

# Survey Questionnaire

## 2016 Survey of Research Data Management at Concordia

### Survey of research data management at Concordia

The aim of this survey is to better understand how Concordia researchers work with their data. The results of this survey and optional follow-up interviews will inform the Concordia Libraries' strategic planning for data services and help PERFORM's Data Science Lab design resources and workshops to meet your needs. In the short term, it will also allow the Data Librarian to provide some assistance to participants.

Participation is voluntary and all published results of this survey will be anonymous. Publications or reports based on the survey or the subsequent interviews will present results in aggregate and will not contain identifying characteristics of respondents.

### DEMOGRAPHICS

- Faculty: (Arts & Science, Fine Arts, Engineering, JMSB, Other)
- Department
- Rank: (Lecturer, Assistant Professor, Associate Professor, Professor)
- Research center/laboratory/team (if applicable):

### RESEARCH DATA

For the following sections of the survey, think of a research project dataset that **you consider to be your most important**. This should be research data collected or created by you or your team, not a pre-existing or secondary dataset. The dataset may include multiple data files from the same research project. If you cannot choose a most important dataset, simply consider your most recently collected data.

### DATA COLLECTION & CURATION

#### 1) Do you collect/create/analyze/store research data?

*Research data is defined very broadly to include all primary data collected during your various research projects at Concordia. This might include survey data, experimental data, simulation data, programming code, qualitative data (interview transcripts, field notes, audio or video files), or other information.*

- Yes

- No (If no, end survey)

**2) How was your research data collection funded? (Choose ALL that apply)**

- NSERC
- CIHR
- SSHRC
- CCA
- FQR-S
- FQR-NT
- FQR-SC
- CALQ
- Internal funding
- Private funding
- Other (please specify)

**3) How did you collect raw/original data? (Choose ALL that apply)**

- Paper and pencil (questionnaires, laboratory notebooks, recording sheets)
- Collect specimens or physical samples (water, tissue, cells, blood, soil, etc.)
- Direct electronic input (iPad/digital forms/online)
- Record images, observations, events
- Other - Specify: \_\_\_\_\_

**DATA MANAGEMENT**

**4) Who is responsible for managing your data? (Choose ALL that apply)**

- Yourself
- Research Coordinator/Project manager
- Graduate student
- Designated person in the research group

- Research assistant
- External Research Collaborator/Project Partner outside of Concordia
- IT staff in the department
- Nobody
- Other (please specify)

**5) How did you convert raw/original data into a structured electronic form/database? (Choose ALL that apply)**

- Data are collected in structured electronic form
- Manual data entry/transcription
- Optical Character Recognition scan to derive data from paper forms (OCR)
- Scan of paper form (PDF)
- I don't put data into structured electronic form/database
- Other - Specify: \_\_\_\_\_

**6) What type of data was collected or created for this specific project? (Choose ALL that apply)**

- Text (e.g. TXT, DOC, PDF, RTF, HTML, XML)
- Numerical or statistical (e.g. CSV, MAT, XLS, SPSS, R, etc.)
- Multimedia, including images, audio, or video (e.g. JPEG, TIFF, MPEG, QuickTime, Bitmap)
- Databases (e.g. MS Access, Oracle)
- Instrument specific (e.g. LC-MS, ECG, QPCR)
- 3D Imaging (e.g., LORIS, MRI)
- Geospatial (e.g. ArcGIS, MapInfo, QGIS, WorldMap)
- Models (e.g. 3D, statistical, similitude, macroeconomic, causal)
- Programming code (e.g. Java, C, Perl, Python, Ruby, PHP, JSON, Dplyr, D3, HTML)
- Other Please specify: \_\_\_\_\_

**7) What type of supporting documentation do you keep on your data? (Choose ALL that apply)**

*Supporting documentation allows someone outside your research group to understand and use the research data or to replicate the methodologies that produce the data. This documentation can be retained in the same file, folder or document as the research data.*

- Data dictionary/Coding manual
- Description of methodology / protocol
- Questionnaire or interview text
- Instrument metadata (manufacture, model, parameters, calibration)
- Programming code (to replicate analysis)
- Additional metadata (e.g., technical, preservation, copyright, access restriction)
- Data citations (e.g., source of merged data with your original data)
- Other - Specify: \_\_\_\_\_
- I do not keep any data documentation

