

国際シンポジウム「大学における研究データサービスの導入と展開」

Carlson, Jacob

ニューヨーク州立大学バッファロー校図書館：研究・コレクション・アウトリーチ担当アソシエート・ユニバーシティ・ライブラリアン

Rice, Robin

エディンバラ大学図書館・大学コレクション部門：データライブラリアン兼研究データ支援サービス部長

Smith, Simon

エディンバラ大学図書館・大学コレクション部門：研究データ支援オフィサー

竹内, 比呂也

千葉大学：副学長

他

<https://doi.org/10.15017/7238303>

出版情報：2024-10-11. University of Edinburgh

バージョン：

権利関係：

DEVELOPMENTS IN RESEARCH DATA SERVICES IN ACADEMIC LIBRARIES OVER TIME

Oct 12, 2024

Jake Carlson

Associate University Librarian for
Research, Collections and Outreach



Agenda

- Background – Why Share Data and Why are Librarians Involved?
- Case Studies on different approaches to data services from my experiences:
 - Data Services Specialist at Purdue University (2007-2014) 
 - Director of Deep Blue Repository and Research Data Services at University of Michigan (2014-2023) 
 - Associate University Librarian for Research, Collections and Outreach at University at Buffalo (2023-) 

What are Data?



Office of Management and Budget

- (i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings...
 - OMB Circular A-110

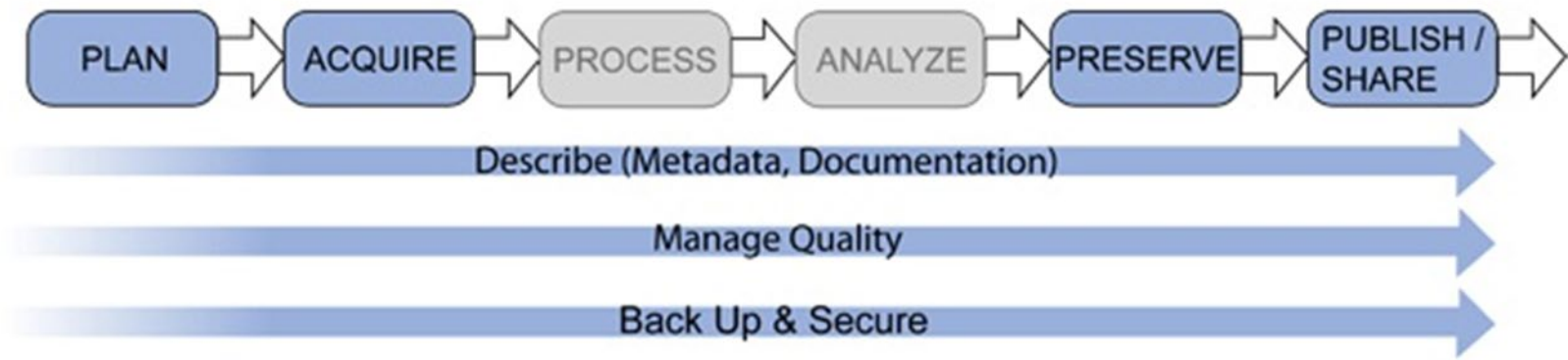
What are Data?

“I prefer to think of research data as information structured by methodology and organized in digital products that are used as evidence in the research process.”

Chuck Humphrey, University of Alberta

<http://preservingresearchdataincanada.net/2013/07/23/data-a-rose-by-any-other-name/>

Data Lifecycle Models



USGS Data Lifecycle Diagram

<https://pubs.usgs.gov/of/2013/1265/pdf/of2013-1265.pdf>

Why are we Talking About Data?

Data are increasingly seen as having value as information objects themselves, not just in the resulting research findings.

- Push for access to results of tax-payer funded research.
 - **Data management / sharing plan requirement**
- Trust / reproducibility issues in research.
 - **Publishers requiring access to data**
- Researchers' recognizing benefits of sharing data.
 - **Scholarly societies issuing statements supporting data sharing**

Data Sharing: Benefits

“We found that cancer clinical trials which share their microarray data were cited about 70% more frequently than clinical trials which do not. ”

Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

“We find strong and consistent evidence that data sharing, both formal and informal, increases research productivity... Data archiving... yields the greatest returns on investment with research productivity (number of publications)...”

Pienta AM, Alter GC, Lyle JA (2010) The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. presented at “The Organisation, Economics and Policy of Scientific Research” workshop, Torino, Italy. <http://hdl.handle.net/2027.42/78307>

Why Librarians? nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾ [Subscribe](#)

[nature](#) > [world view](#) > article

WORLD VIEW | 05 September 2022

Without appropriate metadata, data-sharing mandates are pointless



Funders and investigators must demand appropriate metadata standards to take data from foul to FAIR.

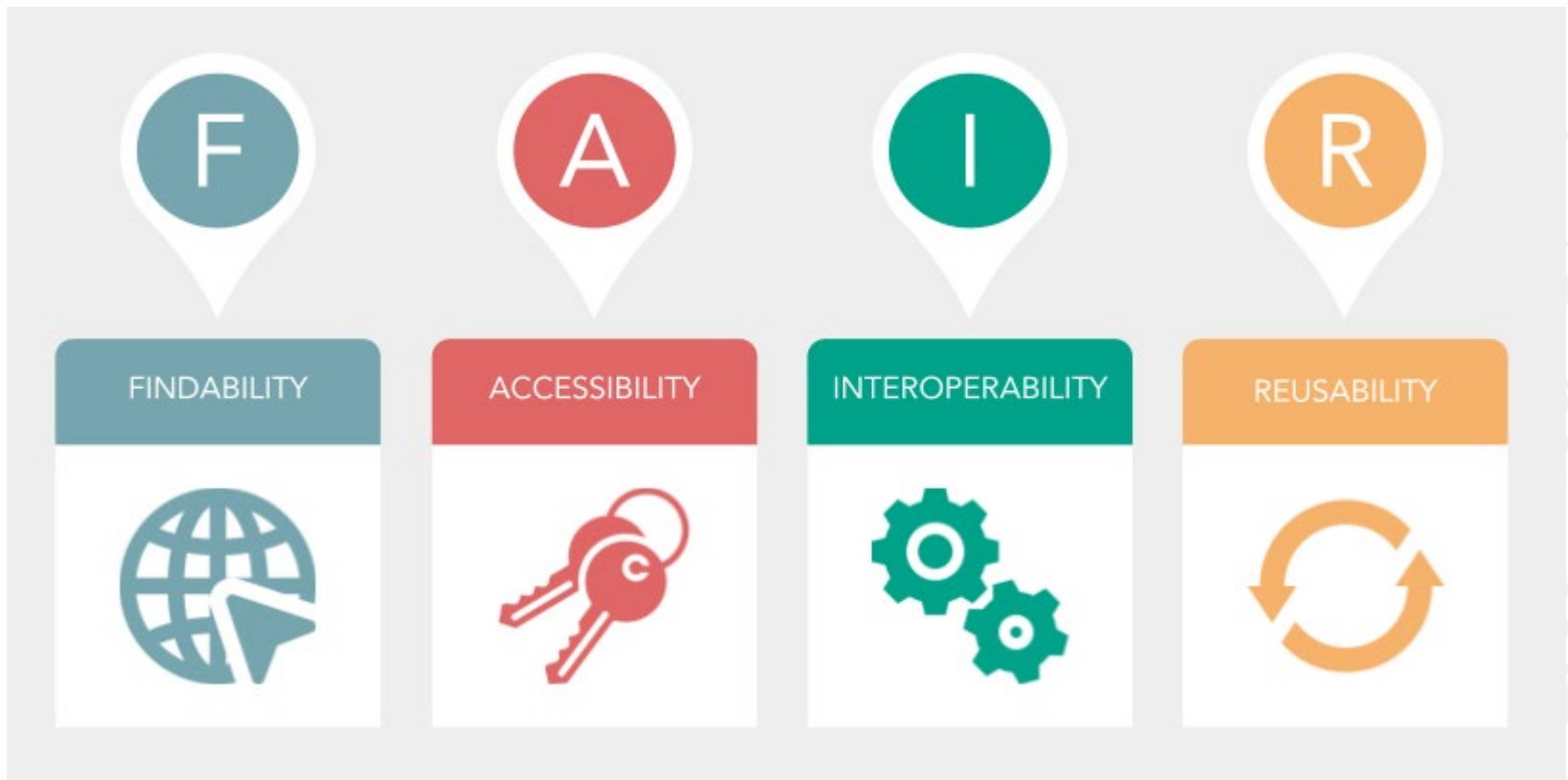
By [Mark A. Musen](#) 



“But just getting those data sets online will not bring anticipated benefits: few data sets will really be FAIR, because most will be unfindable. What’s needed are policies and infrastructure to organize metadata.” 8

<https://www.nature.com/articles/d41586-022-02820-7>

FAIR Data



Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018

PURDUE UNIVERSITY

Exploration and
Research Partnerships

2007 - 2014



Background on the Purdue Libraries and Data

2004: Purdue Interdisciplinary Research Initiative revealed many data needs on campus



What faculty said...

