Workshop Report

Opportunities in Open Science, Metascience, and Artificial Intelligence

Authors
Catherine Aiken
Greg Tananbaum
James Dunham
Ronnie Kinoshita
Erin McKiernan

CSET
CENTER *for* SECURITY *and* EMERGING TECHNOLOGY

OPEN RESEARCH COMMUNITY ACCELERATOR

June 2025

Executive Summary

On March 26, 2025, the Center for Security and Emerging Technology (CSET) and the Open Research Community Accelerator (ORCA), with support from the National Science Foundation, hosted a workshop in Washington, DC. We brought together 35 experts, researchers, and funders to explore opportunities and challenges in open science, metascience, and AI. This report summarizes the conversation that took place over two panel sessions, two guided group discussions, and a concluding synthesis session.

This report also aims to fill a gap identified by participants: the need for a clear agenda to inform and guide efforts at the intersection of metascience research, open science monitoring and impact assessment, and AI for research and science. To that end, we highlight priority research questions and feasible opportunities for collaborative work.

We thank the National Science Foundation—and specifically, the Cyberinfrastructure for Public Access and Open Science program (CI PAOS) within the Office of Advanced Cyberinfrastructure—for the support in planning and hosting this productive workshop.

Table of Contents

Key Takeaways

During the workshop, participants engaged with practical and conceptual questions at the intersection of open science, metascience, and AI.

Some practical questions included:

- What are existing initiatives for monitoring and assessing open science?
- How can we make reliable data and methods more accessible?
- How can we avoid duplicating efforts?
- How can we facilitate collaboration and sustained engagement between open science, metascience, and AI?

Some conceptual questions the group discussed included:

- How should we evaluate the appropriateness of AI for a given research task?
- How can we adapt metascience insights to specific disciplines and contexts?
- Should we be thinking about data and scientific outputs as public goods?
- How do we do research in today's rapidly evolving scientific and technological landscape and ensure responsible deployment of AI in that research?

The workshop surfaced opportunities for leveraging existing data, tools, and knowledge to advance open science and metascience. AI presents numerous opportunities to improve open research outputs (e.g., papers, software, data, protocols, patents), including:

- Transforming open research outputs (e.g., into machine-readable formats)
- Extracting information from open research outputs (e.g., scientific claims)
- Surfacing hidden connections across open research outputs
- Improving information retrieval and search over open outputs
- Enhancing metadata curation (e.g., via persistent identifiers)
- Confirming and enabling the reusability of open outputs (e.g., code review)
- Expanding science communication by making it more accessible (e.g., translation) and diversifying output formats (e.g., blogs, repositories)

However, participants emphasized that incentives, and the ecosystems that respond to them, shape how open science and metascience are done. Participants also emphasized that human review and traditional research methods remain critical to

sound science. In the near term, some high-level priorities surfaced by participants were:

- Evaluating AI models within scientific research contexts
- Fostering greater information-sharing
- Developing shared toolkits
- Collaborating on a feasible path forward

Toward that end, the report summarizes the two panels and group discussions from the workshop and concludes with a list of potential projects and priorities areas the open science, metascience, and AI communities can act on.

Panel: Measuring Open Science Outcomes and Assessing Impact

Our first panel included Dr. Tim Errington (Center for Open Science), Dr. Cassidy Sugimoto (School of Public Policy, Georgia Tech), and Dr. Daniel S. Katz (University of Illinois Urbana-Champaign) and was moderated by Dr. Catherine Aiken (CSET).

The panelists were asked to share current work related to measuring open science and assessing its impact, as well as their ideas for where researchers and the open science community could focus efforts to make progress on open questions and known needs.

Panelist Questions: Measuring Open Science Outcomes and Assessing Impact

1. Please share a current effort you are involved in that aims to measure the outcomes and/or assess the impact of open science policies. What are some successes? What challenges does the effort face?
2. Where can we make progress? What are some measures or methods we could build on or try? What data, tools, or resources do we need? How might AI help?
3. What actions can the community take? What should different actors focus on and prioritize? How might we parcel out next steps to focus on specific outcomes or known gaps?

Existing measurement and impact assessment projects mentioned included incorporation of pre-registration and peer review into more research stages (e.g., Registered Reports), development of open science indicators (e.g., PLOS open science indicators), implementation of FAIR principles (Findable, Accessible, Interoperable, Reproducible) for machine learning (e.g., FAIR4 ML), creation of domain-specific AI readiness checklists (e.g., FARR), researching the impact of requiring open access for some U.S. federally funded research, and tracking the impact and adoption of open source software (e.g., Open Source Software for Science).

