

Teaching Data Management



This project is made possible by a grant from the U.S. Institute of Museum and Library Services and with funds from the National Library of Medicine under Contract No. N01-LM-6-3508.

Lamar Soutter Library UMMS

Developing a RDM Curriculum

- Phase I: Planning
- Phase II: Content Development
- Phase III: Piloting and Implementation
- Phase IV: Evaluation

Phase I: Planning

- Interviews
- Surveys
- Outside Speakers
- Educational Committee
- Planning Board
- Consultants
- Literature search



Student Interviews

- Preference for electronic lab notebooks
- Software used with their data - Perl, GraphPad Prism, Filemaker Pro, SPSS, SAS, Nvivo
- Data backed up and shared in emails, the cloud (Google Drive, Dropbox)
- No standard naming conventions used for directories and/or files.

Student Interviews

- Lab data protocol changes from lab to lab - usually at the call of the PI. (wet labs vs. dry labs)
- Most rely on the network server
- Backups were not consistent.

Faculty Interviews

- RDM Practices are varied
 - No formal RDM training provided for students
 - Lab managers and students come and go
 - Students need a familiar context to learn RDM so that can apply lessons in their own practice
- This led to the idea of developing teaching cases

Customizing Curriculum

- Mix and match modules as needed by discipline/course level
- Provide lesson plans for diverse modes of delivery: online, in person, hybrid
- Case based activities and assessment
- Readings
- Assignments

Research Case Scenarios



Photo: CDC/Taronna Maines

- Aerospace engineering
- Biomedical lab research
- Clinical study on hip replacements
- African American Perceptions of End-Of-Life Care



Gary Meek/NSF



NASA/Michael Soluri

Course Modules

Module 1
Overview of Research Data
Management

Module 2
Data: Types, Stages, and Formats

Module 3
Metadata

Module 4
Data Storage, Backup, and
Security

Module 5
Legal and Ethical Considerations

Module 6
Data Sharing and Reuse Policies

Module 7
Archiving and Preservation

Phase II: Content Development

- 2012-2013 UMMS awarded NN/LM NER grant to develop the frameworks into a course with module content, lecture slides, activities, and teaching cases
- Partners: UMass Amherst, Marine Biological Laboratory and Woods Hole Oceanographic Northeastern University, Tufts University
- Designed for flexibility



Case-Based Learning

- Cases provide the opportunity for instructors and students to explore discipline-specific data management issues
- Course modules provide a context of universal data management issues and best practices

Module 1 Lesson Plan

| Lesson Plan for Module # 1- Overview of Research Data Management | |
|---|---|
| Learning Objectives | <p>By participating fully in this class, student will be able to:</p> <ol style="list-style-type: none"> 1. Explain what research data is 2. Explain the need for managing/sharing research data and identify relevant public policies 3. Explain the lifecycle continuum to manage and preserve research data 4. Understand that data should be managed differently in different phases of the life cycle 5. Be familiar with data management plan (DMP) requirements used to characterize and plan for the lifecycle of research data. 6. Identify the value and relative importance of data management to the success of a research project. |
| Lecture Content | <ol style="list-style-type: none"> 1. Explain broadly what is research data 2. Illustrate need for proper data management practices. Present 3 examples (research data from student project, federally funded project at local institution, and from multi-institutional projects of national scope). 3. Describe funding agency requirements for data management 4. Describe research data lifecycle continuum phases: create, process, analyze, preserve, give access, reuse. Show diagram, and use real life example to illustrate each phase of the continuum. http://www.data-archive.ac.uk/create-manage/life-cycle 5. Show and compare sample data management plan requirements, give examples of DMPs for different funding agencies http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/resources.jsp#a02 6. Present simplified data management plan template. |
| Activities | <ol style="list-style-type: none"> 1. Identify data sets collected and/or generated in examples used in lecture content #2 above 2. Create a data management plan for one of the cases using the simplified data management plan template on page 21. |
| Assessment | <p>Read excerpt from research data management case #A: Outcomes from Orthopedic Implant Surgery (Illustrates the challenges in conducting a multiyear research project with changing personnel each year) and respond to questions.</p> |
| Readings | <ol style="list-style-type: none"> 1. Promoting the Stewardship of Research Data, Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age (2009): pages 95-99 http://books.nap.edu/openbook.php?record_id=12615&page=95 2. Why Share Data? UK Data Archive http://www.data-archive.ac.uk/create-manage/planning-for-sharing/why-share-data 3. Introduction: A Revolution in Science: p.11-13 from Harnessing the Power of Digital Data for Science and Society (2009) http://www.nitrd.gov/about/harnessing_power_web.pdf 4. Steps in the Research Life Cycle, Scientific Data Consulting, University of Virginia Library http://www2.lib.virginia.edu/brown/data/lifecycle.html 5. Data Management and Publishing http://libraries.mit.edu/guides/subjects/data-management/funding.html 6. Funding Agency and Data Management Guidelines: http://www.lib.umn.edu/datamanagement/funding 7. Example Data Management Plan http://www.dataone.org/sites/all/documents/DMP_MaunaLoa_Formatted.pdf |