- Not sure how or whether to share data
- Lack of time to organize data sets
- Need help describing data for discovery
- Want to find new ways to manage data
- Need help archiving data sets/collections

Applying Library Science to Research Data



- Reference
 - How could we develop a deep understanding of researcher needs for managing and sharing their data?
- Information Literacy
 - What training would be needed for researchers to be successful?
- Collection Development
 - How could data be organized, described and discoverable for others to access and use?
 - Could libraries develop and support collections of data in similar ways that we support special collections?

Image: Michael Mandiberg "Database" under CC BY-SA 3.0.
<http://www.mandiberg.com/category/works/page/2/>

Creating a Libraries Based Research Center

2006: Founded the Distributed Data Curation Center (D2C2) to further investigations, organize research, and leverage collaborations



Grant: Data Curation Profiles



2007: Data Curation Profiles Project

Goals of the project:

- A better understanding of the practices, attitudes and needs of researchers in managing and sharing their data.
- Identify possible roles for librarians and the skill sets they will need to facilitate data sharing and curation.
- Develop “data curation profiles” – a tool for librarians and others to gather information on researcher needs for their data and to inform curation services.
- Purdue: Brandt—PI, **Carlson—Proj Mgr**, Witt—coPI;
UIUC: Palmer—co-PI, Cragin—Proj Mgr

Assessing Researcher Needs

Data Curation Profiles Toolkit



A Data Curation Profile (DCP) is a means to determine:

- Information about a particular data set
- What a researcher is doing to manage / curate the data set
- What a researcher would like to do with the data.

Identifying the Lifecycle of a Specific Data Set

Data Stage	Output	# of Files / Typical Size	Format	Other / Notes
Primary Data				
Harvest	Field data – Plant yields	1 / 15-20 kb	Excel	In addition to collecting data points from the field, this stage includes harvesting plant tissues, taking soil samples and other physical specimens for processing.
Lab Work	Initially, data points are prepared for analysis. Ultimately, a "master file" of data is generated.	5 / 125 -175kb	Excel	Data are reviewed and discussed with faculty advisor. Questionable data are re-processed. Accepted data are placed into the "master file"
Statistical Analysis	Results of experiments and computations performed	Minitab – 1 / "large"	Minitab, Sigma Plot, Excel, Word	Multiple experiments and computations live within one Minitab file. Some calculations are done in Excel. Sigma Plot files are generated to represent the data graphically.
Publication	Sigma Plot tables integrated into MS Word	Under development	Sigma Plot, MS Word	The exact methods of representing this data in publications are still a work in progress.

Sections of the Data Curation Profile

- Overview of the Research
 - Focus
 - Intended Audience
 - Funding Source
- Data Kinds and Lifecycle Stages
 - Narrative
 - Target Data for Sharing
 - Re-Use Value
- Intellectual Property
- Organization and Description
- Ingest
- Access
- Discovery
- Associated Tools
- Interoperability
- Measuring Impact
- Data Management
- Preservation

Use of the Data Curation Profiles

- Provided a guide for discussing data with researchers
- Gave insight into areas of attention in data management
- Helped assess information needs related to data collections
- Gave insight into differences between data in various disciplines
- Helped identify possible data services

Human Cell Defense Systems

The categories in the "data stage" column are shaded in gray. The data specifically discussed in this profile are shaded in gray.

Data Stage	Output	File Format
Record-Keeping	Preliminary results of experiments	NA
Processed	Description of experiments / data	NA
Analysis	Consolidated files for statistical analysis	NA
Published	Statistical analysis results of the statistical analysis	Between 20-50 kb .xls, .sas .xls

Note: The data in the "Analysis" column are shaded in gray. The data specifically discussed in this profile are shaded in gray.

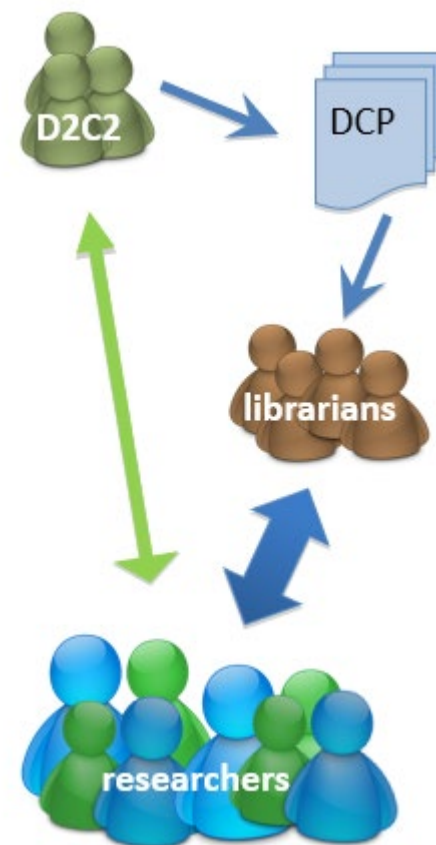
Target data for sharing: The target data for sharing is the summarized data – the data that has been analyzed and the results of the analysis. However, some kind of raw data might also be useful.

Useful value of the data: The scientist notes that the data with the most value to other researchers is the summarized data. However, some kind of raw data might also be useful.

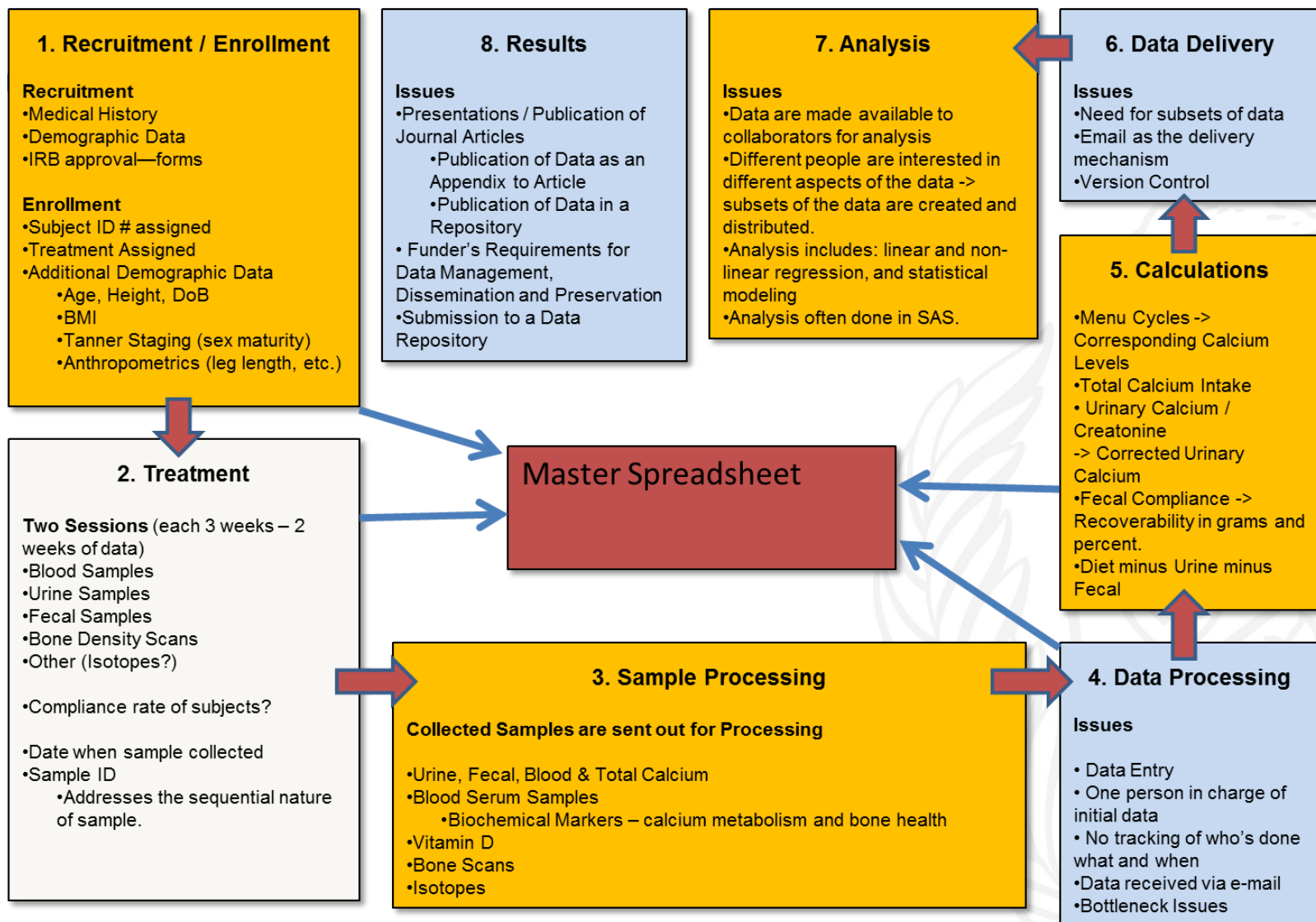
Contextual narrative: The data set is considered to be static and the data set is primarily composed of digital data. The data set is not provide an exact size of from 20 to 50 KB. He analyzed

How did Purdue use the Data Curation Profiles?

