

Chat Transcript

Data Refuge

Justin Schell

4T Data Literacy Conference – July 20, 2017

This is an edited, abridged version of the chat transcript (we've removed things like tech support requests and hellos). Last names have been removed.

Angie (Moderator): <http://www.4tdwvirtualcon.com/>

Kristin (Staff): You can download the book that accompanied last year's conference here --> <http://dataliteracy.si.umich.edu/book>

Amy (Staff): If you have a question you'd like to ask Justin, put them in the chat and they'll be shared with him as we go.

(Justin (presenter) asks who has heard of Data Refuge.)

Claire: I have!

Kristin (Staff): I have!

Angie (Moderator): Go ahead and let us know in chat

Kathy: Nothing at all

Renee: I haven't

Amanda: I have not

Kathryn: I have!

Robbie: I have not.

Peggy: no

Kathy: no

Jole: heard of, not been to

Beth: Nope, very intrigued!

Sheri: No

Kathryn: I have not been to an event.

Nicole: I'm new to this topic.

Kristin (Staff): Wow - more people responding yes than I had anticipated!

Matthew: heard of it familiar

Michelle: I have - they were all over tumblr shortly after they started, they're a brilliant idea.

Angie (Moderator): A lot of new people with this topic!

Dana: ok

Angie (Moderator): More info about Data Refuge here: <https://www.datarefuge.org/>

Angie (Moderator): <https://www.washingtonpost.com/news/energy-environment/wp/2016/12/13/scientists-are-frantically-copying-u-s-climate-data-fearing-it-might-vanish-under-trump/>

Angie (Moderator): A couple of people who are here in the room attended. A lot of people from Ann Arbor and UM came

Shelley: As a previous government documents librarian this was a topic at the beginning of government websites (back in 90s). Glad to see it is still going.

Angie (Moderator): Right!

Angie (Moderator): Justin works with the Gov Docs librarian here at UM as well

Angie (Moderator): He works with her on this project

Kristin (Staff): @Shelley makes a great point -- this is not a new challenge -- there is no mandatory digital archiving system for federal agencies. There have been initiatives to archive the federal web before presidential transitions since at least 2008 (Internet Archive + other partners)

Angie (Moderator): I'm glad it's going too Shelley

Shelley: I really appreciate it their efforts as this is so important!!!!

Angie (Moderator): Brandon: Made a note of your question

Jole: would be great to make our high school students aware of the lack of official infrastructure for archiving info

Angie (Moderator): Hi, Jole!

Jole: hi :)

Shelley: Oops I mean that their efforts are extremely important as websites vanish (not like books).

Kristin (Staff): End-of-term harvest goes back to 2008. Federal end-of-term archiving plan was in place before either of the parties' conventions

Peggy: I'm so happy to be learning about this. I was aware of the loss of data re climate change as it was happening but didn't understand this was part of the data rescue.

Kristin (Staff): @Jole - Hi! Guys, Jole has a fantastic session coming up tomorrow on personal data privacy choices. You're gonna love what she has on tap!

Angie (Moderator): Glad to get that feedback, Peggy!

Angie (Moderator): Kevin says that Question 1 is on point!

Angie (Moderator): Could be an interesting conversation with high school students

Beth: When you rescue data, where do you put it for security and longevity?

Angie (Moderator): Great question, Beth.

Angie (Moderator): I've made a note ... We'll ask about that at the end

Renee: Who is the data accessible to? What about confidential data?

Peggy: these questions are terrific! Will we have access to Justin's slides?

Angie (Moderator): Yes you will have access to Justin's slides

Suzanne: Who will be coordinating all of these various projects so that there is not overlap/duplication?

Jole: go, democracy!

Peggy: thanks!!!

Beth: Thank you!

Suzanne: What about the PEGI project?

Amy: Remember, if you want all participants to see your chat, make you have that option selected with the drop down menu

Peggy: love it! a project with my name :-)

Michael: When selecting materials to undergo a data rescue project, what criteria do you have for selection? You use climate as an example, and that is a popular topic currently, but are there any items on the block that may not necessarily be a "popular topic", but needs to be preserved none-the less?

Peggy: this seems like such an overwhelming project!! but sooooo valuable!!

Angie (Moderator): It is a big project!

Dana: <https://www.crl.edu/preservation-electronic-government-information-peg>i

Angie (Moderator): Thanks, for sharing, Dana!

Karen: Yes, thanks.

Peggy: love using internet archive (wayback machine)!

Kristin (Staff): Love the Wayback Machine, accessible from <http://archive.org>

Angie (Moderator): <https://envirodatagov.org/event-toolkit/submit-a-tool/>

Kristin (Staff): You can use it to find old versions of web pages, blogs that have been deleted, old newspaper articles. Isn't always 100% infallible, but it's AMAZING what you can find ... even U-M's very first website!

Kristin (Staff): By "it," I mean Wayback Machine :)

Peggy: great description!

Angie (Moderator): 19,000 URLs across 2 days!

Amanda: that sounds so cool! Cant wait to share wayback machine with the social studies teachers!

Angie (Moderator): More info about workflow on Github:
<https://datarefuge.github.io/workflow/>

Angie (Moderator): That's an older version but gives you an idea

Angie (Moderator): Github has lots of info about various aspects of the workflow

Angie (Moderator): Good old Wikipedia has more info on Bagit:
<https://en.wikipedia.org/wiki/Bagit>

Angie (Moderator): CKAN: <https://ckan.org/>

Angie (Moderator): King Bagger, Justin Schell!

