Form
Approved
OMB No. 0990-
Approved for use through XX/XX/20XX

Data Management Standard Operating Procedures Survey

Introduction- In this data management survey, data management is defined as "the process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the scientific data to its users." The research lifecycle is the process that a researcher takes to complete a project or study from its inception to its completion. Data management is involved in every step of the research process. These terms may appear in the survey items.

Q1. During the last 5 years, which of the following roles did you hold on a research project you were involved with where there were data management, data processing, data storage, data sharing, or data analysis tasks? (check all that apply)

- ☐ Principal investigator
- ☐ CO-investigator
- ☐ Collaborator
- ☐ Consultant
- ☐ Postdoc
- ☐ Doctoral student
- ☐ Master's student
- ☐ Project manager
- ☐ Other (please specify): _____

Q2. During the last 5 years, was any of your research activity supported by the following agencies? (check all that apply)

- [ ] ACF (ADMINISTRATION FOR CHILDREN & FAMILIES)
- [ ] ACL (ADMINISTRATION FOR COMMUNITY LIVING)
- [ ] AHRQ (AGENCY FOR HEALTHCARE RESEARCH & QUALITY'S)

- [ ] ASA (ASSISTANT SECRETARY TO ADMINISTRATION)
- [ ] ASFR (ASSISTANT SECRETARY FOR FINANCIAL RESOURCES)
- [ ] ASH (ASSISTANT SECRETARY FOR HEALTH)
- [ ] ASPA (ASSISTANT SECRETARY FOR PUBLIC AFFAIRS)
- [ ] ASPE (ASSISTANT SECRETARY FOR PLANNING & EVALUATION)
- [ ] ASPR (ASSISTANT SECRETARY FOR PREPAREDNESS & RESPONSE)
- [ ] ASL (ASSISTANT SECRETARY FOR LEGISLATION)
- [ ] ATSDR (AGENCY FOR TOXIC SUBSTANCES AND DISEASE REGISTRY)
- [ ] CDC (CENTERS FOR DISEASE CONTROL & PREVENTION)
- [ ] CFBNP (CENTER FOR FAITH-BASED & NEIGHBORHOOD PARTNERSHIPS)
- [ ] CMS (CENTERS FOR MEDICARE & MEDICAID SERVICES)
- [ ] CTO (OFFICE OF THE CHIEF TECHNOLOGY OFFICE)
- [ ] DAB (DEPARTMENT APPEALS BOARD)
- [ ] FDA (FOOD & DRUG ADMINISTRATION)
- [ ] HHS (U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES)
- [ ] HRSA (HEALTH RESOURCES & SERVICES ADMINISTRATION'S)
- [ ] IEA (OFFICE OF INTERGOVERNMENTAL & EXTERNAL AFFAIRS)
- [ ] IHS (INDIAN HEALTH SERVICE)
- [ ] IOS (IMMEDIATE OFFICE OF THE SECRETARY)
- [ ] NIH (NATIONAL INSTITUTE OF HEALTH)
- [ ] OCR (OFFICE FOR CIVIL RIGHTS)
- [ ] OGA (OFFICE OF GLOBAL AFFAIRS)
- [ ] OGC (OFFICE OF GENERAL COUNSEL)
- [ ] OIG (OFFICE OF INSPECTOR GENERAL'S)
- [ ] OMHA (OFFICE OF MEDICARE HEARINGS & APPEALS)
- [ ] ONC (OFFICE OF THE NATIONAL COORDINATOR FOR HEALTH INFORMATION TECHNOLOGY)
- [ ] ONS (OFFICE OF NATIONAL SECURITY)
- [ ] SAMHSA (SUBSTANCE ABUSE & MENTAL HEALTH SERVICES ADMINISTRATION)
- [ ] Other (please specify):
- [ ] All of the above

Q3. Were you or your lab involved in the following activities? (check all that apply)

- [ ] Data collection
- [ ] Data processing
- [ ] Data storage
- [ ] Data sharing
- [ ] Data documentation
- [ ] Data reporting
- [ ] Other (please specify):
- [ ] All of the above

