**SRP Trainees/SPAN Leadership Committee/Training Core Leaders Meeting**

**Tuesday, November 2, 2-3:30 pm ET**

**<u>Attendees:</u>**

Ahlam Abuawad (Columbia)

Neel Aluru (BU)

Liana Baconguis (LSU)

Candice Brinkmeyer-Langford (TAMU)

Gabriel Colón (University of Puerto Rico)

Mona Dai (Harvard)

Subham Dasgupta (OSU)

Heidi Dowst (Baylor)

Tammy Dugas (LSU)

Jerika Durham (UKY)

Rocio Estrella (Arizona)

Stephanie Foster (OSU)

Adrian Green (NCSU)

Chuqi Guo (NCSU)

Kristin Hicks (Northeastern)

Jen Irving (LSU)

Susan Korrick (Harvard)

Priyanka Kushwaha (Arizona)

Gabriele Ludewig (Iowa)

Marina Malovichko (Louisville)

Craig Marcus (OSU)

Pamela Mellon (UC San Diego)

Ian Moran (OSU)

Rance Nault (MSU)

Obinna Nwokonkwo (Harvard)

David Purdy (Iowa)

Holly Rudel (Yale)

Krystal Taylor (ECU)

Dylan Wallis (NCSU)

Brian Westra (Iowa)

Jamie Young (Louisville)

J. Zach Hilt (UKY)

Martin Zanaj (OSU)

Luoping Zhang (UC Berkeley)

Ruikun Zuo (LSU)

Sara Amolegbe (NIEHS)

Danielle Carlin (NIEHS)

Brittany Trottier (NIEHS)

Natalie Rodriguez (MDB)

Mali Velasco (MDB)

**Agenda Items:**

**Training in Data Science**

*Dr. Brian Westra, University of Iowa*

- Defining data:
  - Factual material commonly accepted in scientific community.
  - For data sharing and preservation, data does NOT include the things that are in a paper (graphs, visuals, tables, etc.)
- Benefits of good data management practices:
  - Transparency, reproducibility & replicability.
  - Reduce wasted time and effort.
  - Protecting researcher integrity & reputations.
    - Example: *Nature* paper retracted due to mistake in calculations.
    - With good data management, it would be easy to recover from the retraction by finding where the issue was and addressing it. Sloppy data makes this difficult.
- Data lifecycle:
  - Plan → Create → Store → Share
- Enable consistency:
  - Document practices for your future self, your PI, team, etc.
  - [Data Management Guidelines by Mike Harms](#) at University of Oregon.
- Managing data during research
  - During the project:
    - Use preferably non-proprietary tools so that others can work with your data.
    - Be consistent, follow best practices, document your work.
  - Data sharing and publication:
    - May require Data Use Agreements, which Division of Sponsored Programs or Human Subjects Office can assist with.
  - Organizing data:
    - File naming:
      - Should be descriptive and consistent across research team.
      - Track versions of files so you can revert to earlier version if needed.
    - Files and folders:
      - Can be organized by project, researcher, date.
      - Method of organization should be clear in case project is handed off to someone else.
- Documenting data
  - Common options:
    - Lab notebook: can be transcribed to electronic version or scanned to create an electronic backup.
    - Create data dictionary/codebook from the beginning:
      - Spreadsheet that describes each variable by name, description, units.
    - Comments in script/code to narrate research/data collection process.
    - README.txt files:

- Include description of folder hierarchy/file naming, when/where/how data was collected, any notes.
- Data structure
  - Structuring data in spreadsheets:
    - Put variables in columns and observations in rows.
    - Always keep copy of original raw data.
  - Resources:
    - Data organization in spreadsheets
    - Best practices for preparing environmental data sets
    - Tidy data
- Sharing and publishing data
  - Pew research survey found that open access to data and independent review inspire more trust in research by the public.
  - Repositories:
    - Systems where data is described, preserved, and made accessible.
    - Keeps content (data) and context (metadata) safe for future use.
    - Use Disciplinary Data Repositories specific to your research area:
      - NIH Repositories.
      - If domain-specific repository does not exist, use Iowa Research Online.
  - Include data availability statement in text of article and cite in references
- More resources:
  - DataONE
  - ESIP
  - Data Carpentry
  - COPDESS
  - Research data management (RDM) open training materials
  - Data Information Literacy project

- Q&A
  - Liana Baconguis: Do you recommend sharing all raw calculation data (except things that involve protected information), or are there certain types that you think should take precedence?
    - Response:
      - That's a challenging one. There are different schools of thought, and I think it depends on what the accepted practice is in your domain. But having that record preserved in some way can save you in terms of reproducibility or so that you have something to fall back on. There are two ways to look at it:
        1. What's going to provide that other researcher the capacity to reproduce what you're doing?
        2. What would provide the capacity for somebody to reuse and build on your data?

