Conducting Data Interviews

Pathways to Data Services, session 2

October 6, 2025

Nina Exner, PhD, MLS, AHIP. Research data librarian, Virginia Commonwealth University
nexner@vcu.edu

Your presenter

Nina Exner, nexner@vcu.edu

- Research Data Librarian at Virginia Commonwealth University (VCU)
- Supporting both HSL patrons and the nonmedical campus' patrons
- My data *management and sharing* consults are most frequently for medical and health sciences
- My data *discovery* consults are most frequently for the social professions (education, social work, etc)
- VCU is…
  - R1 urban institution, ARL library, ~30,000 students

What is a data interview?

Any reference or consultation interview related to data. Common ones include:

- Pedagogical data interview - dataset search needed for class assignments
- Secondary research data reference interview - dataset search for quantitative analysis where the researcher isn’t also learning methods and terminology
- Data curation intake interview - beginning interview for preparing data for good-practice ingestion into a repository
- Data management and sharing plan interview - consult specific to advising a researcher on how proposed research can best be managed and shared in compliance with grant proposal requirements

Data management and sharing plan (DMSP) interview

I've included some great readings on different data interviews at the end. Today, we'll focus on the DMSP interview. It's …

- Specific, compared to other data interviews
- Based in documents you can cram or spot-check fairly easily
- A good starting place for getting used to talking to researchers
- Likely to come up in any research data librarian job

Goal: Identify key information for a DMSP

1. Data type
    a. Types and amounts of data generated
    b. Which of those data will be preserved and shared
    c. Metadata and associated documentation
2. Related tools, software, and code
3. Standards
4. Data preservation, access, and associated timelines
    a. Repository where the data and metadata will be archived
    b. How the data will be findable and identifiable (e.g. a persistent identifier)
    c. When and for how long the data will be available
5. Access, distribution, or reuse considerations
    a. Factors requiring restrictions to data sharing (ethical, technological, or legal/contractual)
    b. Controlled access, if any
    c. Protections for human research participants, if any
6. Oversight

How hard can a dozen questions be?

Surprisingly hard! If the NIH could just send a questionnaire with those questions, they would

Many researchers crib from colleagues (with varying success) because they don’t really know what the DMSP wants

Elements are from the “Optional format page”

Visit **https://tinyurl.com/dmsp-blank** if you would like to download a copy of the official optional DMSP format page, in Word format

- Not required for these activities, but might add context
- Use the Optional Format Page or else DMPTool for your future consults

Simplified names of the NIH DMSP elements

1. Data and metadata types
2. Code and software
3. Standards
4. Repositories and Sharing
5. Considerations and Restrictions
6. Oversight

Side note: Not every DMSP is an NIH DMSP

The DMPTool is a great place to find lots of DMSP templates in one place.

Even if it's not for the NIH, any DMP or DMSP will need similar basics:

- What kind of data will be shared, and in what formats
- What repository the sharing will happen in
  - For federally mandated data sharing, the Desirable Characteristics are usually required
- How are sharing and ethical restrictions being balanced, if applicable

The FAIR (Findable, Accessible, Interoperable, Reusable) principles are always worth referencing, especially the easy ones like persistent identifier

Common points of confusion to patrons

Sharing

In this context, data sharing means sharing for the long term after a research project is over. But researchers aren't archivists, and usually default to thinking about sharing as how the team shares files during the project!

Repository

In the DMSP, a repository is a digital preservation and archiving platform. A repository must follow both NIH guidance and NSTC guidance on data repositories. But in casual scientific talk, a repository can mean anything!

Important tools: the BMIC Repositories page and the NNLM Finder

Jargon watch 1, some general terms

Using grant and data jargon can help communicate clearly to patrons, in words that they expect. Here are some that might come up in a data interview

- IRB, IACUC, IBC: some of the ethics oversight units on campus
  - IACUC - animal ethics committee
  - IBC - biosafety and biological ethics committee
  - IRB - human ethics committee
- Institute or IC: One among the 27 National Institutes of Health (NLM, NCI, etc)
- PI or PIs: Principle Investigator(s), researchers that are leading the grant team
- R01: the NIH grant code for an independent research project
  - R01 is the most common NIH grant type
  - You can look up other grant codes here. Sometimes patrons start with “I need help with a DMSP for a P50” or “I’m submitting a K99/R00”
- Terms and Conditions (of a grant): Things the PIs have to do and are usually required to report about. The DMSP as finalized during JIT becomes a “Term and Condition” of an NIH grant

Jargon watch 2, the grant process

- Stages of a grant
  - Preaward
    - Proposal writing or grantwriting: Two ways of talking about writing the grant proposal
    - Proposal submission: Sending it to your central office for approval. Then your central office sends it to the NIH to propose the grant.
    - JIT or Just In Time: Stage when promising proposals get the option to submit documents or document revisions that are required but weren't in the proposal (or weren't correct in the proposal). DMSP revisions happen during JIT
  - Post-award
    - RPPR: Research Performance Progress Report
    - Closeout: formal end of a grant, including a final RPPR

Talking to researchers

Researchers are incredibly smart people with a very specific focus and serious time pressures. Most of them are better at talking about their research than about their data.

