THE UNIVERSITY OF BRITISH COLUMBIA

Office of the Vice-President, Research & Innovation

UBC Research Data Management (RDM) Fall Series 2021

Post-Event Summary Report

Submitted by Nick Rochlin on behalf of the RDM Fall Series Team

Event Summary

From September 27th - October 14th 2021, the *UBC RDM Fall Series 2021* (herein after "Fall Series") was collaboratively hosted by UBC Advanced Research Computing (ARC), UBC Library, the UBC Centre for Scholarly Communications (CSC), and the UBC Okanagan Library. The second iteration of the Fall Series, this series virtually offered a broad range of research data management topics, and aimed to provide practical skills and guidance for managing research data throughout the data lifecycle.

Sessions in the series were led by members of UBC ARC, as well as UBC Librarians and associates from both Vancouver and Okanagan campuses. In addition to those who led sessions, there were a variety of discussion panels, featuring speakers from the UBC Okanagan Library, UBC ARC, UBC Behavioural and Clinical Ethics, UBC Privacy and Information Security Management (PrISM), PopData BC, as well as researchers from BC Children's Hospital, the Faculty of Management, the Department of Zoology, and the Department of Linguistics and School of Information. As the Fall Series ran over the course of three weeks, each week had its own theme which is detailed below:

- Week 1: General RDM
    - Introduction to Research Data Management
    - Introduction to Open Science Framework (OSF)
    - RDM Discussion Panel with UBC Researchers
- Week 2: Sensitive Data
    - REDCap
    - Sensitive Data Panel 1: Non-Clinical Human Data
    - Sensitive Data Panel 2: Clinical Data
    - Indigenous Data Panel
- Week 3: RDM with High Performance Computing (HPC)
    - Introduction to the Unix Shell
    - RDM with HPC Part 1
    - RDM with HPC Part

Key metrics	**2021**	**2020**
Unique registrants Total registrants	431 806	365 -
Unique attendees Total attendees	308 592	267 -
Average attendance rate	59%	61%

Table of Contents

Introduction

This report provides a post-session summary of the three-week long *UBC RDM Fall Series 2021* (the "Fall Series"), hosted by UBC Advance Research Computing (ARC), the UBC Library, the Centre for Scholarly Communications (CSC), and the UBC Okanagan Library. The event was attended by UBC students, staff, faculty, and non-UBC individuals from across Canada and internationally. All materials, including slides, materials, and session recordings for the 2021 Fall Series can be found here.

Background

Research Data Management (RDM) has been a growing focus for both researchers and funders for the past several years, and the UBC RDM 2020 Fall Series represented a momentous first step in the development of a comprehensive RDM workshop series that promoted the importance of and best practices in managing research data. In spring 2021, the Tri-Council Research Data Management Policy was officially released, further invigorating the importance of RDM on a national level.

Much like the lead up to the fall 2020 semester, there was a great deal of uncertainty going into fall 2021 with COVID-19 restrictions. Because the planning for the Fall Series began in late spring, it was difficult to gauge what the landscape would look like come fall, and so it was decided that the Fall Series would keep the same three week format as the prior year, with sessions running Monday-Thursday each week (missing one day for the Thanksgiving holiday, and one for The National Day for Truth and Reconciliation). It was also decided that due to limited attendance in 2020, there would be no office hours in the 2021 Fall Series, and all sessions would promote attendees to book consultations outside the Fall Series.

Goals and Objectives

The primary purpose of the 2021 Fall Series was to build on the content and lessons learned from the 2020 Fall Series, and to continue innovating and evolving the delivery of RDM sessions. There were numerous conversations in the planning period discussing pros and cons of various approaches, with the overall goal to put together a series that would be the most beneficial to the research community based on prior feedback and the recent release of the Tri-Council Research Data Management Policy . Below are the lessons learned from the 2020 Fall Series, and how they informed the creation and objectives of the 2021 Fall Series.

More than Just the "Book-Ends"

Noting that the majority of the content in the 2020 Fall Series was situated around data management planning and deposition of data, and acknowledging that RDM includes much more than this, a goal of the 2021 Fall Series was to create more sessions around active data management as well sessions on RDM with high performance computing (HPC). This was accomplished by delivering sessions on REDCap, a data capture tool, the Open Science

Framework (OSF), an active data management tool, as well as the three sessions in the RDM with HPC week, which included an Introduction to the Unix Shell, as well as hands-on training in UBC's Sockeye HPC system.

