

# TDR Steering Committee, Virtual Annual Meeting

Wed., May 27th from 11AM - 3PM

<https://utexas.zoom.us/j/608531643>

## Attendees:

- \*Adrian Shapiro (TWU)
- \*Christina Chan-Park (Baylor) - Vice Chair
- \*Courtney Mumma (TDL)
- Colleen Lyon (UT Austin) - tentative
- \*Dianna Morganti (TXST)
- \*Hammad Khan (UTA) (281-217-3934)
- ~~Jessica Trelogan (UT Austin)~~
- \*Laura Waugh (TXST) - Chair
- Laura Sare (TAMU) - another conference conflict
- \*John Watts (TAMU)
- \*Matthew McEniry (Texas Tech)
- ~~Rogelio Hinojosa (TAMU)~~
- \*Reid Boehm (UH)

## Agenda

### 11AM-11:05

#### Welcome & Nesting (5 minutes)

- Note taker - Lea DeForest (TDL)

### 11:05-Noon

#### Projects updates (55 minutes)

- [Carpentries membership](#) - JW & DM
- [Assessment WG Annual Report/Poster](#) - CP, LW, LS
  - [Report/Poster for each Institution](#)
- [Chronopolis digital preservation](#) - CM
- [Core Trust Seal](#) - CM
- [Larger Datasets](#) - MM
- [R Integration Presentation](#) (Reid)

### Noon-12:10

10 minute Break

### 12:10-12:15

## TDL Executive Director Remarks (5 minutes)

12:15-1

### [Roadmap](#) and priorities discussion

1-1:30

30 minute break

1:30-2:15

### New projects (based on roadmap discussion) CM

- [Code/Software archiving](#)
- Google analytics
- Outreach & Training priorities
- TBD based on Roadmap discussion
- UT Austin and UT iSchool capstone report<sup>1</sup> (Brenna Wheeler and Colleen Lyon)

2:15-2:45

### Other Discussion

- [CARE Principles for Indigenous Data Governance](#) (Courtney)
- Dataverse Annual Meeting Agenda: <https://projects.iq.harvard.edu/dcm2020/agenda>
- Video for Dataverse Community Meeting? (Christina)  
[https://harvard.az1.qualtrics.com/jfe/form/SV\\_ezirKHKbqOMXl6R](https://harvard.az1.qualtrics.com/jfe/form/SV_ezirKHKbqOMXl6R)
- Project management - how's it going?
  - Add unaddressed projects to rolling agenda? Have a separate place for it?

2:45-3

### Election (Mike the TDR Armadillo is still in quarantine at Texas State)



- Nominee(s)
  - Reid - U of H
  - ?
- [TDR Steering Committee Charter](#) (for more info)

---

<sup>1</sup> Abstract of Brenna's report: For the Texas Data Repository, I created a data curation workflow based on the Data Curation Network's CURATE(D) Model to improve the findability and reusability of datasets. The workflow is localized to the Texas Data Repository using needs identified in interviews with academic librarians and my assessment of datasets currently in the repository. My final product is a specialized Data Curation workflow and a list of recommendations that may be used by a team of liaison librarians to curate newly deposited datasets in the future.

## NOTES

### 11AM-11:05

#### Welcome & Nesting

- Note taker - Lea DeForest (TDL)

### 11:05-Noon

#### Projects updates

- [Carpentries membership](#)

*John Watts, Texas A&M University; Dianna Morganti, Texas State University*

- Waiting for approval - following points are contingent upon approval
- Considering timeframe for a start date of Spring 2021 for workshops due to face-to-face component of Carpentries workshops (they aren't set up for online workshops just yet)
  - Once approved, start recruiting the first cohort of six seats for Instructors (from within the TDRSC) or by appointment by cmte members' institution
  - May add more than one person per institution, but a \$1k fee is required per seat
  - Seem to have enough interest to develop full cohort for the spring
  - May decide / need to alter the timeline depending on timing of approval as well as public health situation
- TDL Governing Board is enthusiastic about the membership, even understanding that the pilot year is just Data Carpentries.
  - Uncertainty re: resource allocation difficult during the pandemic; UT Austin (TDL's home institution) is requiring TDL to justify expenditures and these purchasing restrictions are holding up a number of initiatives. (Ex. we're being asked, why do you need this now? How have you survived without it?)
  - Hopeful that the restrictions for purchasing will loosen up in September
  - Expect the TDL Board to approve the expenditure and the budget; also expect some restrictions re: purchasing along the way.
- NOTE: will need to make ample / significant room on the RoadMap