An important skill is picking “what’s data about it” out of the overall discussion on research. This can take a lot of practice, just like other reference interviews!

I’ve included a list of my questions at the end of this slidedeck (slides 25-30), to give you a feel for my approach.

In a real consult, try to look them up!

To make this example more relatable, I didn't read into Dr. Froemke's research. But in a real consult, I would if I had time!

Activity: Take a look at https://med.nyu.edu/faculty/robert-c-froemke

Chat question 1 - what sections of his landing page might be informative? Skim it for 1 minute and share advice to other attendees where we can do closer reading

Chat question 2 .... Next slide (10+ minutes: Depending on time it might be before or after mock consult)

Optional discussion: Cramming about researchers

Pick part of Dr. Froemke's webpage (or lab page) to explore

- Read or explore for 5 minutes
- Add to the chat when you see something that gives you ideas about…
  - The kinds of data they might work with
  - Ethical restrictions that should be especially asked about
  - Whether they're interested in open science practices

Pre-consult preparation discussion

What did we find?

Experienced liaisons, what else do y’all do when prepping for a faculty research consult?

The interview itself

I tried to sketch out my questions so that you have them to refer to.

Unfortunately real reference interviews don’t flow predictably, but this might give you a feel for it. I’ve organized them to follow the DMSP elements, but you don’t have to.

Please take notes during the interview

You can just listen if you want! No grades here. But if you're willing, I'd ask you to:

1. Get a notepad and pen
2. Write down the elements of the DMSP
    a. Data and metadata types
    b. Software and code
    c. Standards
    d. Repository sharing
    e. Considerations for restricting access
3. Listen for parts of the consult that speak to the DMSP and make notes

Bonus: Please jot down terms that are confusing, so that you can ask me about them afterwards.

Interview welcome

Many thanks to Dr. Froemke for helping us!

Post interview discussion

Jargon, then plan elements

Now let’s share what we heard!

First, please start entering in the chat any terms and jargon questions that came up.

- Preferably jargon about data or grants, not the science itself. But if you’re not sure, put it in there!

Writing tasks: about 5 minutes

1. Go to https://tinyurl.com/exner-2025-10
2. Add items you heard in the interview to the element you think they fit in
    - Or you can upvote other additions
    - There are 6 columns - scroll/swipe sideways for more
    - If you have questions you can either put them in the “Other/Not sure” column or add them to chat here

While you do that, I will read the jargon and mystery terms you posted in chat and try to address them.

Let’s see what happens

Like with all reference interviews and consults, there’s a stage where we just have no idea what will happen.

This is that stage! Time to look at the Padlet of what y’all heard and discuss any questions that came up

Thanks for your engagement!

Feel free to contact me later at nexner@vcu.edu too

Continue for a list of my DMSP interview questions

Next 6 slides: a set of DMSP questions

- These are just my best idea of how I approach a DMSP interview
- Adjust to the researcher and your own style
- These are organized according to the DMSP elements, so that you can follow if you have a DMSP blank in front of you
  - In real life I find it easier to
    - start with Data and metadata(element 1)
    - then move to either Repositories and Sharing (element 4) or Considerations for Restricting(element 5)
  - I tend to do Standards (element 3) last because so often that section will focus on using the repository's structure to drive the standards

Questions kit pg 1: startup and Element 1

1. Can you tell me about your research?
   - Optional: I don't know your research topic, can you unpack that part a little?
   - Did I hear you mention an instrument in there? I think I heard… (exs: array, sequencing, MRI)
2. And which IC or Institute do you mostly work with?
3. Tell me about your data files. What does your research look like on your computer, what types of file formats and related stuff do you work with?
   - Optional followup possibilities
     i. When you first hit "save" do you pick a file type? What file type?
     ii. Are there different clinical versus research formats in your field? Which do you use?
     iii. Do you ever convert it between file formats? Like the way you might save a spreadsheet to Excel or to another stats program. I know this isn't just a spreadsheet, but do you have similar conversions?
     iv. If you hand it off to someone else's computer like a new rotator in the lab, or if you imagined handing it off, is there certain software they'd need to read the files? Or would you need to export the data to a different software or format for them to be able to see it on a different computer?