Q4. Do you have a data management standard operating procedures (SOPs) for the project? (FYI, data management SOPs usually describe how, at a practical level, research data from the time of acquisition are: handled, stored, retained, and shared within a research group, with collaborators, and across the broader scientific community.) If "Yes" please check the following areas where data management SOPs are adopted (check all that apply)

- [ ] Yes, if "Yes" please check the following areas where data management SOPs are adopted (check all that apply)
- [ ] Data collection
- [ ] Data processing
- [ ] Data storage
- [ ] Data sharing
- [ ] Data reporting
- [ ] No
- [ ] Not sure what data management SOP is
- [ ] Other (please specify):

Q5. How often do you or your lab develop data management SOPs for your research projects?

- ○ Always
- ○ Often

- ○ Sometimes
- ○ Rare
- ○ Never
- ○ Does not apply
- ○ Often adopt an established one and revise

Q6. What are your or your lab's reasons for having data management SOPs? (check all that apply)

- ☐ Required by funding agency
- ☐ Required by institution
- ☐ Required by supervisor/research group leader
- ☐ Good research practice
- ☐ Other (please specify):
- ☐ Often adopt an established one and revise

Q7. What are your or your lab's reasons for not having data management SOPs? (check all that apply)

- ☐ Not required by project funder
- ☐ Lack of knowledge or experience with creating a data management plan
- ☐ Unaware of any tools or guidance that could help with creating data management plan
- ☐ Not required/appropriate to field of research
- ☐ Time and effort required
- ☐ Lack of training/expertise within research group
- ☐ Lack of local support/guidance
- ☐ Absence of institutional data management policy
- ☐ Do not know
- ☐ Other (please specify):

Q8. Do you or your lab often develop metadata for your research project (if rare or never, skip the next question)? (FYI - Metadata is simply data about data, such as the

description and context of the data. It helps to organize, find, and understand data).

- ○ Always
- ○ Often
- ○ Sometimes
- ○ Rare
- ○ Never
- ○ Do not know what metadata is
- ○ Does not apply

Q9. Is your or your lab metadata usually auto-generated with a software tool or self-developed?

- ○ Auto-generated
- ○ Self-developed
- ○ Not sure
- ○ Does not apply

Q10. Do you or your lab usually develop a record-keeping practice for the following lifecycles of your project? (FYI - Record keeping is the practice of tracking information, including needed documentation, in a systematic way.) (check all that apply)

- ☐ Organizing research projects
- ☐ Planning experiments
- ☐ Recording data
- ☐ Analyzing the results
- ☐ Storing these records for future reference
- ☐ Other (please specify): _____
- ☐ None of the above (note of select "None of the above", all check boxes should be unchecked)

Q11. Which of the following would help you or your lab sustain a record-keeping practice? (check all that apply)

- ☐ Requirement by funding agency

- [ ] Get an online training course (such as a seminar, webinar, etc.)
- [ ] Get a document (publications, white papers, etc.) on record-keeping best practices
- [ ] Buy a book on record-keeping best practices
- [ ] Acquire a software tool that provides record-keeping best practices
- [ ] Other (please specify):

Q12. What kinds of data do the research projects that you or your lab are involved with generate?

- ○ Experimental, e.g., captured in the laboratory, including but not limited to gene sequences, chromatograms, etc.
- ○ Observational, e.g., sensor readings, sensory observations, etc.
- ○ Computational/Simulation, e.g., computer-generated from test models, disease spread production model, etc.
- ○ Derived, e.g., text and data mining, compiled databases, etc.
- ○ Other (please specify):
- ○ Does not apply

Q13. What practices do you or your lab follow in managing your organization of data? (check all that apply)

- [ ] Create a data dictionary to describe data
- [ ] Use data validation to avoid data entry errors
- [ ] Chose self-explanatory file names so that the data files can easily be found
- [ ] File folders are logically organized (such as chronological organization at each level)
- [ ] Maintain data organization with a lab notebook with detailed information on data
- [ ] Maintain data organization with an online lab notebook with detailed information on data
- [ ] Create a “Readme” file to store every command line used for data collection and organization
- [ ] Use a version of control software (such as Subversion) to store all versions of a given collection of related files.
- [ ] Other (please specify):
- [ ] None of these

Q14. Which types of digital files/records/data are generated from your or your lab's research (check all that apply)? (If you select two or more items, please also rank them in the next screen.)

