Data Sharing Seminar Series for Societies

Reproducibility and Open Scholarship: National Academies Efforts and Roles for Societies

3 December Chat Discussion

11:10:32 From Shelley Stall to Everyone:

Fostering Integrity in Research (2017): https://www.nap.edu/catalog/21896/fostering-integrity-in-research

11:11:39 From Shelley Stall to Everyone:

COPE - Committee on Publication Ethics: https://publicationethics.org

11:12:13 From Shelley Stall to Everyone:

Open Science by Design (2018): https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century

11:14:12 From Shelley Stall to Everyone:

Reproducibility and Replicability in Science (2019): https://www.nap.edu/catalog/25303/reproducibility-and-replicability-in-science

11:14:47 From Shelley Stall to Everyone:

Digital downloads are all free from NASEM - see the button on the right side of the screen.

11:18:48 From Todd Carpenter (NISO - he/him) to Everyone:

"making it easy" requires standards that support that systems interoperability and the use of those standards by scientists/systems providers

11:18:57 From Shelley Stall to Everyone:

+1 Todd

11:18:58 From Shelley Stall to Everyone:

https://www.nationalacademies.org/our-work/the-strategic-council-for-research-excellence-integrity-and-trust

11:19:31 From Shelley Stall to Everyone:

https://www.nationalacademies.org/our-work/us-national-committee-for-codata

11:20:43 From Shelley Stall to Everyone:

https://www.nationalacademies.org/our-work/developing-a-toolkit-for-fostering-open-science-practices-a-workshop

Data Sharing Seminar Series for Societies

11:25:31 From Philip Tarrant to Everyone:

The academic make up of the strategic council suggests we may be missing the "voice of the customer". How do we know we are taking steps that will address the increasing public distrust of research results?

11:25:46 From Sarah Nusser to Everyone:

Can you please speak more about how societies can help advance the ways we provide reward and credit for researchers who do a good job of sharing data?

11:26:01 From Shelley Stall to Everyone:

How can we "as societies" work best with international communities to coordinate our work with discipline-specific data sharing practices

11:26:25 From Stuart Chalk to Everyone:

What is the best way that International Unions (I represent IUPAC) can engage with this effort?

11:26:27 From Megan O'Donnell | she/her/hers to Everyone:

There seems to be a lot of uneven education and expectations with data among journal editors, not just publishers (such as societies). Have discussions reached this level yet?

11:26:55 From Plato Smith to Everyone:

USGS FSP - https://www.usgs.gov/about/organization/science-support/office-science-quality-and-integrity/fundamental-science-practices

11:27:12 From Brooks Hanson to Everyone:

marcia—many thanks. Fostering reproducibility will also require real investments in data infrastructure (e.g., supporting domain repositories consistently). What role does you see the academy, particularly in working internationally with other academies and w/ US gov't, to foster this support?

11:27:16 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

Making easier would be to have professional data scientists help work with PI's to get data shared open and FAIR. Making it beyond easy would be Artificial Intelligence that could find information on PI's computing infrastructure and suggest how to make it open and FAIR. Any efforts towards these?

11:27:24 From Brian Westra to Everyone:

This might be of interest: Open Research Toolkit "by Australasian Research Management Society (ARMS) and the Council of Australian University Librarians (CAUL). The Toolkit supports Australasian institutions to implement or further develop open research policy, strategy and practice." https://caul.libguides.com/open-research-toolkit

Data Sharing Seminar Series for Societies

11:27:38 From Zsolt Silberer, ESA to Everyone:

An important part of openness, standards and action is the activities of universities. We can't expect the govt to dictate this. How are you working with universities to ensure that credit, tenure and young researchers are set-up for this future - Zsolt Silberer Director Publishing ESA

11:28:01 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

I agree with Todd Carpenter that standards are important. And those standards need to be easy to use. As someone who manages a data center with a high degree of curation, it's surprising and frustrating how often we run into cases where researchers find getting things compliant with the Climate and Forecasting (CF) convention for netCDF. And the same is true for the frequency where we find significant usability issues with submitted data, like different values for missing data in different files or inconsistencies in variable names.   How can we help researchers make their data more usable by others?

11:28:31 From Ronald Taylor to Everyone:

Your key points — all important — relate to the classic understanding of verifying and validating results in accordance to the traditional scientific method. A lot of emphasis is being placed on results grounded on "evidence-based research." Can you speak a bit on the difference and any concerns that might arise, especially in policy considerations.

11:28:44 From Susan Borda to Everyone:

Here's another good resource: https://the-turing-way.netlify.app/reproducible-research/reproducible-research.html

11:29:35 From Steve Diggs to Everyone:

Now that we've stated our goals, described the current state of things, how do we influence funders to provide resources on the ground so that data curators can build out the infrastructure that will directly address the data challenges before us?

11:30:22 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

+1 Steve Diggs.