Many challenges to such efforts were discussed, though they tended to fall into two bins: shortcomings of current methods and data (e.g., data usability and interoperability, extracting meaningful information from publications, imperfect

measures) or barriers due to existing systems and incentives (e.g., open access to papers, publishing processes, reluctance to view outputs as public goods).

Opportunities for next steps surfaced by panelists included designing and fielding experiments and randomized control trials (RCTs) to assess the impact of open science "interventions" like pre-registration or new funder policies. Other opportunities discussed in the panel included experimenting with AI to refine the review process, extract information from research outputs, identify open science "objects" and monitor compliance, and improve metadata curation and output reusability.

In sum, the panel highlighted the current state of efforts as pertains to assessing the occurrence and impact of open science while also surfacing areas where new research and AI tools could help. Panelists emphasized the need for increasing coordination, reducing duplicated efforts, promoting shared infrastructure, and fostering greater transparency in methodology.

Discussion: Measuring Open Science Outcomes and Assessing Impact

Following the first panel, all participants engaged in a guided group discussion. The following questions were posed to prompt participants:

1. Building on the projects and points raised by the panelists, what data and methods can we leverage to measure outcomes and assess the impact of open science in practical, scalable ways?

2. How can AI and LLMs help? Given known challenges and gaps, what is AI particularly well-suited to address?

3. What immediate actions or collaborations could strengthen existing efforts or initiate new ones?

Over the course of a lively discussion, participants shared additional projects measuring open science, with a focus on where AI can help scale such efforts. One recurring theme was data availability. Open, high-quality, comprehensive data (including bibliometric metadata and full text) was cited as a necessary component for AI and LLMs to execute meaningful information extraction, citation analysis, and claim verification. Even with appropriate data, two concerns emerged from the discussion.

First, questions arose about what we want to measure, what questions we want to ask, and why we want to ask them. Stakeholders have different questions and interests when it comes to open science measurement and evaluation. We should take a targeted approach, with specific projects being informed by these various needs, while building on lessons learned with a shared infrastructure and toolkit.

Second, concerns were voiced about the generalizability, affordability, bias, and ethical use of AI solutions. Participants highlighted AI's potential to improve research accessibility, streamline metadata curation, and identify gaps. However, they cautioned that AI should complement—not replace—existing methods for experimental design and traditional evaluation such as peer review. Discussions also touched on the importance of incentivizing open science when proprietary models dominate; the need for proactive rather than reactive policy responses; and the role of collaborations between and among funders, researchers, industry, and technologists.

Building on the opportunities surfaced in the panel, there was also strong interest in leveraging RCTs to test impact hypotheses, enabling direct and actionable conclusions. This led to discussions on what a well-designed experiment would entail and which actors need to be involved to ensure rigor and relevance.

Panel: AI for Metascience

Our second panel included Prof. Maria Liakata (Queen Mary University of London, Alan Turing Institute), Dr. Tom Pollard (Massachusetts Institute of Technology), Caleb Smith (University of Michigan), and Dr. Lucy Lu Wang (University of Washington), and was moderated by Dr. James Dunham (CSET).

The panelists were asked to share their current work incorporating AI in their metascience research and their ideas for where researchers and the open science community should go next.

Panelist Questions: AI for Metascience

1. Please share an example where you've used (or are using) AI in your metascience research or tool development. How did you approach incorporating AI into your work? How did you evaluate its performance and utility? What challenges or limitations did you encounter?
2. What next? What is possible now and in the near future? What don't we know yet and what ethical and/or technical challenges arise when considering AI-enabled methods?
3. How can this group take action? What research, methods, data, and/or tools can we develop and contribute?

The discussion explored the use of AI—particularly large language models (LLMs)—in research and scientific workflows, addressing their benefits, limitations, and future directions. Key themes included AI's ability to enhance knowledge synthesis, automate content creation, and support interdisciplinary collaboration. Recent metascience research leveraging AI that panelists discussed included research on LLM evaluation, AI applications to remove personally identifiable information from clinical data, aggregating and describing bibliometric data, making science more accessible for broader audiences, and applications for reviewing and summarizing informed consent documents.

In terms of challenges and concerns, two themes emerged. The first involved gaps in evaluation and in assessing AI-generated content. Evaluation challenges ranged from

benchmark saturation and (in)applicability, to confidence (mis)calibration to lacking incentives for context-specific model and output evaluation. While LLMs can accelerate research by summarizing vast amounts of information and generating hypotheses, their outputs require rigorous validation. Hallucinations, overconfidence, and bias amplification limit appropriate use. Evaluation frameworks must account for domain-specific knowledge and needs, balancing automation with human oversight. The research process could better account for—and the research community could better prioritize—the development of benchmarks, evaluation frameworks, and output validation. Concerns around evaluation are particularly relevant in high-stakes applications like clinical research, public services, and scientific communication.

