducibility and code/data availability played?

Quality of reviews : High-quality constructive reviews are the backbone for selecting good papers to publish in our venues, yet authors often complain about the quality of reviews their papers receive. What steps have we taken to ensure that our reviewers are providing such high quality reviews? What role does the choice of program committee have on review quality? What can we learn from other communities who have experimented with ideas such as open reviewing?

Community fatigue : Our experiments to improve the publication culture has demanded more time commitments from authors (to revise papers to address detailed requests from reviewers), reviewers (reviewing papers year around, often for multiple publication venues simultaneously), associate editors and program chairs. Has this led to community fatigue? What can we do to reduce the cost significantly while keeping many of the benefits that have emerged from these experiments?

Our goal is to have an open discussion with the data management community on these topics, so we will solicit questions from the community ahead of time as well as during the panel for our panelists to discuss.

3 PANELISTS

We have a diverse set of panelists who have played leading roles in the major data management publication venues in recent years, and who have thought deeply about our publication culture and review processes. They are listed below in lexicographical order of their last names, along with a summary of their contributions to the community.

Sihem Amer-Yahia is a CNRS Research Director and Deputy Director of the Laboratoire d'Informatique de Grenoble in France. Sihem is Vice-President of the VLDB Endowment, has chaired prestigious database conferences including ICDE 2021 and VLDB 2018, and is chairing SIGMOD 2023. She leads the Diversity & Inclusion initiative of the database community whose charter is to guide our community to adopt a more inclusive mindset in general toward different individuals regardless of their differences.

Sourav S Bhowmick is an Associate Professor at the Nanyang Technological University, Singapore. He is interested in data-driven solutions to detect and manage conflicts-of-interest and bias in peer-review systems. He is the inventor of CLOSET, a comprehensive software for detection and management of COI that is currently being used in multiple venues. He is serving or has served as PC co-chair for several venues such as EDBT 2023, CODS-COMAD 2022, CIKM 2020, and DASFAA 2014.

Xin Luna Dong is the Head Scientist at Facebook AR/VR Assistant. Prior to joining Facebook, she was a Senior Principal Scientist at Amazon, leading the efforts of constructing Amazon Product Knowledge Graph, and before that one of the

major contributors to the Google Knowledge Vault project, and has led the Knowledge-based Trust project, which is called the “Google Truth Machine” by Washington’s Post. She has co-authored books “Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases” and “Big Data Integration”, was awarded ACM Distinguished Member, and VLDB Early Career Research Contribution Award for “Advancing the state of the art of knowledge fusion”. She serves in the VLDB endowment and PVLDB advisory committee, and is a PC co-chair for WSDM 2022, VLDB 2021, KDD 2020 ADS Invited Talk Series, and SIGMOD 2018. She introduced multiple changes to VLDB 2021 when she was a PC co-chair, such as introducing the Scalable Data Science Category and reducing the page limit to 8 pages for this track.

Stratos Idreos is an associate professor of Computer Science at Harvard University where he leads the Data Systems Laboratory. Stratos is chair of the ACM SoCC Steering Committee and has served as PC co-chair of ACM SIGMOD 2021 and IEEE ICDE 2022. Stratos was also chair of ACM SIGMOD Reproducibility from 2015 to 2020. Stratos believes that the long term health of the community depends on a publication culture and review processes that prioritize quality, transparency, reproducibility, integrity and fairness as the means to curate the best papers and also crucially as a feedback mechanism for the growth of the entire community.

Wolfgang Lehner is full professor for Database Systems at Technische Universität Dresden, Germany. He is the Managing Editor of the Proceedings of the VLDB (PVLDB) and has been PC-Co-Chair of many research and industrial tracks. Regarding publication culture, Wolfgang Lehner pursues a comprehensive approach that can be characterized by the governing slogan “PDF and more”, i.e. including additional material like code, data, but also additional videos, slides, demos, discussions that are vital elements of the development process/history of a publication.

The panel will be moderated by Divesh Srivastava, who is the head of Database Research at AT&T and the President of the VLDB Endowment. He has served as the program committee co-chair of many data management conferences including, most recently, SIGMOD 2021, VLDB 2020 (Industrial), SIGMOD 2020 (Industrial) and ICDE 2019. As a former Managing Editor of the Proceedings of the VLDB (2013–2018), he oversaw many of the PVLDB experiments to improve our publication culture and review processes.

ACKNOWLEDGMENTS

We would like to thank the SIGMOD 2022 panel chairs, Mirek Riedewald and Xiaofang Zhou, for inviting us to organize a panel on this important topic at the SIGMOD 2022 conference.