



A Case Study: Data Management in Biomedical Engineering

Glenn R. Gaudette¹ and Donna Kafel²

¹ Worcester Polytechnic Institute, Worcester, MA, USA

² University of Massachusetts Medical School, Worcester, MA, USA

Abstract

In a biomedical engineering lab at Worcester Polytechnic Institute, co-author Dr. Glenn R. Gaudette and his research team are investigating the effects of stem cell therapy on the regeneration of function in damaged cardiac tissue in laboratory rats. Each instance of stem cell experimentation on a rat yields hundreds of data sets that must be carefully captured, documented, and securely stored so that the data can be easily accessed and retrieved for papers, reports, further research, and validation of findings, while meeting NIH guidelines for data sharing. After a brief introduction to the bioengineering field and stem cell research, this paper focuses on the experimental workflow and the data generated in one instance of stem cell experimentation; the lab's data manage-

ment practices; and how Dr. Gaudette teaches data management to the lab's incoming graduate students each semester. The co-authors discuss the haphazard manner by which engineering and science students typically learn data management practices and advocate for the integration of formal data management instruction in higher education STEM curricula. The paper concludes with a discussion of the Frameworks for a Data Management Curriculum developed collaboratively by the co-authors' institutions – the University of Massachusetts Medical School and Worcester Polytechnic Institute – to teach data management best practices to students in the sciences, health sciences, and engineering.

Biomedical Engineering

Currently the fastest growing field in engineering, biomedical engineering advances knowledge in engineering, biology, and medicine, and improves human health through cross-disciplinary activities that integrate principles of engineering sciences with biomedical sciences and clinical practice (Bureau of Labor Statistics, 2012). Biomedical engineering applications include the development of new medical devices and processes that improve medical practice and

advance the quality of life (Imperial College London, N.D.). Examples of biomedical research areas include biomechanics, bioinstrumentation, imaging, genetic and tissue engineering, and stem cell therapy. Stem cell therapy has been a key focus of biomedical engineering research as it has shown promise in treating multiple clinical conditions, including the repair of damaged heart tissue resulting from coronary artery disease. Damaged heart tissue results in decreased mechanical contraction of the heart. Restoration of function to damaged heart

Correspondence to Donna Kafel: donna.kafel@umassmed.edu

Keywords: biomedical engineering, data management, case study, stem cells, coronary heart disease, regeneration, heart, paper lab notebooks, instruction, experiment

tissue is the goal of this stem cell therapy.

Coronary Heart Disease and Stem Cell Therapy

As the leading cause of death in the world, coronary heart disease is a major focus of biomedical engineering research. A common aftermath of coronary heart disease is cardiac ischemia, a condition in which there is decreased blood flow and oxygen to the heart muscle. Cardiac ischemia affects approximately 16 million Americans. Heart attacks, or myocardial infarctions, are triggered by ischemia and occur when the blood flow that normally brings oxygen to the heart muscle is severely reduced or cut off completely (American Heart Association, 2012). When a heart attack occurs, the heart muscle that has lost blood supply dies. While other areas of heart muscle may continue working, the heart muscle affected by a heart attack, i.e. the area of infarct, cannot self-repair. Researchers are exploring ways to save additional lives by using replacement cells for dead or impaired cells so that the weakened heart muscle can regain its pumping power. Stem cell therapy has been shown to have potential for treating and preventing heart failure resulting from heart attacks. When delivered to injured cardiac tissue, stem cells have the potential for differentiating into cardiac muscle cells and restoring function to areas of damaged heart tissue.

In the research laboratory, investigators use animal models to study the effects of stem cell therapy on heart tissue damaged by heart attacks. A plethora of data is generated from each animal experiment. The management of these data is essential to publishing and reporting the results. In addition, as tissue samples can be stored for years, new data (e.g. staining for a protein recently reported to be important in tissue regeneration) can be generated years after the termination of the experiment. It is essential to have an organized storage system to link new data to data acquired and analyzed

years prior. Herein we provide an example of how data files are saved and linked to a specific experiment. These steps are a beginning to a data management system that continues to evolve, especially as new analysis techniques are developed.