- [ ] Audio files (e.g., interviews, instructions for the experimental procedures, etc.)
- [ ] Data automatically generated from or by computer programs
- [ ] Data collected from sensors/instruments (e.g., microscopes)
- [ ] Databases (e.g., Excel, Access, SQL, MySQL, Oracle)
- [ ] Digital photographs and other images
- [ ] Documents or reports (e.g., Word, PDF, etc.)
- [ ] Genomic data
- [ ] GIS (Geographic Information Systems)
- [ ] Laboratory notebooks (digital)
- [ ] Observational data
- [ ] Spreadsheets
- [ ] Standard operating procedures and protocols
- [ ] Survey results & interview transcripts
- [ ] Text files (e.g., .txt)
- [ ] Video files
- [ ] Websites and blogs
- [ ] Non-digital research data (e.g., notebooks, physical samples, field notes, etc.)
- [ ] Other (please specify):

Q14. This is a follow-up question to the previous question: Please rank your choices from the previous question to reflect how frequently these types of digital files/records/data are generated as part of your or your lab's research. (Items at the top of the list are those that are generated more frequently, those at the bottom are generated less frequently.)

- » Audio files (e.g., interviews, instructions for the experimental procedures, etc.)
- » Data automatically generated from or by computer programs
- » Data collected from sensors/instruments (e.g., microscopes)

- Databases (e.g., Excel, Access, SQL, MySQL, Oracle)
- Digital photographs and other images
- Documents or reports (e.g., Word, PDF, etc.)
- Genomic data
- GIS (Geographic Information Systems)
- Laboratory notebooks (digital)
- Observational data
- Spreadsheets
- Standard operating procedures and protocols
- Survey results & interview transcripts
- Text files (e.g., .txt)
- Video files
- Websites and blogs
- Non-digital research data (e.g., notebooks, physical samples, field notes, etc.)
- Other (please specify):

Q15. What kinds of non-digital research data do you or your lab store? (check all that apply)

- [ ] Specimens
- [ ] Samples
- [ ] Paper records/portfolios
- [ ] Consent forms
- [ ] Questionnaires
- [ ] Notebooks/Lab books
- [ ] Sketches
- [ ] Films

- [ ] Videos
- [ ] Other (please specify):

Q16. Do you or your lab create digital copies of these data?

- ○ Yes
- ○ Sometimes
- ○ No

Q17. Who would you usually expect to access and use your or your lab's research data, apart from yourself? (check all that apply)

- [ ] Only myself
- [ ] Other researchers at my institution
- [ ] Researchers at other academic institutions
- [ ] Funders
- [ ] Publishers
- [ ] General public
- [ ] Other (please specify):

Q18. Does your or your lab's data usually contain any of the following? (check all that apply)

- [ ] Personally identifying data about living individuals
- [ ] Sensitive personal data
- [ ] Patient-identifiable data
- [ ] Commercially sensitive data
- [ ] Other types of confidential/restricted data
- [ ] None of the above
- [ ] Not sure

Q19. Do you or your lab always retain data on at least two different types of storage media or at another location?

- ○ Yes
- ○ Sometimes
- ○ No

Q20. How many years do you or your lab usually maintain older data for your research project?