Second, panelists raised concerns about the yet-unknown impact of AI use in the scientific process, including how it reshapes research workflows and scholarly communication. While AI can speed up tasks such as reviewing and drafting, it also may dilute scientific rigor. Ensuring responsible AI integration will require new incentives beyond traditional publication impact metrics and the development of norms for AI use and disclosure in research.

Looking forward, the discussion emphasized the need for dynamic evaluation benchmarks and interdisciplinary collaboration. Opportunities surfaced by panelists include leveraging AI for extracting more information from scientific outputs (e.g., claims, images and tables, hypotheses) and other transformations of existing scientific outputs for broader accessibility and reuse (e.g., translation, reproduction). Panelists also discussed the difference between incorporating AI-assisted tooling for discovery and retrieval tasks (e.g., checking citations, navigating literature, reviewing code, comparing pre-registration with submissions) and generation tasks (e.g., communicating scientific findings, multi-document summarization, drafting blogs).

In sum, the panel showcased how AI presents opportunities to revisit underexplored research areas and enhance research accessibility, but scientific rigor should be paramount, and practical and ethical considerations must guide its adoption. The panel called for adaptive frameworks that assess when and how AI should be used, ensuring it complements rather than replaces human expertise. Ultimately, the responsible deployment of AI in science will depend on transparency, accountability, and ongoing refinement of evaluation methodologies.

Discussion: AI for Metascience

Following the second panel, all participants engaged in a guided group discussion. The following questions were posed to prompt participants:

1. Building on ideas shared by panelists, how can metascience and AI help us advance open science processes (e.g., preregistration), outputs (e.g., data, code), and outcomes (e.g., accelerate pace, reproducibility)?

2. Where do you think this group can make a meaningful contribution to advancing metascience, especially for open science measurement and evaluation, in the next 1 to 2 years?

The group discussion explored the role of AI in advancing open science processes and outputs, emphasizing its potential to enhance reproducibility, streamline workflows, and enable faster experimentation and broader dissemination. The discussion also highlighted several limitations, given the current state of AI. For example, participants noted trade-offs between using commercial AI models and open-source alternatives. Several participants emphasized the need for community-driven, open-source AI infrastructure that allows for greater transparency and access. Institutional or consortium-based approaches to hosting such models were proposed as potential solutions. Participants also noted an agentic trend in AI for science and discussed limitations with current offerings, including a lack of transparency in reasoning, high computational cost, and limited points of intervention for human researchers.

The conversation also explored AI's broader impact on the scientific process and research incentive structures. AI's ability to generate narratives and synthesize information could, among other things, speed up discovery and dissemination, and expand research impact metrics beyond traditional citation-based approaches to include more use cases and impact stories. However, there were concerns that overreliance on AI tools for research and scientific communication might distort research priorities, encouraging researchers to optimize for AI-friendly outputs and focus on repurposing existing analyses at the cost of genuine scientific advancement.

Ultimately, the discussion underscored the importance of human judgment in setting research priorities and defining evaluation criteria, with AI serving as a tool to implement these priorities rather than dictate them. The need for clear research agendas, articulation of shared goals and priorities, and more interdisciplinary collaboration between metascientists and AI experts were identified as critical gaps.

A Path Forward

The workshop brought together key actors in the open science, metascience, and AI communities to establish a shared understanding of current efforts and gaps. Another success was identifying viable, near-term paths to advance our understanding of open science processes, outcomes, and impact given the current state of metascience and AI.

Moving forward, participants would like to establish and maintain networks and shared infrastructure to enable coordination and collaboration and avoid duplication of efforts. The group agreed that a common goal is evidencing and communicating the value of open science through a targeted distribution of the research agenda, accumulation of the evidence, and communication of what is learned.

Informed by community input, we propose the following research priorities that can be tackled in a coordinated manner based on the interests of different parts of these communities.

Metascience Priorities

Project	Description
LLM Use in Science	Survey how researchers and scientists are using LLMs now.
Research Output Metadata Curation	Improve, standardize, and share metadata for scientific outputs, including publications, software, data, and models, etc.
Making Sense of Research Outputs	Synthesize open outputs to build new knowledge and better understand the scientific process.