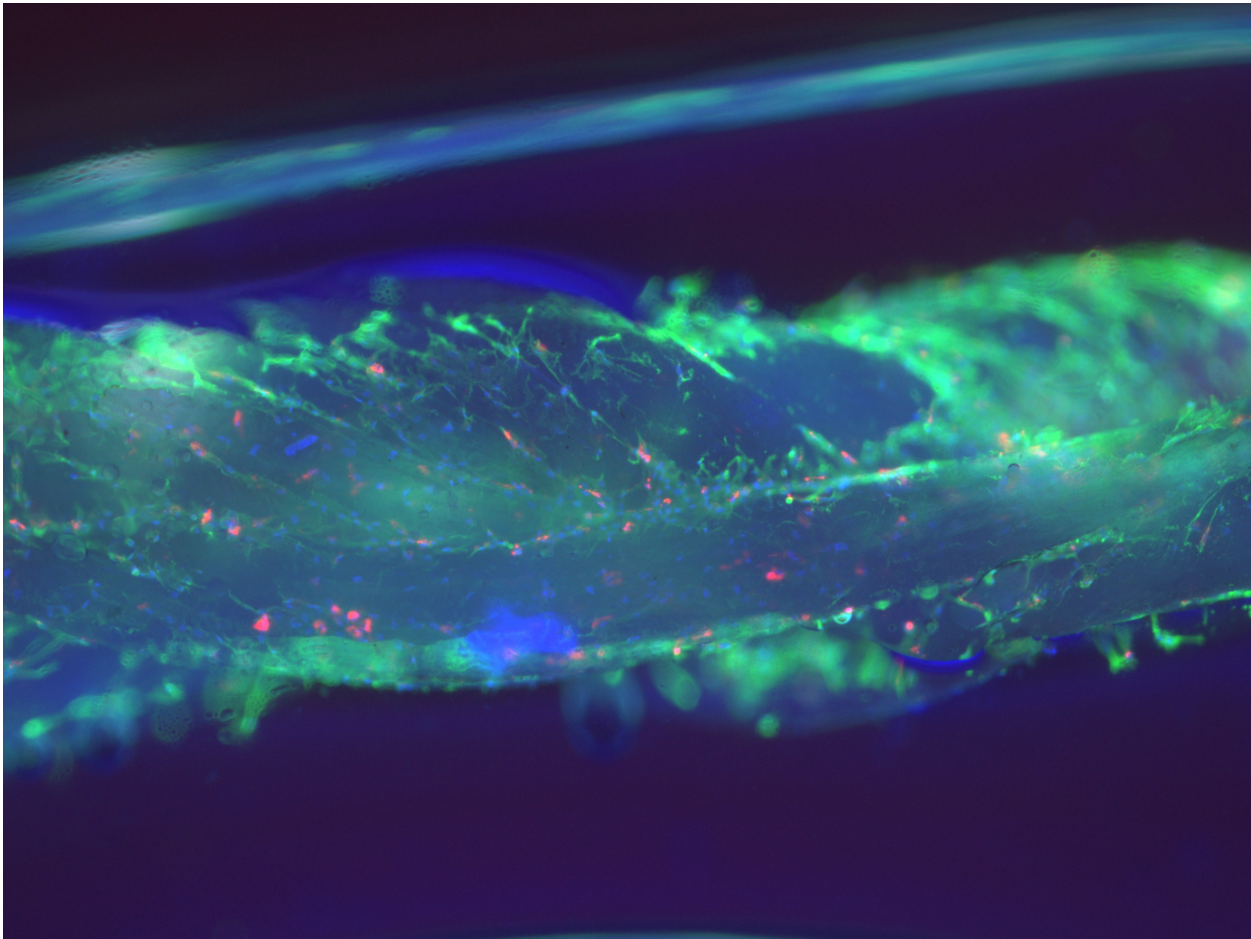
Case study: Regeneration of Functional Heart Tissue in Rats

In the following case study, a research team at Worcester Polytechnic Institute, led by coauthor, Dr. Glenn Gaudette, attempts to regenerate functional heart tissue in rats by delivering stem cells to areas of damaged heart tissue. To successfully perform the delivery of stem cells to an area of infarct, capture key data, and evaluate subsequent changes in the heart tissue, all members of the research team must closely adhere to established procedures and consistent data management practices.

In addition to this particular experiment in this study group, there are multiple control groups of rats that are also studied and multiple time points that are studied. Experimental procedures involving the control groups include:

- Delivering fibrin microthreads to the non-infarcted heart;
- Delivering fibrin microthreads without stem cells to the infarcted heart;
- Delivering fibrin microthreads without cells to the non-infarcted heart;
- Delivering cells via injections (current standard of care) to the infarcted heart;
- Delivering cells via injections (current standard of care) to the non-infarcted heart;
- Delivering cell solution without cells via injections (current standard of care) to the infarcted heart;

Figure 1A: Quantum Dot (red) loaded mesenchymal stem cells on fibrin microthread suture.



- Delivering cell solution without cells via injections (current standard of care) to the non-infarcted heart.