Discipline/Concept-Specific Sessions

A focus of the 2020 Fall Series was to make the content as discipline agnostic as possible, with the goal of making sessions broadly accessible and valuable. However, we realized that in doing so we inevitably ran into the issue of not being specific enough or missing key concepts or practices of certain research types. When planning the 2021 Fall Series, we had numerous conversations about how we could potentially have discipline-specific sessions, or some type of forked series for specific fields, but the most feasible option was to have three separate themed weeks (General RDM, Sensitive Data, RDM with HPC), in which researchers could pick and choose sessions.

Additional Partners/Portfolios

While the UBC Library and ARC play major roles in RDM, there are many other groups and portfolios that bring valuable knowledge and experience to this field. In addition to individuals from ARC and the Library, below is a list of groups who were represented as speakers at the 2021 Fall Series (in alphabetical order):

- BC Children's Hospital Research Institute
- Behavioural Ethics
- Department of Zoology (PhD candidate)
- Faculty of Management (Associate Professor), UBCO
- Indigenous Research Support Initiative (IRSI), UBCO
- Population Data BC
- Privacy & Information Security (PrISM)
- Research Ethics (Director)
- School of Education (Assistant Professor), UBCO
- School of Education (Director), UBCO
- School of Information / Computational Linguistics (Assistant Professor)

Attendance

With uncertainty about what the return to campus would look like when planning for the 2021l Fall Series, and recognizing the benefits of having participants from both UBC campuses, it was decided to hold the Fall Series in a completely virtual format again. This also provided an opportunity for national and international attendees.

Attendance Rate

Overall average attendance rate dropped 2% from the 2020 Fall Series, from 61% to 59%. The REDCap session had a much higher rate than others (79%), indicating a very strong demand for this type of training.

Table 1. Average Attendance Rate.

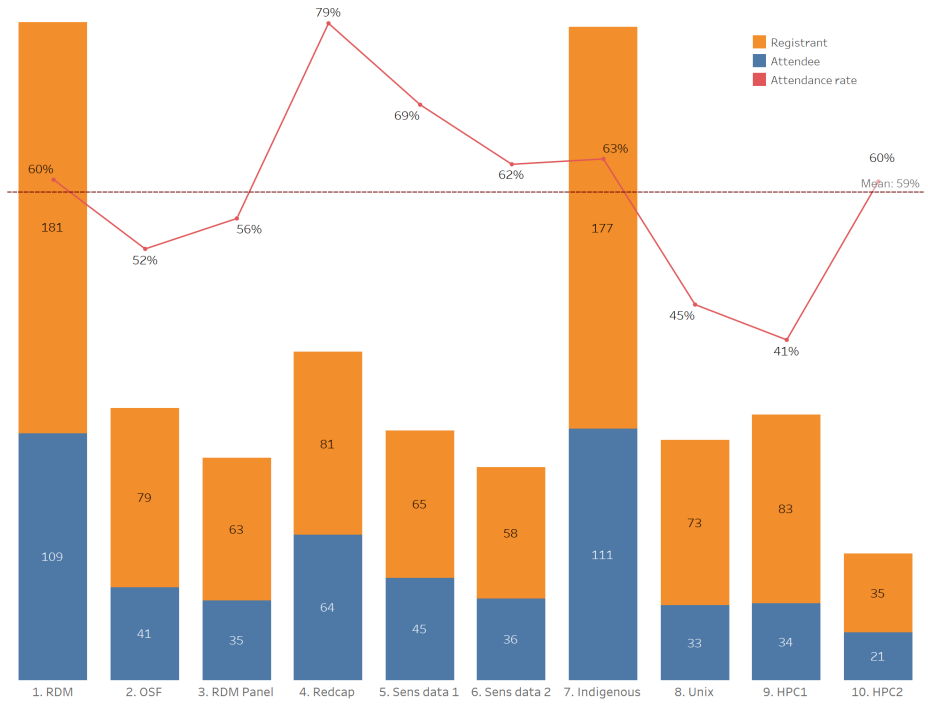
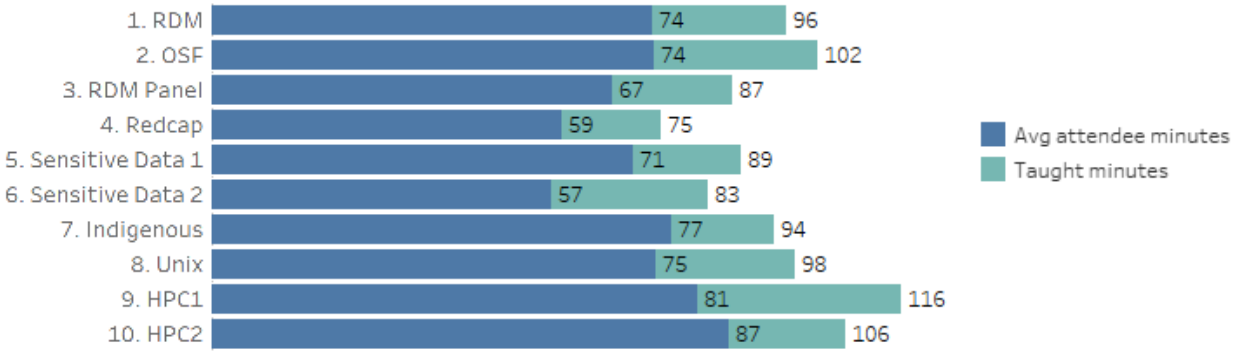


