Data Sharing Seminar Series for Societies

The Data You Document are the Data We Love

5 November Chat Discussion

10:11:47 From Shelley Stall to Everyone:

> I love this…80% metadata and 20% PIDs

10:14:41 From Shelley Stall to Everyone:

> Metadata like toothbrushes. 😆

10:15:06 From Elisha Wood-Charlson to Everyone:

> Best quote about metadata ever!

10:15:48 From Megan O'Donnell | she/her/hers to Everyone:

> The more granular "universal metadata" goes... the hardest it gets to use.

10:16:48 From Shelley Stall to Everyone:

> +1 Megan

10:18:37 From Shelley Stall to Everyone:

> Eva, that was great!!

10:19:50 From Michelle Heacock to Everyone:

> And how does the scientist who has figured out what the good metadata is, best communicate this for adoption of others? Toothbrush problem :-)

10:20:21 From Megan O'Donnell | she/her/hers to Everyone:

> We give folks a codebook/metadata template for tabular data and that seems to help with the basics at least

10:20:47 From Carly Huitema to Everyone:

> Can organizers of domain specific conferences begin to include more domain data specific sessions? Speakers, workshops, statements of agreed upon standards?

10:20:47 From Biru Zhou, Dr. to Everyone:

> +1 Megan

10:21:18 From Shelley Stall to Everyone:

> Later today we'll share the slides and recording. thank you all for asking.

10:22:00 From Yuhan (Douglas) Rao to Everyone:

> With the increasing interdisciplinary/transdisciplinary research, how feasible is it to maintain disciplinary metadata or how to best use these metadata to enable the research that is really hard to categorize into a single discipline?

Data Sharing Seminar Series for Societies

10:22:25 From Bonnie Crotty Nelson (RTI Intl, RTP, NC, USA) to Everyone:

I'm seeing a black bar across the slides (and the link). Help?

10:22:55 From Shelley Stall to Everyone:

Happy to add more repositories to this list for Earth, space, and environmental sciences: https://data.agu.org/resources/useful-domain-repositories

10:23:02 From Murray, Seth C to Everyone:

@Yunan that is a fantastic question!

10:24:05 From Megan O'Donnell | she/her/hers to Everyone:

the structured metadata requirements are low

But no one may be able to use it /shrug

10:25:18 From Bonnie Crotty Nelson (RTI Intl, RTP, NC, USA) to Everyone:

+1 Yuhan!

10:25:42 From Plato Smith to Everyone:

A scientist asked that question yesterday. What are the core elements should we used for their large grant funded project? Referenced the 16 core element set from Data Audit Framework (Jones et al, 2009) and ACM SIGMOD Readme reproducibility template. +1 Megan.

10:28:17 From Murray, Seth C to Everyone:

re: how to do it - My lab was the first to make an agricultural drone field data set availible and FAIR (on Cyverse) the only way this was possible was working with an actual open data scientist (PhD level). It was a lot of work but now a template is set for others. However, the first user of this data did not appropriately cite the data even though the citation was clear. That is an issue not often discussed about open FAIR data, but it is also a minor problem (for now).

10:29:07 From Shelley Stall to Everyone:

@Seth - Good point. We really do need the publishers to encourage accurate citations.

10:29:13 From Megan O'Donnell | she/her/hers to Everyone:

\o/ - on getting a schema established and :-(for no citaton

10:29:45 From Megan O'Donnell | she/her/hers to Everyone:

I like the role drop-down in Zenodo!

10:29:55 From Carolina Peralta to Everyone:

@Seth and @Shelly I totally agree!

10:30:31 From Murray, Seth C to Everyone:

@Shelly I don't think there is knowledge in the reviewer and Editor community to enforce this even if publisher requires it is the rubber meets the road challenge.