Module 1 Content

- Use agreed upon, documented, and descriptive information about project or specific experiment (link it to your experiment number in your lab notebook!).
- Use ISO 8601 for date formats: YYYYMMDD; some software programs are known to reformat dates if they use special symbols to separate the years, months, and days.
- Include a version number.

Here is an example of a name for an image file created by a biomedical researcher.

Data Naming Format

Histology

- Apex of the heart - A
- Basal side (base of the heart) - B
- Experiment #, A or B, and Section #
- Example: 0231A_216
- When sample is stained, an abbreviation for the type of stain used is added to the slide
- Example: 0231A_216_act (actinin)

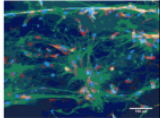


Figure 4 (Gaudette 2013)

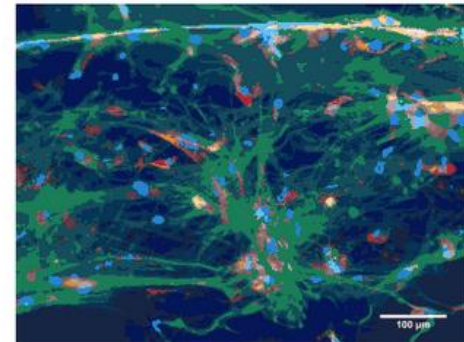
Notice in Figure 4 that his name for this image file contains information regarding the experiment number. This number corresponds with its record in the appropriate laboratory notebook. The name is concise; it allows for quick reference and he can add additional metadata within the lab notebook (names of personnel involved in the experiment, the date and time of the experiment, the version, etc.) without having to make the file name unnecessarily longer than 25 characters. Also, notice there are no special characters or spaces and he uses underscores between elements. While it might seem short, the file name communicates a lot of agreed upon standards: the experiment it is associated with, the section of heart tissue, the location of the heart tissue, and the stain used on the tissue in the image.

Module 1 Slides

Data Naming Format

Histology

- Apex of the heart - A
- Basal side (base of the heart) - B
- Experiment #, A or B, and Section #
- Example: 0231A_216
- When sample is stained, an abbreviation for the type of stain used is added to the slide
- Example: 0231A_216_act (actinin)



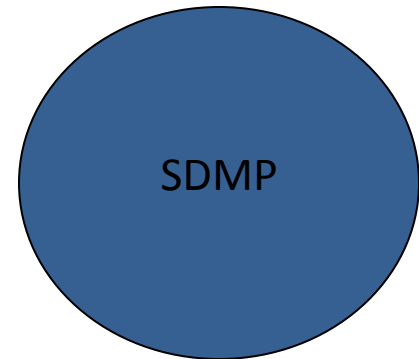
Regeneration of functional heart tissue with stem cell delivery



