Data Sharing Seminar Series for Societies

Public Access to Research Data: Guidance, Culture, and Practice

2 April Chat Discussion

10:08:22 From Richard Nakamura to Everyone : All are invited to a public virtual meeting at NASEM on April 28-29 on "Changing the Culture of Data Management and Sharing - A Workshop"

10:11:59 From Shelley Stall to Everyone : AAU-APLU Public Access Working Group Report and Recommendations (2017), https://www.aau.edu/key-issues/aau-aplu-public-access-working-group-report-and-recommendations

10:12:37 From Brooks Hanson to Everyone : @richard—presume this is the link: https://www.nationalacademies.org/event/04-29-2021/changing-the-culture-of-data-management-and-sharing-a-workshop

10:14:17 From Shelley Stall to Everyone : When the guide is released, we'll share with all the registrants of the seminar.

10:16:36 From Shelley Stall to Everyone : I love how hydrology is used to show institutional change!!

10:17:00 From Amy Nurnberger (she/her) [MIT] to Everyone : In terms of changing metrics by which excellence in academia is assessed, HumetricsHSS is an interesting effort https://www.ssrc.org/programs/component/digital-culture/humetricshss/

10:22:32 From Shelley Stall to Everyone : @Amy - I really like the HumetricsHSS metrics recognizing there is a human being is at the center with: Collegiality. Quality. Equity. Openness (that include OPEN ACCESS). Community.

10:26:31 From BZ to Everyone : Can we have the presentation slides after the seminar?

10:27:24 From Shelley Stall to Everyone : The slides and recording will be placed here later today: https://zenodo.org/communities/data-sharing-seminar-series-for-societies/

10:27:31 From BZ to Everyone : Thank you

10:28:06 From Shelley Stall to Everyone : We'll also share this link as soon as its ready.

10:33:21 From Shelley Stall to Everyone : @ 57% Change Behavior — might be too low a number in reality!!

10:42:53 From Clarissa Russell to Everyone : Is ASU sharing data outside of the university?

10:43:36 From susan borda to Everyone : Yes https://dataverse.asu.edu/

10:46:08 From Amy Nurnberger (she/her) [MIT] to Everyone : On Developing training/curriculum & not recreating the wheel: these exist!

10:47:30 From Shelley Stall to Everyone : @Amy - Wondering how to connect researchers with what is available. Does AAU/APLU have a role here?

10:47:43 From plm to Everyone : Agreed @Amy. Including lots of this being done internationally.

10:48:07 From Philip Rubin to Everyone : I would like to hear more about combatting bias in data, and also more about increasing public trust in data sharing.

10:48:13 From Shelley Stall to Everyone : @plm and @Amy - Maybe RDA can help here.

10:48:13 From BZ to Everyone : Also wondering where can we find these success stories?

10:48:19 From Amy Nurnberger (she/her) [MIT] to Everyone : Check out: https://www.rd-alliance.org/groups/education-and-training-handling-research-data.html, https://elixir-europe.org/, https://dmtclearinghouse.esipfed.org/

10:48:48 From susan borda to Everyone : Research data documentation continues to be a huge struggle, huge % of datasets deposited without any documentation.

10:48:49 From Dennis Brown to Everyone : Who should decide which data to share, and when shoaled it be shared. The day of the experiment, a week after, after 3, 4 or 5 repetitions, when it is in pre-print ready format?

10:48:54 From plm to Everyone : Hooray for strategic discarding!

10:48:55 From Dennis Brown to Everyone : should

10:49:13 From Amy Nurnberger (she/her) [MIT] to Everyone : @Shelley - I think AAU/APLU could have a role in this space!

10:49:14 From Brian Westra to Everyone : I would love to do more collaboration. Canada has a great model: https://portagenetwork.ca/

10:49:23 From Susan Ivey to Everyone : +1 to strategic discarding!

10:49:27 From Amy Nurnberger (she/her) [MIT] to Everyone : @Brian absolutely!

10:49:37 From Sarah Nusser (she/her) to Everyone : Philip, do you have loops associated with evaluating the quality of the data or that information is protected? If so, how does that work?

10:49:57 From Sarah Nusser (she/her) to Everyone : *sensitive information is protected

10:50:39 From Amy Nurnberger (she/her) [MIT] to Everyone : @BZ Part of the APARD report will be including these success stories

10:50:57 From Joana Beja to Everyone : @DennisBrown some data are not finalised for years, e.g. if they have to undergo lab analysis which include transport to a specific lab and expertise to analyse them. Shouldn't be a blanket approach, but I would say before the project that collected the data, most, if not all data should be shared/submitted to a repository for long term management

10:52:21 From Danie Kinkade (she/her) to Everyone : Agreed Philip and Ian: Ecological data may be small but often heterogeneous in form and type.