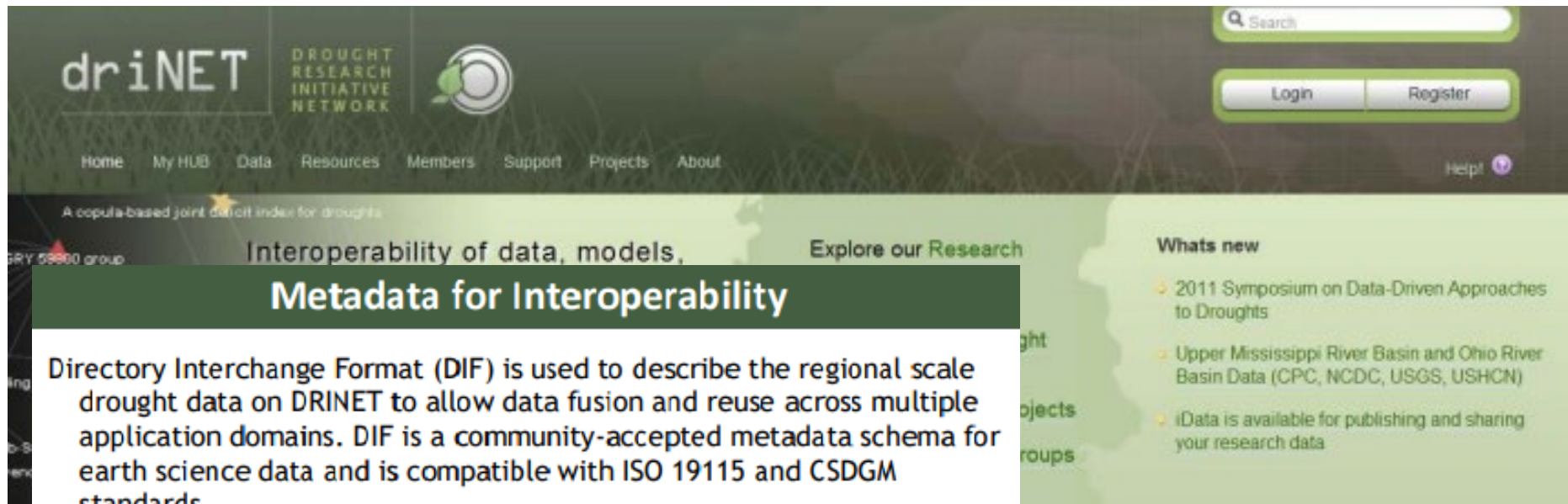
- **Strategy 1:** D2C2 provides expert consultant/collaborator reputation with researchers.
- **Strategy 2:** D2C2 creates tools to help liaison librarians engage researchers; Data Curation Profiles help give structure to conversations about data and facilitate information gathering.



Camp Calcium – Current Data Workflow



Grant: Drought Research Initiative Network



The screenshot shows the driNET website interface. At the top, there is a navigation bar with links: Home, My HUB, Data, Resources, Members, Support, Projects, and About. A search bar and Login/Register buttons are also present. The main content area features a large banner with the text "Interoperability of data, models, Metadata for Interoperability" and "Explore our Research". To the right, a "Whats new" section lists recent updates, including a 2011 Symposium on Data-Driven Approaches to Droughts and Upper Mississippi River Basin and Ohio River Basin Data.

Directory Interchange Format (DIF) is used to describe the regional scale drought data on DRINET to allow data fusion and reuse across multiple application domains. DIF is a community-accepted metadata schema for earth science data and is compatible with ISO 19115 and CSDGM standards.

```
<Entry_ID> 243912 </Entry_ID>
<Entry_Title>Real time data collection of water quality parameters for the Massie Creek,
Ohio</Entry_Title>
<Data_Set_Citation>
  <Dataset_Title>Real time data collection of water quality parameters for the Massie Creek,
Ohio</Entry_Title>
  <Dataset_Release_Date>02/22/2011</Dataset_Release_Date>
  <Dataset_Release_Place>auto - Purdue University, West Lafayette, IN</Dataset_Release_Place>
  <Dataset_Publisher>driNET</Dataset_Publisher>
</Parameters>
<Category>EARTH SCIENCE</Category>
<Topic>TERRESTRIAL HYDROSPHERE</Topic>
<Term>WATER QUALITY/WATER</Term>
<Variable_Level_1>Water Level</Variable_Level_1>
```



Directory Interchange Format (DIF) Writer's Guide, 2010.
Global Change Master Directory. National Aeronautics and Space
Administration. [<http://gcmd.nasa.gov/User/difguide/>].



National Science Foundation
WHERE DISCOVERIES BEGIN

1846

Findings from the DCP Research

I: Is there a need for education in data management or curation for graduate students...?

Fac: Absolutely, God yes...

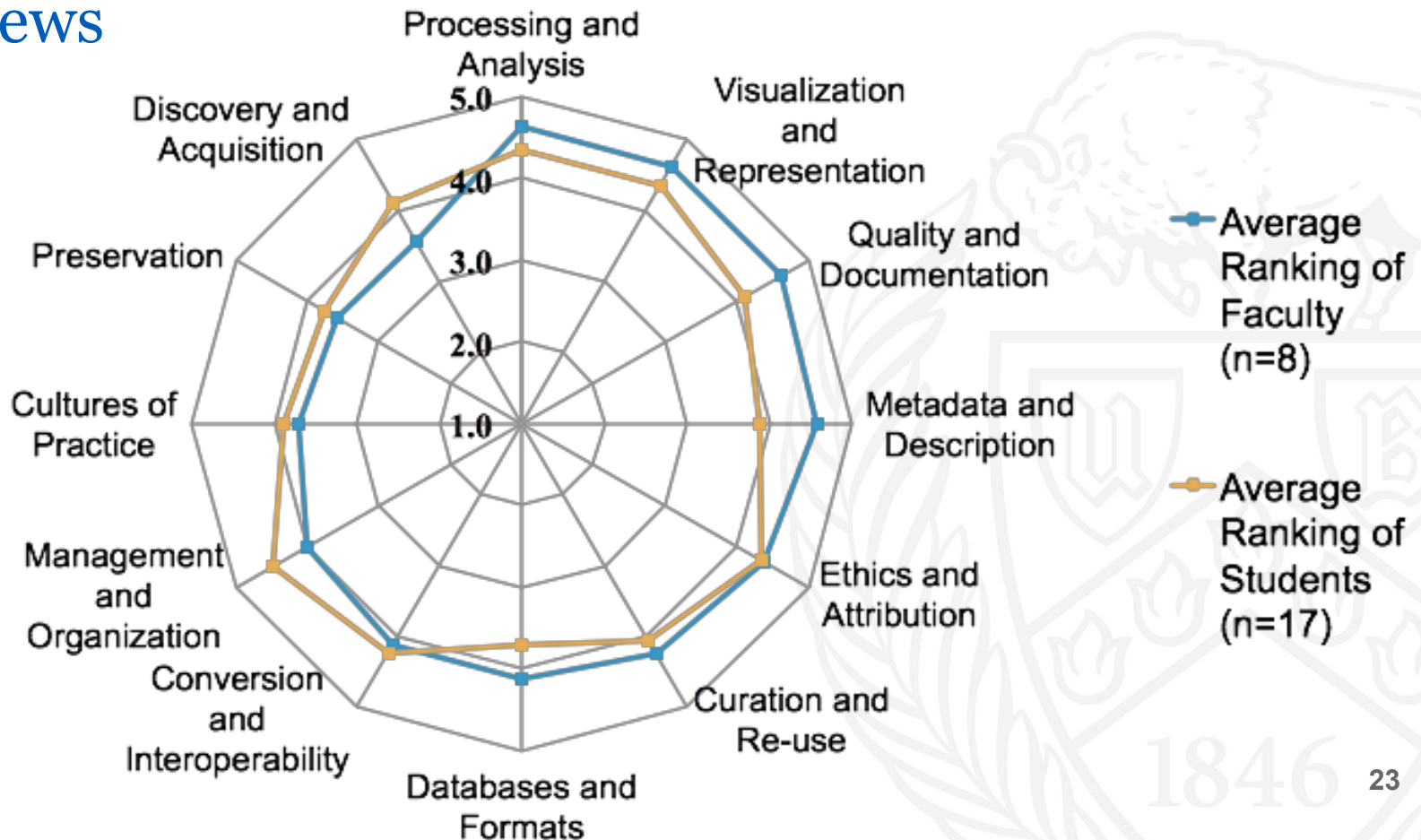
I: So, what would that education program look like... What kind of things would be taught?

Fac: Um, I don't really know actually, just how to do you manage data? Or you know, confidentiality things, ethics, probably um...I'm just throwing things out because I hadn't really thought that out very well.