Table 2. Average Attendee Minutes by Session



Registration by Faculty

Similar to the 2020 Fall Series, Medicine (which includes both the Faculty of Medicine and the Southern Medicine Program) represented the top registered discipline. This finding from last year inspired the clinical data discussion panel, and could potentially explain the high

attendance rate of the REDCap session as it is a frequently used electronic data capture tool in health research.

Table 3. Registration by UBC Faculty.

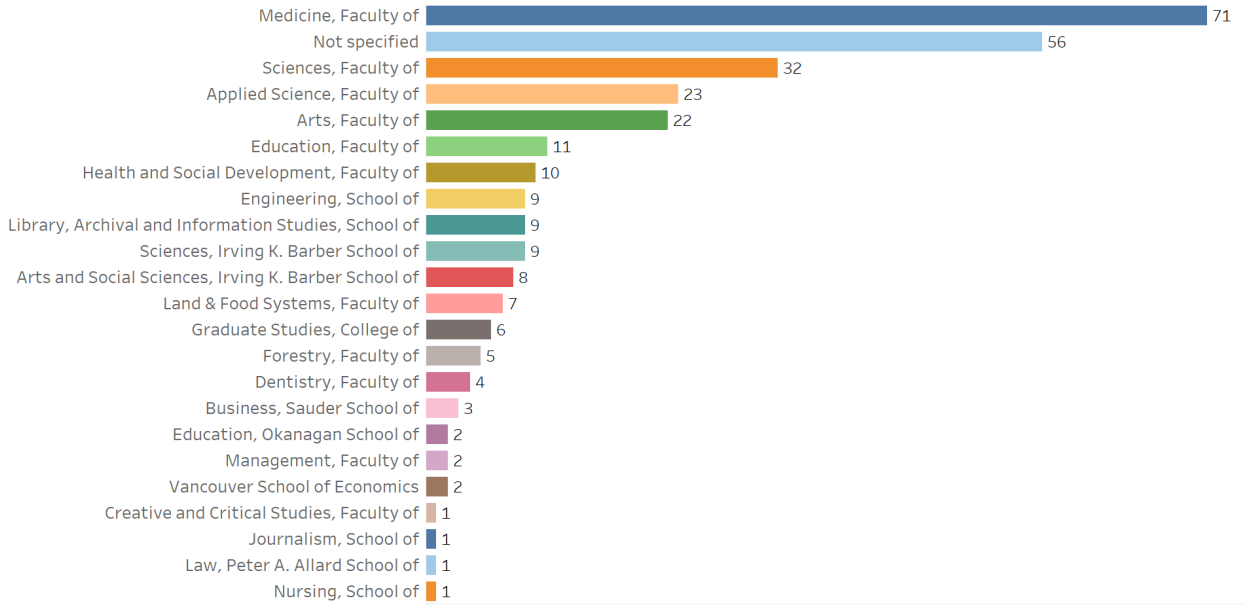
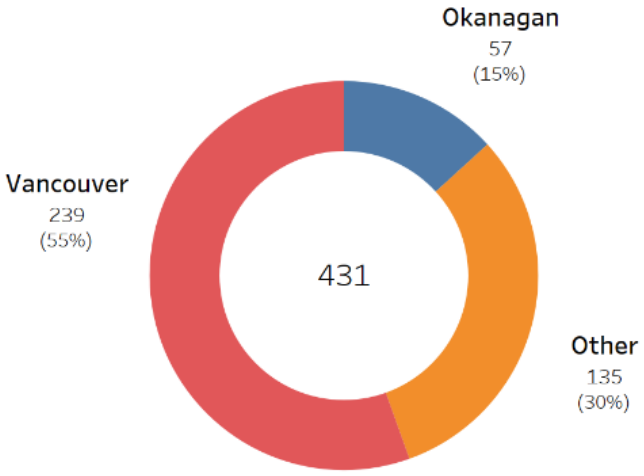


Table 4. Registration by UBC Campus.



Non-UBC Attendees

The percentage of non-UBC attendees doubled this year, from 15% in the 2020 Fall Series to 30%. This can be attributed to the earlier promotion of this series on a national listserv (CANLIB-DATA), but may also indicate that the Fall Series is gaining a national reputation for high-quality sessions.

Okanagan Context

Much like the 2020 Fall Series, the 2021 Fall Series was a collaboration across the UBC Okanagan and Vancouver campuses and featured significant contributions from the Okanagan campus. Below are some Okanagan-specific highlights for the event:

- 5 of 10 sessions featured a presenter from the Okanagan campus
- 4 of 5 speakers from the Indigenous Data Panel were from the Okanagan
- 15% of unique registrants were from the Okanagan
    - This was a decrease from the 2020 Fall Series, although the Okanagan campus represented 20% of total UBC registrants, which is still a strong showing relative to the size of the Okanagan campus (1/10 of graduate students and faculty compared to Vancouver).

Effort to Create Series

Following on the success of the 2020 Fall Series, we approached the same instructors to gauge their interest in developing and delivering the 2021 Fall Series. At an initial meeting in late May, the group decided that they wanted to shift the focus of the 2021 Fall Series away from higher-level, abstract sessions toward more concrete sessions that support the needs of researchers.

UBC ARC Involvement

UBC ARC played a considerable role in the planning, preparation, and execution of the Fall Series. Highlights of UBC ARC's involvement are listed below:

- Preparing and delivering 6 of 10 Fall Series sessions
    - 2 workshops on RDM for HPC (content developed for session)
    - 1 workshop on REDCap (revamped from previous year)
    - 1 general data discussion panel (preparing questions, soliciting panelists, and moderation session)
    - 2 sensitive data discussion panels (preparing questions, soliciting panelists, and moderation session)
- Organizing and arranging the Indigenous Data Panel over the course of several months (see below for full details)

- Sending communications to all faculties on both campuses, as well as associated research hospitals and health authorities
- Providing technical support during all sessions via Zoom chat
- Analyzing and visualizing participant and session data
- Preparing final report

Series Planning

After reviewing the lessons learned from the 2020 Fall Series, a variety of approaches and possible sessions were discussed. While the 2020 Fall Series was deemed to be a success, the attempt to illustrate the entire data lifecycle made the presumption that attendees would attend every session, which can be a challenge over the course of three weeks. A lot of the discussions in the planning phase were around the possibility of discipline-specific sessions, and various ways in which this could be achieved and attendees wouldn't be expected to attend every session. This led us to the final decision to put together two weeks of themed sessions that attendees could pick and choose from (General RDM and Sensitive Data), and the final week, RDM with HPC, consisted of three sessions that were required to be taken in succession.

Communications

Colleagues in the Research Commons and Library Communications created the communications visualizations that we used throughout the Fall Series. The Library also arranged the details of the event registrations and Zoom webinars. In addition to the communications handled through the Library and ARC's general communication channels, Fall Series promotions were sent twice to every faculty on the Vancouver and Okanagan campuses, as well as to associated research hospitals and health authorities. ARC staff sent additional targeted emails to researchers and narrative blurbs to faculty communication representatives to those faculties that had a member presenting in a session. There were also two sets of communications sent out through the CANLIB-DATA listserv, to promote the Fall Series on a national level.

Week 1: General RDM

The group decided to retain the Introduction to RDM session in more or less the same format. The content for that session, however, was refreshed by Jentry Campbell, the Librarys' RDM graduate student, in July and August. This brought the content for the session into line with the workshop that the Research Commons offers monthly. The other workshop featured in this week was the Introduction to the Open Science Framework (OSF), which was adapted from the 2020 Fall Series to put a stronger focus on RDM and how the OSF can be leveraged throughout the data life cycle. The final session in this week was the Research Data Panel, which featured discussion points ranging from general RDM, to various types of data sensitivities, as well as managing data in HPC environments (effort to create discussion panels described below).

Week 2: Sensitive Data

This week began with a presentation on REDCap, a data collection tool geared for sensitive data and clinical trials. In the planning discussions, it was thought that a hands-on workshop for REDCap would be a strong adaptation from the 2020 Fall Series, which featured a presentation

on REDCap. However, after looking at various options for training, and acknowledging that ARC's instance of REDCap is not set up for hands-on workshops, a similar presentation from 2020 was offered. The other sessions in this week, Sensitive Data Panel 1: Non-Clinical Human Data, Sensitive Data Panel 2: Clinical Data, and Indigenous Data Panel, were all discussion panels, and are described below.

Discussion Panels

Because of the success of the researcher discussion panel in the 2020 Fall Series, this session was recreated with a new group of researchers, and an additional three discussion panels were added. The creation of each panel involved a great deal of outreach and emailing, as well as the development of questions for the panels. Once panelists were finalized, a meeting was set up with all panelists to discuss the questions and format for the session, to ensure that everybody felt comfortable and would be given an opportunity to speak to topics they felt were important.

Indigenous Data Panel

The creation of this panel was fundamentally different from the other discussion panels, in that discussions began in late 2020 with Sandra Fox, who is the UBC Okanagan representative for the Indigenous Research Support Initiative (IRSI). Sandra was very kind and keen to discuss RDM in the context of Indigenous research and knowledge, but mentioned that a key aspect of working respectfully with Ingidenous groups is to slow down the pace of work, and to focus on building reciprocal relationships. After a few more conversations with Sandra over the course of several months, we formally approached her about the possibility of an Indigenous data panel as part of the 2021Fall Series, being clear that we were not prescribing any format or content, but would merely be providing a platform for a community-lead session if there was interest. Sandra was very interested in this, and provided the names of researchers from and working with Indigenous communities to be contacted, and she was CC'd on the outreach emails. An initial meeting was arranged with the Indigenous Initiatives Librarian (UBCO), the Indigenous Programs and Services Librarian (UBCV), Sandra, and an Indigenous researcher from the Okanagan School of Education. This meeting was a free-form discussion about topics of interest and potential formats for the session, and some additional names of potential panel members were brought up. After synthesizing notes from this meeting, and reaching out to additional panelists, the Director from the Okanagan School of Education joined the panel, and another meeting was set up to finalize the format for the session.

RDM with HPC

The RDM with HPC week was formulated around giving people hands-on training in HPC, with a particular focus on building skills for those who have little to no computing background but are interested in using these resources in their research. The pre-existing Unix Shell lesson that is taught regularly by the Research Commons was a great way to springboard into this mini-series by providing fundamental Unix commands, which were further used and elaborated upon in subsequent lessons.

There were several discussions and meetings leading up to the RDM with HPC 1 and 2 sessions, many of which surrounded the scope, content, and differences between pure HPC

training and RDM with HPC training. Out of these discussions, RDM with HPC 1 began by giving a hands-on demonstration of the DMP (data management plan) Assistant, and then walking through the ARC DMP template to identify the considerations and resources to support working in an HPC environment. The second half of this session introduced UBC ARC's services, including UBC ARC Sockeye ("Sockeye") high performance cluster, and did some light-weight demonstrations to prepare for the following session. RDM with HPC 2 focused on developing data management fundamentals when working in Sockeye, and featured hands-on activities to walk people through organizing folders and preparing documentation, transferring data to and from Sockeye, preparing software, and running analyses then managing outputs.

Feedback

Event promotion and registrations were handled through the Library's event booking platform, LibCal. We used the same basic outline for each session that we used for the 2020 Fall Series. An hour after the close of each session, LibCal sent an email to every registrant to direct them to the Library's standard workshop feedback form on Qualtrics.

In total, we received only six responses via the feedback form. Two responses were for the Introduction to OSF session and the other four were for the REDCap session. Respondents appreciated the pacing of the sessions and valued the discussions that occurred through the Q&A mechanism. The single suggestion for future development was for the REDCap workshop, to use "a scenario to take the users through the process of building a project and survey. It would be helpful to have a real-world context." Regular workshops could potentially be offered by highly qualified personnel in ARC through the Library Research Commons.

Lessons Learned

Planning for Discussion Panels

While the planning for the Fall Series began in early May 2021, which was sufficient time to plan workshop sessions, booking speakers for the discussion panels throughout the summer proved to be quite tricky. Going forward, earlier planning and booking speakers by April may prove to remove a lot of the headache of having to recruit people who may be off on vacation and away from their computers.

Possible revamp of the ARC DMP template

The RDM with HPC 1 session featured a walk-through of the ARC DMP template and its questions. This seemed to be an effective way to demystify the DMP Assistant and introduce researchers to the questions and considerations when starting research in an HPC environment. With that said, it was noted by the presenters when preparing for this session that the questions in the current ARC DMP template are a bit unclear, repetitious, and could be restructured/reordered to be more user-friendly. As Sockeye currently has a set of technical questions in the application form, it would certainly be worth exploring whether there could be

some sort of aligning with these two documents to avoid duplication of researcher efforts and to build RDM principles into the application.

REDCap Training

As mentioned earlier, the attendance rate for the REDCap session was much higher compared to other sessions. There was also a great deal of audience interaction throughout the entire presentation. It was noted in the planning stage of the Fall Series that having some sort of REDCap sandbox that would allow for hands-on training would be tremendously beneficial for the research community, although this is something that UBC currently does not have. Going forward, offering regular REDCap training sessions should be explored, as well as the possibility of setting up a training sandbox to offer more hands-on training.

Training vs. Showcase Event

The 2021 Fall Series contained a great balance of introductory presentations, hands-on workshops, as well as discussion panels to illustrate the considerations and challenges when working with various data types. This balance of pure training and showcase sessions provided a very even-keeled approach to an event like this, but future iterations of the Fall Series should consider if this balance is the preferred format, or if sessions should lean one way or another. Another area worthy of exploration is how discussion panels might be paired more closely with workshops, as a way to stir up audience interest in topics via the discussions, and then cover the topics of the discussion in a hands-on workshop session.

National Exposure and Interest

With the doubling of non-UBC attendees from 2020 to 2021, the Fall Series is developing a reputation for delivering high-quality content outside of UBC. Looking forward to a 2022 Fall Series, it should be considered whether there would be a benefit in taking this work nationally, and partnering with other universities as well as the Digital Research Alliance of Canada.

Appendix A. Summary Statistics of Sessions

- Session 1 Introduction to RDM
    - Unique registrants: 181
    - Unique attendees: 109
    - Attendance rate: 60%
    - Taught hours: 1.6
    - Attendee hours: 134
    - Average attendee hours: 1.2
- Session 2 Introduction to OSF
    - Unique registrants: 79
    - Unique attendees: 41
    - Attendance rate: 52%
    - Taught hours: 1.7
    - Attendee hours: 51
    - Average attendee hours: 1.2
- Session 3 RDM panel
    - Unique registrants:63
    - Unique attendees: 35
    - Attendance rate: 56%
    - Taught hours: 1.5
    - Attendee hours: 39
    - Average attendee hours: 1.1
- Session 4 REDCap
    - Unique registrants:81
    - Unique attendees: 64
    - Attendance rate: 79%
    - Taught hours: 1.3
    - Attendee hours: 63
    - Average attendee hours: 1.0
- Session 5 Sensitive data panel 1
    - Unique registrants:65
    - Unique attendees: 45
    - Attendance rate: 69%
    - Taught hours: 1.5
    - Attendee hours: 53
    - Average attendee hours: 1.2
- Session 6 Sensitive data panel 2
    - Unique registrants:58
    - Unique attendees: 36
    - Attendance rate: 62%
    - Taught hours: 1.4
    - Attendee hours: 34

    - Average attendee hours: 0.9
- Session 7 Indigenous data panel
    - Unique registrants:177
    - Unique attendees: 111
    - Attendance rate: 63%
    - Taught hours: 1.6
    - Attendee hours: 142
    - Average attendee hours: 1.3
- Session 8 Introduction to Unix shell
    - Unique registrants:73
    - Unique attendees: 33
    - Attendance rate: 63%
    - Taught hours: 1.6
    - Attendee hours: 41
    - Average attendee hours: 1.2
- Session 9 RDM with HPC 1
    - Unique registrants:83
    - Unique attendees: 34
    - Attendance rate: 41%
    - Taught hours: 1.9
    - Attendee hours: 46
    - Average attendee hours: 1.4
- Session 10 RDM with HPC 2
    - Unique registrants: 35
    - Unique attendees: 21
    - Attendance rate: 60%
    - Taught hours: 1.8
    - Attendee hours: 30
    - Average attendee hours: 1.4

Appendix B. Course Descriptions

Session 1: Introduction to Research Data Management

Session length: 1.5 hours
Presenters: Sheryl Adam, Susan Atkey, Allan Cho
Host: Doug Brigham
In-person delivery: Monday, September, 27, 10:00am
Description:
In this first session of the Fall Research Data Management series, we'll focus on the early stages of the research process. We'll introduce the core concepts involved in research data management; discuss planning for the success of your research projects; building a data management plan (DMP) to support your work and meet granting agency requirements; and finally investigate the FAIR principles and unpack some of the complexities of the research process.

Session 2: Introduction to OSF

Session length: 2 hours
Presenters: Mathew Vis-Dunbar, Sarah Parker
Host: Doug Brigham
In-person delivery: Tuesday, September 28, 10:00am
Description:
RDM touches every aspect of the research life cycle and good RDM includes connecting research data to the processes and inquiries that generate a research question, the tools used to collect, clean, and analyze data, and the myriad of outputs used to disseminate these findings. However, working with multiple platforms, services, and tools makes discrete connections between these objects and across the research life cycle challenging. OSF allows you to draw these linkages and identify where each facet of the research life cycle and your data interact. Bibliographies, protocols, data collection tools, scripts or code, data, posters, manuscripts etc - OSF offers one landing place to connect every aspect of your research project.

Session 3: RDM Discussion Panel with UBC Researchers

Session length: 1.5 hours
Panelists: Barb Marcolin, Faculty of Management (UBCO), Libby Natola, Department of Zoology (UBCV), Muhammad Abdul-Mageed, Linguistics & School of Information (UBCV)
Host & Moderator: Nick Rochlin
In-person delivery: Wednesday, September 29, 10:00am
Description:
As a way to illustrate Research Data Management (RDM) in practice, this session will feature UBC researchers discussing their experiences with data management, with a focus on

unforeseen events that lead to a disruption, near disruption, or horror story in their research. The session will be led by a moderator to facilitate a panel discussion in which UBC researchers will answer questions relating to data and RDM, and we encourage you to bring your own questions for the panel!

Session 4: REDCap

Session length: 1.5 hours
Presenter: Michael Tang
Host: Doug Brigham
In-person delivery: Monday, October 4, 10:00am
Description:

The REDCap (Research Electronic Data Capture) platform is a secure web application for building and managing research data collection instruments. The platform is specifically designed to support online or offline data capture for research studies. We'll cover the REDCap basic features that you may need to meet your research projects' data collection requirements, that includes: building data instruments and surveys, preparing data set exports, and managing user access control. We'll also share some creative ways REDCap can be set up to help facilitate your research data collection workflow.

Session 5: Sensitive Data Panel 1: Non-Clinical Human Data

Session length: 1.5 hours
Panelists: Luc Letarte, Wendy Bond, Marjorie Mitchell
Host & Moderator: Nick Rochlin
In-person delivery: Tuesday, October 5, 10:00am
Description:
As a way to illustrate Research Data Management (RDM) in practice, this session will feature a panel of those who support researchers working with sensitive data, with a particular focus on non-clinical human data. The session will contain a series of guided questions for the panel, lead by a moderator, to illustrate multiple perspectives of how to best work with sensitive data to avoid unanticipated headaches and pitfalls. Audience questions will be encouraged throughout, so please bring questions if you have any!

Session 6: Sensitive Data Panel 2: Clinical Data

Session length: 1.5 hours
Panelists: Laurel Evans, Dr. Phillip Richmond, Susan Johnson, Jim Mintha
Moderator: Elizabeth Kinney
Host: Doug
In-person delivery: Wednesday, October 6, 10:00am
Description:

As a way to illustrate Research Data Management (RDM) in practice, this session will feature a panel of those who support researchers working with sensitive data, with a particular focus on clinical data. The session will contain a series of guided questions for the panel, lead by a moderator, to illustrate multiple perspectives of how to best work with sensitive data to avoid unanticipated headaches and pitfalls. Audience questions will be encouraged throughout, so please bring questions if you have any!

Session 7: Indigenous Data Panel

Session length: 2 hours
Panelists: Kayla Lar-Son, Christian Isbister, Dr. Bill Cohen, Dr. Margaret Macintyre Latta
Moderator: Sandra Fox
Host: Nick
In-person delivery: Thursday, October 7, 10:00am
Description:
With researchers showing an increased interest in engaging with Indigenous community, and UBC looking to grow and support its Indigenous faculty, it is critical we have conversations around how to do this work meaningfully and appropriately. One of the most interesting and challenging topics is around Indigenous data as historically, Indigenous peoples have not seen any benefits or have even been negatively impacted by research and how data is used. It is then very important UBC works towards repairing this damage and establishing reciprocal and decolonial relationships where Indigenous community leads the research process and is in control of their data. Through sharing of their thoughts and experiences this panel of Indigenous and non-Indigenous researchers will explore some of the many facets of Indigenous data in relation to research, and how Indigenous research and data management can better align with Indigenous values and priorities.

Session 8: : Introduction to the Unix Shell

Session length: 2 hours
Presenter: Dr. Liam Doherty (supported by: Chelsea Palmer, Eka Grguric)
Host: Nick
In-person delivery: Tuesday, October 12, 10:00am
Description:
This workshop will introduce the Unix shell, a powerful way to communicate with your computer more directly through a command line interface. There are many ways to interact with a computer. Most of the time we click on things and select options with a cursor through a graphical user interface (GUI). A command line is exactly what it sounds like, a way of writing commands to your computer line by line and is incredibly powerful. In this workshop we will show you how to navigate through different parts of your system and introduce some of the most useful commands to be aware of.

Session 9: RDM with High Performance Computing (HPC) Part 1

Session length: 2 hours
Presenters: Nick Rochlin, Megan Meredith-Lobay, Jiarui Li
Host: Doug Brigham
In-person delivery: Wednesday, October 13, 10:00am
Description: The first part will focus on the Tri-Agency RDM Policy requirement of completing a Data Management Plan (DMP), and will introduce the DMP Assistant as a valuable tool in creating a DMP. After providing a walkthrough of the DMP Assistant, the remainder of the session will be spent reviewing and discussing the questions and considerations of the DMP template for Advanced Research Computing (ARC), to help you thoroughly prepare and plan for research projects in a high performance computing (HPC) environment.

The second part will begin by introducing UBC ARC Sockeye, which is an on-premise HPC system available to UBC researchers. This supercomputer provides an environment to accelerate their research timeline - if you struggle to analyze data on a laptop because of data size, this system is for you. We will outline the technical details of the system, how you can access this resource, as well as use cases. The session will conclude by getting those who have signed up for RDM with HPC 2 to log on to Sockeye, in preparation for that session.

This session is part 1 of 2 sessions in the RDM with HPC mini-series. There is no obligation to attend the second session, although if you are able to attend both, we think it will be all the more valuable!

Session: RDM with High Performance Computing (HPC) Part 2

Session length: 1.5 hours
Presenter: Jiarui Li (supported by Nick Rochlin, Jeff Gardner)
Host: Doug Brigham
In-person delivery: Thursday, October 14, 10:00am
Description: The first part will be geared towards setting up your working environment in UBC ARC Sockeye, to promote sound data workflows and for your research to be reproducible. There will be hands-on activities in which you will use the command line to navigate through different spaces within Sockeye, create folders, identify and implement considerations and best practices when naming large amounts of files, and create the proper documentation to ensure that your work is easily interpreted by others.

The second part will focus on efficiently using storage space, as well as transferring data to and from Sockeye. There will be hands-on activities in which you will identify and implement best practices in maintaining your project space so you do not max out your storage quota, transfer files to and from Sockeye with storage platforms, as well as download data directly from public databases.