Trusting in Science: Automating Quality Checks in Publishing and Beyond

MODERATOR:
Heather Staines
Delta Think
Trumbull, Connecticut

Leslie McIntosh
Ripeta
St. Louis, Missouri

SPEAKERS:
Cynthia Hudson Vitale
Association of Research Libraries
Washington, DC
and
Ripeta
St. Louis, Missouri

Chris Graf
SpringerNature
London, England

Gerardo Machnicki
Independent Professional
Buenos Aires, Argentina

Leslie McIntosh

REPORTER:
Ruth Isaacson
Genetics Society of America
Rockville, Maryland

Discussions around open science and reproducibility have taken center stage these past years with more focus on the importance of trust in science. This session offered an opportunity to look at initiatives involved in upholding and supporting our trust in science.

Leslie McIntosh, CEO of Ripeta, moderated and spoke at the CSE session on Trusting in Science and shared how Ripeta is using a combination of automated and manual quality checks of submitted manuscripts.

Ripeta focuses on indicators of "Trust" to evaluate research articles along axes of professionalism, integrity, and reproducibility.

Professionalism and integrity checks examine the structure of the research article itself. Is the research article in a standard format (clear hypothesis, sections, etc.)? Does it contain the content and declarations you would expect to see in a valid research article (data availability statement, citations, data availability, declarations of conflicts, etc.)?

Automated checks also focus on authors of scholarly work, with an eye toward verifying identity and qualifications. The natural assumption about a peer-reviewed publication is that the author is a scientist who possesses knowledge beyond that of a lay person and is qualified via their education, training, or experience.[1] Among many items, the system looks to see if the author has an ORCID ID, legitimate affiliations, or has used an institutional email address. If Ripeta flags the submission, then manual investigation takes place to confirm if the flags are warranted. McIntosh emphasized the need for manual verification as incorrect flags and subsequent rejections or institutional reporting could lead to serious consequences for the author.

Checks also focus on the reproducibility of the research paper—a topic that has been in the forefront for the past decade. Reproducibility allows other researchers to replicate the original study and achieve the same results. Automated checks look for indicators of trust, such as:

- Analysis software
- Software citations
- Statistical analysis methods
- Availability of biological materials, code, and data
- Code and data availability statements with data location clearly identified

Why should these checks occur prior to publication? McIntosh mined the RetractionWatch database on April 11, 2021, and looked at data for a 10-year span (2010–2019). She found there were a total of 2,772 retractions, with an average of 277 and a median of 266 retracted manuscripts per year. These retractions were for various reasons, including authorship concerns, ethical violations, fake peer review, paper mills, and rogue editors. Preventing the publication of papers that go on to be retracted is just one step in preventing further erosion of the public's trust in science.

Following McIntosh's presentation about automating checks, Gerardo Machnicki spoke about reproducibility and trust in science as it relates to researchers in Latin America.

For Machnicki, the topic of trust in science falls under the umbrella of open science. Trust is important for scientists to have confidence in their findings, which in turn are grounded on reliability and trust in the scientific process. However, in many areas of the world, democratic and ethical principles need to be considered first, with trust being the end goal.

Machnicki explained that among researchers in the south, questions are often raised about who benefits from the open sharing of data. Reservations about data misuse, patient privacy, exploitation concerns, or fear of undermining research careers are just some of the challenges to overcome before FAIR (Findable, Accessible, Interoperable, Reusable) principles are met.

CONTINUED

An overall reticence to share data has led to research being siloed and data/outcomes provided in nonuniversal formats and models. Machnicki advocates for tools to facilitate science at scale, including repositories that allow researchers to access data and information in a standardized format. Collaborations and knowledge bases result in shared analysis, reporting, etc., which generate reproducible and transparent content.

Machnicki shared examples of initiatives that have been successful, including the WorldWide Antimalarial Resistance Network (WWARN).[2] This research alliance of more than 250 researchers is linked to many recognized institutions and is working to promote data sharing and data reuse around clinical trials for malaria. Machnicki is engaged in a community seeking to expand the growth and use of the Observational Health Data Sciences and Informatics (OHDSI)[3] program within Latin America. Since its founding in 2014, OHDSI has provided large-scale analytics using health data. Worldwide, the program has over 2,000 collaborations from 74 countries and contains health records for around 810 million global patients.

Next, Chris Graf, SpringerNature's Research Integrity Director, shared 5 bite-sized insights about trust.

Insight 1: Net trust by the public in scientists and professors is higher than that of business leaders[4]; it's a good job that we put scientists and scholars in charge of what research is published.

Insight 2: Peer review is not perfect, and researchers have subtle and sophisticated ways in which they work out trust. Their weighing of indicators and decision making is analog and personal; this works for individual researchers, but may not scale-up well in our digital world.

Insight 3: There remains a collective skepticism in the research communities. Decades of concerns have resulted in questions of research reliability and reproducibility.[5]

Insight 4: There are reasons to be optimistic. In a 2019 article,[6] Dorothy Bishop, an experimental psychologist at the University of Oxford, shared her view that threats to trust in science may be brought under control through innovations such as meta-science, social media, registered report formats, and funder requirements for open science.

Insight 5: Collective action and collaboration across the publishing sector is underway! Launched in 2022, STM Integrity Hub[7] is a cloud-based environment for publishers to check submissions for research integrity issues. This safe and confidential hub respects privacy and competition/

antitrust laws and allows publishers to collaborate with each other while identifying, for example, simultaneous submissions and paper mill publications.

Finally, Cynthia Hudson Vitale, Director of Scholars and Scholarship with the Association of Research Libraries and cofounder of Ripeta, spoke about trust in science from an institutional/library perspective.

The framework for research integrity within institutions and higher education associations requires not just trust in science but also community engagement (listening/learning) and open science.

Institutions and libraries have 3 primary challenges to consider when it comes to research integrity:

1. Instead of looking at peer-reviewed, published articles as the primary source of trust in science, institutions and libraries should start to focus upstream. A culture shift is needed and could occur via innovations including workshops or materials about data literacy, research conduct, ethics, and compliance.
2. The burden on researchers and institutional units needs to be balanced against what is necessary for quality research and publishing.
3. Scalability and equitable access to solutions and services is a must. Institutions have many competing demands; thus, solutions ideally should meet multiple needs.

Institutions should ask how they can best support research integrity to increase trust in science, and how, with their decentralized setup, they can best collaborate. Organizations and campus units need to use policies, infrastructure, and services to lower the burden on researchers while continuing to advance quality and trustworthy research.

References and Links

1. Levine, AH, Yates PJ. Taking a science expert deposition to set up a Daubert motion. [accessed May 14, 2022]. https://www.arnoldporter.com/en/perspectives/publications/2016/01/20160120_taking_a_science_expert_deposit_12607.
2. https://www.wwarn.org/
3. https://www.ohdsi.org/
4. Ipsos. Ipsos MORI veracity index 2021: trust in professions survey. https://www.ipsos.com/sites/default/files/ct/news/documents/2021-12/trust-in-professions-veracity-index-2021-ipsos-mori_0.pdf.
5. Bastian, H. Reproducibility crisis timeline: milestones in tackling research reliability [accessed June 8, 2022]. https://absolutelymaybe.plos.org/2016/12/05/reproducibility-crisis-timeline-milestones-in-tackling-research-reliability/.
6. Bishop, D. Rein in the four horsemen of irreproducibility. Nature. 2019;568:435. https://doi.org/10.1038/d41586-019-01307-2.
7. https://www.stm-assoc.org/stm-integrity-hub/