- ○ 1-2 years
- ○ 3-4 years
- ○ 5 or more years
- ○ Indefinitely
- ○ Other (please specify) _____

Q21. Which of the following security measures do you or your lab use to protect your files and data? (check all that apply)

- ☐ Access logging
- ☐ Anonymization
- ☐ Encryption
- ☐ Password protection of files
- ☐ Physical security (e.g., locked room, controlled access to premises)
- ☐ Not sure
- ☐ Other (please give details): _____
- ☐ None

Q22. Please indicate roughly how much of your or your lab's digital research data is held in each of the following locations. (The answer options for each item are: a) None, b) Some, c) Substantial, d) All), e) n/a

	a) None	b) Some	c) Substantial	d) All	N/A
Hard disk drive of a computer owned by my institution	○	○	○	○	○
Hard disk drive of a privately-owned computer	○	○	○	○	○
External hard drive	○	○	○	○	○

	a) None	b) Some	c) Substantial	d) All	N/A
Memory sticks	○	○	○	○	○
Institution-managed network storage	○	○	○	○	○
CD/DVD	○	○	○	○	○
Cloud service – Dropbox	○	○	○	○	○
Cloud service - Google Drive	○	○	○	○	○
Cloud service – OneDrive	○	○	○	○	○
Cloud service – Box	○	○	○	○	○
Cloud service – Other	○	○	○	○	○

Q23. If you or your lab have collected research data and recorded it on paper, how do you make sure the data are entered into an electronic format accurately? (check all that apply)

- ☐ Double entry (data are entered twice in order to check for mismatches and other data entry errors)
- ☐ Single data entry – data entered by one staff.
- ☐ Software scan
- ☐ Other (please describe): _____

Q24. How do you or your lab currently track and manage your data during the active phase of your research project? (check all that apply)

- ☐ Dedicated data management software (please specify)
- ☐ In a local database (e.g., within research group)
- ☐ In a spreadsheet
- ☐ In an electronic logbook
- ☐ In a paper logbook
- ☐ Other (please specify): _____
- ☐ None of the above

Q25. How often is your or your lab's data during the active phase of the research project backed up by you,or your colleagues, or yourinstitution?

- ○ I do not know
- ○ Immediately upon creation
- ○ Daily
- ○ Weekly
- ○ Monthly
- ○ Yearly
- ○ On request
- ○ At the end of the project
- ○ Never

Q26. What is your or your lab's primary backup solution for your digital research data?

- ○ Institution-managed backup storage
- ○ Cloud Drive (e.g., Google Drive, Dropbox, OneDrive, etc.)
- ○ External hard drive or memory stick/USB/Flash drive
- ○ Hard disk drive of a computer owned by the university
- ○ Hard disk drive of a privately owned computer
- ○ Third party (including commercial data storage)
- ○ A discipline-specific or generalist repository
- ○ CD/DVD
- ○ I do not know
- ○ Other (please specify): _____

Q27. Have you or your lab ever lost any research data?

- ○ Yes
- ○ No

Q28. If you or your lab ever lost research data, what was the cause of the loss? (check all that apply)

- ☐ Notebook damage
- ☐ File deleted by mistake and was unable to retrieve for some reason

- [ ] Environmental disaster such as fire or flood
- [ ] Equipment failure
- [ ] Equipment stolen
- [ ] I do not know
- [ ] Disk crash
- [ ] Other (please specify):

Q29. What was the impact of the loss of person-days? (please estimate how many person-days)

	person-days					
	1-5	6-10	11-20	21-30	> 30	N/A
Wasted research effort:	○	○	○	○	○	○
Delay to publication:	○	○	○	○	○	○
Reputational damage:	○	○	○	○	○	○
Failure to meet funder requirements:	○	○	○	○	○	○
Reduction in quality of research outputs:	○	○	○	○	○	○
Failure to meet regulatory requirements:	○	○	○	○	○	○
Other (please specify):	○	○	○	○	○	○

Q30. At the minimum, what information should a data management standard operating procedure contain? (check all that apply)

- [ ] Types of data
- [ ] Samples
- [ ] Data collections
- [ ] Software
- [ ] Metadata
- [ ] Data security
- [ ] Data distribution policy
- [ ] Other (please specify):

Q31. Do you or your lab follow any guidelines to ensure good documentation of your data?