- For example, if you use interview data, you can't share the original recording because that would include identifying information, but you could possibly share a transcript.
  - o Martin Zanaj: How does one go about creating useful metadata (philosophy, and what are some implementation tools)?
    - Response:
      - Metadata is one of the most challenging things because it can make a huge difference, but it takes up a lot of the front work.
      - I can point to a good blog post from the Harms lab about always creating README files in chemistry research to keep track of what they did, why, and what the machine settings were.
      - In mass spectrometry, there's a lot of information about machine settings that are embedded in the files themselves. Need to be aware of what type of data is embedded in the file
      - As much as possible, you should be using vocabulary that are established/standardized for your field.
      - Recommend contacting a data librarian as well.
  - o Susan Korrick: For very old datasets that are still actively used but are not organized optimally, are there resources you would recommend for improving the dataset organization/structure/naming conventions, etc.? It seems like a huge up-front effort...
    - Response:
      - Some tools can be used for reviewing and "cleaning" data, such as OpenRefine. It works well but for tabular data that is generally well organized. But you can use it to split columns (if, say, a column has city and state).
      - If there is some kind of pattern that a program could recognize, you might be able to use scripts to identify and extract the pieces you want. Python, R and other programs, including OpenRefine, can use Regular Expressions (RegEx) to identify a string with a particular pattern, and then you could have extract that information. There are different "flavors" of RegEx, depending on the software you are using.
      - Unfortunately, if the structure is extremely variable, it might require a lot of manual work.
      - The "return on investment" may mean that you decide it's worth the work to get the information into a better structure. Or not!

**K.C. Donnelly Externship**

*Brittany Trottier, MPH, SRP Health Specialist*

- What is the K.C. Donnelly Externship?
  - o Meant for SRP trainees (graduate students/postdoctoral researchers).
  - o All about new skill development that can be applied to trainee research.

- Example: if there's a skill that a trainee wants to learn but it's not available to them at their P42, but another center has an expert in it on their team, the trainee can do an externship with them.
    - o Includes support for up to 3 months (supplies, travel, housing) but cannot exceed $10,000.
        - Travel funds should include attending the following SRP Annual Meeting.
- Eligibility
    - o Graduate student or postdoctoral researcher.
    - o Conducting research in any SRP R01 grant or project/core in SRP center.
    - o Externship must be with another current SRP-funded grantee, government lab or agency.
    - o Centers under a no-cost extension are not allowed to apply.
- Application Information
    - o Letter of intent due **January 14, 2022.**
    - o Applications accepted **February 1-28, 2022.**
    - o P42 Centers limited to 2 applications per year, R01s limited to 1.
    - o Applications evaluated via review panel.
    - o 8-10 applications expected to be awarded each year.
    - o More information on the application process after the Notice of Special Interest (NOSI) comes out!
- Helpful hints
    - o SF 424 Application
        - Question 4b: NOSI notice number goes here.
        - Question 8: Type of application = revision.
        - Question 11: KC Donnelly Externship – Trainee's name.
        - Question 14: Center director information (not applicant).
    - o Research strategy section
        - Summary of current research:
            - Write this part as if a scientist that is not an expert in your field is going to read it .
        - **NEW PART** Include a description of how KCD project fits within scope of your Center's parent grant.
        - Be clear and concise about the mentoring plan.
- Helpful resources:
    - o [K.C. Donnelly website](#)
    - o [K.C. Donnelly guidelines](#)
- For questions about KCD application:
    - o Brittany Trottier, MPH, [Brittany.trottier@nih.gov](mailto:Brittany.trottier@nih.gov) and your Program Officer:
        - Danielle Carlin, PhD, [Danielle.carlin@nih.gov](mailto:Danielle.carlin@nih.gov)
        - Heather F. Henry, PhD, [henryh@niehs.nih.gov](mailto:henryh@niehs.nih.gov)
        - Michelle Heacock, PhD, [heacockm@niehs.nih.gov](mailto:heacockm@niehs.nih.gov)
- Q&A

- o Jamie: Last year I had asked about the possibility of the other funding group being a nonprofit or private laboratory that's not a government funded group but may have similar interest. Is there a possibility of them being included in the future?
- o Response:
    - ▪ Brittany: It is up for discussion, currently is not available for this round of application.
    - ▪ Danielle: One of our goals is to increase collaboration among the centers, so that's the reason why they're not currently included.

**SRP Annual Meeting**

- Save the date! **December 15-17, 2021,** Virtual events
- Will feature opening remarks for NIEHS leadership, past K.C. Donnelly winners, and new 2021 Wetterhahn awardee.
    - o Past KC Donnelly winners who will be presenting:
        - ▪ Jessica Ewald
        - ▪ Katlyn McGraw
        - ▪ Paige Varner
        - ▪ Yvonne Rericha
        - ▪ Anna Robuck
        - ▪ Ahlam Abuawad
        - ▪ Maya Spaur

**Other Agenda Items:**

- **Are your Data Collection Form Entries Up to Date? CareerTrac?**
    - o Trainee awards, publications, and honors are submitted through Data Collection Form
    - o Send us photos!
    - o Helps with Annual Update process
    - o [Data collection tool submission form](#)
    - o Keep [CareerTrac](#) updated with trainee information
- [Karen Wetterhahn Memorial Award](#)
    - o Winner will be announced at SRP Annual Meeting
- **Reminder to send Updates to SPAN Leadership Committee**
    - o Taken September/January of each year
    - o Email nominations to [Danielle.carlin@nih.gov](mailto:Danielle.carlin@nih.gov) and indicate current status of new member
    - o Membership on SPAN Leadership Committee is a 2 year term
- [SRP Website for SPAN/Training](#)
    - o [Notes from previous SPAN call discussions](#)
- **SRP e-Posted: Trainee Highlights**
    - o Send to [Danielle.carlin@nih.gov](mailto:Danielle.carlin@nih.gov) and [Natalie.rodriguez@nih.gov](mailto:Natalie.rodriguez@nih.gov) any news items of interest to trainees, such as:
        - ▪ Trainee success stories
        - ▪ Pictures and videos
        - ▪ Job announcements

- [LinkedIn](#)
    - Job announcements, research briefs, SRP news, and more!
    - For SRP Trainees only

**See you December 16th at the virtual SRP Annual Meeting!**