- [Assessment WG Annual Report/Poster](#)

*Christina Chan-Park, Baylor University; Laura Waugh, Texas State University; Laura Sare, Texas A&M University*

- Same as poster from last year, just in slide format :)
- Users by month: steady increase
  - Almost 600 registered users in the TDR
  - UT = highest number of users with TAMU closely following
- Depends on the model the school has:
  - Example: TX State users not creating Dataverses but adding at the top level
  - Some schools make Dataverses, but don't add sets to them
  - Different naming conventions and research focus area by school
- Creating more datasets than are being opened; archive space after project is done
  - Example: only half of UT's sets have been published; rest of schools show that nearly all sets are published
- Could we add a slide to the annual report that includes list of outreach activities?
  - TDR SC has edit access to the slides; customize colors and add what info is necessary to customize for your institution
  - add to Annual report template - "Projects and plans for the coming year, and opportunities for individual institutions"
- [Report/Poster for each Institution](#)
  - Your own ppt with just the numbers from your institution
    - These will automatically update when the spreadsheet is updated
    - You'll just need to change the YEAR; otherwise, data in charts is automatically updated.
    - Includes a link to a s/s with a list of users w/ contact info; allows you to see activities of users and identify users who may benefit from training materials, touching-base, etc.
    - Spreadsheet is in the folder.

- [Chronopolis digital preservation](#)

*Courtney Mumma, Texas Digital Library*

- CM is talking with budgeting office to make it a contract or PO for Jim Meyers (PO is more straightforward)
  - Jim can make time this summer, but project will be more spread out than originally expected due to Covid-times/wfh

- Expect Jim to finish by August
  - While Jim works, CM will publish documentation developed by Reid, Santi, and Laura W.
  - Promote to faculty afterwards
  - CM will share the justification with the SC once it's written
- Jim's role in the Chronopolis plan:
  - Jim works for Harvard (quantitative data)
  - Hired to expand Dataverse storage by the Global Consortium
  - When DPN existed, Jim had been tapped to attach additional storage; he actually created the ability to attach to the backend
  - This project will give him an opportunity to update code to 4.20 version
  - Benefit of getting functionality for TDL consortium, but to contribute code back to Dataverse community

- [Core Trust Seal](#)

*Courtney Mumma, Texas Digital Library*

- Dataverse has announced their annual community meeting agenda; discussion around Core Trust Seal is on the agenda
- Application wasn't ready, then pandemic, now budget issues: CM will attend that meeting and learn more about the original core trust
- Cost is only \$1500, but TDL will face delays with the imposed purchasing restrictions and budget justifications
- Attached storage will not interfere with main Dataverse, but will complicate the application

- [Larger Datasets](#)

*Matt McEniry, Texas Tech University*

- ("TDR Plus")
- Met last Friday (May 22, 2020); answered questions that were needed and will add to documentation in Confluence.
- Each institution will need to create their own policies
- May require addendum to MOU to utilize large data storage
- Hiatus on the pilot until August
- Updating policies and procedures
- Cross-institutional data storage usage: Researchers deciding to store their data based on cost.
  - Could TTU offer storage usage / 50/50 with another institution

- Possible to make that off-limits?
  - Workarounds: Easy enough to call up a friend at another institution to find a cheaper storage solution.
    - How would that function?
    - How could we avoid abusing the larger data storage
    - Could impact faculty approach to grant proposals
    - May need to have an answer in place so prepared for the question(s) that will come up.
- UT Austin / TACC pilot showed the challenge of working with researchers on this
  - Expect long lead time to make it work
  - TX State users want to know that the TDR is there, but they are having trouble getting users to take part
  - UH - different, cultural understanding of what to do with their data; not necessary using the data but it's living in their computational space
  - No one else can find that un-indexed data; even though it's not "dying," it's still not "out there."
- [R Integration Presentation](#)

*Reid Boehm, University of Houston*

- What is the integration, and are we interested in continuing?
  - R client for Dataverse 4 repos
  - Gives access to Dataverse APIs and integrate public data sharing
  - Search by name, files - all pulled up in the R environment and can do computational analysis
  - Example: search for "tree frogs" and perform computational analysis in R
  - Push metadata and data into Dataverse using R
- Reasons to keep R
  - Crosses disciplines
  - Language is popular, highly used, focused on reproducibility (shows dedication to that)
  - Direct channel for access to datasets; can search "from your world" and reduce barrier to entry for some users
  - API might increase functionality for ingest and management
- What's involved?
  - Nothing on the backend
  - No purchasing (R is free)
  - Front end
    - Learning how three actions work

- Testing to assure functionality
- Provide clear documentation for TDR SC and local institutions
  - Can provide to users and tweak language depending on their level of proficiency with R
- Discrepancies in documentation in package
  - Github had newer versions that were different than what was in R
  - Reid has yet to test and rectify inconsistencies; also some gaps in knowledge around APIs and programming languages
  - RB created an outline for the documentation and is starting to fill some things in
    - Use cases, testing functions
- Comments
  - Good candidate for a webinar (Spring 2021?)
  - Call out to Dataverse Community: who is using beyond IQSS?
    - Would lighten the burden of documentation, etc.
    - Been around since at least 2014!
    - Mention at Dataverse community meeting (review agenda)
  - Bring Brenna into discussion / ask for documentation
  - Sense that implementing R would bring in more users?
    - UH need is growing very slowly; if users are working in R often, this would be helpful but it's hard to know who is using it
    - Baylor has statistical consulting center; new director could be an in: Ask users new to R to plug TDR
      - Applied sciences would be likely users over social sciences; maybe business researchers?
      - Add to workflow at the end
  - Good candidate as we focus on projects for the coming year!

**Noon-12:10**

*10 minute Break*

**12:10-12:15**

**TDL Executive Director Remarks**

(5 minutes), *Kristi Park, Texas Digital Library*

**12:15-1**

## [Roadmap](#) and priorities discussion

### Outreach

- Tidying up of documentation complete (Thanks, Matt @ TTU)
- Slide / message for new members hasn't been tested out yet; Dianna @ TX State and Hammad @ UTA will collaborate
  - Mentor pairing: outreach team will look for a TDR SC member with a similar role / compatibility when new member joins
    - Hammad willing to be a part of the project; mentorship would be valuable to get involved on projects and have another person to talk with
    - Onboarding new members becomes a “documentation dump” so having a person to work more closely and interpret info would be helpful
  - Any prospective TDR members?
    - UTSA (will be a while!)
    - No one queued up at the moment:
      - Trinity not interested
      - TAMUCC working on restructuring their staff
      - Will see new members joining from TAMIU and TAMUG (new liaisons not yet hired)
  - Encouraging TDR SC participation locally: TX State is a great example!
    - Main liaison has voting rights, can chair
    - Ancillary members are invited to participate
    - Local liaisons can combine their efforts
    - Does SC agree to allow other local liaisons to join the cmte?
      - Yes and Carpentries work could be a great way to include other liaisons from your institution
  - Welcome message for new users and early-adopters of TDR
    - Develop template
    - TAMU system schools all use single Shib authentication. Makes it appear that TAMUG (for example) datasets are TAMU on the back end.
    - Add to future meeting agenda
- Build on strategies for intra-campus partnerships (e.g. with grant networks, sponsored projects) / Establish mechanism for TDR SC to assist each other with outreach projects and do collective outreach
  - TDL's DDP grant project
  - Brainstorming meeting? Develop messaging around grant funding & open data for faculty (may be depend on institutional protocols around grant proposal process)



