



RDAP ASSOCIATION

RDAP Happenings

Hi RDAP members,

Welcome to the July, er... August edition of RDAP's quarterly newsletter! Summer can be a quieter time, a season of vacations in the sun and recharging for the next academic year. It is also a time of transition for RDAP as incoming and outgoing members of Leadership and the Action Committees find their places, transfer their Google Docs, and chart the course of the Association's next year.

We know you are all on the edge of your seats/email clients waiting to learn when the next RDAP Summit will take place—we are too! In the meantime, check out the Annual Report, which offers an update on RDAP's strategic goals, amendments to the bylaws, and reports from each Action Committee on their progress for the year. We are also happy to share the activities of the Executive Board and several Action Committees in this newsletter.

In our feature article, outgoing chair of the RDAP Publishing Committee Ali Krzton dives into the issues and debate surrounding generative artificial intelligence (AI). As generative AI improves and makes it more difficult to delineate between human- and AI-generated content, ethical questions around data use, ownership, and attribution loom. Thank you to Ali for writing this article. We are always searching for guest contributors to write our feature article and welcome those who are interested to review our [guidelines](#) or email us at publications@rdapassociation.org. The next writer to grace the pages of a future newsletter could be you!

Updates from the Board

Happy Fiscal New Year to all, and welcome—to our incoming board and committee members, and new RDAP members, too!

As always, July brings changes and updates from the Board. The recently-released [2023 Annual Report](#) gives a great summary of the work we accomplished over the past year. It also includes a link to our newly-approved [2023-2026 Strategic Plan](#), which will help guide our organization into the future. The four overarching plan goals are: 1) Establish RDAP's position in the research data management landscape; 2) Foster sustainable organizational diversity and inclusion; 3) Improve

and refine RDAP's business practices and policies; and 4) Continue to iterate and add to RDAP's membership value.

Outgoing 2022-2023 board members include Jonathan Petters (outgoing Past President) and Patricia Condon (outgoing Treasurer). Our new board members are Reid Boehm (incoming Treasurer) and Tess Grynoch (incoming President-elect). A huge thank you to Jon and Patti for all their hard work, and we are excited to collaborate with Tess and Reid in their new roles! (Remember, if you think you may be interested, it's never too early to consider stepping up for a [board](#) or [committee](#) position next spring. We have many folks who have been involved for years, but we're always glad to welcome new faces and perspectives.)

As we move into my term as President, a few specific efforts I see as likely to define this next year include:

- Implementing feedback from the UNITE report on accessibility.
- Providing organizational DEIA training (more on this as we solidify our plans).
- Navigating the process of hybridizing future Summits. (Results from the Hybrid Summit Town Hall are now available on [OSF](#).)
- Figuring out our place in the research landscape—for example, responding to calls for public comment on new federal agency plans and policies.

I'm excited to traverse the next twelve months with all of you and see what we can build together. One of the main reasons I was willing to take on the role of President is that over the past several years of my involvement with RDAP leadership, I've seen firsthand how collaborative our work is; I know I can rely on those around me to contribute great ideas and experience beyond what any one person brings to the table.

As always, the RDAP Executive Board welcomes your input and feedback. Please don't hesitate to reach out to us at execboard@rdapassociation.org or to me, directly, at president@rdapassociation.org...I think I've got my email forwarding successfully set now!

The RDAP Board,

- Rachel Woodbrook (President)
- Tess Grynoch (President-elect)
- Lindsey Quinn (Secretary)

- Lindsay Gypin (Secretary)
- Reid Boehm (Treasurer)
- Amy Koshoffer (Past President)

RDAP Feature Article

Editorial: We Are All Data Now

Ali Krzton, Research Data Management Librarian, Auburn University



Image generated with the Wonder app on iOS by Ali Krzton

An abiding concern for the responsible use of data is nothing new for those of us

among the ranks of data professionals. Whether from the perspective of research, education, commerce, or policy, careful consideration of data ethics is an important aspect of our work. The recent increase in the accessibility and prevalence of “generative AI” technology adds a new dimension to the digital landscape and a new level of difficulty for would-be facilitators of information literacy. Briefly, generative AI uses complex, impenetrable statistical models to produce novel combinations of text or images in response to user prompts. The best-known example may be ChatGPT, which can in mere seconds write a poem in the style of Edgar Allen Poe, a tailored marketing pitch, or a research paper that may or may not be accurate. What does this mean for us?

As always, the story begins with the data. Rapid advances in generative AI tools such as ChatGPT depend on access to a truly massive corpus of data to scale up their learning. Thus, it is no surprise that material scraped and aggregated from the open web for free would be a critical source of training data. However, the people who created and published the articles, blogs, and forum posts used to shape large language models like those that power ChatGPT could not have foreseen, much less intended, their work would be used this way.

These concerns are now echoed by a broader community of content creators including authors, artists, and coders. Recently, some fanfiction writers expressed dismay at compelling evidence that the Archive of Our Own repository was used to train the GPT-3 LLM [1]. This example, while seemingly obscure, shows that nearly anything on the web might end up fueling the next AI iteration.

Researchers skeptical of open data sharing have long objected that putting their data online without restriction could allow bad actors to misuse it, or at the very least avoid attributing the source. The increasing emphasis on machine-readability and automated processing of research data [2] makes the attribution problem even more acute. In response to a lawsuit brought by developers of open source code on GitHub alleging that GitHub’s AI-powered Copilot tool violated their licenses, Microsoft and OpenAI defense relies in part on the lack of particular examples of misappropriated code [3].