Grant: Data Information Literacy (DIL)



Interviews



DIL Project



- Embedded Librarians



- Workshop Series



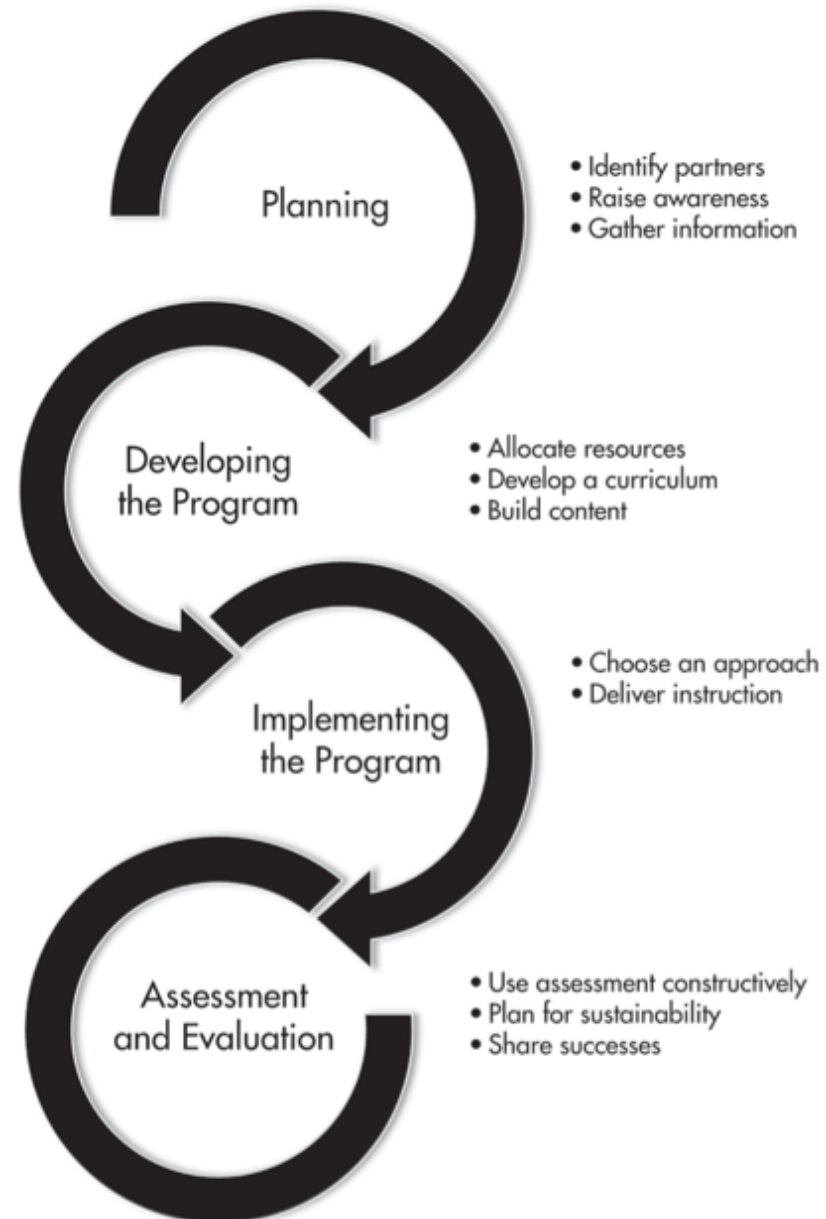
- 6 Week Course



- Educational Modules



- One-Shot Workshop



UNIVERSITY OF MICHIGAN

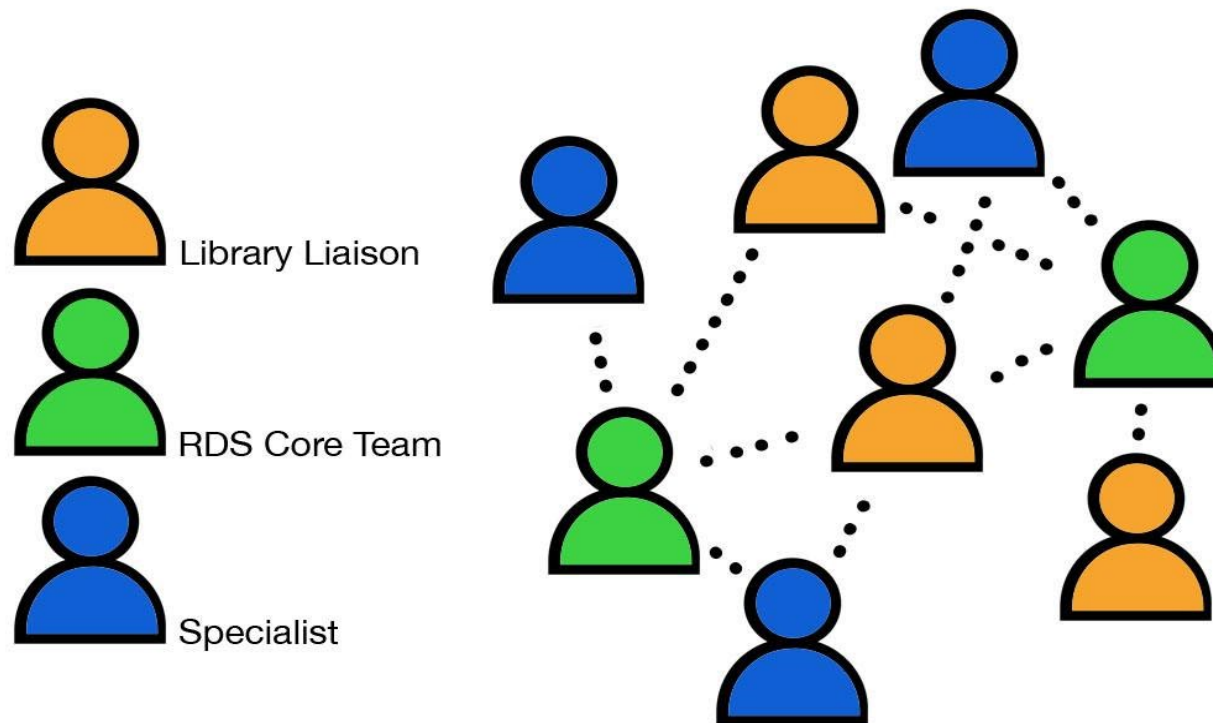
Program Development &
Building Infrastructure

2014 - 2023



Research Data Services Framework

The initial approach of the University of Michigan was to engage everyone in the library in developing its data services.



Developing U-M's Data Services



2014 – 2016: I led or supported several initiatives as a part of developing our data services program.

- Data Interviews
 - Coded to Identify Themes: “Researcher (x) needs (y) in order to accomplish (z)”
- Data Education Sessions & Workshops
 - Primarily for Training for Librarians and Staff to Work with Data
- “Data Bites” Discussion Forums
 - Information sharing
 - Opportunities to Delve into Deeper Issues
- Data Profile Statements (Do / Collaborate / Refer)

Research Data Services



Data Management Planning: helping plan for managing, sharing and curating data and develop Data Management Plans (DMPs) that meet funder requirements.



Discovery & Access: assisting in discovering, accessing, and acquiring different types of research materials, including data.



Data Organization & Management: helping researchers to understand, develop and apply strategies for organizing and managing their data.



Metadata & Documentation: locating standards for documentation that capture the details of generating, processing and analyzing data so it can be discovered, understood and reused.



Data Sharing & Publication: helping disseminate research data for discovery, access and reuse in ways that enable researchers to receive credit for their work.



Preservation: taking action to sustain the accessibility and scholarly value of data over time.

Deep Blue Data
grew out of the
RDS Initiative and
launched in 2016

Now Available!

**Deep Blue
Data**

U-M Library
Data Repository

Related Services at U-M



- Launched in 2006
- Over 150,000 works
 - Open Access Publications
 - Original Works
 - Archival Materials



Clark Library

- GIS, Maps
- Gov Docs
- Data Visualization



- Data Repository for the Social Science community

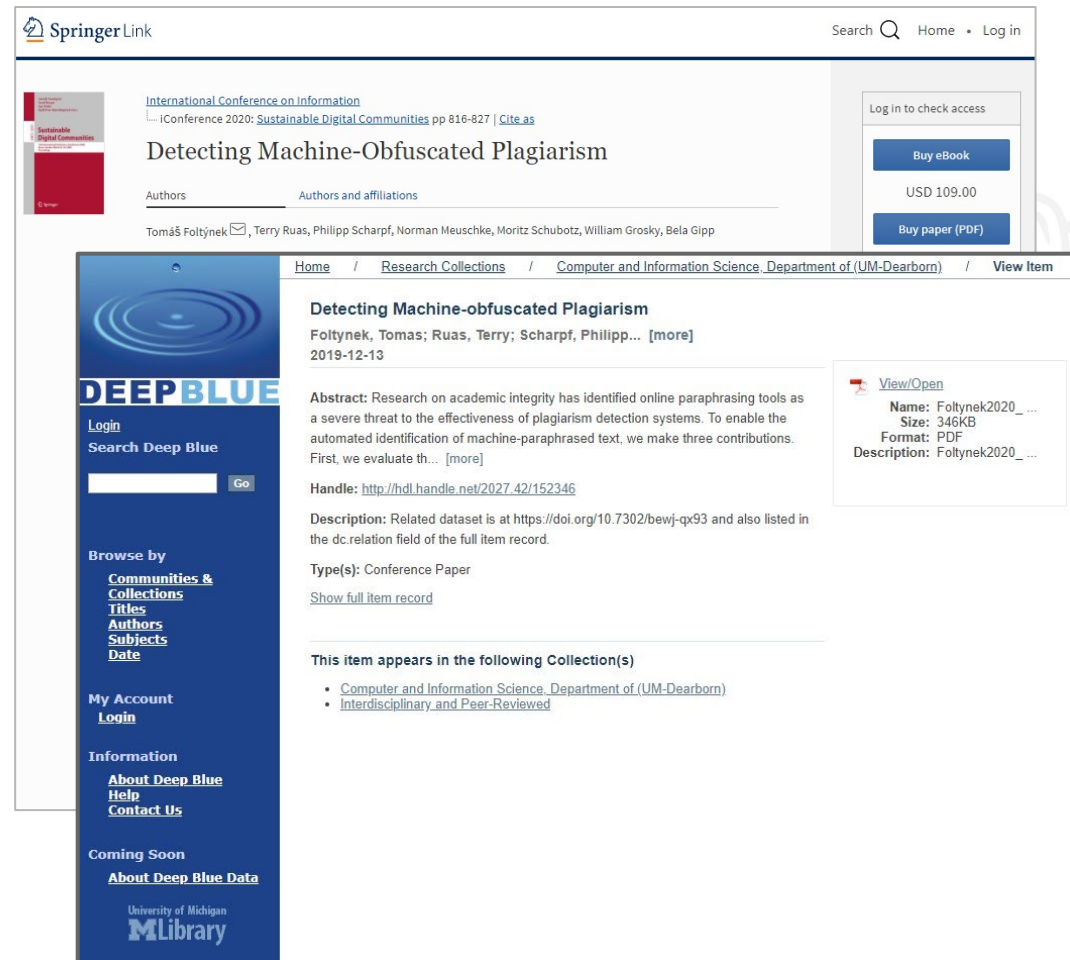
Bringing the Deep Blue Repositories Together