Open Science Priorities

Project	Description
Open Science Indicators	Build on existing efforts to monitor occurrence, compliance, and impact of open science practices.
Expand Access	Modify and update systems to enable greater access to data, software, outputs, and models.
Case Studies and Impact Stories	Develop narratives and communicate successful cases of open science.
Experimental Impact Assessment	Design and conduct RCTs and responsive research to assess impact of open science interventions.

Applied AI Priorities

Project	Description
Improving AI for Research Evaluation	Develop benchmarks and evaluation metrics specific to AI use in research tasks.
Disseminating and Communicating Science	Develop AI tooling to communicate science outputs and their novelty, use, and impact.
Enabling Open Science With AI	Develop AI tooling to speed up, improve, and scale the compilation and review of open science outputs.
Research Output Information Extraction	Improve extraction of information from research outputs.

Appendix A: List of Attendees

Name	Affiliation
Adam Russell	University of Southern California
Alan Tomkins	University of Nebraska-Lincoln
Bobbie-Jo Webb-Robertson	Pacific Northwest National Laboratory
Caleb Smith	University of Michigan
Cassidy R. Sugimoto	School of Public Policy, Georgia Tech
Catherine Aiken	Center for Security and Emerging Technology (CSET)
Christine Custis	Institute for Advanced Study
Daniel S. Katz	University of Illinois Urbana-Champaign
David Barker	National Science Foundation
Dr. Jwan Khisro	University of Maryland
Dwight Kravitz	National Science Foundation
Erin McKiernan	Open Research Community Accelerator (ORCA)
Eamon Duede	Purdue University & Argonne National Lab
Greg Tananbaum	Open Research Community Accelerator (ORCA)
Helen Toner	Center for Security and Emerging Technology (CSET)
Irene Pasquetto	University of Maryland
James Dunham	Center for Security and Emerging Technology (CSET)
Jean-Paul Chretien	Renaissance Philanthropy
Jennifer Hansen	Open Data Policy & Strategy, Microsoft
Jessica Polka	Astera Institute
Josh Greenberg	The Alfred P. Sloan Foundation
Juan Pablo Flores Cortes	GitHub
Karthik Ram	The Navigation Fund

Name	Affiliation
Kate Hertweck	Chan Zuckerberg Initiative
Katija Ra'oof	Department of Defense, Defense Technical Information Center
Kyle Demes	OpenAlex
Lucy Lu Wang	University of Washington
Maria Liakata	Queen Mary University of London, Alan Turing Institute
Mark Greaves	Schmidt Sciences
Matt Marx	Cornell University, NBER, and the Innovation Information Initiative
Plato Smith	National Science Foundation
Ronnie Kinoshita	Center for Security and Emerging Technology (CSET)
Steve Crawford	National Aeronautics and Space Administration
Stuart Buck	Good Science Project
Tim Errington	Center for Open Science
Tom Pollard	Massachusetts Institute of Technology
Vani Mandava	Scientific Software Engineering Center, University of Washington

Appendix B: Projects Shared by Participants

Projects
Emerging Technology Observatory
Innovation Information Initiative
Map of Science
MIMIC
Data Commons Initiative
PhysioNet
ML Carpentries (e.g., Responsible ML)
UW SSEC Projects – eScience Institute
Addressing sociotechnical limitations of LLMs (AdSoLve)
Building a Measurement, Evaluation, & Learning Framework
(Re)designing AI for Diverse Disciplines
Syntheos PathFinder
InnoRate
Research Software Alliance (ReSA)
FAIR4ML
FARR
Journal of Open Source Software (JOSS)
Insights and Impact From Five Cycles of Essential Open Source Software
Adoption of Generative AI by Academic Biomedical Researchers
GitHub — docling-project
MOSS — OSSci
Lifecycle Journal
Registered Revisions
Predicting Replicability Challenge

Authors

Catherine Aiken is the director of data science and research at CSET.

Ronnie Kinoshita is the deputy director of data science and research at CSET.

James Dunham is the NLP engineer at CSET.

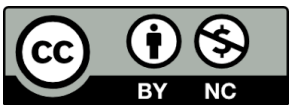
Greg Tananbaum is the executive director at the Open Research Community Accelerator.

Erin McKiernan is the director of programs and strategies at the Open Research Community Accelerator.

Acknowledgements

We thank the National Science Foundation, Office of Advanced Cyberinfrastructure for supporting this workshop (award number 2444480).

For report review, feedback, and editorial assistance, we thank workshop participants, Lauren Lassiter, Rai Hasen Masoud, and Shelton Fitch.



Document Identifier: doi: 10.51593/20250008