An Industry-Wide Effort to Ensure Research Integrity

MODERATOR:
Tony Alves
Senior Vice President
HighWire Press
@highwirepress

SPEAKERS:
Tony Alves

Renee Hoch
Managing Editor leading the
  Publication Ethics team
PLOS
@PLOS

Heather Staines
Director of Community
  Engagement and Senior
  Consultant
DeltaThink
@heatherstaines; @DeltaThink

REPORTER:
Emilia (Emily) Arturo
Managing Editor, *Blood*
American Society of Hematology
@moonlighterturo

Tony Alves, Senior Vice President at HighWire Press, and Treasurer of CSE, moderated a session at the 2024 CSE Annual Meeting focused on the STM Integrity Hub.[1] The STM Integrity Hub is a relatively new initiative developed by STM, the International Association of Scientific, Technical, and Medical Publishers. The invited speakers included Renee Hoch, currently the Managing Editor leading the PLOS Publication Ethics team at PLOS,[2] and Heather Staines, Director of Community Engagement and a Senior Consultant at DeltaThink.[3]

The session began with Staines's account of the earliest research, conversations, and plans that took place, mainly in member discussions called "Futurelab," that would set the stage for the eventual development of the STM Integrity Hub. This collaborative effort was to be an industry-wide effort to tackle the burgeoning stressors faced by all stakeholders in the worldwide scientific publishing endeavor.

STM Trends

Each year, the founding members of 2 of STM's earlier initiatives, SeamlessAccess[4] and GetFTR,[5] which stemmed from an initiative called the Resource Access in the 21st Century (RA21) Project[6] that concluded with a NISO recommendation in 2019, meet to collate and articulate the factors that most stress the scientific publishing industry. Each spring, the group releases their findings as infographics called "STM Trends"[7] that highlight predictions for the industry over the upcoming 3–5 years. Staines discussed the STM Trends for 2024 through 2027, and she highlighted common themes among these: paper mills and general trust in published research.

The STM Trends for 2028 were recently released (see Figure). Staines explained that this newest prediction envisions a workforce where researchers, authors, publishers, and artificial intelligence (AI) tools can coexist and work in harmony. The STM Trends for 2028 also highlights the need to maintain appropriate human involvement to assure integrity and trustworthiness of AI machine-derived content.

STM Integrity Hub

Close examination of the common themes among the Trends led to the resolve to establish an industry-wide effort to protect the scientific record, current and future, from "bad actors" who might infest the literature with fraudulent and unreliable data. Thus, the STM Integrity Hub was born with the mission to "equip the scholarly communication community with the data, intelligence, and technology to protect research integrity."[8] Alves reviewed how the STM Integrity Hub is organized as a team under a Governance Board, chaired by Chris Graf, 4 working groups (Expert Group, User Group, Image Alteration & Duplication, and a Watch List), and 3 task forces that respectively focus on legal aspects, communication, and cybersecurity.

As a member of 1 of the original task forces, Alves led the effort to evaluate the occurrence of simultaneous submission of manuscripts to 2 or more journals, which can be a signal of paper mill activity. The initial estimate was that as much as 4% of all submissions are duplicate submissions, and, as submission to multiple journals is not the norm among scientific researchers and peer reviewed science journals, this suggested the need to inspect this signal more closely.

To help the STM Integrity Hub continue developing manuscript evaluation tools, the initial participating publishers gave the team access to proprietary publisher content. These data were used to test different article and data evaluation and reporting systems, and they allowed for the development of policies on the ethical, legal, and efficient use of pooled data to notify publishers and researchers, when needed, about investigations of the Hub.

At least 15 volunteer publishers have since piloted the Integrity Hub's Paper Mill Checker Tool and Duplicate Submission Checker Tool. The Paper Mill Checker Tool packages multiple tools aimed at detecting signals of fraudulent submissions, and takes input from sources such as PubPeer,[9] Clear Skies,[10] and Retraction Watch.[11] This initial pilot screened approximately 20,000 manuscripts per month. As these tools are honed, more signals of potential