Data Sharing Seminar Series for Societies

10:30:52 From Eva Méndez to Everyone:

    @Yuhan very good question. If your research is really inter/transdisciplinary the metadata schema that you Will use is probably a domain-agnostic schema/standard and several specific schemes (value-content vocabularies) that helps to describe the disciplinary richnest of your interdisciplinary Research data!

10:31:26 From Murray, Seth C to Everyone:

    Here is the link if anyone is interested in an example "CyVerse Data Commons" https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/tamu_corn_2017_CS17_G2F_20190305

10:31:33 From Yuhan (Douglas) Rao to Everyone:

    In recent two reviews for AGU journals, I did not authors are not citing data correctly and have pointed that out in the review. I think the data citation can be improved by better integrating the dataset citation with reference management software since that is what authors really rely on.

10:31:36 From Eva Méndez to Everyone:

    *richness I meant

10:31:57 From Megan O'Donnell | she/her/hers to Everyone:

    +1 Eva

10:31:58 From Shelley Stall to Everyone:

    @Seth…So true. At AGU we are working hard on this awareness and expectation. We have a new reviewer/editor checklist for data and software that will be unveiled in the next week (-ish).

10:32:10 From Megan O'Donnell | she/her/hers to Everyone:

    <3 Shelley

10:32:14 From Carolina Peralta to Everyone:

    One good thing is to add a recommendation about how the data must be cited: https://obis.org/manual/policy/

10:32:46 From Shelley Stall to Everyone:

    +1 Carolina

10:33:09 From Shelley Stall to Everyone:

    Woo! for the Related Outputs field!!

10:33:18 From Yuhan (Douglas) Rao to Everyone:

    Thanks, Eva!

10:34:03 From Biru Zhou, Dr. to Everyone:

👍 @Carolina

10:34:05 From Megan O'Donnell | she/her/hers to Everyone:

oh wow, the relationship metadata is... complicated

10:34:22 From Murray, Seth C to Everyone:

@carolina thanks for the link, I think this is clear. In my reuse case they put the data DOI's in data availability statement but did not cite it. I am no sure a reviewer or Editor would even catch this.

10:34:51 From Shelley Stall to Everyone:

ah..yes @Megan…thinking we need that to give some guidance there.

10:35:30 From Megan O'Donnell | she/her/hers to Everyone:

@Seth Have you seen the latest recommendations about data citation vs availability statements? It's not standardized and "data citation" is not "done" by some journals

10:35:48 From Megan O'Donnell | she/her/hers to Everyone:

it's going to take time to change this

10:36:24 From Shelley Stall to Everyone:

Data citations are needed in order for automated attribution and credit to work. We do need to encourage that in papers.

10:37:09 From Murray, Seth C to Everyone:

@Megan, this is exactly the info I came here looking for, can you send a link to that recommendation?

10:37:28 From Bonnie Crotty Nelson (RTI Intl, RTP, NC, USA) to Everyone:

Yes -- I'm continually telling staff that metadata makes it discoverable, not accessible!

10:38:00 From Amy Nurnberger (she/her) [MIT] to Everyone:

Data citations: so necessary!

10:38:19 From Megan O'Donnell | she/her/hers to Everyone:

here you go @Seth: https://osf.io/z4py9/

And yes - citations are needed for more than "papers"

10:38:20 From Murray, Seth C to Everyone:

@Shelly thanks, that is well put. Was not sure that is a common belief, so appreciate you saying that

Data Sharing Seminar Series for Societies

10:38:51 From Shelley Stall to Everyone:

@Seth. On this page, under the "Data and Software Sharing" are references

10:38:53 From Shelley Stall to Everyone:

https://www.agu.org/Learn-About-AGU/About-AGU/Data-Leadership

10:39:12 From Shelley Stall to Everyone:

Also, most publishers have signed the Joint Declaration on Data Citation.

10:39:29 From Shelley Stall to Everyone:

That's the last reference in the list.