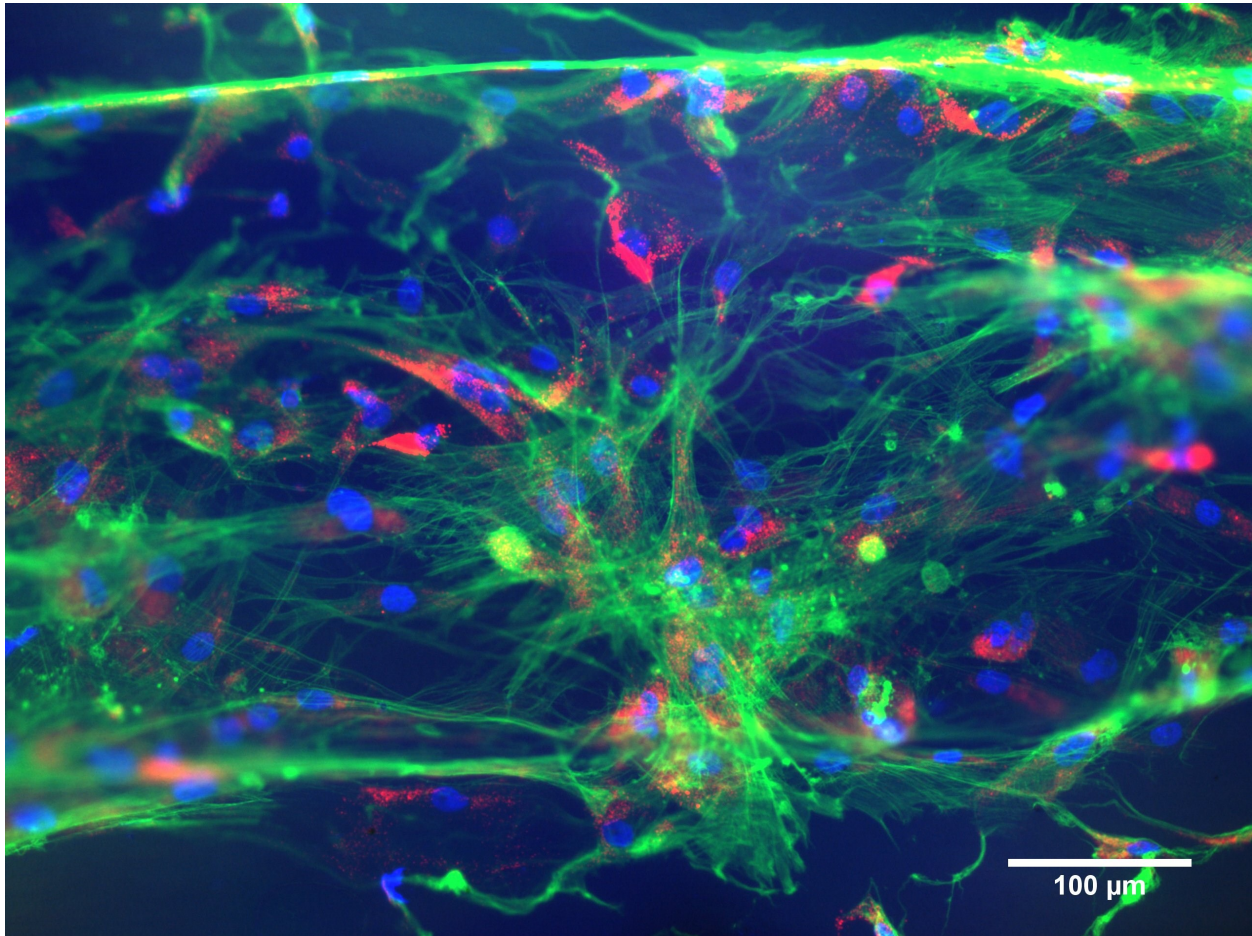
Time points studied may include one hour, one week, and one month after cell delivery. Each group usually consists of 6-8 animals. The large amount of data collected requires all members of the lab involved in data acquisition and analysis to be consistent in data labeling. The following paragraphs explain the steps involved in one experiment for delivering stem cells to damaged rat tissue, the types of data sets collected, and the protocols for team management of the generated data sets.

Delivering Stem Cells to Damaged Heart Muscle

The process of inducing heart attacks in rats is done surgically. Two days prior to this surgery, the research team takes adult stem cells and incubates them with Q dots – fluorescent nanoparticles that serve as cell markers – for 24 hours (Rosen et al, 2007). Then the stem cells are put into a solution and injected into a tube that contains a biological suture. The stem cells cluster around the fibrin microthreads of the biological suture (Figure 1A & B).

The tube is incubated for another 24 hours, and then the team performs surgery on the rat. After opening the rat's thoracic (chest)

Figure 1B: Image is higher magnification of stem cells on a different fibrin microthread.



cavity, the researchers occlude the left anterior descending coronary artery, thus stimulating a heart attack and cutting off oxygen to the heart muscle (Figure 2A). The researchers keep the tissue ischemic for an hour, not allowing any blood to flow through the artery. Thereafter they reperfuse the artery and let the blood return to the surrounding heart tissue. Five minutes afterwards the cell-loaded biological suture is placed through the area of infarcted tissue (Figure 2B), the thoracic cavity is closed up, and the rat is placed back in its cage for a week.

