Research data management and sharing support for chemistry researchers

Shannon Farrell, Research Data Services Lead
University of Minnesota Libraries

Research Data Services (RDS)

- Coordinate education, engagement, and networking around data management, curation, and sharing
- Develop, advise on, and implement data services and education to faculty, students, and staff
- Ensure University of Minnesota Libraries continue to be engaged in and informed by national and international best practices around data management, and be leaders in this space
- Collaborate with Liberal Arts Technology & Innovation Services (LATIS)

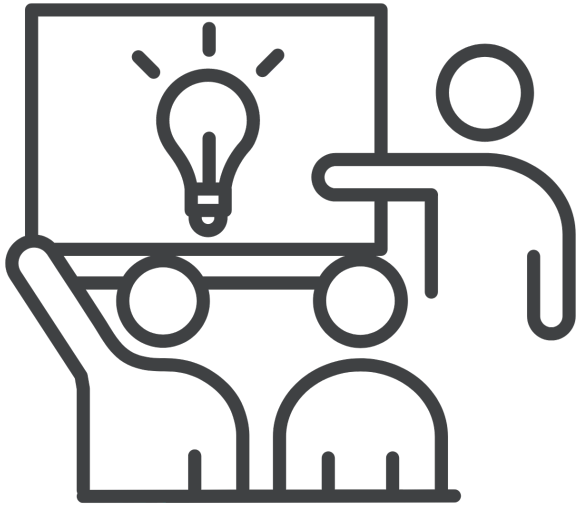
https://www.lib.umn.edu/services/data

What are data services?

- Touch points
  - Data management
  - Data curation
  - Data sharing
  - Long-term data stewardship options
- Activities
  - Consultations
  - Reviewing data management and sharing plans
  - Instruction
  - Annual events (data management boot camp, love data week, research ethics week, graduate your data)

Data management instruction

- Large-group instruction (webinars, panels, seminars)
  - Writing data management & sharing plans
  - Data sharing expectations for various funders
- Small-group instruction (labs, research teams, one-on-one)
  - Managing research data as a group
  - Storage and back-up for various scenarios
  - Managing data during your research project
- Course-integrated instruction
  - General data management principles
  - Human subjects research data management
  - And so on!



Data management bootcamps

- For early-stage graduate students
- Focused on data management basics
  - file and folder organization
  - risk management
  - storage and back-up
  - documentation
- Also covered
  - citation managers
  - workflows and tools for backing up and versioning
  - federal data sharing mandates
  - publishing with data.
- Offered annually since 2013

Data management bootcamps

Data management bootcamps for physical sciences

- 42 attendees, all types of grad students, but skewed to earlier years
- Topics covered included
  - Why we should care about data management
  - What data curators do & what researchers need to know
- Breakout sessions
  - Citation managers
  - Cleaning up existing data
  - Backing up data
  - File naming, versions, version control (and when to delete something!)
  - Taking Better Care of Your Lab Notebooks (& physical samples)

Data Repository for U of M

drum.umn.edu

Search the Data Repository

Go

The Data Repository for University of Minnesota (DRUM)

DRUM is a publicly available collection of digital research data generated by U of M researchers, students, and staff. Anyone can search and download the data housed in the repository, instantly or by request.

The Data Repository accepts submissions from University affiliates for digital archiving and access. Learn more about depositing to the Data Repository and other services to manage your data.

Upload to the Data Repository >

*U of M affiliates only | How to submit

How to Upload

1. Prepare Data

Data should be free of identifying or sensitive information and include adequate documentation. Not sure? Contact us for help!

Features

Access Options

Choose to make your data immediately accessible to everyone, or restrict access to your data for up to two years.

Our Services

Data Management Plan Assistance

We offer personalized assistance for drafting your next grant's Data Management Plan. Contact us for assistance during your planning process.