10:39:29 From Plato Smith to Everyone:

https://graph.openaire.eu/

10:39:36 From Murray, Seth C to Everyone:

@Megan and @Shelly, I think you both just made my day! 🙏

10:41:00 From Eva Méndez to Everyone:

@Bonnie interesting reflection… Metadata are important for all FAIR principles but for sure INTEROPERABILITY and REUSABILITY, and also Findability (when you know that the data exist) and DISCOVERABILITY (when you do not know that the data exist)

10:44:32 From Amy Nurnberger (she/her) [MIT] to Everyone:

And not just data availability statements, please, also data citations!

10:44:45 From Amy Nurnberger (she/her) [MIT] to Everyone:

Data citations absolutely necessary for driving the academic incentive system

10:44:53 From Biru Zhou, Dr. to Everyone:

+1 Amy

10:45:03 From Murray, Seth C to Everyone:

+2 Amy

10:45:06 From Shelley Stall to Everyone:

+1000 Amy — must have both availability statement and citation in the references.

10:45:13 From Megan O'Donnell | she/her/hers to Everyone:

Yes - there has to be a structured way to track use.

10:45:37 From Biru Zhou, Dr. to Everyone:

Love the +1000 Amy. I was too moderate.

10:45:49 From Shelley Stall to Everyone:

   :-)

10:45:49 From Richard Nakamura to Everyone:

   This makes clearer that planning for the metadata presentations should be accomplished before starting to collect data, not at the end of the study.

10:46:05 From Megan O'Donnell | she/her/hers to Everyone:

   you rebel Richard!

10:46:24 From Shelley Stall to Everyone:

   @Richard - oh yes…planning

10:46:25 From Yuhan (Douglas) Rao to Everyone:

   Eva's response with legal issue seems to just open another can of worms about proper license for data...

10:46:48 From Eva Méndez to Everyone:

   EXACTLY...

10:48:34 From Murray, Seth C to Everyone:

   +1 Richard, that's a long row to hoe when most researchers don't even plan how to analyze thier data until its collected😉

10:48:59 From Megan O'Donnell | she/her/hers to Everyone:

   I had to have some very real talks with grad students about how the data they want to use for a project... isn't available or may not exist. Not everything is "at their finger tips"

10:49:13 From Amy Nurnberger (she/her) [MIT] to Everyone:

   To answer the question, I like this quote from Dremio as to where we should be now, ""It's mission critical for data consumers to access data from a single pane of glass and aggregate different data products on the fly, especially when data ownership is spread across multiple data domains."

10:49:15 From Eva Méndez to Everyone:

   But I was referring the regulations and mandates to deposit your data, ej in Europe the directive 1024/2019

10:49:24 From Biru Zhou, Dr. to Everyone:

   +1 Seth, that is certainly a reality needs to be considered

10:50:37 From Amy Nurnberger (she/her) [MIT] to Everyone:

   …and it's the large gap between what researchers expect, the behaviors they're willing to engage in, and the behaviors that incentivized and prioritiezed

Data Sharing Seminar Series for Societies

10:50:49 From Shelley Stall to Everyone:

    NIH has a new policy requiring data sharing staring Jan 2023

10:51:41 From Shelley Stall to Everyone:

    NASA is embracing Open Science that includes open collaboration and open sharing to include data and software.

10:51:53 From Megan O'Donnell | she/her/hers to Everyone:

    The NIH policy is going to be a bit of a mess - lack of guidelines and infrastructure for safe sharing for some of the data....

10:52:04 From Bonnie Crotty Nelson (RTI Intl, RTP, NC, USA) to Everyone:

    +1 Shelley. That's going to drive adoption of all of this at our org.

10:52:37 From Frank Krause to Everyone:

    Trying to prepare lifesciences researchers for the NIH Data Sharing policy is a key driver of the FASEB DataWorks! program.

10:52:39 From Shelley Stall to Everyone:

    NSF is updating their Public Access repository to include datasets. That's coming soon.

