**Case: Regeneration of Functional Heart Tissue in Rats**

**Summary of Teaching Points**

**Module 1: Overview of Research Data Management**

* Paper Lab notebook inconsistencies across users
* Lack of synchronization between lab notebook entries and surgical log

**Module 2: Types, Formats, and Stages of Data**

* Data sources are linked together on an Excel spreadsheet

**Module 3: Contextual Details**

* Naming conventions for data sets

**Module 4: Data Storage, Backup, and Security**

* Lack of consistent plan to keep track of slides (in refrigerators) containing tissues and of stages of tissue processing
* Storage issues for large number of optical and electronic images
* Lack of backup for lab notebooks
* Backup plan for data but obviates usefulness of naming conventions

**Module 5:** **Legal and Ethical Issues**

* Home grown analysis software ownership and preservation

**Module 6: Data Sharing and Re-Use**

* Home grown analysis software ownership and preservation

**Module 7: Plan for Archiving and Preservation of Data**

* None

**Research Data Management Case: Regeneration of Functional Heart Tissue in Rats**

The goal of the study is to try to regenerate functional heart tissue in a rat. Unlike other organs and tissues which regenerate themselves, the heart does not have the ability to regenerate, so we intend to regenerate it by delivering stem cells to the heart. The hope is that in generating heart tissue, we generate tissue that is actually functioning and contracting and doing mechanical work.

Two days before we operate on the rat, we take adult stem cells and we incubate them for 24 hours with our marker for cells [fluorescent nanoparticles]. We then put them in a solution and inject them into a tube that has a biological suture in it, so the cells sit down on the outside of the biological suture. We incubate it for 24 hours, and then do the surgery. During the surgery, we open up the thoracic cavity of the rat and create a myocardial infarction by occluding the left anterior descending coronary artery. At this point it is ischemic; we keep it ischemic for 1 hour, not letting any blood flow go through, and then we reperfuse it and let the blood go back. About a minute after that, we put the biological suture with the cells on it through the infarcted region. We then close the rat up and put it back in the cage for a week. We go back a week later, open the rat up again, and use our camera system to acquire images of the heart. We take images with two cameras simultaneously and we’ll also have a pressure transducer which syncs automatically with the pictures inside the left ventricle cavity to measure left ventricle pressure. Then we reposition our cameras and take another data set and we usually do that about 4 or 5 times to look at different regions around that infarct.

Then we euthanize the animal. We isolate the heart. We fix the heart in a fixative and then put it in the freezer for about 24 hours. Then we start cutting sections of the heart and putting them onto slides – about 3 sections of the rat heart per slide. We generate about 200 slides per rat heart. At any time, some tissue that was sectioned and on slides may be in one freezer, and some tissue that had not been sectioned yet but was embedded and ready to be sectioned is in another and still other tissue that may be sitting in a container someplace in another freezer. It should be entered into the excel spreadsheet saying what was done and where it is, but that doesn’t always happen. Then we stain some slides and then sometime after – anywhere from a day to a couple months - we stain some of them with trichrome. That tells us what tissue is dead. We stain some of them for specific markers in looking to find out exactly where the stem cells are in that cross section. Then we take images on our microscope, which is an epifluorescent microscope and, if we are happy with the staining and the way they look, then we make an appointment to use the confocal microscope which takes much better quality pictures and take those images on the confocal. At the same time, we also look at the data we acquired and use our home-grown custom software to track particles on the surface of the heart to see how far and how fast those particles are moving. The software was written in C and MATLAB (C runs the code faster, but MATLAB is easier to work with; usually we develop the code in MATLAB and then convert it to C so that it runs faster).  We use this software to analyze the optical images of the heart. That tells us what the function is like in that region of the heart. We do that for several heartbeats in different data sets. And then we save that data. That is everything we do for one heart.

**Data sets:**

1) We have the optical images after the first surgery to insert the cells – on average for one experiment we probably have about 10,000 images. We acquire images of the heart at about 250 frames/second. We acquire 4 seconds worth of data, so we have 1,000 images for each data set. The images are initially stored on the hard drive of the acquisition computer, then are transferred to a Drobo backup system and the hard drive of a network computer that is backed up by the institution.

2) The second data set is where we measure the left ventricle pressure at the same time we are acquiring those images so that we know that image # 127 correlates with the pressure at time point # 127 milliseconds. For this measure we use an analog to digital (A/D) board and a Millar pressure transducer. Both the camera and the acquisition system are computer controlled to synchronize them. These data are stored in the same way as the optical images, although they are separate files.

3) Electronic data sets are used to acquire images from the different stained tissue sections after the second surgery. We may stain on average 4-5 different markers and we will have different data sets for the different stains. So we will have images taken from the epifluorescent scope, and, usually in a cross-section of heart, we may take some high resolution images in zoomed in regions and some low magnification images. On average for each section, we take about 20 pictures with the epifluorescent and then with the confocal, we probably take about another 20; if we take a z-stack with the confocal, that can be an additional 200 images. They are both taking the same images except one is a much better resolution than the other. Long term storage for these datasets is on the Drobo and DVD backups.