Questions kit pg 2: elements 1 (cont.) and 2

4. Are you collaborating with other campuses? How do you share files?
5. What kinds of documentation do you keep, and out of that, what do you feel comfortable to share?
6. Is there any other metadata or descriptive and support files that you work with? Either from instruments, or lab documents like protocols or codebooks.
7. And then code and software. Do your team program in R or python, and make code or scripts for analysis?
    - (if yes) Do you plan to share that, or is there proprietary content there?
8. Do you use other specialized software? It could be like AI image analysis tools, or patient activity apps analytics, or really any software that could be important to understanding where your data and analysis come from.

Questions kit pg 3: element 3

9. If you've read the DMSP rules at all, you probably know that it asks about Standards. Standards can really be all sorts of things. Are there any common standards you can think of that you use?
    - Optional: Tell it to me like I'm a brand new grad student rotating in please! Some standards are really complex, but others are so basic to a field that we stop thinking about them.
10. Do you use any Common Data Elements or CDEs?
11. What about Minimum Information or Minimum Data Set rules, are there things like that baked into your protocols to make it easier to compare data?
    - Optional: Or have you ever had to change data to make it easier to harmonize uploads to a data platform?
    - Genomics only: GEO's MINSEQE rules are an example, if that helps you think of any others

Questions kit pg. 4: element 4

12. Is there a certain archive or platform that you use to preserve and share data. Personally or just generally in your field?
    - Optional if genetics are involved: I'm sure you use GEO or dbGaP or both. Right? Is there anything you can tell me about how you link that to your other data?
13. Is there anywhere where you download shared data to incorporate into your research?
    - If yes: Could that be a place to preserve and share your data? Why or why not
    - Optional if genetics are involved: Anywhere other than GEO and dbGaP? Where do you put nongenomic data, just there in dbGaP or elsewhere?
14. What about journals, do any journals that you use favor a certain data sharing place? Some common ones are Dryad, Figshare, or Zenodo, do those sound familiar?
    - If yes: Are there any of them you especially like? / Was that a place you'd be comfortable using again?

Questions kit pg. 5: element 5

15. How about restrictions, what can you tell me about reasons you'd want to reduce or restrict post-closeout sharing of your data?
    - What are you thinking of to respond to those restrictions? Restricted access to the shared data, for example. Or deidentification or aggregation of the data.
16. Do you have to get special permissions from any committees on campus, like the IRB, IACUC, IBC, radiation safety, and so on?
    - Does the protocol that was approved have restrictions? For example, does the consent form say that data will only be shared with the research team?
17. Is there patent potential or tech transfer plans with yoru studies?
18. Is any of this under a data use agreement, materials transfer agreement for antibodies or technologies, anything like that your team signed off on?
19. Are you working with big data? Do you have to load to a high performance computing server for any of your analyses?

Questions kit pg 6: Element 6 and closing

20. Do you have your own data manager? Or share up to a data coordinating center that tracks final compliance?
    a. I guess the buck stops with you then! My sympathies
21. That's the end of my zillion questions. Is there anything you can think of that I've missed, that came up on past DMSPs?
22. Thank you so much for talking to me about your research!

Notes

- Element 6 may have a certain way your central office of research wants it phrased
- It's good to have a campus favorite generalist repository (Figshare, Zenodo, etc. Mine is OSF) as a go-to solution during Element 4 for parts that don't fit anywhere else

Resources

Note: general good reads on data, mostly outside of medical librarianship, are JeSLIB and IASSIST Quarterly

Carlson, J. (n.d.). Data Curation Profiles Toolkit Interviewer's Manual and Worksheet. Retrieved September 30, 2025, from https://docs.lib.purdue.edu/dcptoolkit/

Carlson, J. R. (2012). Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about their Data. *Reference Services Review, 40*(1), 7–23. https://docs.lib.purdue.edu/lib_research/153/

Kubas, A., & McBurney, J. (2019). Frustrations and roadblocks in data reference librarianship. *IASSIST Quarterly, 43*(1), Article 1. https://doi.org/10.29173/iq939

MacDonald, G. (2017). Top tips for a data reference interview. CILIP Blog. https://www.cilip.org.uk/news/482813/Top-tips-for-a-data-reference-interview.htm

Partlo, K. (2010). The Pedagogical Data Reference Interview. *IASSIST Quarterly, 33*(4), Article 4. https://doi.org/10.29173/iq884

Read, K. B., Surkis, A., Larson, C., McCrillis, A., Graff, A., Nicholson, J., & Xu, J. (2015). Starting the data conversation: Informing data services at an academic health sciences library. *Journal of the Medical Library Association : JMLA, 103*(3), 131–135. https://doi.org/10.3163/1536-5050.103.3.005