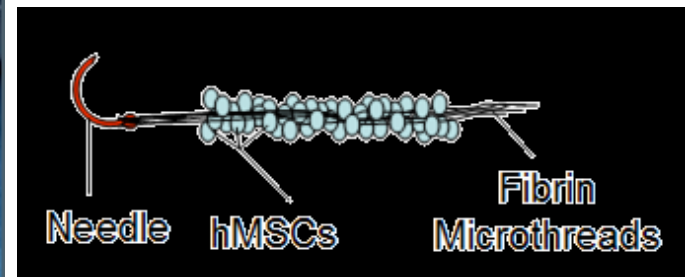
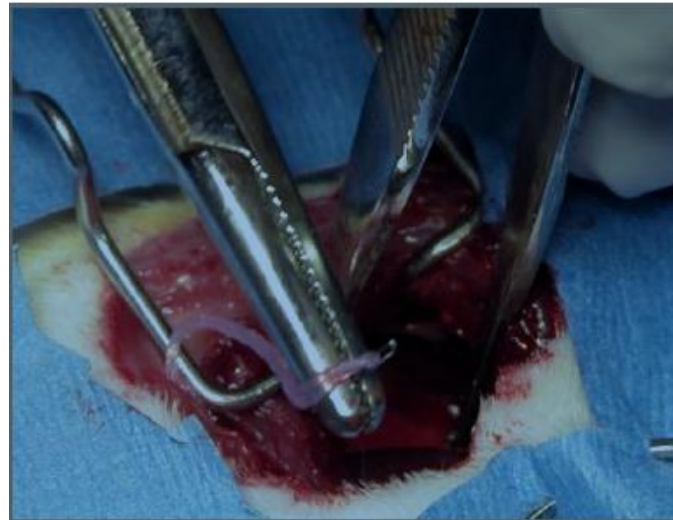
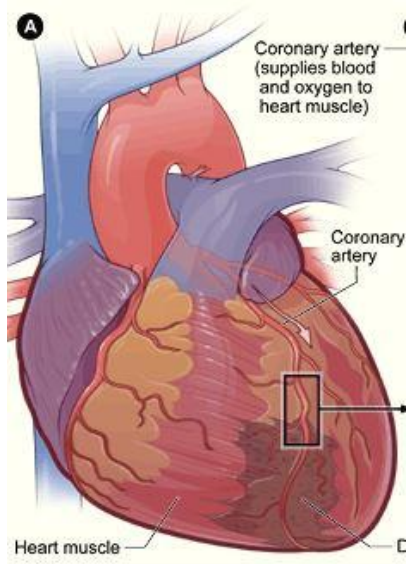
A data management case study in biomedical engineering research

Case Analysis using Simplified Data Management Plan

1. Types of data
2. Contextual details (metadata) needed to make data meaningful to others
3. Storage, Backup, and Security
4. Provisions for Protection/Privacy
5. Policies for re-use
6. Policies for access and sharing
7. Plan for archiving and preservation of access