- If engagement is nurtured locally, could leverage that at other campuses w/ lower engagement levels
- Annual activities:
  - Present on/promote the repository at a minimum 3 conferences (disciplinary and library) - hit that goal!
  - Include on Annual Report outreach template - separate project

#### Larger Datasets

- Consider goal of implementation of an additional institution (UT Austin is currently the pilot institution for this project)
  - Assumption is that UT will move from pilot to established in the next year
  - Goal for coming year: one complete, one in progress
- Any others aside from UT and A&M?
  - UH - not ready yet
  - Baylor - at least three years away (priority is HIPAA-compliant because it's a newer idea to store health data)
  - UTA had a recent researcher w/ TBs of data; their OIT dept is working on a solution. Not too much interest yet, but interested in offering it. May be able to connect the OIT option in the future?

#### Training

- Note: merge elements of Training section w/ Outreach section
- Create discipline-specific guidance and templates for researchers: the Outreach and Training Group emerged from this work
- Annual activities:
  - Webinar series
    - TDL policy (new): two webinars per month (Monthly Member Forum "counts" as one webinar 😊)
    - Looking to host one to two webinars each semester (Fall, Spring, Summer)
  - in-person event
    - Carpentries workshop
  - Review and update Confluence user documentation, including:
    - DMPTool language - Dianna will make sure this is included in the new member onboarding slide.
    - Policies - "done" but will need to update w/ Digital Preservation once that's ready

## Assessment

- Tools and reporting: nearly done
- Follow Dataverse development towards Making Data Count compliance - committed to turning on all available functionality once it is made available
- Core Trust - done
- Annual Report - done (thanks, Christina @ Baylor!)
- Google Analytics - how to prioritize / make a project

## Accessibility

- Not prioritized in 2020; not sure if we can prioritize in 2021
- May need to wait on this - watching progress in DSpace and it moves slowly, despite everyone knowing it needs to be done.
- Will stay on the roadmap.

## Systems Integrations

- Chronopolis in progress (see above notes re: Jim Meyers)
- R (see above notes from Reid)
- Wait on Vireo and DSpace until upgrades are done and stable.
- Investigate ORCID integration (maybe contact Jon Crossno and Jane Scott at UTSW)
- Request feedback on need for integrations:
  - OJS group
  - DUG

## Future projects

- R
- Software containers

## 1-1:30

30 minute break

## 1:30-2:15

## New projects

(based on roadmap discussion)

- [Code/Software archiving](#)
  - Important project; people tend to use Github so it would be hard to switch.
    - Github doesn't give you a DOI - create dataset w/ DOI to link to Github; hard to see changing workflow from putting code in git.

- TTU is a member of Software Preservation Network (SPN) and could borrow resources from them re: metadata
- Interest at UH, not always archiving code but would like to share the full environment of their work
  - DH person would like to put a docker container in Dataverse
  - Visualization
  - Github -and- ... github isn't a replacement for the archive and it's hard to convince users of that, so git would need to be used in conjunction with another resource
- Information-gathering project before documentation project?
  - Environmental scan: recon different uses by dept / research focus?
  - Moving forward:
    - Name project in June meeting
    - Discuss priorities and see who volunteers at the next meeting
- TBD based on Roadmap discussion
  - Google Analytics - YES: capstone project for Fall 2020
    - Need: Unknowns around who looks, opens a Dataverse; learn about traffic to the site
      - We know what's in TDR, but not necessarily who is using it
      - Are people finding the datasets through Google first?
    - Leverage expertise at libraries: Do we know anyone who could implement this for us?
      - GA already set up in Dataverse, just need help organizing, configuring.
      - Good project for a capstone student
  - Data Curation in the TDR - YES: contingent upon UTL findings
    - Background: Jessica Trelogan attended a conference last year that sparked interest at UTL. From there, she and Brenna Wheeler developed a process for working with the liaison librarians.
      - Now that the report is out, a small group of UTL librarians is meeting to develop an implementation plan. Especially interested in establishing limits of librarian involvement.
    - UTL will share workflows and findings with TDR SC.
      - TDR SC will take and develop into generic recs for TDR)
  - Outreach projects:
    - Mentor pairings: YES - asap
      - Might be tricky with a smaller group? How to manage logistics
      - Short term: Hammad needs a mentor :)