2018: The Research Data Services Department Took on Oversight of the Deep Blue Repository and Moved into the Publishing Department. We became Deep Blue Repositories and Research Data Services (DBRRDS).

- Better Support for Open Access (and Open Access Requirements)
- A Desire for Stronger Connections between the Repositories
- Long-Term Plans for Merging Repository Platforms
- “Publishing” was a Better Model for Data Sharing than “Collection Development”

Supporting Open Access in Deep Blue

Provide a means for the U-M community to make their work openly accessible to anyone in the world with an internet connection.



The screenshot displays a SpringerLink article page for the paper "Detecting Machine-Obfuscated Plagiarism" by Foltýnek, Tomas; Ruas, Terry; Scharpf, Philipp et al. The page includes a search bar, navigation links, and a sidebar with options to buy the eBook or paper. A blue overlay for the "DEEPBLUE" repository is positioned in the foreground, featuring a search bar, login options, and a list of collections where the item is available, including "Computer and Information Science, Department of (UM-Dearborn)" and "Interdisciplinary and Peer-Reviewed".

SpringerLink Search Home Log in

International Conference on Information
... IConference 2020: Sustainable Digital Communities pp 816-827 | Cite as

Detecting Machine-Obfuscated Plagiarism

Authors Authors and affiliations

Tomáš Foltýnek, Terry Ruas, Philipp Scharpf, Norman Meuschke, Moritz Schubotz, William Grosky, Bela Gipp

Log in to check access

Buy eBook

USD 109.00

Buy paper (PDF)

Home / Research Collections / Computer and Information Science, Department of (UM-Dearborn) / View Item

Detecting Machine-obfuscated Plagiarism

Foltýnek, Tomas; Ruas, Terry; Scharpf, Philipp... [more]
2019-12-13

Abstract: Research on academic integrity has identified online paraphrasing tools as a severe threat to the effectiveness of plagiarism detection systems. To enable the automated identification of machine-paraphrased text, we make three contributions. First, we evaluate th... [more]

Handle: <http://hdl.handle.net/2027.42/152346>

Description: Related dataset is at <https://doi.org/10.7302/bewj-qx93> and also listed in the dc.relation field of the full item record.

Type(s): Conference Paper

[Show full item record](#)

This item appears in the following Collection(s)

- Computer and Information Science, Department of (UM-Dearborn)
- Interdisciplinary and Peer-Reviewed

[View/Open](#)

Name: Foltýnek2020_...
Size: 346KB
Format: PDF
Description: Foltýnek2020_...

DEEPBLUE

Login

Search Deep Blue

Go

Browse by

- Communities & Collections
- Titles
- Authors
- Subjects
- Date

My Account

Login

Information

- About Deep Blue
- Help
- Contact Us

Coming Soon

About Deep Blue Data

University of Michigan
MLibrary

Supporting Open Data in Deep Blue Data

Publish content of scholarly or educational value, research data in particular, that may not have a defined path for dissemination and might otherwise be unavailable or lost.



Detecting Machine-obfuscated Plagiarism



Creator: [Foltýnek, Tomas](#), [Ruas, Terry](#), [Scharpf, Philipp](#), [Meuschke, Norman](#), [Schubotz, Moritz](#), [Grosky, William](#), and [Gipp, Bela](#)

Description: This data set is comprised of multiple folders. The corpus folder contains raw text used for training and testing in two splits, "document" and "paragraph". The Spun documents and paragraphs are generated using the SpinBot tool <https://spinbot.com/API...> [\[more\]](#)

Keyword: [paraphrase detection](#), [plagiarism detection](#), [document classification](#), and [word embeddings](#)

Citation to relate... Foltýnek, T. & Ruas, T. & Scharpf, P. & Meuschke, N. & Schubotz, M. & Grosky, W. & Gipp, B., "Detecting Machine-obfuscated Plagiarism," in Sustainable Digital Communities, vol. 12051 LNCS, Springer, 2020, pp. 816–827. [\[C\]](#)

Files (Count: 6; Size: 2.8 GB)

	Title	Original Upload	Last Modified	File Size	Access	Actions
	README.txt	2019-12-04	2019-12-13	4.77 KB	Open Access	Select an action ▾
	corpus.zip	2019-12-04	2019-12-10	213 MB	Open Access	Select an action ▾
	Human_judgement.zip	2019-12-16	2019-12-16	177 KB	Open Access	Select an action ▾
	models.zip	2019-12-04	2019-12-04	330 MB	Open Access	Select an action ▾
	vectors.test.zip	2019-12-04	2019-12-10	718 MB	Open Access	Select an action ▾

Challenges of Data Curation

It's not enough to deposit data into a repository. Data need to be curated to enhance their value to others through making them as FAIR as possible. But...

- How to Scale Local Data Curation Services?
- What Kind of Data Curation Expertise is Needed:
 - To Handle Different Types / Formats of Data? (Tabular, Survey, GIS, Images, Audio, Code etc.)
 - To Handle Different Fields? (Genomic Sequences, Chemical Spectra, Bioinformatics, etc.)
- Could we Better Leverage Our Expertise to Develop Specializations in Curating Certain Kinds of Data?



Grant: Building a Data Curation Network



2016: Six institutions form the “Data Curation Network” to explore how we could collaboratively share data curation staff across member institutions with financial support from the Alfred P Sloan Foundation

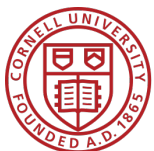


UNIVERSITY OF MINNESOTA



PennState

I ILLINOIS



Cornell University



Washington
University in St. Louis



The CURATED Steps

Checklist of CURATED Steps Performed on Deposited Data by DCN Members

- 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)

Data Curation Workflow (U-M Local)