- ○ Yes
- ○ Sometimes
- ○ No

Q32. When you describe your research data, what do you usually consider? (check all that apply)

- ☐ How the data was generated?
- ☐ The format of the data
- ☐ The number of records and the number of files
- ☐ The stages the data pass through (e.g., raw, processed, analyzed, etc.)
- ☐ What software tools you used to collect data
- ☐ What hardware tools you used to collect data

Q33. What standard format(s) do you or your lab usually use for **managing** or **maintaining** data? (check all that apply)

- ☐ CSV
- ☐ Excel
- ☐ Access
- ☐ ASCII (text)
- ☐ SAS
- ☐ SPSS
- ☐ Stata
- ☐ R
- ☐ XML
- ☐ Database (SQL/MySQL, Oracle, etc.)
- ☐ JPEG (or JPG) - Joint Photographic Experts Group
- ☐ PNG - Portable Network Graphics
- ☐ GIF - Graphics Interchange Format
- ☐ TIFF - Tagged Image File

- [ ] PSD - Photoshop Document
- [ ] PDF - Portable Document Format
- [ ] EPS - Encapsulated Postscript
- [ ] AI - Adobe Illustrator Document
- [ ] INDD - Adobe Indesign Document
- [ ] RAW - Raw Image Formats
- [ ] Other (please specify)

Q34. What standard format(s) do you or your lab usually use for **disseminating** data? (check all that apply)

- [ ] CSV
- [ ] Excel
- [ ] Access
- [ ] ASCII (text)
- [ ] SAS
- [ ] SPSS
- [ ] Stata
- [ ] R
- [ ] XML
- [ ] Database (SQL/MySQL, Oracle, etc.)
- [ ] JPEG (or JPG) - Joint Photographic Experts Group
- [ ] PNG - Portable Network Graphics
- [ ] GIF - Graphics Interchange Format
- [ ] TIFF - Tagged Image File
- [ ] PSD - Photoshop Document
- [ ] PDF - Portable Document Format
- [ ] EPS - Encapsulated Postscript
- [ ] AI - Adobe Illustrator Document
- [ ] INDD - Adobe Indesign Document
- [ ] RAW - Raw Image Formats

- [ ] Other (please specify)

Q35. How do you or your lab currently share data with others such as your colleagues, collaborators, or others who are interested in your data? (check all that apply)

- [ ] By emailing data files
- [ ] Using a cloud storage service, e.g., Dropbox, Google Drive, etc.
- [ ] Using portable storage, such as CDs, DVDs, memory sticks, etc.
- [ ] By uploading to a website or FTP server accessible to other researchers
- [ ] Institutional file-sharing service
- [ ] Deposit in a public repository/data center
- [ ] Deposit in institutional repository
- [ ] Publish it on a website
- [ ] Publish it in a data journal or other formal publication
- [ ] Share it on an academic social network
- [ ] Include it as supplementary data to a published article
- [ ] Deposit on a code-sharing platform (e.g., GitHub)
- [ ] Other (please specify):

Q36. Are you or your lab willing to share your research data publicly, once the research is complete?

- ○ Yes, I already have done so
- ○ I do not currently, but expect to do so in future
- ○ No
- ○ Not sure

Q37. Why would you or your lab be unwilling to share your data publicly? (Please check all that apply)

- [ ] I do not want others to see my data
- [ ] I do not know how to share it easily
- [ ] I want to keep the data to do further research

- [ ] I do not have funding to cover the costs involved
- [ ] It is impractical to share the data due to its size
- [ ] It is impractical to share the data due to its format
- [ ] Never considered it
- [ ] No data center for the discipline
- [ ] No re-use potential
- [ ] Not required by funder
- [ ] Sensitive/confidential data
- [ ] Waiting for the university to set up a data archive
- [ ] I do not have permission to share the data
- [ ] I want to patent/commercialize my research
- [ ] I do not want my data to be used for commercial purposes
- [ ] Other (please specify):

Q38. When you or your lab report your data, do you validate or cross-validate the results?