Capturing data post stem cell delivery

The following week the team opens up the rat again, and images of the heart are acquired with a camera system. Two cameras

are used to simultaneously photograph the outside wall of the heart's left ventricle. These cameras use a high frame rate (about 30 frames per cardiac cycle at a rate of 125 frames per second; the cardiac cycle is the cycle of events that occur as the heart is contracting). Figure 3 shows an example of a single frame. A pressure transducer that measures the pressure inside the left ventricle is also used and is synced with the photographing. The digitized pressure data is stored in a file separate from the image files. Four or five times during this process the cameras are repositioned around the area of infarct and additional data sets (images and ventricular pressures) are taken.

The rat is then euthanized. Its heart is removed, put in a fixative, and put into the

Figure 2A: Microthread implantation.

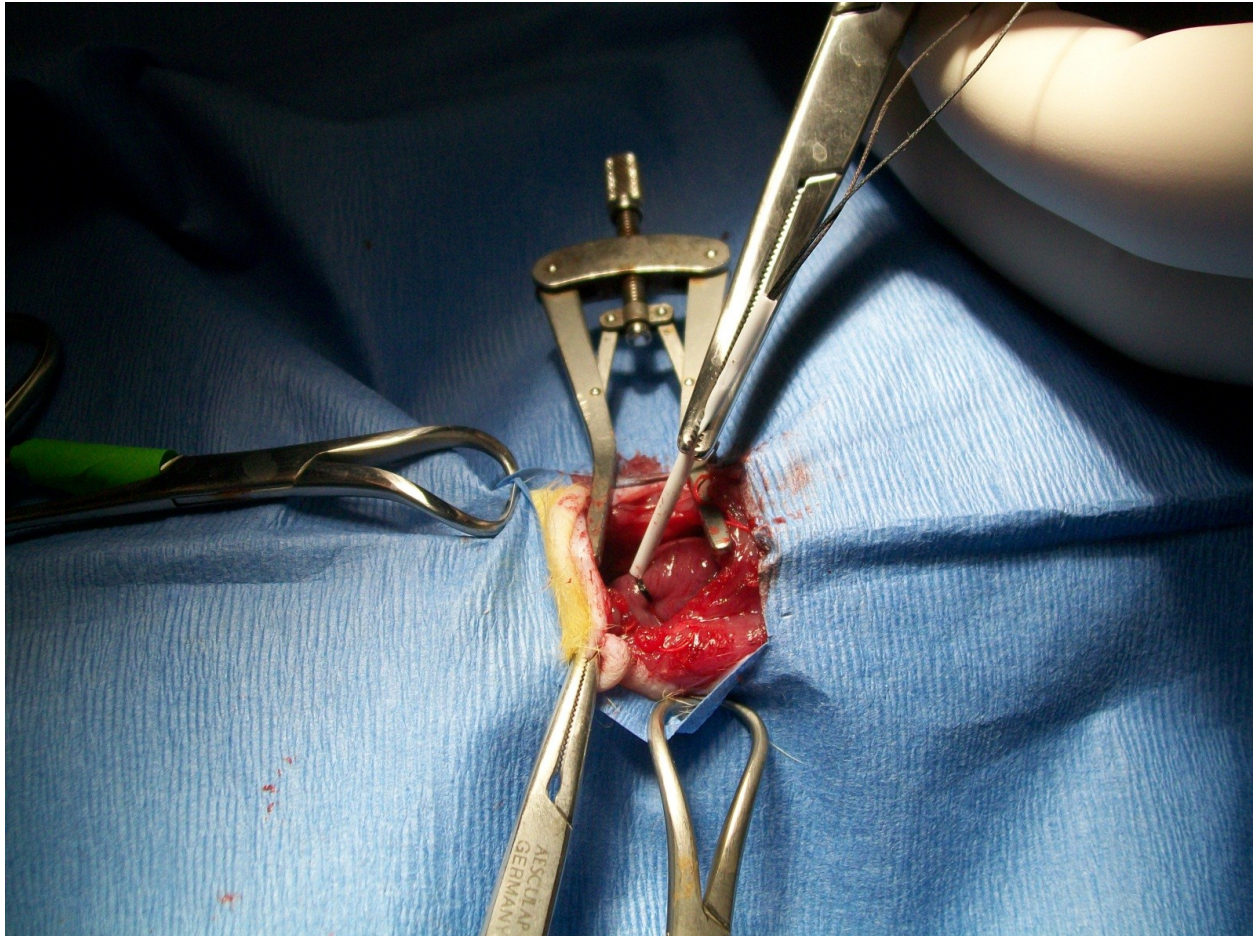


Figure 2B: The cell loaded fibrin biological suture is delivered to the heart.