CONTINUED
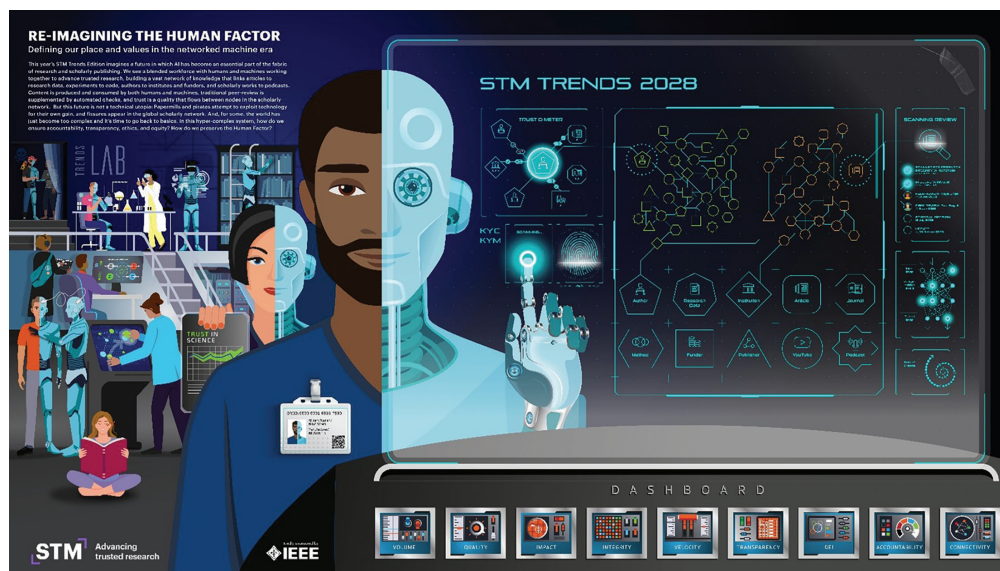
Figure. Re-imagining the human factor. The STM Trends for 2028 were published in the spring of 2024, predicting that AI tools would become an integral part of the scientific publishing industry and will likely lead to exploitation by paper mills and other fraudsters.

paper mill activity will be added, and all tools will be incorporated into an Ambient Screening tool that will allow publishers to select which tools they would wish to deploy. Additional future objectives include bringing on additional volunteer publishing partners, integrating these tools with commonly used submission platforms, and addressing the occurrence of false positives.

User Perspective

Renee Hoch spoke on behalf of PLOS to describe the experience of a volunteer publisher that has been involved in testing the tools of the STM Integrity Hub. Hoch explained that PLOS decided to become involved in the industry-wide movement to battle issues related to paper mills and fraudulent submissions. "If we continue to take siloed approaches, we are really limiting ourselves." In isolation, an article generated by a paper mill is too difficult to detect. "[All publishers] are battling some of the same problems; let's do this together."

One journal of the PLOS portfolio, *PLOS One*, began piloting the STM Integrity Hub Duplicate Submission Checker tool in December 2023, and all other PLOS journals were added in March 2024. The tool notified PLOS of 209 duplicate submissions involving 22 journals over a 4.5-month period. These flagged submissions were investigated by PLOS staff, false positives were removed from consideration, and the result was 150 desk rejections, none of which were successfully appealed. PLOS also piloted the Paper Mill Checker Tool: of the PLOS submissions for which this tool flagged issues, 75% were rejected prepublication, 4% were ultimately published, and 20% are still being monitored.

Takeaways

Several attendees asked about the signals identified by the STM Integrity Hub as indicators of potential paper mill activity. In response, the panelists emphasized that these details will be kept confidential, from both users of the tools and from authors, to help protect the tools from actors that would aim to evade them.

In response to questions from the audience about the economic sustainability of a multi-publisher collaboration, Alves mentioned that an economic model is being developed. Alves emphasized that the prediction is that using the packaged tools of the STM Integrity Hub (i.e., bringing multiple vendors under one roof) would make the program more economically feasible and equitable.

Finally, several members of the audience wondered what would or could be done with the potentially large volume of information that could be generated by the STM Integrity Hub tools. They expressed that costs may be prohibitive for smaller publishers and journals, and that some may find it difficult to staff papermill checks and follow-up investigations. Hoch emphasized the value of the investment in the STM Integrity Hub, asking "If we're not dealing with these issues prepublication, then what is the cost later on?"

References and Links

1. https://www.stm-assoc.org/stm-integrity-hub/
2. https://plos.org/
3. https://deltathink.com/
4. https://seamlessaccess.org/
5. https://www.getfulltextresearch.com/