10:53:05 From Anne Cambon-Thomsen to Everyone:

    some years ago we published a guideline of a standard for biorepository citation (not only data, but the physical bioresource in the citation :;it may help; https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-015-0266-y

10:53:10 From Shelley Stall to Everyone:

    YES!! We need our libraries!!!

10:53:22 From Amy Nurnberger (she/her) [MIT] to Everyone:

    +1 libraries

10:53:31 From Megan O'Donnell | she/her/hers to Everyone:

    @NSF is just for metadata, not data files though right?

10:53:36 From Frank Krause to Everyone:

    These presentations have been exceptional and very much appreciate the panelists' time and expertise. Thank you very much! Have to run, actually have an in-person meeting to get to :)

10:53:45 From Shelley Stall to Everyone:

    @Megan - it will be an index. Right.

Data Sharing Seminar Series for Societies

10:53:50 From Richard Nakamura to Everyone:

　　Agree!

10:54:15 From Yuhan (Douglas) Rao to Everyone:

　　I have to confess that as a researcher, I am somewhat disengaged with libraries until I need literature and help from them. So that is another problem need to address as well.

10:54:21 From Megan O'Donnell | she/her/hers to Everyone:

　　@Shelley - it's a start!

10:54:25 From plm to Everyone:

　　Incentive extends to libraries (resources, support) for building and maintaining infrastrucgture

10:54:58 From Yuhan (Douglas) Rao to Everyone:

　　I think @plm is absolutely right that libraries is not well supported

10:55:12 From Amy Nurnberger (she/her) [MIT] to Everyone:

　　+1 pls: infrastructure and the people that make it work

10:55:22 From Megan O'Donnell | she/her/hers to Everyone:

　　full disclaimer - I'm a librarian and I specialize in helping people publish and share data - please do connect with your librarians. If your library doesn't support this, tell them you're interested.

10:55:26 From Murray, Seth C to Everyone:

　　tenure can't be granted based on datasets because these are not peer evaluated, until that happens (routinely, I have a couple data papers), it won't be considered in tenure

10:56:15 From Yuhan (Douglas) Rao to Everyone:

　　@Megan, yes! Until I became friends with some librarian I fully realized how much I missed by not taking full advantage of the library

10:56:22 From Shelley Stall to Everyone:

　　@Seth…what you say is really important. We need that conversation to be more prominent.

10:56:29 From Megan O'Donnell | she/her/hers to Everyone:

　　Pass that on to students :-)

10:56:46 From Shelley Stall to Everyone:

　　We will share the chat as well. :-)

10:56:48 From Anne Cambon-Thomsen to Everyone:

in France there is a large push towards open science through a "National Open science plan" with one pillar being about evaluation.

10:56:49 From plm to Everyone:

Chicken and egg of infrastructure and participation

10:56:49 From Amy Nurnberger (she/her) [MIT] to Everyone:

+1000 Megan & libraries: Please make friends with your librarians

10:57:20 From Shelley Stall to Everyone:

Yes, indeed

10:57:36 From Yuhan (Douglas) Rao to Everyone:

There should be a requirement for grad school party to + 1 librarians as your guest…

10:57:57 From Megan O'Donnell | she/her/hers to Everyone:

😂

10:57:59 From Bonnie Crotty Nelson (RTI Intl, RTP, NC, USA) to Everyone:

Yes, I'm a librarian and many of us are collecting datasets metadata in our repositories. Thank goodness for DOIs -- which everyone now [mostly] understands.

10:58:38 From Shelley Stall to Everyone:

Eva and Joy - if there is a moment…what is FAIRsFAIR doing with data peer-review?

10:58:53 From Murray, Seth C to Everyone:

Thanks all!!!!

10:59:10 From Lindsay Callaghan to Everyone:

Thanks all, that was really interesting!

10:59:11 From Megan O'Donnell | she/her/hers to Everyone:

Thank you for such a wonderful session!

11:00:06 From Anne Cambon-Thomsen to Everyone:

data papers peer review in specialised journals (data papers)