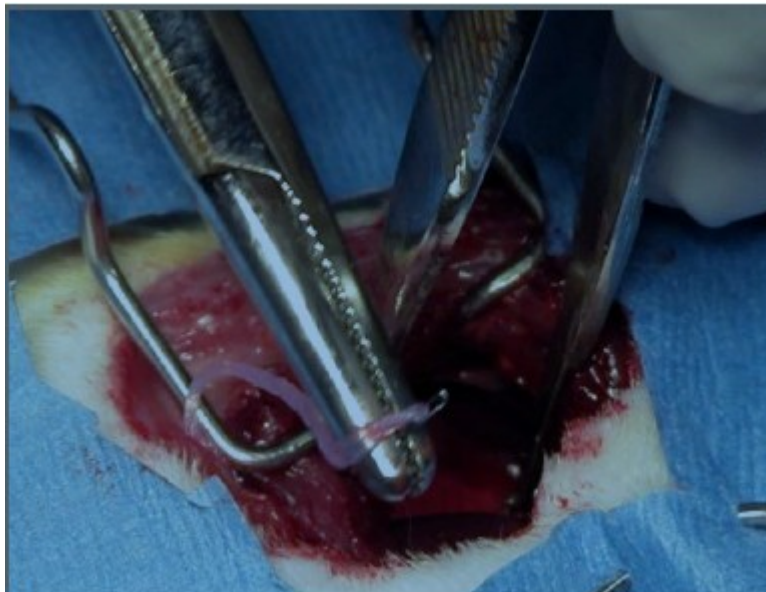


Figure 3: An example of a single image that is acquired to determine regional function in the rat heart. For each data set approximately 1,000 images are acquired at 125 frames per second. The pressure transducer, which is inserted into the left ventricle, is shown on the left (white cable like structure).



freezer for 24 hours. Then the research team begins cutting sections of the heart and putting them onto slides. Three sections of the heart are placed on each slide. The research team generates approximately 500 slides for each rat heart. At any time during this process, tissue that was sectioned and placed on a slide is kept in one lab freezer; tissue that had not been sectioned but embedded and ready to be sectioned is kept in another, and still other tissue that is left in a container may be placed in yet another freezer. This data should be recorded on a spreadsheet to inform all members of the lab

research team.

Some of the slides are stained immediately, and then anywhere from a day to a couple of months after that, other slides may be stained. One of the commonly used stains is Masson's trichrome stain, which enables the research team to identify areas of dead tissue (Figure 4A). Some of the slides are stained for specific markers that indicate exactly where the stem cells are in a cross section of heart tissue, whereas other images document the cardiac myocytes present in the heart (Figure 4B). Images of these are

Figure 4A: Image shows heart tissue with areas of trichrome staining. Blue areas represent mostly dead tissue. Red areas represent healthy tissue. This image is stitched together from 107 individual images. An example of one of these images is shown in Figure 4B. Figure 4C shows healthy heart tissue stained for cardiac specific actinin (a protein found specifically in heart tissue).