- Targeted outreach: YES - cross-committee project
  - Review current user roster and identify users for targeted outreach
  - Brainstorm other intra-campus opportunities with wiki to store ideas (what worked, what didn't / some concerns there w/ candid responses and capturing delicate info in wiki)
  - Identify opportunities for local teams to participate
    - Need to alter the charter? No, just allowed only one voting member per institution.
    - Talking point for TDL to solicit members to join TDR (esp. Share example of TX State where duties are split - no single data librarian)
    - No requisite staff level - anyone on the staff who is interested / tasked may participate
- Training:
  - Use Christina's model for DMPtool
  - Identify and define metadata schema by focus area. Different metadata for different disciplines: offer discipline-specific trainings?
    - Related to curation work
    - Link to Data Dictionary w/ keywords
  - Parameters for ingest: document workflows for ingest from different TDR members
  - Webinar series:
    - Inward-facing: Share workflows
    - Round tables
      - Invite campus partners outside the libraries
    - Goal of two per semester
- UT Austin and UT iSchool capstone report (Brenna Wheeler and Colleen Lyon)
  - Please send your feedback to Brenna by June 12. You can email her at [brennawheeler@utexas.edu](mailto:brennawheeler@utexas.edu).

**2:15-2:45**

Other Discussion

- [CARE Principles for Indigenous Data Governance](#)
  - Courtney would like to take a unified voice from the TDR to the Dataverse Community w/ recommendation to adopt / add publicly

- Promotion:
  - State endorsement on TDL wiki
  - TDL platforms: forum, blog, newsletter, social media
- Dataverse Annual Meeting Agenda <https://projects.iq.harvard.edu/dcm2020/agenda>
  - June 17, 18, 19
  - Shared note-taking doc (CM will share her notes doc)
- Video for Dataverse Community Meeting? (Christina)
  - [https://harvard.az1.qualtrics.com/jfe/form/SV\\_ezirKHKBqOMXl6R](https://harvard.az1.qualtrics.com/jfe/form/SV_ezirKHKBqOMXl6R)
  - DC is asking different Dataverse teams to send along a video with updates (2 mins max)
  - Roll call and list of accomplishments this year
    - Assessment & Activities reports
    - Logo slide w/ all members
  - CM will send email to SC to gauge interest
  - Video is due (unknown): will shoot to send in June 10

**2:45-3**

### Election

(Mike the TDR Armadillo is still in quarantine at Texas State)



- Incoming chair: Christina!
- Nominee(s): Cannot be UT, TX State, or Tech
  - **Reid Boehm, UH**
- [TDR Steering Committee Charter](#) (for more info)

Next meeting: June 24, 2020

### LD'S SIDEBAR NOTES

- Replicate the reporting template for other services (new DSpace stats?) -- *added to DUG agenda planning meeting notes*
- Google Analytics for TDR: Add GA project to capstone wishlist -- *added to TDL's project wishlist in Confluence*
- Send webinar stats to TDR SC -- *sent*
- Alex - create a Doodle for the TDR SC video -- *done*

- Send Dataverse info to Benson -- *sent*
- Update tdl.org and member group list w/ new chair & vice-chair -- *done*
- Brenna follow up: -- *done*
  - Send thank you email w/
    - Dataverse Community meeting info  
<https://projects.iq.harvard.edu/dcm2020/agenda>
    - Capture BW's slides