Our naming convention is that we name our files EXP (for experiment) and then usually 4- or 5-digit codes like 2001, and then we have several data sets so it is DS # and then it could be image 1. Then we need to link these data sources together via an excel spreadsheet. The sections are all linked by the same experiment number. They are linked with the digital images just based on what number section they are.

Multiple research staff may be analyzing the same heart, and one person will be doing the mechanical function of the heart, one will be doing the trichrome staining, another will be doing the actinin staining and maybe another will be doing the imaging. The data sets should all be linked in the excel spreadsheet. 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. We have an excel spreadsheet basically, and it says in DS1 – the mechanical function in this area was xyz, in DS2 it was this. In DS1 the tissue section showed this, and we try to link them all up together, but the tissue sections are on conventional microscope slides that are stored someplace. Even that – the location of where they are stored is a problem. We have the usual places where we store things but we have 3 or 4 freezers and if it is not in 1, we look in 2, and so on. The slide box is labeled with the experiment number and the individual slides are labeled with the slide number & experiment number.

The types of data we use are mostly images and numeric measures in addition to the lab notebook which may have some observational notes. Some of it is number-crunching but a lot of it is images.

The content of a lab book relates to a particular experiment and is used by all staff working on that experiment. There is a format they are all supposed to follow, which they don’t always do. There could be on average 5-6 people using the notebook. The paper lab notebook basically performs the function of being an index into the actual datasets and it should record all the information the PI specifies. We also have a paper surgical log that is kept with the animal and whatever project staff writes down in that surgical log should be transferred into the lab notebook – so it has to be in 2 places. It has to be down there in case there is a problem with the animal, but the PI also needs it in the lab notebook to be able to write papers. The older lab notebooks are in the PI’s office, but the ones that are currently in use are in the lab. Older Lab notebooks are only in the PI’s office of lab with no backup. The lab notebook has to be in pen on specific paper because this paper is supposed to be good for 100 years.

We backup the data sets on an external hard drive someplace. The optical and electronic images are both backed up. The current backup system we are using truncates the data set name to 6 digits then puts a tilde sign and number starting from 001.

The files are not password protected or anything.  The lab notebook is either in the PI’s office or (most often) in the lab, which has key card access (although it is an open floor plan).

**Module 1: (Overview module) discussion question:**

What issues need to be addressed on this project related to the 7 segments of the data management plan components?**Discussion Questions for Other Modules:**

1. Types of data

a. What types of data are being collected for this study?

* 1. How will you ensure all research staff used the same data sources and data definitions?
  2. What would be needed in a data management plan to describe use of novel equipment?
  3. What needs to be in the plan related to the data capture for the various data sets?
  4. What analytical methods and mechanisms will be applied to your data either prior to or post integration?
  5. What type of outcome data will be generated?

1. File Formats and Contextual details
   1. What file formats and naming conventions will be used for the separate data sources and for the integrated file used for analysis?
   2. What impact would the naming conventions and the use of homegrown software have on later data access?
   3. What other contextual details would you specifically need to document to make your data meaningful to others?
   4. In what form will you capture these details?
2. Data Storage, Backup, Security
3. Where and on what media will the data from each data source be stored?
4. How, how often and where will the data from each source be backed up?
5. How will you manage data security across research staff on the study for each data source?
6. How long following the completion of your study will you store the data?
7. Data protection/privacy
8. How are you addressing any ethical or privacy issues?
9. What mechanism are you using to identify individual animals or hearts?
10. Who will own any copyright or intellectual property rights to the data from each source?
11. How will the dataset be licensed if rights exist?
12. How will the data be associated with a study ID?
13. Policies for reuse of data
    1. Will you need to create a de-identified copy of the data?
    2. Will the data be restricted to be re-used only for certain purposes or by specific researchers?
    3. Are there any reasons not to share or re-use data?
14. Policies for access and sharing
    1. Will some kind of contribution or fee be charged for subsequent access to this data?
    2. What process should be followed to gain future access to your study data?
15. Archiving and preservation
16. What is the long-term strategy for maintaining, curating and archiving the data?
17. What data will be included in an archive?
18. Where and how will it be archived?
19. What other contextual data or other related data will be included in the archive?
20. How long will the data be kept beyond the life of the project?

Creative Commons License This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 United States License](http://creativecommons.org/licenses/by-nc-sa/3.0/us/). You are free to re-use part or all of this work elsewhere, with or without modification. In order to comply with the attribution requirements of the Creative Commons license (CC-BY), we request that you cite:

* **Editor**: Lamar Soutter Library, University of Massachusetts Medical School
* **Title**: New England Collaborative Data Management Curriculum
* **URL**: <http://library.umassmed.edu/necdmc>

Revised June 12, 2015