In fact, it would be impossible for anyone to say which specific passage, image, script, or dataset was used to produce any given generative AI output by design. These models can only be built by digesting inputs down to vectors that do not translate to human-intelligible dimensions. This is how things produced by AI can be at once novel and derivative. Technologist Jaron Lanier warned of the attitude that such developments can only improve human understanding in a prescient passage from 2010: “A fashionable idea in technical circles is that quantity not only turns to quality at some extreme of scale, but also does so according to principles we already understand. Some of my colleagues think a million, or perhaps a billion, fragmentary insults will eventually yield wisdom that surpasses that of any well-thought-out essay, so long as sophisticated secret statistical algorithms recombine the fragments [4]” For those who hold improvements in AI capability as an end in

the fragments [1]. For those who hold improvements in AI capacity as an end in itself, the necessary stripping away of all context is an acceptable cost.

What are we to make of this massive disruption in our ability to control digital representations created either by us or of us? Continuing to learn everything we can about it is essential. Sharing our knowledge of how AI systems work with those around us is crucial so that we can all curate our personal data more intentionally. Helping more people be aware of possible interactions between AI and what they publish (or perhaps even enter) online could have a greater impact than any of us currently appreciate.

References

[1] Eveleth, R. (15 May, 2023). The fanfic sex trope that caught a plundering AI red-handed. *Wired*. <https://www.wired.com/story/fanfiction-omegaverse-sex-trope-artificial-intelligence-knotting/>

[2] Huerta, E. (14 February, 2022). FAIR and AI-ready scientific datasets. *SpringerNature Research Data Community*. <https://researchdata.springernature.com/posts/fair-and-ai-ready-scientific-datasets>

[3] Brittain, B. (27 January, 2023). OpenAI, Microsoft want court to toss lawsuit accusing them of abusing open-source code. *Reuters*. <https://www.reuters.com/legal/litigation/openai-microsoft-want-court-toss-lawsuit-accusing-them-abusing-open-source-code-2023-01-27/>

[4] Lanier, J. (2010). *You Are Not a Gadget*. New York: Vintage Books, p. 49.

RDAP Summit Updates

The Conference Planning Committee has begun meeting regularly, and planning is already underway for the next Summit. Stay tuned as dates for the RDAP Summit 2024 will be announced soon!

Action Committee Updates

DEIA

No updates at this time.

Education and Resources

The Education and Resources Committee asks RDAP members to keep an eye out for the call for webinar proposals in mid-August.

Marketing

The RDAP Marketing Committee has two new co-chairs this year: Jennifer Chaput from UMass Amherst, and Helenmary Sheridan from the University of Pittsburgh. The committee will have its first meeting in August, and we look forward to promoting and sharing all the great things happening with RDAP!

Membership

The Membership Committee would like to remind everyone to consider nominating someone for the [RDAP Member Awards](#): Volunteer of the Year and Work of the Year. Nominations will close on December 31, 2023.

Publishing

A new publication opportunity is available through our partners at the *Journal of eScience Librarianship*, who have shared the following information with us:

“We are happy to announce the first *Curatorial Connections Column* has been published in *JeSLIB*!

Hadley, H. & Narlock, M. 2023. Librarians and Research Integrity Committees: Finding professional joy and fulfillment in practice and service. *Journal of eScience Librarianship* 12(2): e680. <https://doi.org/10.7191/jeslib.680>

The *Curatorial Connections Column* is a partnership between [Data Curation Network](#) and the *Journal of eScience Librarianship*. In this recurring column, to be published in every regular issue of *JeSLIB*, the editors request articles on how

published in every regular issue of *VEEDL*, the editors request articles on how information professionals are implementing data curation practices, workflows for institutions and repositories, and ideas for improving data curation practices, and more. This column will be edited by Hannah Hadley, Princeton University, and Mikala Narlock, Data Curation Network.

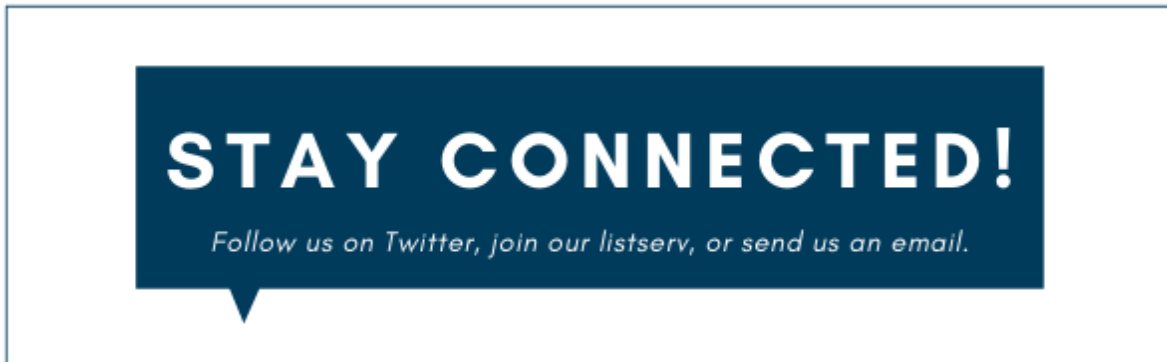
[Read more](#) about this new article type, and how you can submit your own ideas, experiences, and expertise for the *Curatorial Connections Column*.”

Sponsorship

No updates at this time.

Website

No updates at this time.



Visit us!



[Unsubscribe](#)