10:52:23 From plm to Everyone : Cost of storage is a red herring. Is cost of software and file maintenance, migration as software changes, versioning, readability

10:53:20 From susan willner to Everyone : And non-US govt funder requirements!

10:53:35 From Amy Nurnberger (she/her) [MIT] to Everyone : @ palm +1 Also, the active cost of infrastructure, data stewardship & documentation

10:53:48 From susan willner to Everyone : foundations, other government rqts, etc

10:54:05 From Dennis Brown to Everyone : @JoanaBeja understood, but are "data" statistical summaries of larger collections, such as thousands of micrographs, or does "data" refer to all of the individual images that were analyzed to come to a conclusion?

10:54:06 From Shelley Stall to Everyone : And there is significant investment by the European Commission on open and FAIR data sharing.

10:54:22 From susan willner to Everyone : Agree with @amy re: data stewardship, infrastructure and documentation!

10:54:38 From Greg Tananbaum to Everyone : Cost of data hoarding, irreproducibility, etc. can't be discounted when we think about the status quo.

10:54:38 From Doug Schuster to Everyone : +1 on discarding data once it loses value. Need formalized processes to evaluate what and when data meets a TBD threshold to be discarded.

10:54:59 From Danie Kinkade (she/her) to Everyone : NSF Geoscience directorate has invested in a small number of domain repositories to provide data curation to their awardees at scale vs charging individual awards for data management. This model has been working well in some domains.

10:55:13 From Amy Nurnberger (she/her) [MIT] to Everyone : On the topic of funders and discussions, the Research Data Alliance does host a group that brings funders together to discuss these issues https://www.rd-alliance.org/about-rda/governance/rda-funders-forum

10:55:14 From Joana Beja to Everyone : @DennisBrown... raw vs processed :) both important for different reasons, no?

10:56:23 From Doug Schuster to Everyone : Measurements that can't be recreated need to be preserved indefinitely. Model outputs are great candidates to be discarded as they tend to lose value over time.

10:56:53 From Susan Ivey to Everyone : +1 Amy— funding infrastructure is such a challenge

10:57:08 From Amy Nurnberger (she/her) [MIT] to Everyone : On discarding data, as frequently heard, if the data isn't documented it's not usable, has minimal value and further resources should not be expended on it

10:58:14 From Joana Beja to Everyone : I agree @AmyNurnberger... Should we call it data if it's not properly documented?

10:58:41 From Joel Gershenfeld to Everyone : The infrastructure of repositories and campus data facilitators is all crucial. Equally important and even more challenging are the deep cultural changes. A classic model for culture change developed by Ed Schelin focuses on deeply

embedded assumptions in a culture. We have held a few "assumption wrangling" workshops on data sharing and reuse. This surfaced embedded assumptions about data being proprietary, research being competitive, priorities in spending on the science rather than the infrastructure, and others. These are deep challenges that we all face.

10:59:41 From Shelley Stall to Everyone : Thank you everyone…clearly this is a really important topic that needs further discussion.

10:59:52 From Amy Nurnberger (she/her) [MIT] to Everyone : Making data responsibly open :) As open as possible, as restricted as necessary.

11:00:05 From Jonathan Petters to Everyone : Amy beat me to the tagline :)

11:00:06 From Raul Miranda to Everyone : algorithms that generate the data are as (or more) important than the data and have longer survival rate. Part of the open metadata?

11:00:19 From Matthew Harp to Everyone : In my experience many of these challenges include a disparity between funding the technology (such base storage and repository platforms) but not necessarily the people to support the processes, culture change, and general decision making

11:00:28 From Chris Erdmann (AGU) to Everyone : On the horizon, the Biden administrations plans to invest in science funding, particularly infrastructure.

11:00:30 From Amy Nurnberger (she/her) [MIT] to Everyone : The CARE principle are an important partner to the FAIR principles: https://www.gida-global.org/care

11:00:33 From Dennis Brown to Everyone : @joanbeja Can you imagine the effort involved in uploading literally thousands of individual micrographs (perhaps both raw and processed) that contributed to a single scatterplot, for example. This would be totally different to uploading a single excel sheet containing thousands of individual values for gene expression levels, for example. Anyway, we need to think about what "data" really means in the context of rah current discussion.

11:00:54 From Sarah Nusser (she/her) to Everyone : @raul - yes, many researchers think their methods are the valuable area of reuse

11:00:55 From plm to Everyone : +1 @AMy FAIR and CARE principles

11:01:18 From Joana Beja to Everyone : @DennisBrown... I can... maybe we still need to find the technical solutions to make that work easier

11:01:25 From Laura Lyon to Everyone : Registration for the May seminar: https://forms.gle/YA7LktQqy5u2DWpR8

11:01:38 From Amy Nurnberger (she/her) [MIT] to Everyone : Also, Congratulations to Shelley Stall, New co-chair of the RDA Organizational Advisory Board!