Deposit Your Work



Metadata Entry



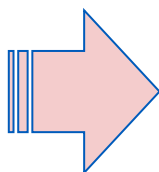
Upload Data Files



Terms of Use



Submit



Check files

Understand documentation

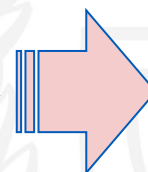
Request missing information

Augment the submission

Transform the format

Evaluate for FAIRness

Document throughout



Deep Blue Data

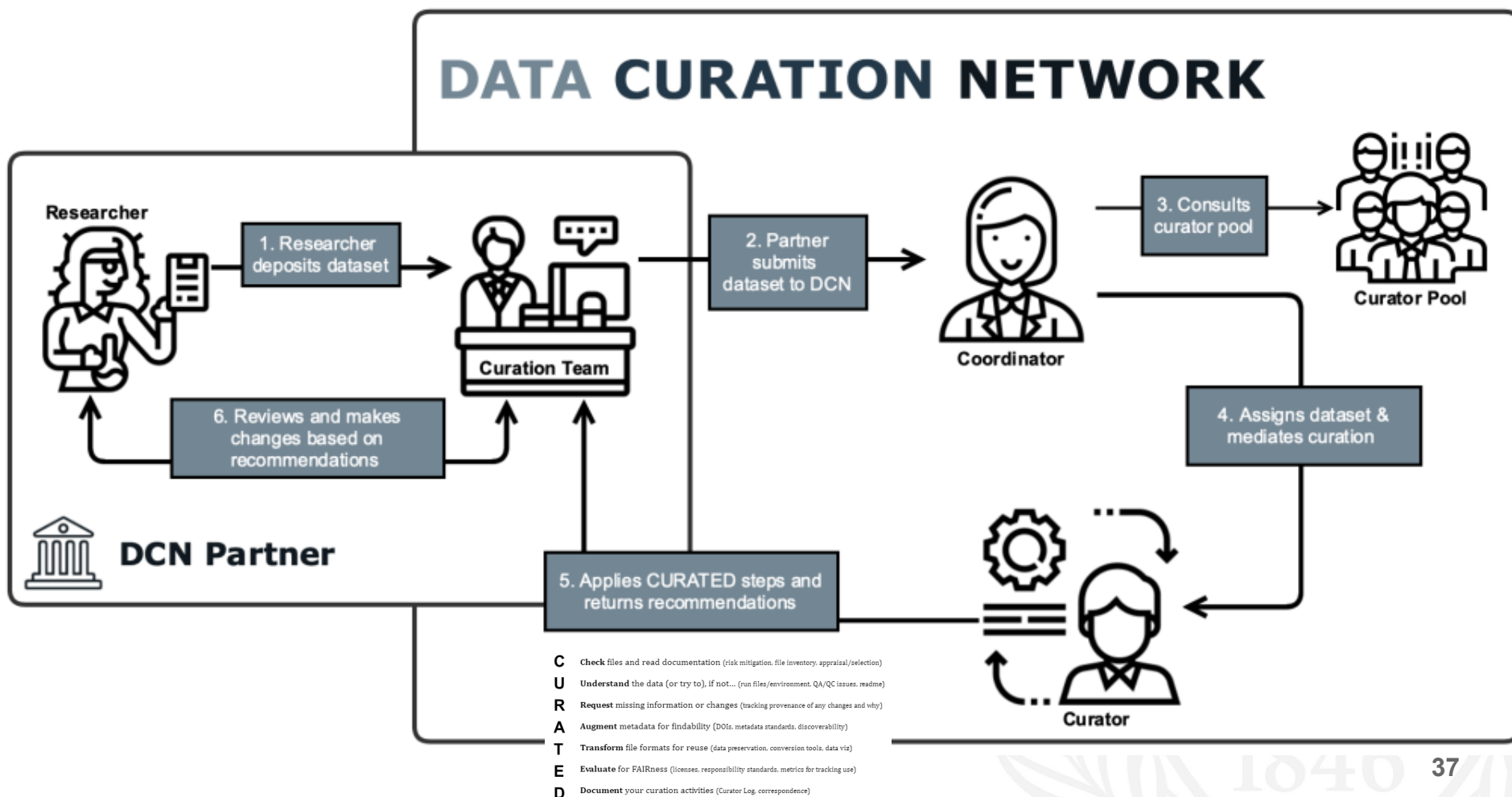
Enter search terms

Work Description

Title: Climate Action Plan adoption for 176 U.S. cities, 2010-2019

Attribute	Value
Methodology	A manual review of municipal government adoption of climate action plans was performed by checking individual city government websites. Google search terms "city name?" + "climate action plan" were used as the primary search method. The full list of cities included in the search comprise over 1,000 ... (more)
Description	Time series dataset of adoption by year of climate action plans by 177 U.S. cities, 2010-2019, with links to plans included. This dataset is intended for use by researchers and practitioners investigating both individual climate action plans and time series patterns of adoption at the municipal level. ... (more)
Creator	Benjamin Leffell
Depositor	bluff@umich.edu
Contact information	bluff@umich.edu
Discipline	Social Sciences Government, Politics and Law
Keyword	climate change climate action plan municipal

Data Curation Workflow (DCN)



DCN's Data Curation Primers

DATA CURATION NETWORK

About ▾ Resources ▾ Community ▾ Education ▾ Research ▾

[Human Participants Data Essentials Primer](#)

Creators: Jen Darragh, Alicia Hofelich Mohr, Shanda Hunt, Rachel Woodbrook, Dave Fearon, Jennifer Moore and Hannah Hadley

[Interdisciplinary and Highly Collaborative Research \(IHCR\) Data Primer](#)

Creators: Inna Kouper, Andrew M. Johnson, Jordan Wrigley, and Aditya Ranganath
Mentors: Neggin Keshavarzian and Mikala Narlock

[ISO Images Primer](#)

Creators: Kate Barron and Jonathan Bohan
Mentor: Cynthia Hudson Vitale

[Jupyter Notebooks Primer](#)

Creators: Daina Bouquin, Sophie Hou, Matthew Benzing and Lee Wilson
Mentor: Susan Borda

[Mass Spectrometry Primer](#)

Creators: Brian Westra, Ye Li, Nick Ruhs, and Leah Rae McEwen
Mentors: Lisa Johnston and Wendy Kozlowski

[Matlab Primer](#)

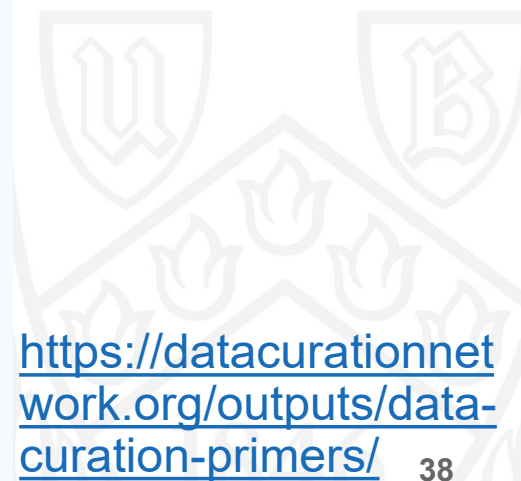
Creators: Sam Sciolla and Susan Borda

[Microsoft Access Primer](#)

Creators: Fernando Rios
Mentor: Dave Fearon

[Microsoft Excel Primer](#)

Creators: Greg Janée, Sandra Sawchuk and Ho Jung Yoo
Mentor: Wendy Kozlowski



<https://datacurationnetwork.org/outputs/data-curation-primers/>

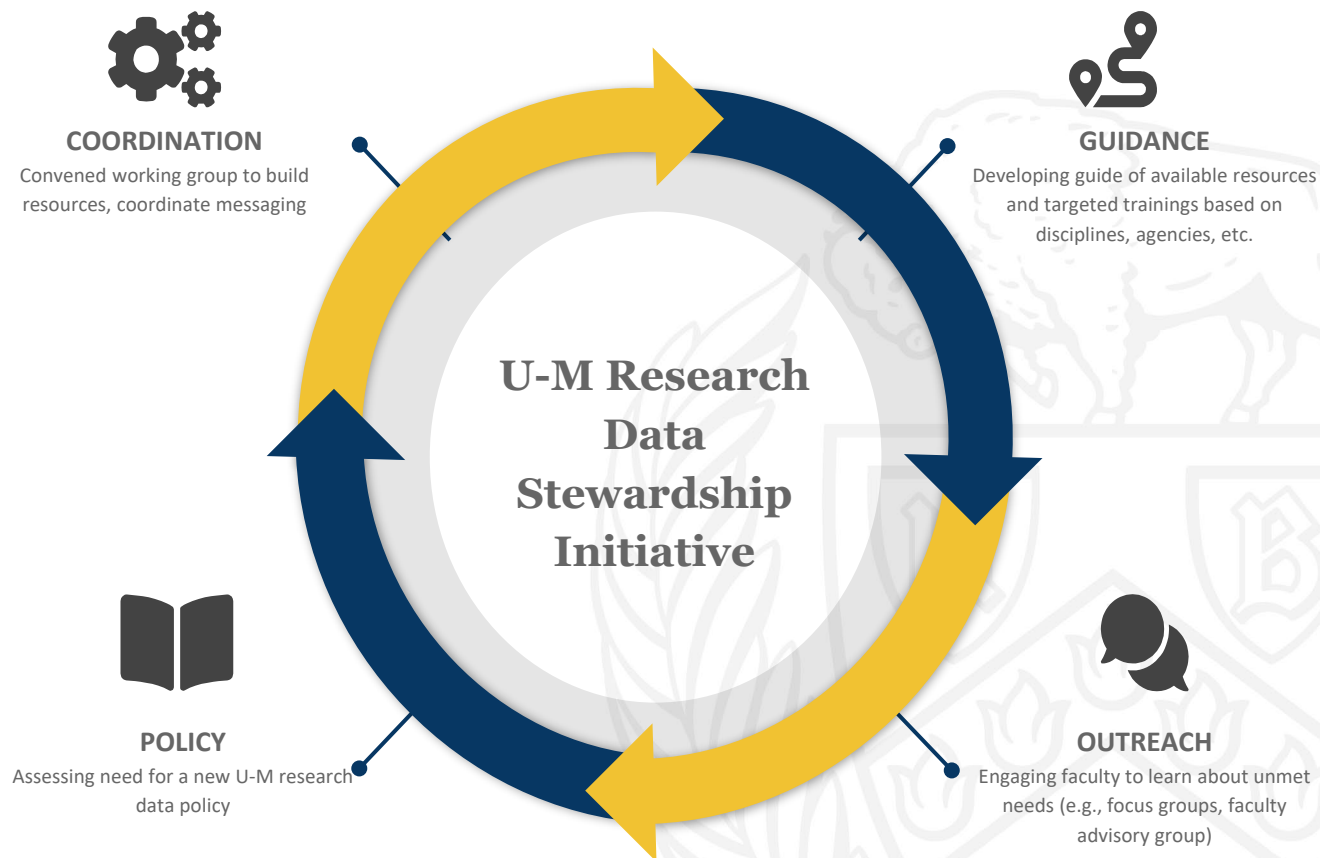
DCN Research Projects

- The Value of Data Curation Surveys (2021)
 - [From the Perspective of Repository Staff and Directors](#)
 - [From the Perspective of Researchers and Data Depositors](#)
- The Realities of Academic Data Sharing (RADS)
 - What service and cost models do institutions use to support research data management and sharing policies?
 - What are the direct expenses for institutions, particularly academic libraries, in implementing federally mandated data-sharing policies?
 - What costs do researchers incur to comply with funded research data-sharing policies?
 - [RADS Phase 1](#) (2021 – 2023) was funded by the National Science Foundation (NSF)
 - [RADS Phase 2](#) (2023 – current) is funded by the Institute of Museum and Library Services (IMLS)