Jole: impressive

Jole: can we get a benchmark statistic for context? something to compare those numbers to?

Michelle: Wow!

Angie (Moderator): Jole: Does that 20 million number give you a good comparison?

Angie (Moderator): justin said 300+?

Angie (Moderator): For 2012 it was 20 million at the end of the fiscal year

Angie (Moderator): it shows a real increased effort

Angie (Moderator): Questions about workflow?

Angie (Moderator): Just giving you a chance here ...

Angie (Moderator): A chance to ask questions

Angie (Moderator): Moving on ...

Angie (Moderator): Right, Jole!

Angie (Moderator): Laura: Justin has info about future plan in the next few slides

Peggy: That's why this project is essential! The Trump administration doesn't seem to believe that this data belongs to everybody!

Angie (Moderator): Great question, Brandon ...

Angie (Moderator): Making a note of your question which went out to just the panelists

Shelley: Yes. One minute the site is there and the next it is gone.

Angie (Moderator): I'll make a note so that Justin gets it

Stephen: It's great that the data storage is happening now. Had the national archives been in place earlier, the valuable 1890 census might have been properly stored, and not destroyed by fire and water damage in a warehouse.

Angie (Moderator): I think of that flood all of the time, Stephen!

Layla: This is so mind blowing! Never thought about the data rescue heroes out there! Such a vital role, but I guess I never thought that this is something that is being done right now.

Stephen: That flood created a 20 year gap, which makes family research frustrating at tiems.

Angie (Moderator): Laura: Made a note of your question ...

Amy: We'll get your questions in front of Justin when the opportunity arises.

Peggy: we take so much for granted about the availability of data!

Angie (Moderator): Zooniverse: <https://www.zooniverse.org/>

Claire: Justin, can you provide links for data gathering tools to be used by individuals in service of a larger project?

Renee: Is the data stored in native language or English?

Angie (Moderator): SciStarter: <https://scistarter.com/>

Suzanne: How do we identify these specific projects that we might contribute to (big cats, etc.)?

Amanda: @peggy- agreed!

Karen: Zooniverse sounds great.

01:18:09 Susan Wizinsky: This fits very well within the ACRL Information Literacy Threshold Concept of information creation. Exciting new ideas to get students thinking about data and information.

Angie (Moderator): Suzanne: There's a great webinar tomorrow about choosing citizen science projects

Angie (Moderator): True, Susan!

Peggy: these are great suggestions for involving students in these research efforts and making them more aware of why they are important!

Angie (Moderator): Claire: Zooniverse and Sci Starter are a good place to start

Peggy: being searchable is critical!!!

Claire: But Zooniverse and Scistarter are for analyzing other people's data. Justin mentioned tools for collecting data yourself as well.

Angie (Moderator): Ok, Claire

Beth: Searchable - yes! I was wondering how they'll deal with this since data may be coming off of websites that had high traffic and may become buried????

Claire: Just thinking of projects we could do for Open Data Day--we don

Claire: don't have the resources to do a data rescue, but we could collect data.

Peggy: so happy to connect with Justin on Twitter! <https://twitter.com/612to651>

Angie (Moderator): Beth: Searchability has been a problem for federal data all along though

Angie (Moderator): There were people in Data Refuge who just did metadata

Beth: No doubt!

Angie (Moderator): Here's an example of UM's data repository
<https://deepblue.lib.umich.edu/data/>

Peggy: Open Data Day resources from March 2017 <http://opendataday.org/>

Beth: Makes me realize how important it is to have a fundamental understanding of how the internet works!

Angie (Moderator): Other data repositories:
http://oad.simmons.edu/oadwiki/Data_repositories

Susan: So true. Beth!

Angie (Moderator): <http://www.downthemall.net/>

Peggy: wow! never heard of DownThemAll! Fantastic!

Beth: Cool resource!

Claire: Was thinking more of the rainmeter-type tools you mentioned.

Nicole: Thank you for an informative presentation!

Justin (Presenter): CoCoRaHs

Peggy: is there a Chrome Extension for DownThemAll?

Angie (Moderator): <https://www.cocorahs.org/>

Beth: Cool! That would fit nicely into my school garden project!

Angie (Moderator): Cool, Beth!

Angie (Moderator): We would love to hear how your project goes

Karen: Thanks.

Angie (Moderator): Any other questions?

Robin: Thank you!

Peggy: Thanks Justin for this very informative session!! Learned so many new things!
Appreciate all of the links shared in the chat!

Kathy: Exploding with info - Thank you J!!

Beth: I feel so enlightened from hearing all of this for the first time!!! Thanks a mil!!!

Layla: Thank you so much, Justin!

Michelle: Thanks Justin! This explained a lot of things I was curious about and added to my interest in the topic

Karen: Thanks, Justin. This is my first exposure to data refuge/rescue. Didn't even realize it existed.

Shelley: I appreciate all the efforts.

Shelley: This is such a massive undertaking but an important one.

Debra: wow... I truly had no idea . thank you

Angie (Moderator): <shares link for evaluation form, now closed>

Matthew: Thank you Justin!

Justin (Presenter): Thank you everyone!

Justin (Presenter): The UCLA Project is here: <http://newsscape.library.ucla.edu>

Sheri: Thanks.

Shannon: Thank you.

Dana: Thank you.

Angie (Moderator): Kathy: A link to what in particular?

Angie (Moderator): Eval link: *(no longer active)*

Kathy: yes thank you

Amanda: Thank you! :)

Stephen: Wonderful presentation.

#