**8) What is the approximate total size of this dataset?**

- Less than 10GB (Small USB stick is 8GB)
- Between 10 and 100GB (Portable Pocket Drive)
- Between 100GB and 4TB (Large External Hard Drive)
- More than 4TB (Server)
- I don't know

**9) DURING the research project and data collection, where do you store your working dataset/electronic files? (Choose ALL that apply)?**

- Cloud storage (Dropbox, Google Drive, MS OneDrive, etc.)
- Computer hard drive
- External hard drive/Portable storage/Flash drive (USB)
- Hard drive of the instrument which generates the data
- Department or lab server
- IITS server
- CDs/DVDs/Tape

- Data Repository (ICPSR, GenBank, PubChem, figshare, DRYAD, GitHub, LORIS, Harvard's Dataverse, Scholars Portal's Dataverse, etc.)
- Physical hardcopy retained (in boxes, cabinets, etc.)
- Other (please specify)

**10) How do you secure your electronic data? (Choose ALL that apply)**

- Password protected
- Data is de-identified (i.e., only anonymous data stored)
- Data is encrypted
- Data is not secured
- Other. Specify \_\_\_\_\_

**11) How often do you back up your working data?**

- Daily
- Weekly
- Monthly
- Ad Hoc
- Never
- Other (please specify)

**12) Where do you store your back up copies (additional copies of your working datasets kept in case of data loss?) (Choose ALL that apply)**

- Cloud storage (Dropbox, Google Drive, MS OneDrive, etc.)
- Computer hard drive
- External hard drive/Potable storage/Flash drive (USB)
- Hard drive of the instrument which generates the data
- Department or lab server
- IITS server
- CDs/DVDs/Tape
- I don't store backup copies

- Other (please specify)

**13) What software or programs do you use when collaborating with others? (Choose ALL that apply)**

- Word Track Changes
- Google Documents
- Shared Folders (DropBox)
- Overleaf, Authorea, DocEar (LaTeX)
- GitHub
- Referencing (Refworks, BibTex, EndNote, Papers)
- None
- Other - Specify: \_\_\_\_\_

**14) How do you deal with different versions of your working data files?**

- I save multiple files and update the files with different names/naming convention
- This is done automatically (e.g., Dropbox)
- I use a versioning control program (NAME: \_\_\_\_\_)
- I have not used versioning control, but I am keen to learn more.
- I only use one dataset version.
- Other (please specify)

**DATA STORAGE AND ARCHIVING AT THE END OF PROJECT**

**15) Where do you store your dataset at the conclusion of your research project? (Choose ALL that apply)**

- Cloud storage (Dropbox, Google Drive, MS OneDrive, etc.)
- Computer hard drive
- External hard drive/Portable storage/Flash drive
- Hard drive of the instrument which generates the data

- Department or lab server
- IITS server
- CDs/DVDs/Tape
- Physical hardcopy retained (in boxes, cabinets, etc.)
- I do not keep a copy of my dataset **[skip to 17]**
- Other

**16) How long do you retain your dataset after the completion of your research project?**

- 1 year
- 5 years
- 10 years
- Indefinitely
- I do not know
- Other - Specify: \_\_\_\_\_

**17) After your research project is completed –the data has been cleaned and analyzed— in what data repository would you consider archiving this dataset? (Choose ALL that apply)**

- SPECTRUM (Concordia’s institutional repository)
- General or discipline-specific repository (ICPSR, Protein Data Bank, Cambridge Structural Database, GitHub, Dryad, Figshare, Harvard’s Dataverse, Scholars Portal’s Dataverse, etc.)
- Other – Specify  
\_\_\_\_\_
- I do not know about data repositories, but I am keen to learn more.
- Not interested in archiving **[skip to 19]**

**18) How long do you think your dataset should be preserved in a data repository after the completion of your project?**

- 1 year
- 5 years
- 10 years

- Indefinitely
- I do not know
- Other - Specify: \_\_\_\_\_

**19) Are you aware of any discipline-specific data repositories in your area of research?**

- Yes  
Please list them. [Free text]
- No

**20) How do you share your research data with others? (Choose ALL that apply)**

- Physically hand over data (USB, CD, Portable laser drive)
- Email data files
- Link to data on academic website
- Data repository, open-access (anonymous dataset available to public)
- Data repository, controlled access (approved researchers only/data sharing agreements)
- I have not been sharing my data so far, but would like to learn more about different options
- I prefer not to share my data
- Other (please specify)