U-M's Research Data Stewardship Initiative (UMDSI)

Membership:

- Office of Research
- Library
- IT
- Med School
- Sponsored Programs
- General Counsel
- Others



UNIVERSITY AT BUFFALO

Institutional Coordination &
Community Building

2023 - current



About the University at Buffalo (UB)



New York State's flagship university

19,118

undergrads

11,263

grads

\$474

million

Annual spending
on research

A few of our biggest grants

\$47.5 million

Science and Technology
Center Awards*

\$36.7 million

Clinical and Translational
Science Awards**

33 years

Continuous funding for
Women's Health Initiative**

**National Science Foundation | **National Institutes of Health*

Getting Started

2023: “UB recognizes that the university needs to proactively plan for and guarantee the storage, security, and access to research supported by both federal and state funding.”

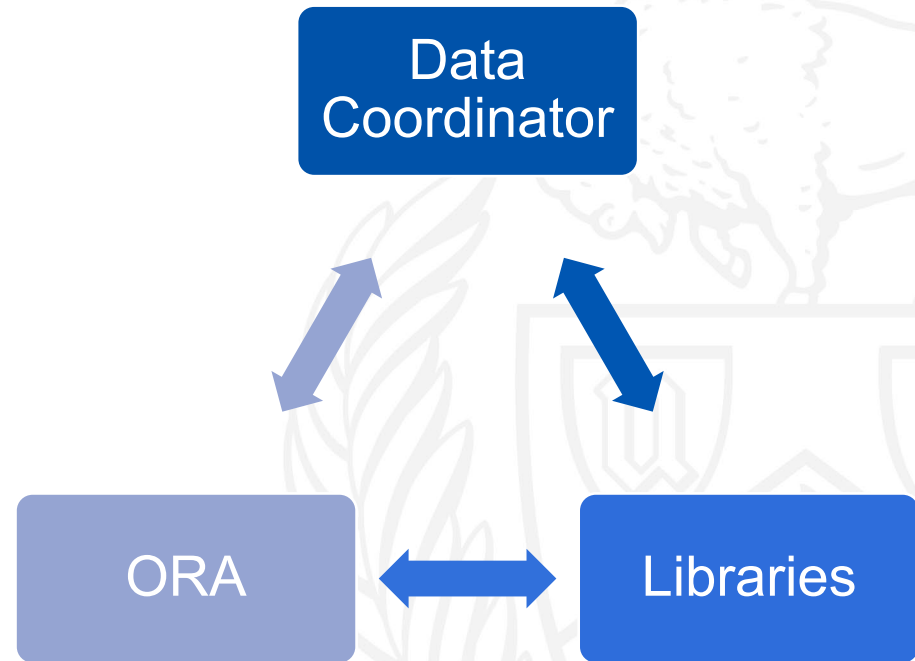
Committees

- Data Repository & Services
 - Chaired by Evviva Lajoie (University Librarian)
- Graduate Student Needs & Services
 - Chaired by Graham Hammill (Dean of the Graduate School)
- Medical Research Data Safety & Security
 - Chaired by Tim Murphy (Senior Associate Dean for Clinical and Translational Research)
- IT Needs
 - Chaired by Brice Bible (CIO)

Coordination Across the University

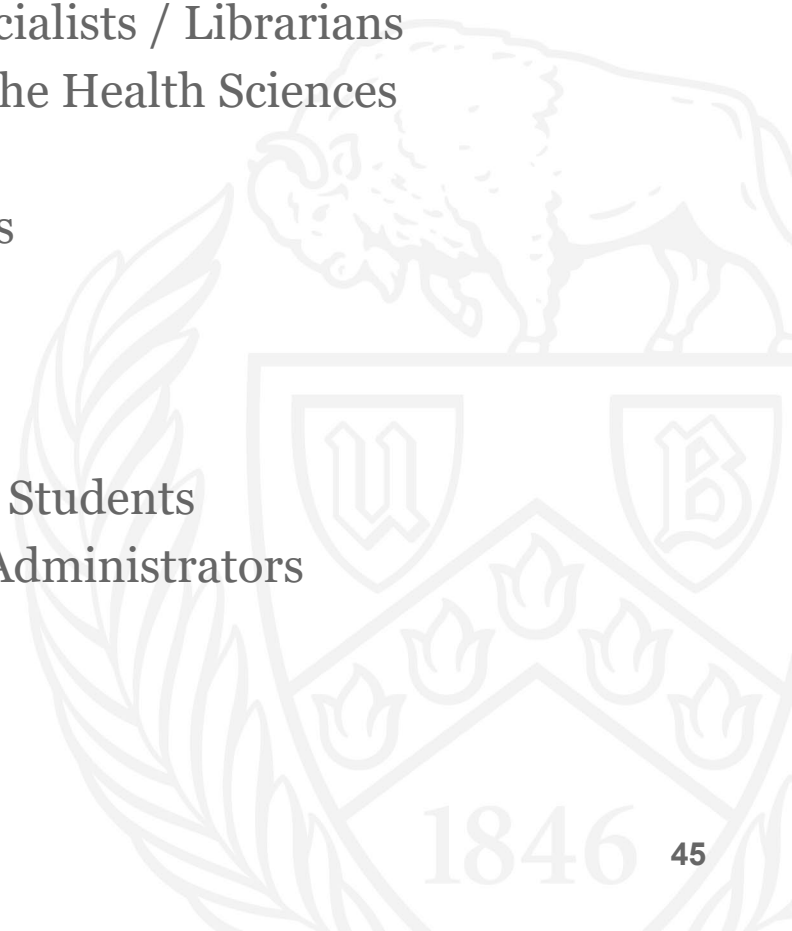
Pre-award Resources and Services

- Provides seamless, efficient support
 - Data Coordinator triages needs
 - Data Librarians offer templates and materials for data management
 - ORA provides direct personalized support to faculty for their specific proposal



Data Librarians at UB

- Planning for a department of three Data Specialists / Librarians
 - We currently have a Data Librarian for the Health Sciences
 - One for Sciences / Engineering
 - One for the Social Sciences / Humanities
- Responsibilities
 - Training, Consulting, and Outreach
 - Partnerships with Faculty and Graduate Students
 - Building Relationships with University Administrators
 - Policy Development
 - Strategic Planning



“All of Us” Training and Engagement Program



The Dataset



All of Us Data Training and Engagement for Academic Libraries Program

What is the *All of Us* Data Training and Engagement for Academic Libraries Program?

The ***All of Us* Data Training and Engagement for Academic Libraries Program** (hereafter referred to as the “Academic Libraries Program”) will encourage academic libraries at participating institutions to learn about the *All of Us* [Researcher Workbench \[researchallofus.org\]](https://researchallofus.org)  dataset and may enhance their skills in biomedical and public health data, as well as their library's research capacity, through engaging in an exciting community-learning cohort. In addition to these learning opportunities, participating institutions will be eligible for capacity building funding awards up to \$40,000 to support their data-related infrastructure and research capacity needs.

Research focuses on the intersection
of three factors

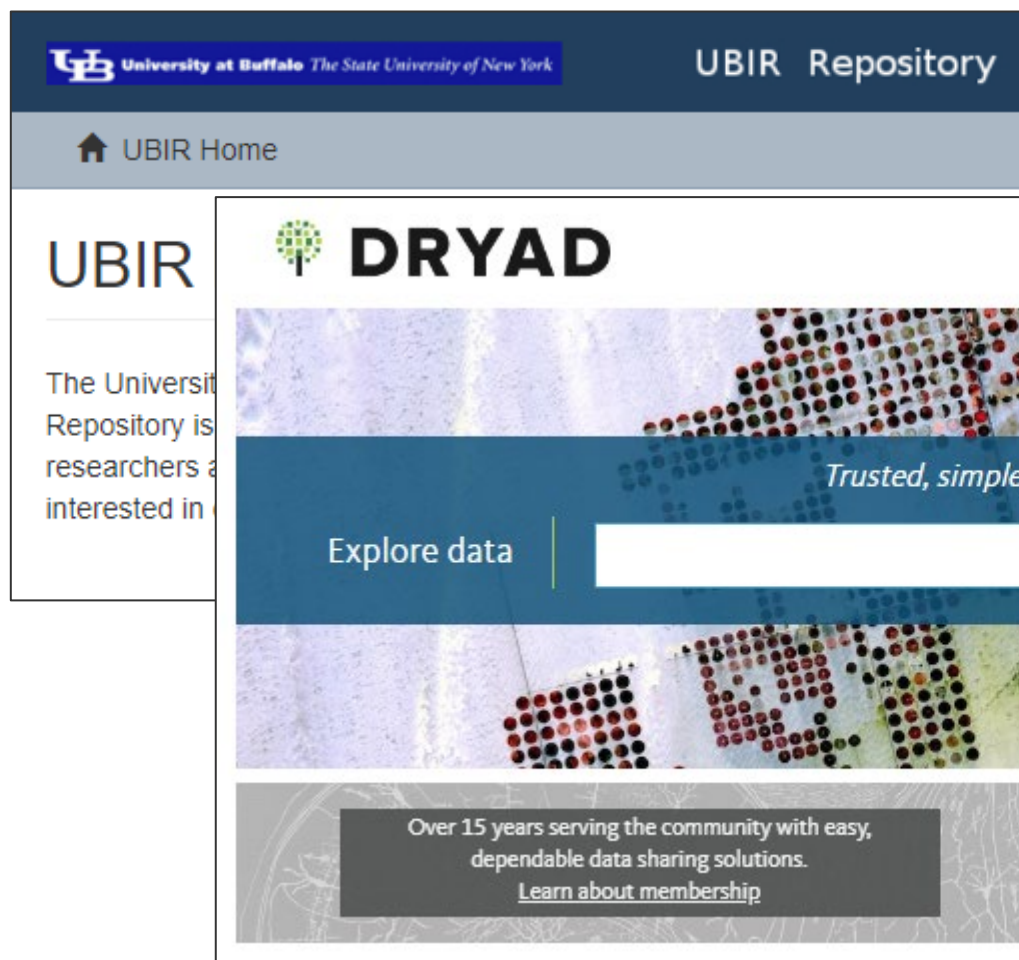
Data Lab

One of the recommendations from the committees was to create a Data Lab for graduate students

- 60% of all survey respondents indicated strong desire for a study / data analysis room.
- Access to software is often a challenge. Software is expensive and limited grants are available to graduate students to support purchases.
- Many software programs will only grant student level licenses to individuals, which do not always have the full functionality available to university entities.
- UB's Health Sciences Library was selected as the location for the first Data Lab
- This will provide a space for our Data Librarian for Health Sciences to consult and teach



Data Repository



The screenshot shows the UBIR Repository website. At the top is the University at Buffalo logo and name. Below it is a navigation bar with 'UBIR Home'. The main content area features the 'UBIR' logo and a description: 'The University at Buffalo Repository is a place where researchers and interested individuals can...'. A large banner for 'DRYAD' is prominent, with the text 'Trusted, simple' and 'Explore data' followed by a search bar. At the bottom, a grey box states: 'Over 15 years serving the community with easy, dependable data sharing solutions. [Learn about membership](#)'.

Why use Dryad?



Any field. Any format. Submit data in any file format from any field of research. Share all of the data from a project in one place.



Quality control and assistance. Our curators will check your files before they are released, and help you follow best practices.



Straightforward compliance. Submit your data to satisfy publisher and funder requirements for preservation and availability with a minimum of effort. We work directly with many publishers -- including Wiley, The Royal Society, and PLOS -- to streamline the process.



Community-led. Dryad is a nonprofit membership organization that is committed to making data available for research and educational reuse now and into the future. Modest Data Publishing Charges help ensure our sustainability.

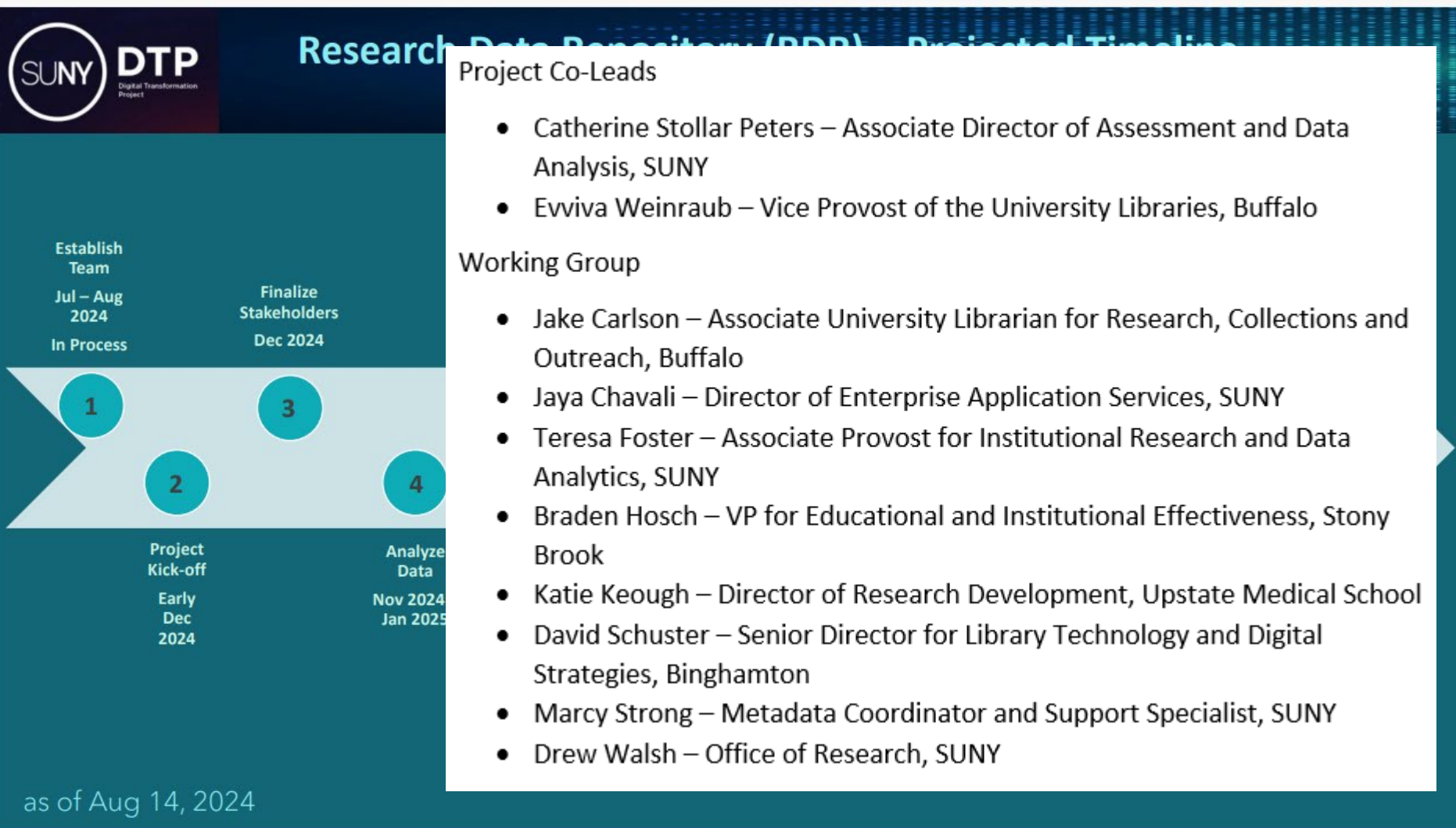
© 2011 Pearson Education, Inc.

NY State authorized \$200 million for “Digital Transformation Projects” in part to facilitate advanced research.

Our proposal for building a Research Data Repository with supporting services for all SUNY institutions was accepted

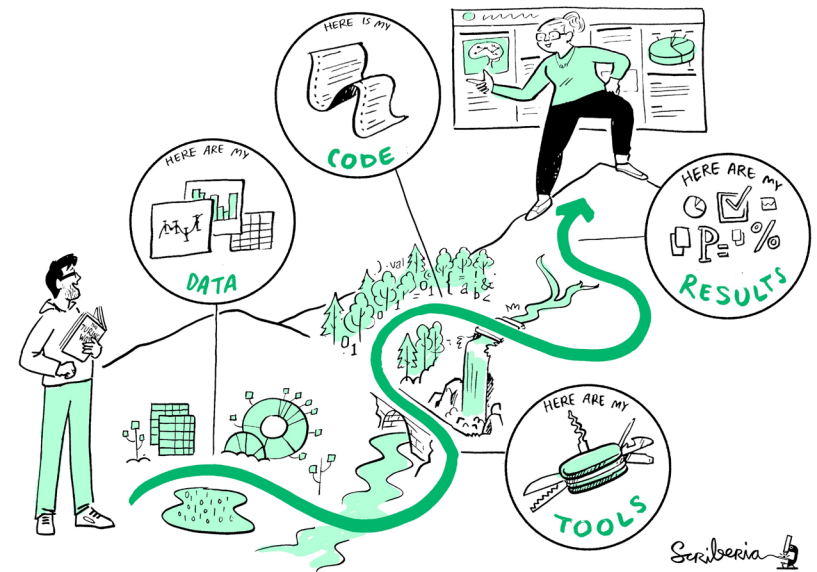


Timeline for a SUNY Data Repository



Final Thoughts

- This is still very much an evolving space
 - Challenges
 - Opportunities
 - Incentives
- “It takes a village”
 - Libraries are an important stakeholder but are only one part of what is needed as institutional support.
 - Costs
 - Capacity



THANK YOU!

Jake Carlson

Associate University Librarian for
Research, Collections and Outreach

University at Buffalo

jakecarl@buffalo.edu