11:30:46 From Susan Borda to Everyone:

+2 Steve Diggs

11:30:49 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

@Zsolt, in my experience Universities are most interested in their own faculty and not in making data that helps disciplines or science move forward. I think our scientific societies should take the lead and the federal government provide the resources needed to implement. I still see NCBI being the best example of federal leadership in supporting open science and FAIR data

11:30:51 From Nancy Ritchey to Everyone:

+1 Steve Diggs and +1 Bruce Wilson

Data Sharing Seminar Series for Societies

11:30:59 From Lisa M Raymond to Everyone:

   +4 Steve Diggs

11:31:44 From Jonathan Petters to Everyone:

   Brooks Hanson asked a similar question to Steve Diggs, and +1 to both of them

11:33:52 From Paul Guinnessy (AIP, Physics Today) to Everyone:

   arXiv is a wonderful platform but it still hasn't got a long term viable business model. As we've seen just today with the purchase by Wiley of Knowledge Unlatched, non-profit OA is hard and can have long term risks without strong backing.

11:34:41 From Zsolt Silberer, ESA to Everyone:

   @Seth I agree and entirely support societies taking the lead. Yet, academics are part of NAS and other committees and do not pressure for change in merit, credit and tenure, Societies can help, but the Provosts, Deans and Presidents really need to step up.

11:36:13 From Jing Liu to Everyone:

   Another infrastructure question is how universities can be motivated to invest in developing institutional support for data sharing, both in terms of investing in research scientists who develop reproducible workflows and collaborate with faculty researchers to adopt best tools for data sharing and open science, and in terms of supporting faculty effort to improve reproducibility and replicability. Also, there are many, many tools / platforms out there, but the usability / interoperability, etc vary greatly and many researchers still don't know which to use.

11:36:59 From Jing Liu to Everyone:

   +1 Zsolt Silberer

11:38:11 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

   @Zsolt. Yes! I agree to that (and their organizations like e.g. APLU) I read it as meaning that university should be these repositories (e.g. libraries) and that was what I wish to discourage. Thanks for the clarification.

11:38:34 From Brian Westra to Everyone:

   +1 Steve Diggs

11:38:53 From Brian Westra to Everyone:

   +1 Zsolt Silberer

11:39:42 From Todd Carpenter (NISO - he/him) to Everyone:

   "Another flaw in the human character is that everybody wants to build and nobody wants to do maintenance."

   — Kurt Vonnegut

11:39:56 From Megan O'Donnell | she/her/hers to Everyone:

    🔗

11:39:58 From Biru Zhou|McGill to Everyone:

    +1 @Todd

11:40:03 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

    +1000 Todd Carpenter. Particularly for getting funding.

11:41:00 From Jing Liu to Everyone:

    +1000 Todd Carpenter.

11:41:58 From Dr. Donna J. Nelson to Everyone:

    One challenge is that open data are at risk of being reported by other researchers, who claim the data as theirs. It is easy to show and report who actually owns and generated the data, but little is done to stop the illegal reports. This discourages openness.

11:42:26 From Brian Westra to Everyone:

    There is a growing population of data curators who are actively involved in raising the bar.

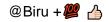11:42:50 From Biru Zhou|McGill to Everyone:

    The danger of having data professionals using soft money (funding) is that findings are periodic, this may bring discontinuity to research data maintenance. This may also put significant barriers to young researchers and certain disciplines that simply can not effort a data professional in their research program. This will make the rich richer and the poor poorer.

11:43:38 From Brooks Hanson to Everyone:

    survey Marcia is referring to: https://www.nature.com/articles/d41586-020-03191-7

11:43:45 From Jake Carlson to Everyone:

    +1 Biru. Infrastructure funded on soft money is not a real commitment to open science

11:44:01 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

    @Biru +💯 👍

11:45:08 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

    Biru -- agreed. This is part of the ongoing discussion within the Research Software Engineering forums, such as the US-RSE (https://us-rse.org).

11:46:07 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

    @Dr Donna J. Nelson - this essentially just happened to me. They used our open dataset but then only mentioned it in the Data Availability statement. But I already whined about this on a past session with this group. 🙁

Data Sharing Seminar Series for Societies

11:46:40 From Megan O'Donnell | she/her/hers to Everyone:

> @Seth - This goes back to needing to educate the editors on issues, not just scientists in my book.

11:46:47 From Sylvie Brouder to Everyone:

> Many funders don't fund "infrastructure"

11:46:48 From Megan O'Donnell | she/her/hers to Everyone:

> *data issues

11:47:20 From Todd Carpenter (NISO - he/him) to Everyone:

> Funding bodies are keen on supporting new infrastructure, but have explicit policies that prohibit supporting ongoing operational costs. There is a need for viable and sustainable business models that support these infrastructure costs and diverse buy-in to those service models. Marcia is spot on that we can't get to where we need to be without this funding.