Data Curation Network

- National collaboration
- 19 institutions
- 50+ curators
- UMN is the financial “house”

Mikala Narlock, Director
mnarlock@umn.edu

Data Curation Network

Mass Spectrometry Primer

Authors: Brian Westra, Univers[illegible]; Ye Li, Massachusetts Institute of Technology; Nick Ruhs, Florida State University; Leah Rae McEwen, Cornell University

DCN Mentors: Lisa Johnston and Wendy Kozlowski

Affiliate Contributors (external peer reviewers): Meghan Lafferty

https://github.com/DataCurationNetwork/data-primers/blob/master/Mass%20Spectrometry%20Primer/massspec.md

Table 1. Proprietary mass spectrometry file formats.[5]

Topic	Description	Description
Agilent	.AEV, .ASR	ASCII Report format (for Analytical Studio Reviewer)
Bruker	.BAF	instrument data format
Agilent Bruker	.D (folder)	Agilent MassHunter, Agilent ChemStation, or Bruker BAF/YEP/TDF data format
Chromtech Finnigan*** VG	.DAT	Finnigan ITDS file format; MAT95 instrument data format MassLab data format
Bruker	.FID	instrument data format
ION-TOF	.ita	analysis data
ION-TOF	.itm	raw measurement data
Shimadzu	.lcd	QQQ/QTOF instrument data format
Finnigan***	.MS	ITS40 instrument data format
Waters	.PKL	MassLynx peak list format
Shimadzu	.qgd	instrument data format
Shimadzu	.QGD	GCMSSolution format
Physical Electronics/ULVAC-PHI	.raw*	raw measurement data
Thermo PerkinElmer	.RAW*	Thermo Xcalibur PerkinElmer TurboMass
Micromass**/Waters	.RAW* (folder)	Waters MassLynx
Bruker/Varian	.SMS	instrument data format
Shimadzu	.spc	library data format
ABI/Sciex	.t2d	4700 and 4800 file format
Physical Electronics/ULVAC-PH	.tdc	spectrum data
Bruker	.TDF	timsTOF instrument data format
ABI/Sciex	.WIFF	instrument data format
Bruker/Varian	.XMS	instrument data format
Agilent/Bruker	.YEP	instrument data format
ACD/Labs	*.spectrus	Imports LC/MS and GC/MS data from most major instrument vendors listed here

Data Curation Network

Mikala Narlock, Director
mnarlock@umn.edu

CURATE(D) Model for Data Curation

C **Check** files and read documentation (risk mitigation, file inventory, appraisal/selection)

U **Understand** the data (or try to), if not… (run files/environment, QA/QC issues, readme)

R **Request** missing information or changes (tracking provenance of any changes and why)

A **Augment** metadata for findability (DOIs, metadata standards, discoverability)

T **Transform** file formats for reuse (data preservation, conversion tools, data viz)

E **Evaluate** for FAIRness (licenses, responsibility standards, metrics for tracking use)

D **Document** your curation activities (Curator Log, correspondence)

DCN CURATE(D) Steps

Curated checklist

Example: UMN Materials Research Science & Research Center

Data Repository for University of Minnesota (DRUM)

Each **MRSEC-supported publication for Award DMR-2011401** requires the archival of data sets to DRUM. The slideshow below (from the May 15, 2023 MRSEC data archival and DRUM training) gives an overview of why and how MRSEC requires data to be submitted.

Download the MRSEC data archival and DRUM training slides (pdf).

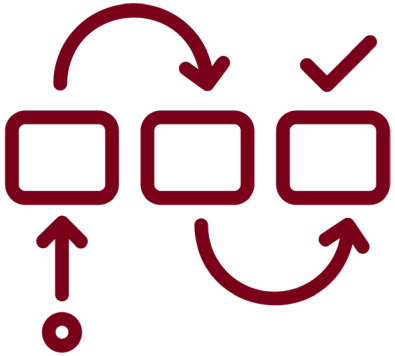
Download detailed instructions on uploading data sets (pdf).

Download the *readme.txt* template (pdf).