- ○ Yes
- ○ Sometimes
- ○ No
- ○ Does not apply

Q39. Are you or your lab always willing to share your raw data with other colleagues? (may add a follow-up question)

- [ ] Yes, if yes, please answer the following (check all apply)
- [ ] Who recorded the data?
- [ ] When did they record it?
- [ ] Where was the measurement preformed (at which facility)?
- [ ] What system and instrument(s) were used?
- [ ] Why were they measuring it?
- [ ] Was it part of a standard operating procedure (SOP)?
- [ ] What was the name of the SOP step that they were performing?

- [ ] No

Q40. If you or your lab share the data, are you concerned that others may validate the results you reported and may find results that may suggest your original reports might have errors?

- ○ Yes
- ○ Sometimes
- ○ No
- ○ Does not apply

Q41. When you or your lab share your data, do you prepare a codebook or user guide for your others?

- ○ Yes
- ○ Sometimes
- ○ No
- ○ Does not apply

Q42. When you or your lab share your data, do you prepare the methodology about your data?

- ○ Yes
- ○ Sometimes
- ○ No
- ○ Does not apply

Q43. Does your institution have a data management policy document?

- ○ Yes
- ○ No
- ○ Does not apply

Q44. Does your institution provide a protocol on procedures, standards, and guidelines for data management?

- ○ Yes
- ○ No
- ○ Does not apply

Q45. Do you expect to make use of institutional services designed to support data management and sharing?

- ○ I already use these services
- ○ I do not currently use these services, but I expect to in the future
- ○ I do not expect to use these services
- ○ I do not know what services are available
- ○ There are no services available
- ○ Not sure

Q46. How concerned are you about the following data management issues? (The options for each item are: a) Not at all concerned, b) Slightly concerned, c) Somewhat concerned, d) Moderately concerned, e) Quite concerned)

	a) Not at all concerned	b) Slightly concerned	c) Somewhat concerned	d) Quite concerned	c
Delays in getting access to data	○	○	○	○	
Disputes over ownership of research data, e.g., conflicts over intellectual property rights	○	○	○	○	
Inability to access data due to obsolescence, expired software license, etc.	○	○	○	○	
Insufficient storage space for research data	○	○	○	○	
Insufficient security over confidential data	○	○	○	○	
Inability to interpret data (e.g., due to poor/lost documentation or inadequate descriptions)	○	○	○	○	
Inability to maintain control of my data and understand how it is used	○	○	○	○	
Lack of file-naming/metadata conventions	○	○	○	○	

making it difficult to retrieve data

Q47. Would you value training on any of the following? (check all that apply)

- [ ] Citing research data
- [ ] Collaboration and sharing of data
- [ ] Citing software
- [ ] Copyright and intellectual property rights within a data context
- [ ] Creating or working with .xml documents (including text-mining)
- [ ] Data licensing
- [ ] Developing a research data management plan for a funding application
- [ ] Ethics, consent, and legal issues with research data
- [ ] Funder requirements for research data management
- [ ] Guidance on costing data management in grant applications
- [ ] Long-term storage of your data
- [ ] Publishing research data
- [ ] Security of data
- [ ] Support in data selection, metadata creation, and licensing for preservation
- [ ] Technical support for data processing (e.g., database design, High Performance Computing (HPC))
- [ ] Other (please specify):
- [ ] None

Q48. Would you be interested in a data management standard operating procedures (DMSOP) online toolkit where you could point and click to generate your own DMSOP document?

- ○ Yes
- ○ No
- ○ Not sure

Q49. Please provide any further comments or observations you would like to make on support for research *data collection* procedures:

Q50. Please provide any further comments or observations you would like to make on support for research *data processing* procedures:

Q51. Please provide any further comments or observations you would like to make on support for research *data storage* procedures:

Q52. Please provide any further comments or observations you would like to make on support for research *data sharing* procedures:

Q53. Please provide any further comments or observations you would like to make on support for research *data reporting* procedures: