Preparing Medical Librarians to Understand and Teach Research Data Management

Kevin Read, Catherine Larson, Karen Yacobucci, So-Young Oh, Suvam Paul, Colleen Gillespie, Alisa Surkis
Background
NIH BD2K Initiative on Research Education

RFA: Open Education Resources for Sharing, Annotating and Curating Biomedical Big Data

“curriculum modules that can be used by librarians and other information specialists to prepare researchers, graduate students and research staff”
Grant Contributors

**PIs:** Kevin Read & Alisa Surkis

**Video creation:** Catherine Larson & Karen Yacobucci

**Instructional design:** So-Young Oh

**Data analysis:** Suvam Paul

**Evaluation and assessment:** Colleen Gillespie
Do you see a role for your library in teaching research data management?

N = 118
Do you currently teach research data management at your institution?

N = 118
Research Data Management @ NYUHSL

Experience providing:

- Continuing education classes for medical librarians
- Classes for researchers

How to avoid a data management nightmare

Data Sharing and Management Snafu in 3 Short Acts
Approach
Online Librarian Training

Web-based Curriculum
- Practice of research
- Story of data
- Fundamentals of RDM

Piloting
- Cohort of biomedical librarians
- Varied teaching experience

Librarian Evaluation
- Knowledge gain & satisfaction
- Comfort level/Intent to teach
- Qualitative follow-up

Dissemination to Biomedical Library Community

Librarians Teaching Researchers

In-person Curriculum
- Slides + script
- Evaluation methods
- “Edutainment” videos

Piloting
- Observing librarians teach
- Follow-up interviews with librarians

Researcher Evaluation
- Knowledge gain
- Satisfaction
- Intent to practice
## RDM Best Practices

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Field type</th>
<th>Description</th>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject_id</td>
<td>Text</td>
<td>Unique identifier for each subject</td>
<td></td>
</tr>
<tr>
<td>last_name</td>
<td>Text</td>
<td>Subject last name</td>
<td></td>
</tr>
<tr>
<td>first_name</td>
<td>Text</td>
<td>Subject first name</td>
<td></td>
</tr>
<tr>
<td>birth_date</td>
<td>Date</td>
<td>Subject birth date</td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>Integer</td>
<td>Subject sex</td>
<td></td>
</tr>
</tbody>
</table>
### RDM Best Practices

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Field type</th>
<th>Description</th>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject_id</td>
<td>Text</td>
<td>Unique identifier for each subject</td>
<td></td>
</tr>
<tr>
<td>last_name</td>
<td>Text</td>
<td>Subject last name</td>
<td></td>
</tr>
<tr>
<td>first_name</td>
<td>Text</td>
<td>Subject first name</td>
<td></td>
</tr>
<tr>
<td>birth_date</td>
<td>Date</td>
<td>Subject birth date</td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>Integer</td>
<td>Subject sex</td>
<td></td>
</tr>
</tbody>
</table>

### Understanding Research Context

[Image with DNA and lab flasks]
Online Librarian Training

Web-based Curriculum

Practice of research
Story of data
Fundamentals of RDM
Online Librarian Training: Compass Learning System
Online Librarian Training

- Piloting
- Cohort of biomedical librarians
- Varied teaching experience
Online Librarian Training

Librarian Evaluation

- Knowledge gain & satisfaction
- Comfort level/intent to teach
- Qualitative follow-up
Librarians Teaching Researchers

In-person Curriculum

- Slides & Script
- Evaluation materials
- “Edutainment” Videos
Librarians Teaching Researchers

- Piloting
- Observing librarians teach
- Follow-up interviews with librarians
Librarians Teaching Researchers

Knowledge gain
Satisfaction
Intent to practice
Six month follow-up surveys

Librarians
Intent to teach
Did you teach?

Researchers
Intent to use
Change in RDM practice?
Online Modules Pilot
Online Modules: Story of Data

The story can be simple

The story can be complex

Create → Process → Analyze

Patient ID | Tumor Area (m²) | Tumor Bounding Box (m²) | Number in Treatment
---|---|---|---
1234 | 4567 | 8901 | 2
1235 | 3658 | 9120 | 3

Process

Create → Process → Analyze

Model 1 → Model 37 → Model 300

<table>
<thead>
<tr>
<th>X (m)</th>
<th>Y (m)</th>
<th>Z (m)</th>
<th>Thickness (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>56</td>
<td>56</td>
<td>1.2</td>
</tr>
<tr>
<td>65</td>
<td>65</td>
<td>65</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Create → Process → Analyze

<table>
<thead>
<tr>
<th>True (m)</th>
<th>Model (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>4.2</td>
</tr>
<tr>
<td>0.2</td>
<td>4.9</td>
</tr>
<tr>
<td>0.3</td>
<td>4.8</td>
</tr>
</tbody>
</table>
Online Modules: Data Lifecycle

[Diagram showing the data lifecycle with stages: Idea, Process Methods, Create Collect Assemble, Processing Analysis, Findings, Disseminate Data, Supplemental Files Repository Data Publication Author mediated, Disseminate Findings, Publication Presentation Poster]
Online Modules: Data Lifecycle

Data Lifecycle: Data reuse + reproducibility

- Process Methods
- Create Collect Assemble
- Disseminate Data
- Supplemental Files Repository Data Publication Author mediated
Online Modules: Understanding Researchers
Online Modules: RDM Climate

Timeline of **funder data regulations**
### Data to do list

<table>
<thead>
<tr>
<th>Task</th>
<th>Create Documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>➜ Determine how, what, where, who will work with data (the workflow)</td>
<td></td>
</tr>
<tr>
<td>➜ Develop a system for naming files and folders</td>
<td></td>
</tr>
<tr>
<td>➜ Select and name variables to be collected</td>
<td></td>
</tr>
</tbody>
</table>
Online Modules: Data Standards

What do standards cover?

Research

Standard A

Standard B

Standard C-1

Standard C-2

Standard C-3

Standard D

Standard E

Standard F

Standard G
Online Modules: Data Standards

What do standards cover?

NO STANDARD FOR STANDARDS
Online Modules: Preservation

storage ≠ preservation
Online Modules: Getting Started

**Who are the contacts at your institution related to managing research data?**
- Contacts for questions about retaining or destroying data
- Contacts for questions about sponsored research data
- Contacts for questions about de-identifying data
- Contacts for questions about storing, backing up, and securing data
- Contacts for questions about archiving and preserving data
- Contacts for questions about depositing data in a repository
- Contacts for questions about describing data
- Contacts for questions about sharing data
- Contacts for questions about licensing data
- Contacts for questions about data ownership
- Contacts for questions about analyzing data
- Contacts for questions about visualizing data

**What local resources and tools related to data management at your institution?**
- Data storage options
- Describing and annotating data tools
- e-Lab notebooks
- Data backup tools
- Resources and tools for sharing data
- Resources and tools for de-identifying data
- Tools for analyzing data
- Tools for visualizing data
- Resources and tools for publishing data
- Resources and tools for archiving and preserving data
- Resources and tools for citing and licensing data
Online Modules Pilot
Online Modules: Pilot Data

- Satisfaction w/ modules
- Change in understanding
- Change in comfort level
Learner Rating of Effectiveness of Presentation

### Story of Data (n=67)

- Not effective: 5
- Somewhat effective: 10
- Mostly effective: 25
- Very effective: 27

### Data Lifecycle (n=59)

- Not effective: 5
- Somewhat effective: 7
- Mostly effective: 25
- Very effective: 12

### Understanding Researchers (n=45)

- Not effective: 5
- Somewhat effective: 5
- Mostly effective: 20
- Very effective: 15

### RDM Climate (n=36)

- Not effective: 5
- Somewhat effective: 5
- Mostly effective: 20
- Very effective: 16

### RDM Best Practices (n=32)

- Not effective: 5
- Somewhat effective: 10
- Mostly effective: 10
- Very effective: 7

### Data Standards (n=30)

- Not effective: 5
- Somewhat effective: 10
- Mostly effective: 10
- Very effective: 5

### Storage, Preservation, and Sharing (n=29)

- Not effective: 5
- Somewhat effective: 10
- Mostly effective: 10
- Very effective: 4
Learner Rating of Module Level

Story of Data (n=67)

Data Lifecycle (n=59)

Understanding Researchers (n=45)

RDM Climate (n=36)

RDM Best Practices (n=32)

Data Standards (n=30)

Storage, Preservation, and Sharing (n=29)
Learner Rating of Module Length

- Story of Data (n=67)
- Data Lifecycle (n=59)
- Understanding Researchers (n=45)
- RDM Climate (n=36)

- RDM Best Practices (n=32)
- Data Standards (n=30)
- Storage, Preservation, and Sharing (n=29)
Take Home

We’re in good shape!
Online Modules: Story of Data

The story can be simple

The story can be complex

Create → Process → Analyze

Patient ID | Treatment Area (cm²) | Target Area (cm²) | Number of Treatments
---|---|---|---
1001 | 404 | 317 | 4
602 | 254 | 42 | 3

<table>
<thead>
<tr>
<th></th>
<th>X (mm)</th>
<th>Y (mm)</th>
<th>Z (mm)</th>
<th>Thickness (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>55</td>
<td>75</td>
<td>1.2</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>56</td>
<td>75</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Model 1 → Model 37 → Model 300
Online Modules: Story of Data

The story of data change in understanding:

- Initial No Understanding
- Initial Minimal Understanding
- Initial Moderate Understanding
Online Modules: Data Lifecycle
Online Modules: Data Lifecycle