Collection:
https://conservancy.umn.edu/handle/11299/214080

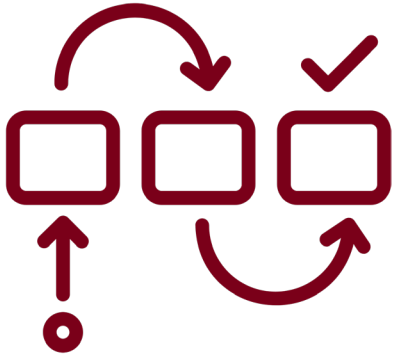
Workflow:
https://mrsec.umn.edu/research/resources/

Workflow on researcher's side



- Confirm you have permission to publish all the data contained in the files.
- Locate data files for each figure, supporting figure, table, scheme, etc. in the paper (this includes .jpgs of Chemdraw files used in figures).
- Open each of the files to make sure they are operable and review contents.
- Make sure the data files are thoroughly described.
- If applicable, export a .jpg of each Chemdraw file to include with the submission.
- Organize the files into a logical directory structure.

Workflow on researcher's side



- Make a copy of the README.txt template and fill it out.
  - Spell out all acronyms/abbreviations.
  - Include your data collection methodology in the README.
  - Describe any important relationships between files.
  - Name each piece of software that would be required to access your data.
  - Generate the directory structure, copy into README template.

- Fill out the DRUM upload form and upload your data.
- Work with the DRUM Curator to finalize your dataset. Once completed, you'll receive a DOI to add to your manuscript.

Workflow on DRUM side

- Author/submitter deposits dataset
- Coordinator looks over, accepts, assigns to data curator
- Curator runs through C-U-R steps of the CURATE(D) steps
  - Check files/code & read documentation
  - Understand data (or try to)
  - Request missing information or changes
- Curator emails author with questions and recommendations
  - Curator waits to hear back from author before proceeding
  - This can go back and forth quickly or slowly

Workflow on DRUM side

- Curator runs through A-T-E-(D) steps of the CURATE(D) steps
  - Augment metadata
    - Curator makes changes, uploads changes, edits README
    - Curator mints DOI
  - Transform file formats
  - Evaluate for FAIRness

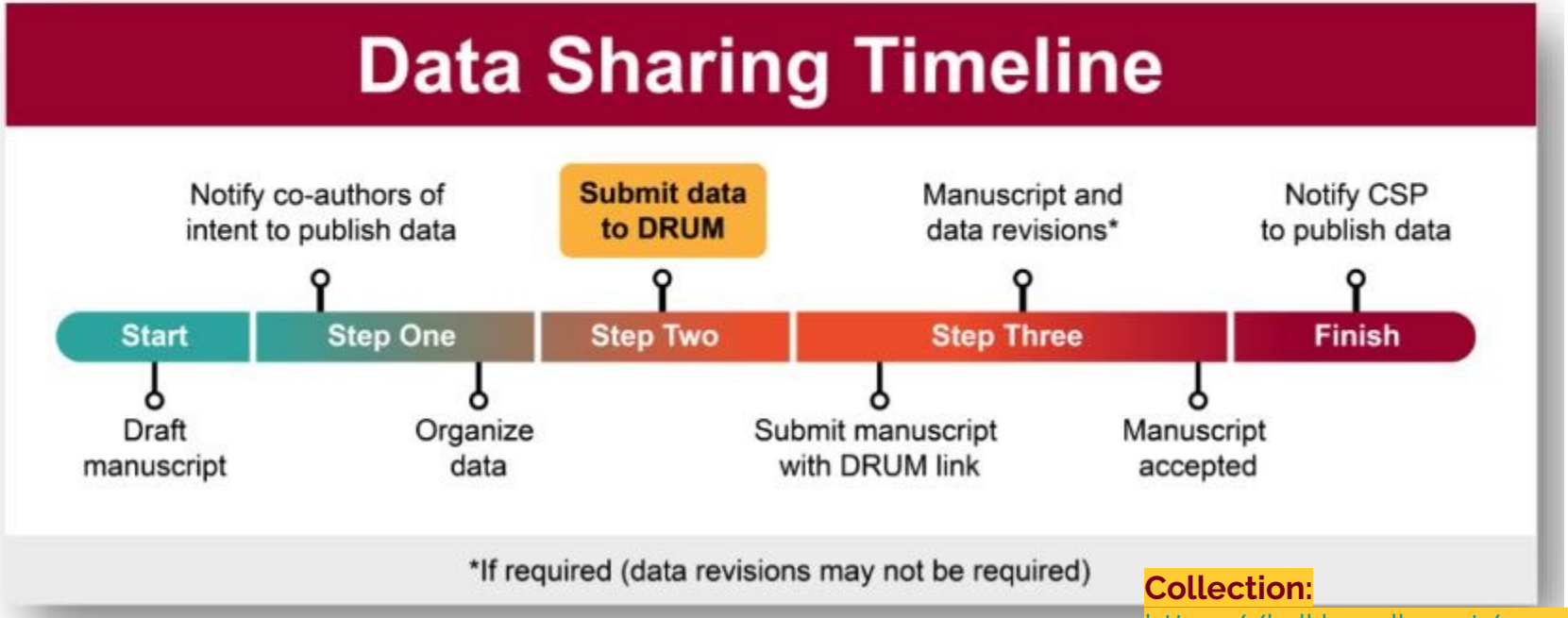
- Curator sends author email with list of changes made during curation & saying that curation is complete

Example: Center for Sustainable Polymers

CSP Data Curation Process

Document Updated: 09, December 2022

NSF Center for Sustainable Polymers



Collection: https://hdl.handle.net/11299/193178

Example: Center for Sustainable Polymers

Supporting data for Chemically Recyclable Linear and Branched Polyethylenes Synthesized from Stoichiometrically Self-balanced Telechelic Polyethylenes

Readme.txt	Description of the files	24.15Kb	Text file
Chemdraw.zip	Reaction schemes	166.2Kb	application/zip
Raw Differential Scanning Calorimetry Data.zip	DSC data	3.749Mb	application/zip
Raw Infrared Specta.zip	IR data	159.7Kb	application/zip
Raw Matrix Assisted Laser Desorption Ionization Data.zip	MALDI data	673.2Kb	application/zip
Raw Nulear Magnetic Resonance Spectra.zip	NMR data	6.900Mb	application/zip
Raw Size-Exclusion Chromatography Traces.zip	SEC data	1.950Mb	application/zip
Raw Small Angle X-Ray Scattering Data.zip	SAXS data	267.4Kb	application/zip
Raw Tensile Data.zip	Tensile data	2.925Mb	application/zip
Raw Thermogravimetric Analysis Data.zip	TGA data	714.6Kb	application/zip
Raw Water Vapor Permeability Data.zip	WVP data	73.01Kb	application/zip
Raw Wide Angle X-Ray Scattering Data.zip	WAXS data	221.8Kb	application/zip
Raw Extensional Rheology Data.zip	Extentional rheology data	93.71Kb	application/zip

Creating buy-in

- Started with a high profile center on campus and they served as advocates
- Make it as easy as possible
- Create detailed instructions and a logical workflow
  - Is there admin that can help?
- Working on training the trainer to create a culture of sharing

What we've learned

- We are all figuring this out together
- It is possible to share and disseminate your work openly
- It gets easier once we figure out a workflow, becomes habitual
  - Less work for both of us
  - Less back-and-forth and unexpected questions
  - Sets expectations for everyone, including new hires/graduate students
- We are growing the next generation of scholars

Challenges

- “Big data”
  - Storage & storage cost
  - Access for end-user
- Restricting access
  - Patents & intellectual property concerns
- Expectation of growth
  - Only working with a small amount of chemistry dept
  - DRUM staff
  - Staff fielding repository/curation consultations

Thank you to our UMN team!

- **Alicia Hofelich Mohr**, Research Support Services Coordinator
- **Shanda Hunt**, Research Data Education & Outreach Librarian
- **Kent Gerber,** University Data Archivist
- **Melinda Kernick**, Spatial Data Analyst and Curator
- **Meghan Lafferty,** Librarian for Chemistry, Chemical Engineering, & Materials Science
- **Allison Langham-Putrow**, Scholarly Communications Librarian
- **Wanda Marsolek**, Data Curation Librarian
- **Mikala Narlock**, Director, Data Curation Network