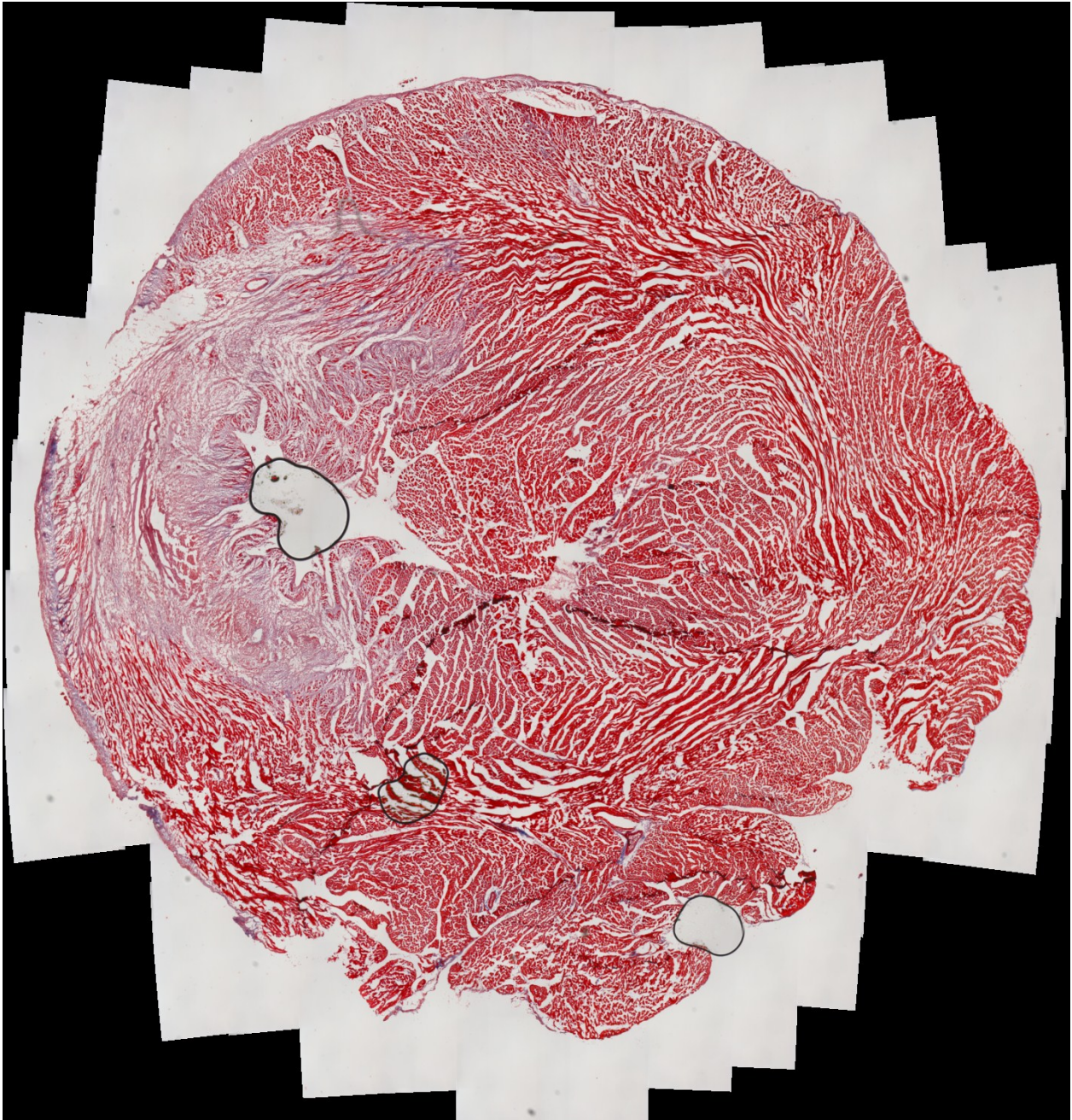
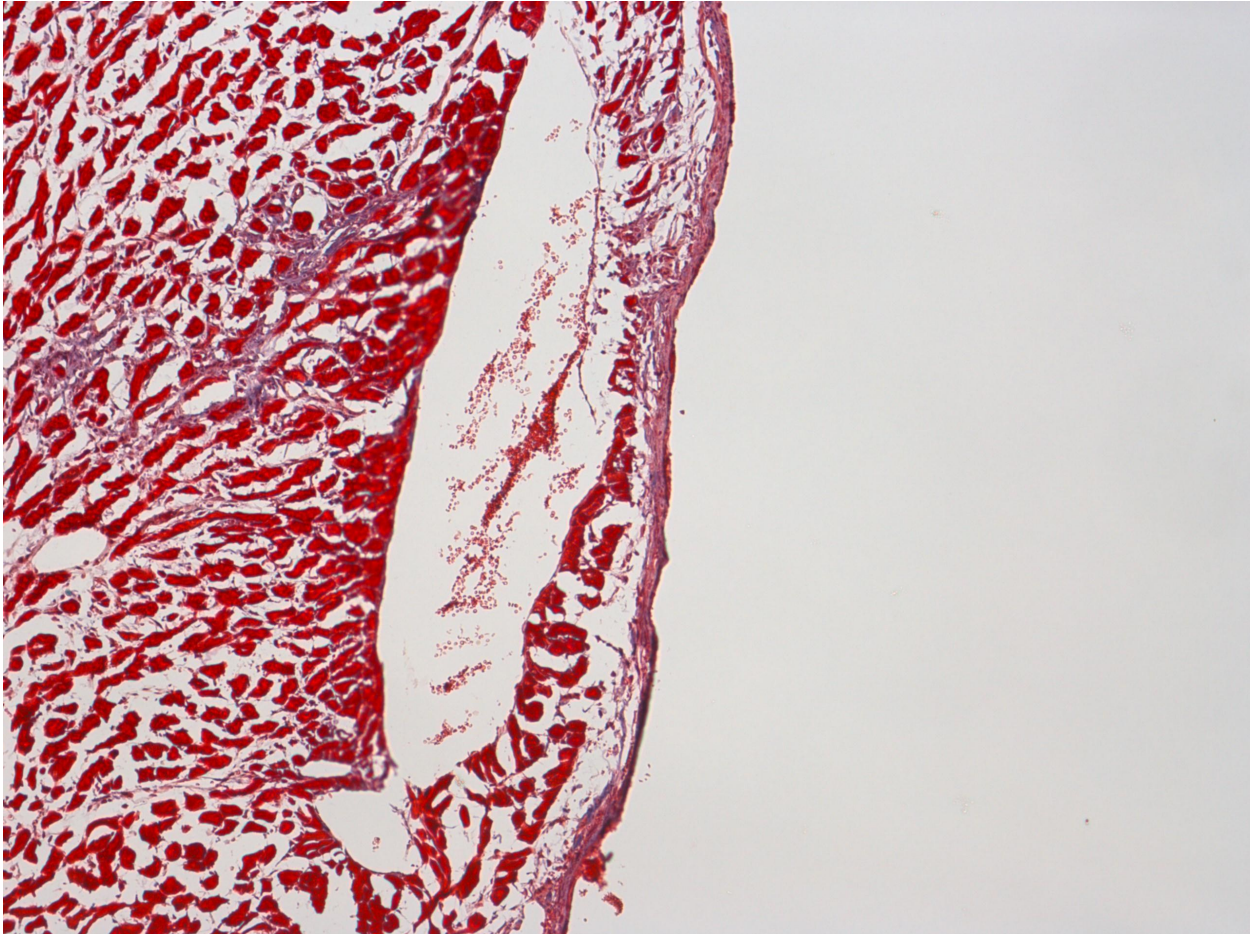