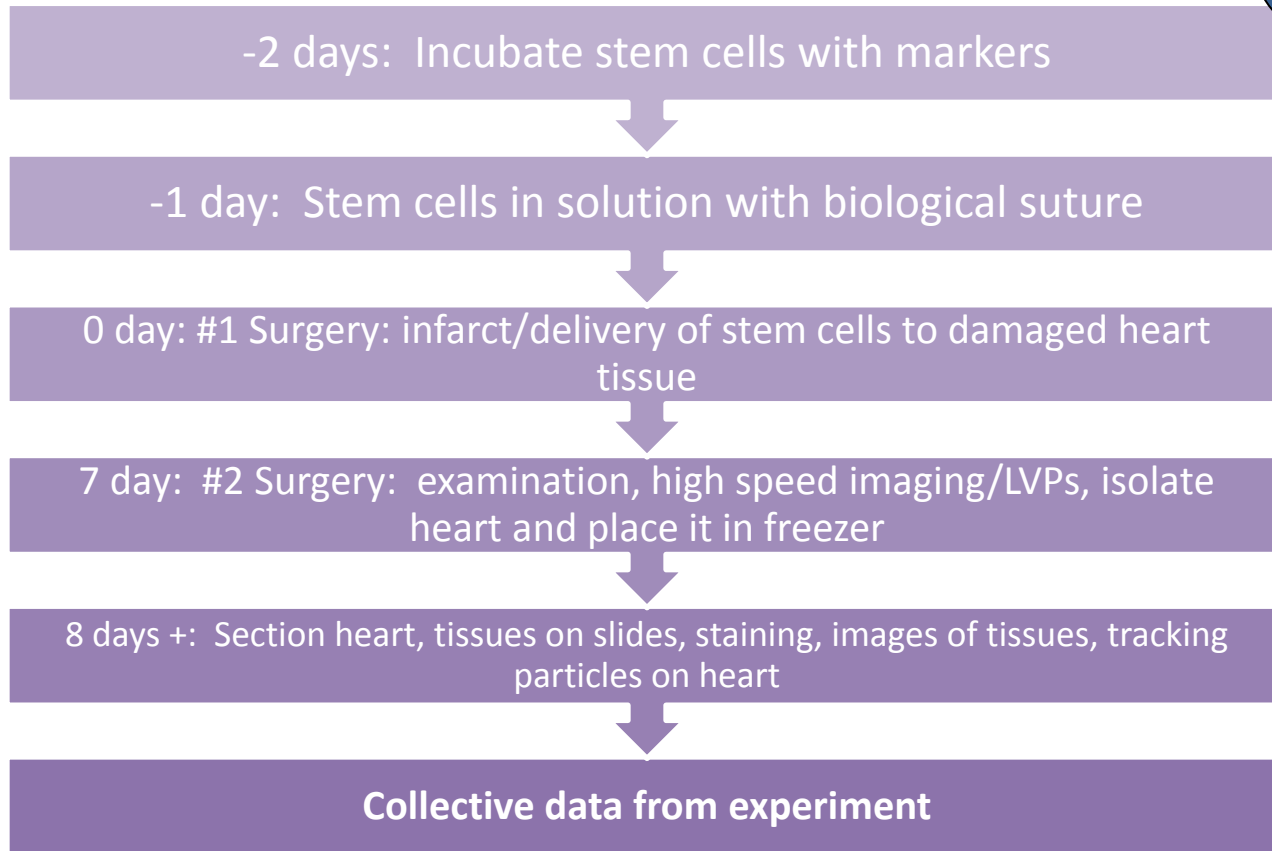


Experiment: delivery of stem cells on a biological fibrin microthread to areas of damaged heart tissue in one rat's heart
Purpose: to restore mechanical function of damaged heart tissue



Timeline of Experiment:

SDMP #1
Creating
data



File Formats

| Data | File Format |
|--|---|
| Images | ? |
| Left ventricular pressure measurements | ? |
| Home made software | MATLAB or C |
| Histology sections | Slides—file name based on stain—example .act is actinin stain |
| Contextual | Paper lab notebook, animal log |

SDMP #2
File formats

SDMP #3
Storage

SDMP #4
Ownership

Directory that links data sets together: Excel spread sheet

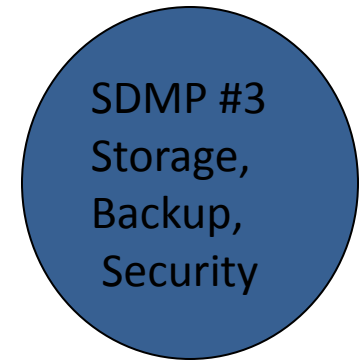
Analyzing the data

“There could easily be up to 10 people involved in data analysis and we have not yet found a good way to link all the data.”