Data Lifecycle Change in Understanding

Initial No Understanding

Initial Minimal Understanding

Initial Moderate Understanding
Online Modules: Understanding Researchers
Online Modules: RDM Climate

Timeline of funder data regulations:

- **2003**
  - NIH Data Sharing Policy
    - Research >$500,000
    - No enforcement

- **2008**
  - NIH Public Access Policy
    - No enforcement

- **2011**
  - NSF Data Management Plan
    - Required w/ grant application

- **2013**
  - NIH Public Access Policy
    - Restrict grant funding

- **2015**
  - OSTP Memo
    - "Funded...research are made available to the public...including digital data"

- **2016**
  - *Coming soon* NIH Data Management Requirements
    - What created the data?
    - What protocols were used?
    - Preservation?
    - Access to data?

- **2016**
  - NIH Genome Wide Association Studies
    - Required to de-identify data
    - Required to deposit in dbGap

- **2016**
  - NIH Genomic Data Sharing Policy
    - All genomic data must be shared

- **2016**
  - Rigor & Reproducibility
    - Results reporting
Online Modules: RDM Climate
### Online Modules: RDM Best Practices

#### Data to do list

<table>
<thead>
<tr>
<th>Task</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determine how, what, where, who will work with data (the workflow)</td>
<td>CREATE DOCUMENTATION</td>
</tr>
<tr>
<td>Develop a system for naming files and folders</td>
<td>CREATE DOCUMENTATION</td>
</tr>
<tr>
<td>Select and name variables to be collected</td>
<td>CREATE DOCUMENTATION</td>
</tr>
</tbody>
</table>
Online Modules: RDM Best Practices

RDM Best Practices Change in Understanding

Initial No Understanding

Initial Minimal Understanding

Initial Moderate Understanding
Online Modules: Data Standards

What do standards cover?

- Standard A
- Standard B
- Standard C
  - Standard C-1
  - Standard C-2
  - Standard C-3
- Standard D
- Standard E
- Standard F
- Standard G
Online Modules: Data Standards
Online Modules: Preservation

storage ≠ preservation
Change in Understanding by Initial Level
“I answered "mostly comfortable" - but that's just my personality. I don't feel confident with new material until I've thoroughly researched all aspects of a subject.”
Change in Comfort Level by Initial Level

Initially Not Comfortable
- Not comfortable
- Somewhat comfortable
- Mostly comfortable
- Very comfortable

Initially Somewhat Comfortable
- Not comfortable
- Somewhat comfortable
- Mostly comfortable
- Very comfortable

Initially Mostly Comfortable
- Not comfortable
- Somewhat comfortable
- Mostly comfortable
- Very comfortable
Sufficient Knowledge to Teach RDM
Other Use of Knowledge Gained
In-Person Pilot
In-Person Pilot

Visit pilot institutions to observe teaching

Challenge of those wanting to pilot vs. those who were able to

3 pilots completed to date
Researcher Satisfaction

**Level of Presentation**
- Too low
- Just right
- Too advanced

**Length of Presentation**
- Too short
- Just right
- Too long

**Effectiveness of Presentation**
- Not
- Somewhat
- Mostly
- Very

**Would Learner Recommend**
- No
- Reservations
- Recommend
- Highly
Use of Knowledge Gained
In-Person Pilot Data: Take Homes

It works!

Script is helpful but dangerous

Customization is an ongoing process

More classes on the horizon
Next Steps
Online Modules

Decoupled modules

Removed satisfaction surveys

Encourage follow up

Teaching Toolkit for Librarians


This teaching toolkit is designed to be used for a one hour introductory data management class for biomedical researchers. It consists of an instructional guide for teaching the material, a PowerPoint presentation with a script in the accompanying notes to each slide, and a separate evaluation form. This material is built on training material provided in the BD2K funded online research data management educational modules that are freely available here: http://bit.ly/RDM_Modules
Regional Programs

Middle Atlantic Programs

2017 Proposals Funded by the MAR

New York University Health Sciences Library -- Facilitating the Development of Research Data Management Services at Health Sciences Libraries

The proposed Research Data Management program is designed to provide training, tools and strategies for medical libraries seeking to develop research data management services. This program provides a holistic approach to developing data services that focuses on building the required knowledge base, understanding and connecting with researchers, promoting effective outreach strategies, and integrating with the broader institutional data community. The program consists of online training modules to provide the necessary background in research data management, and three components to assist libraries in implementing services at their institution:

- A data interview template with outreach guidelines
- A teaching toolkit consisting of slides, script, and evaluation materials
- Strategies and promotional materials for developing a data class series
Questions?

kevin.read@med.nyu.edu

alisa.surkis@med.nyu.edu

Supported by grant R25 LM012283-01 from NIH