11:47:52 From Brian Westra to Everyone:

> + 100 Todd Carpenter

11:48:05 From Biru Zhou|McGill to Everyone:

> +1 Sylvie and the door of funding infrastructure is getting tighter and tighter.

11:49:02 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

> @Megan I agree. However, as a society journal Editor myself I am almost positive my TE's and AE's don't actually review or follow the voluminous existing policies, not sure adding more will help. Everyone feels overwhelmed. It can hurt but it will take time as a culture change.

11:49:08 From Todd Carpenter (NISO - he/him) to Everyone:

> It's worth noting that the Director of ArXiv announced her departure this week as well for a position in the corporate world

11:49:22 From Zsolt Silberer, ESA to Everyone:

> +1 Brookes. There is a report in Washington Post Young Americans are really down on democracy and reeling from the pandemic. Also speaks to a younger generation that wants to engage differently.
> https://www.washingtonpost.com/politics/2021/12/01/young-americans-are-really-down-democracy-reeling-pandemic/

11:49:29 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

> It is interesting and frustrating to see the problems that remain with citing data. My particular data center was pretty early in this, getting DOIs for data in 2006. And we're still struggling with journals when we point out articles where authors failed to cite data.
> https://doi.org/10.1029/2020EO151665 (I am a co-author)

11:50:44 From Biru Zhou|McGill to Everyone:

I shared your frustration!!! @Bruce

11:51:29 From Susan Borda to Everyone:

@BruceWilson we also need CrossRef and DataCite to better together!!

11:52:02 From Susan Borda to Everyone:

To *work better together!

11:52:18 From Sarah Nusser to Everyone:

Two great partners in academic progress are AAU and APLU - https://www.aau.edu/accelerating-public-access-research-data

11:52:47 From Peter Schiffer to Everyone:

An issue for numerous university PIs is that they run small groups, they take data on home-built instruments or in specialized ways with output that is not in standard formats, and their grants are limited in total dollars so that funds invested in data curation are funds not spent on creating new results.   Much of this data may not be impactful enough to justify retaining for the long-term.   Can you talk about the tradeoff between cost and effort to make such data widely accessible and meeting FAIR standards, and the value of accessibility to the research community and broader society?

11:52:55 From Todd Carpenter (NISO - he/him) to Everyone:

@susanBorda => CrossRef and DataCit do work pretty closely together, but editors need to require that data be cited. CrossRef and Datacite can't fix citations that don't exist.

11:52:57 From Biru Zhou|McGill to Everyone:

+1 for AAU and APLU

11:53:05 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

" I walked to school uphill barefoot against the wind, so you should too" — too many university faculty.

11:53:28 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

MOOCs are going to replace universities we were told 🤣

11:53:47 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

@SusanBorda — Thanks, agreed.

11:54:38 From Steve Diggs to Everyone:

Dr. McNutt makes an excellent point about the rate of change (socially) in academia. Perhaps the solution is to have parallel non-academic institutions that get federal funding to assume proper full-stack data curation?

Data Sharing Seminar Series for Societies

11:55:12 From Susan Borda to Everyone:

> @ToddC CrossRef's citation matcher does not DOI's that are not CrossRefs

11:55:20 From Megan O'Donnell | she/her/hers to Everyone:

> @Peter - this is something institutional repositories are grappling with, the "what is long-term?" question

11:56:24 From Susan Borda to Everyone:

> @BruceWilson if you have DOIs in DataCite that are not explicitly in the Refs of an article you can attach the ref later: https://support.datacite.org/docs/contributing-data-citations#

11:56:37 From Biru Zhou|McGill to Everyone:

> @Peter, I think another important conversation to have is that what (data) actually deserves long-term retention. A lot of research can be re-conducted with minimal cost that may be cheaper than long-term preservation.

11:57:05 From Bruce Wilson (he/him; ORNL DAAC) to Everyone:

> @Susan — thanks. I didn't know that and I'll dig into what we can do with this. It at least makes some important data relatively public, even if it's not in the article itself.

11:58:21 From Brian Westra to Everyone:

> Inconsistent approach to data citation by journals (or no standardized approach), as well as differences between online and print are a significant problem. Can funders impose standards to improve this?

11:58:44 From Biru Zhou|McGill to Everyone:

> Sorry for my terrible typos, bad habits! 😂

11:58:56 From Todd Carpenter (NISO - he/him) to Everyone:

> @brianWestra - Funders can't impose those standards, but societies and publishers can

11:59:43 From Seth Murray (Soil Crop Sci. - AG) to Everyone:

> That is the UofA NSF funded Cyverse's model!

11:59:55 From Peter Schiffer to Everyone:

> @Biru Yes, agreed.   Also, much data on reproducible systems becomes irrelevant because better samples or techniques become available.