Figure 4B



taken on the lab microscope, and if satisfactory, are then taken with a confocal microscope for higher quality photos. At the same time, the team looks at the data they acquired and uses custom-made software to track particles in the heart to see how far and fast they were moving. This software is used to analyze the optical images of the heart and relays what the function is like in specific regions of heart tissue. This tracking of the particles is captured for several heartbeats in different data sets.

Summary of Experiment Timeline

Time = -2 days

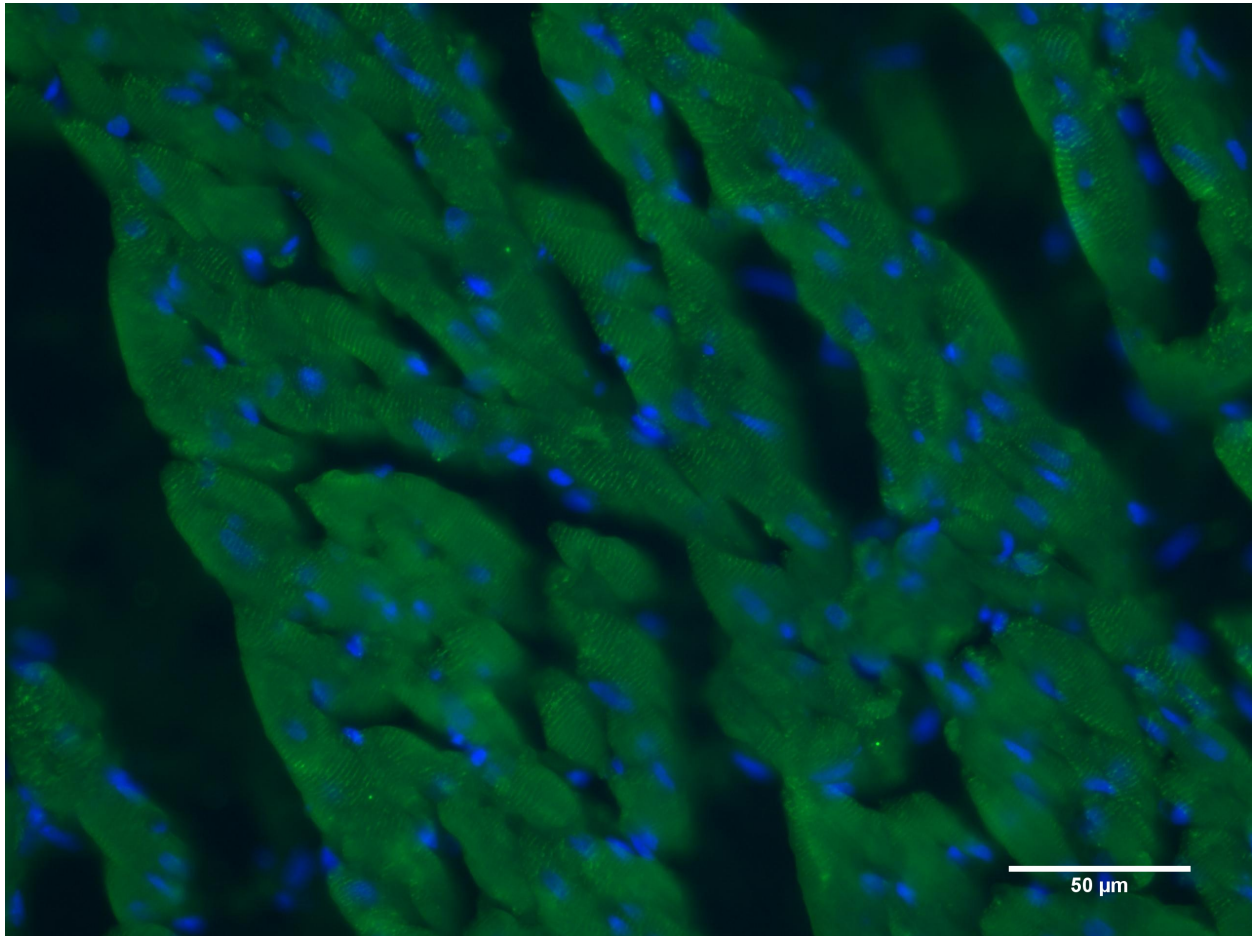
- Incubate mesenchymal stem cells (MSCs) in quantum dot (qDot) solution
 - Lot # for MSCs

