

Title: The Data Science Lifecycle Ethos - Infusing Ethics into the Research Process

## Sept Data-Facing Call

Date: 2021-09-07 Time: 10am PST / 11am MST / 12pm CST / 1pm EST

Call coordinators: Deb McCaffrey (UMich Med) & Galen Collier (Rutgers Univ) & Amy Koshoffer (UCincinnati)

## *CaRCC Code of Conduct*

### Announcements:

- About CaRCC
- Code of Conduct
- Steering committee
  - email [df-coordinators@carcc.org](mailto:df-coordinators@carcc.org) if interested

**Title:** The Data Science Lifecycle Ethos - Infusing Ethics into the Research Process

**Presenter:** Steve Van Tuyl

**Description:** As an emerging field, data science, and data scientists, have the opportunity to reframe how we think about ethics in the research process. This presentation presents a set of tools, developed through the Academic Data Science Alliance's Ethics Working Group, to build ethical considerations into each step of the research process, rather than address ethics ex post facto. These tools provide a case-study based framework, with conceptual lenses through which to consider the ethical questions presented at each stage of the research lifecycle. Working group members have been piloting the tools in the classroom and are looking for adopters, additional case studies, and new applications of the tools.

**Biography:** Steve Van Tuyl is the Community Manager for ADSA. Before joining ADSA, he was variously a digital repository librarian, research data librarian, and computer science librarian at Oregon State University and Carnegie Mellon University. In this previous life, he was an active member of the Samvera open source repository community, helping coordinate resources for and development of the Hyrax repository solution bundle. Steve was trained as an ecologist at Colorado College and Oregon State University, and worked for many years with the US Forest Service researching the effects of natural and human disturbances on terrestrial carbon cycles. Steve likes to bake bread and tries to run far.

### Slides:

### Zoom details:

<https://utah.zoom.us/my/carcc?pwd=TjFuR3VVM2d5eE5zWnEvWWxDTFBCUT09>

Meeting ID: 824 051 8198

Password: 31415926

**CaRCC YouTube Channel** (today's call will be added soon):

[https://www.youtube.com/channel/UCMU1PEMM7V4X\\_KtPDcfB6HA](https://www.youtube.com/channel/UCMU1PEMM7V4X_KtPDcfB6HA)

**NOTE:** This document is publicly accessible. The link to the Google Drive folder containing this document is posted on the Track Webpage. Please keep that in mind when entering information.

## Sign-in (Name / Affiliation / E-mail):

1. Deb McCaffrey / UMich Med School / [debmccaf@med.umich.edu](mailto:debmccaf@med.umich.edu)
2. Amy Koshoffer / UC / [koshofae@ucmail.uc.edu](mailto:koshofae@ucmail.uc.edu)
3. Mathew Vis-Dunbar / University of British Columbia / [mathew.vis-dunbar@ubc.ca](mailto:mathew.vis-dunbar@ubc.ca)
4. Steve Van Tuyl / ADSA / [steve@academicdatascience.org](mailto:steve@academicdatascience.org)
5. Laura Harris / ICER MSU / [oesterei@msu.edu](mailto:oesterei@msu.edu)
6. Chris Reidy / UArizona / [chrisreidy@arizona.edu](mailto:chrisreidy@arizona.edu)
7. Cyd Burrows-Schilling / University of California San Diego / [cburrows@ucsd.edu](mailto:cburrows@ucsd.edu)
8. Jenny Muilenburg / U of Washington / [jmuil@uw.edu](mailto:jmuil@uw.edu)
9. Sue Oldenburg / Rutgers / [sue.oldenburg@rutgers.edu](mailto:sue.oldenburg@rutgers.edu)
10. Don Brower / U. Notre Dame / [dbrower@nd.edu](mailto:dbrower@nd.edu)
11. Carrie Iwema / University of Pittsburgh / [iwema@pitt.edu](mailto:iwema@pitt.edu)
12. Sajesh Singh / American Museum of Natural History / [ssingh@amnh.org](mailto:ssingh@amnh.org)
13. Melissa Cragin / SDSC (UCSD) / [mcragin@sdsc.edu](mailto:mcragin@sdsc.edu)
14. Bala Desinghu/Rutgers University/bala.desinghu@rutgers.edu
15. Holly Mercer / University of Tennessee / [hollymercer@utk.edu](mailto:hollymercer@utk.edu)
16. Susan Ivey / NC State University / [slivey@ncsu.edu](mailto:slivey@ncsu.edu)
17. Anna Klimaszewski-Patterson / California State University, Sacramento / [anna.kp@csus.edu](mailto:anna.kp@csus.edu)
18. Claudia Engel / Stanford University [cengel@stanford.edu](mailto:cengel@stanford.edu)
19. Claire Mizumoto / University of California San Diego / [claire@ucsd.edu](mailto:claire@ucsd.edu)
20. Trina Marmarelli / Reed College / [marmaret@reed.edu](mailto:marmaret@reed.edu)
21. Robin Donatello / Cal State Univ Chico / [rdonatello@csuchico.edu](mailto:rdonatello@csuchico.edu)
22. Jayshree Sarma / George Mason Univ / [jsarma@gmu.edu](mailto:jsarma@gmu.edu)
23. Melissa Ratajeski/ U of Pgh/ [mar@pitt.edu](mailto:mar@pitt.edu)
24. Bryan Raney / Rutgers University / [bryan.raney@rutgers.edu](mailto:bryan.raney@rutgers.edu)
25. Paula Lackie / Carleton College / [plackie@carleton.edu](mailto:plackie@carleton.edu)

**Max join: 34**

Session Notes

Link to ADSA website - <https://academicdatascience.org/>

Developing a tool to infuse ethics throughout the research lifecycles.

- Conceptual lenses
  - Positionality
  - Sociotechnical Systems
  - Power
  - Narrative
- Case studies
  - Introductory - Cell phone records
  - Complex - property tax assessment project
  - Complex - clean drinking water

Where is ADSA now

- Working with UC Berkeley
- Piloting the tool
- Tested tool in a workshop

Seeking funding

Contribution framework for case studies

Curriculum

Future adopters

## Questions

Q: Who at the institution gets invited to the leadership summit?

A: ADSA attempts to identify who is a head of institute at an organization and invite them. It can be tricky to identify who is the lead given that data science is something shared across many disciplines.

Q: Data Science vs Data Management - where is the pain point in the cycle - is the data good and then how does it impact analysis

A: Challenge is how to have case studies to cover the spectrum of problems that could occur. So far cases focus on implementation. Try to talk about data management issues, but not always easy to separate them out.

Q: Are you working with IRB to develop criteria.

A: Would want to think more broadly. IRB are already strapped to reach beyond and this project wants to cover those areas not reached by IRB.

Q: What about the missing data sets

A: Library of missing data sets: <https://mimionuoha.com/the-library-of-missing-datasets>

Q: Who should we point data science faculty to?

A: Check the website. And more information coming