**WORKING WITH DATA: STATISTICS, PROGRAMMING, & VISUALIZATION**

**21) What software do you use for statistical analysis? (Choose ALL that apply)**

- SAS
- R
- SPSS
- Stata
- Matlab
- Excel
- NVivo / Qualitative software
- Other - Specify: \_\_\_\_\_



- I do not conduct statistical analyses

**22) Where do you go for help with unfamiliar statistical procedures? (What resources are available to you when you have a challenging statistical issue/problem?) (Choose ALL that apply)**

- Colleagues within department
- Faculty outside of department
- Online resources
- Specialized research centers
- Google/YouTube
- Workshop - external
- QICSS Consultation
- I do not need help with statistics
- Not applicable
- Other (please specify)

**23) Do you use programming or write code for any of the following research data activities? (Choose ALL that apply)**

- Data collection
- Data manipulation and management
- Statistical and computational analysis (including simulations)
- Graphics and manuscript preparation
- Webpage creation
- Other. Specify\_\_\_\_\_
- I have no/little experience programming, but I am keen to learn more.
- Not applicable

**24) In what languages do you program? (Choose ALL that apply)**

- Python
- Perl

- Java
- C++
- HTML/CSS
- Not applicable
- Other. Specify \_\_\_\_\_

**25) What tools do you use for creating graphs, figures, or infographics? (Choose ALL that apply)**

- Excel
- Statistical Software (SPSS, Stata, R)
- Adobe Creative Suite
- Java Script (D3)
- Other - Specify: \_\_\_\_\_
- Not applicable

**26) Thinking about peer-review journals in your discipline, or other respected sources for research dissemination, what is your experience with the following: (Choose ALL that apply)**

- I have read articles with interactive graphics
- I have submitted manuscripts with interactive graphics
- I have submitted manuscripts to journals that required data files/statistical code be made accessible
- I avoid journals that have an open-access requirement to link my data files

**DATA WORKSHOPS/ASSISTANCE**

**27) For each of the following, please indicate your level of interest in data management assistance at Concordia or training/development opportunities (workshops/resources):**

	Very interested	Somewhat interested	Not interested	Not applicable
Data management workshop (examples of topics: setting up electronic files, file formats, IP and licensing, de-identification of data)				
Data visualization (interactive graphics, etc.)				
Data management plan (organization and curation of data throughout research project)				
Data management activities that will be required by grant funding agencies				
Data documentation and metadata (i.e. data dictionaries, data codebooks containing description of dataset and variables; other supporting documentation)				
Data archiving (e.g. finding appropriate data repository, understanding how I can control access to my data)				
Research collaboration software tools (e.g. Latex, Overleaf, versioning software or other tools that allow secure storage and/or online collaboration between members of research team <i>during the research project</i> )				
Digitization of paper records (lab or field notebooks, paper questionnaires, etc.)				
Data sharing and access (copyright, patents, licenses, ownership of data, data sharing agreements, etc.)				
Help in ensuring the confidentiality/anonymity of data on human participants				
Finding and accessing existing data sources (data scraping, secondary datasets)				
Other (please specify)				

**28) Are you aware of any resource or project in your department or faculty aimed at facilitating research data management?**

- Yes. Please describe.
- No

Thank you for completing this survey! Your participation is sincerely appreciated. If you would like to receive information about planned data science activities or workshops, please provide your contact

information below. Also indicate if you would be willing to be contacted to participate in a follow-up interview.

- Name:
- Email:
- Phone:
- Department:
- Willing to participate in follow up interview? (Y/N)
- Keep me informed of workshops or data science activities for me or my students? (Y/N)

If you have any questions or comments about this survey, please use the text box below or contact one of the researchers:

Alex Guindon ([alex.guindon@concordia.ca](mailto:alex.guindon@concordia.ca))

Danielle Dennie ([danielle.dennie@concordia.ca](mailto:danielle.dennie@concordia.ca))

Jennifer McGrath PERFORM Data Science ([data.science@concordia.ca](mailto:data.science@concordia.ca))