Dr. Glenn R. Gaudette, PI

| Data Set | Storage |
|---|--|
| Optical images taken during Surgery: (~10,000 images) 1000 images for each data set | Hard drive acquisition computer > Drobo backup>hard drive of network computer backed up by institution |
| Left ventricular pressures (numeric) correlating with specific images | Same as above |
| Tissue sections | Slide boxes—could be in any of 3 or 4 freezers |
| Software | ? |
| Images from different stained tissue after second surgery | Drobo > DVD backups |
| Contextual data | Paper lab notebook (lab, PI's office), surgical log (with animal) |



Case B Analysis using Simplified Data Management Plan

1. Types of data: Images of heart, LVP measurements, histology slides (tissues), images of slides, software

2. Contextual details (metadata) needed to make data meaningful to others:

- experiment #
- dates of experimental activities
- stem cell line
- details about animal (species, age, identifier)
- area of infarct
- area where stem cells implanted
- type of stain used
- instrumentation

3. Storage, Backup, and Security:

- Hard drive of acquisition computer (initial)
- Drobo
- networked hard drive (backed up by institution)
- DVDs
- Slide boxes in freezers

Case B Analysis using Simplified Data Management Plan

3. (continued):

- Lab notebooks/animal surgical logs (some backup when data is transcribed from animal surgical log to paper lab notebook)

4. Provisions for Protection/Privacy:

- Files are not password protected
- Paper lab notebooks are kept in lab/older ones in PI's office (key card access to these rooms)

5. Policies for re-use:

- Not addressed – need to ask researcher (possibilities: reuse with permission of PI, institutional policies)

6. Policies for access and sharing:

- Not addressed—need to ask researcher (possibilities: PI will make accessible after paper publication, will share them immediately, may depend on funding agency reqs.)

Case B Analysis using Simplified Data Management Plan

7. Plan for archiving and preservation of access
 - Not addressed—ask researcher (possibilities: convert files in proprietary formats to generic formats, appraisal of data and data versions, decision on where data should be archived (IR or DR))

Phase III: Piloting

- UMMS CTSA Pilot Summer 2013
- Piloting Partners: Oregon State, Tufts, University of Tennessee, VCU, Colorado State
- Looking for additional partners
- Looking to add cases to curriculum
- Expand RDM educational community

Train-The-Trainer

- Webinar 10/31 on Module 1: Overview of Research Data Management and writing data management plans
- On-site Professional Development Day 11/8: Regional Data Management Education Course: How to Teach RDM Using the New England Collaborative Data Management Curriculum

Educating Next Gen-Librarians

- Partnered with Simmons College Graduate School of Library and Information Sciences
- 15-week course covering librarian roles in research data management
- Students conduct data interviews with researchers, develop teaching cases, and write data management plans

532G-01 Scientific Data Management

- Students learn from researchers' cases
- Scientists share workflows and their data management practices, challenges
- Data Management Plans (DMPs)
- Data repositories, Open Science, Open Data
- Annotating data sets
- Preserving and archiving data
- Developing library data services and data policies
- Research Informationists

Phase IV: Evaluation

Challenges

- Different Modes of Teaching
- Different Modules
- Different Cases
- Different Audiences
- Different Instructors
- Different Locations
- Different Institutions

...Focus on the content and its usefulness

References

Gaudette, G. R. 2011. “Keeping a lab notebook in the Gaudette lab.” Unpublished.

Gaudette, G. R. 2012. “Data management in biomedical engineering: needs and implementation.” Power point presentation. University of Massachusetts Medical School, Shrewsbury, MA 4 April 2012.
http://escholarship.umassmed.edu/escience_symposium/2012/program/9/

Gaudette, G.R. and Kafel, D. 2013. “ A case study: data management in biomedical engineering.” *Journal of eScience Librarianship*: 1(3): 159-172. <http://dx.doi.org/10.7191/jeslib.2012.1027>

University of Massachusetts and WPI (2011). Frameworks for a Data Management Curriculum. University of Massachusetts. http://library.umassmed.edu/imls_grant

For More Information.....

Contact:

Elaine R. Martin, DA

Director of Library Services

Lamar Soutter Library

University of Massachusetts Medical School

Elaine.Martin@umassmed.edu