- Passage # for MSCs
- Confirm qDot loading
 - ◆ Fluorescent images on inverted scope
 - * Phase contrast image
 - * Fluorescent image
 - * Overlap image (analyzed)
 - * Other
 - ◇ Objective used
 - ◇ Scale bar on image
 - ◇ Multiple images per well

Time = - 1day

- Load MSCs on biological suture
 - Concentration of cells in solution
 - Volume delivered
 - Prepare extra biological suture for imaging to confirm cell loading

Figure 4C: Healthy heart tissue stained for cardiac specific actinin.



- ◆ Stain with Hoechst (nuclei) & Phalloidin (f-actin)
- ◆ Fluorescent images on upright scope
 - * Phase contrast image
 - * Fluorescent images (green, blue, red)
 - * Overlap image (analyzed)
 - * Other
 - ◇ Objective used
 - ◇ Scale bar on image
 - ◇ Multiple locations on each suture
 - ◇ Antibodies used (lot#)
- Gross images of heart and suture placement
- Deliver cell loaded biological suture
- Fix extra cell loaded biological suture in paraformaldehyde
- Prepare extra biological suture for imaging to confirm cell loading
 - Stain with Hoechst (nuclei) & Phalloidin (f-actin)
 - Fluorescent images on upright scope
 - ◆ Phase contrast image
 - ◆ Fluorescent images (green, blue, red)
 - ◆ Overlap image (analyzed)
 - ◆ Other
 - * Objective used
 - * Scale bar on image
 - * Multiple locations on each

Time = 0 day

- Surgical record (pre-op: intra-op and post-op)

- suture
- * Antibodies used (lot#)

Time = 7 days

- Surgical record (post-op)
- Gross images of heart
- Assessment of function
 - Left ventricular pressure
 - Oxygen saturation levels
 - High speed imaging of heart
- Images of heart after bisected (just prior to fixing in paraformaldehyde)

Time = 7+ days

- Analyzed high speed images
 - Displacements between images
 - Displacements over entire heart cycle
 - Correlation of pressure with displacements
 - Displacements within small regions of the heart
- Histological analysis
 - Heart section #
 - Stained images
 - Images on upright scope
 - ◆ Phase contrast image
 - ◆ Color or Fluorescent images (green, blue, red)
 - ◆ Overlap image (analyzed)
 - ◆ Other
 - * Objective used
 - * Scale bar on image
 - * Multiple locations within each heart
 - * Antibodies used (lot#)

Confocal images

Management of Project Data Sets

Paper lab notebooks

The data generated in these experiments are mainly images and numeric measures that are recorded and stored digitally. However, some data – such as observational data – are stored in a paper lab notebook and a paper surgical log. The paper surgical log

is kept with the animal. Lab staff is charged with transferring all information noted in the surgical log into the paper lab notebook.

The content of a lab notebook relates to a particular experiment and is used by all the staff working on that experiment. Keeping a lab notebook enables the preservation of experimental data and observations, proof of discovery (useful for patent applications). The lab notebook is a source of information for reports, white papers, and theses, and serves as an index into the actual datasets. Given the transient nature of graduate students' participation in research labs (they typically work in the lab while pursuing their graduate degree, leaving the lab once they graduate), data management instruction is regularly provided to new team members. To each new graduate student, detailed instructions on what needs to be recorded in lab notebooks and the proper formatting of lab notebook content are provided. Rules for lab notebook formatting include the following:

- The title of the project is to be noted on the notebook's exterior;
- Notebook needs to be bound and numbered;
- There must be a sign-out page;
- Instructions about the project must be noted;
- Table of contents (that must be updated with each new entry);
- Preface;
- Table of abbreviations;
- On each page, the date, time of entry, signature of author and signature of witness, and a clear descriptive heading must be posted;

- Any attachments such as a computer printout must be permanently mounted. A signature or initials should go from this insert to the notebook page, and a line should be made from the printout to the page.

(Gaudette, 2011)

Each experiment in the lab generates huge amounts of digital data. The following outline indicates the raw and analyzed digital data generated in one experiment.

Raw Data

The raw data sets that result from experimentation on one rat heart include the following:

- Images of cells loaded with fluorescent particles (to confirm cell uptake);
- Images of cells loaded onto biological sutures (from quality control sutures not used in the experiment);
- Surgical records;
- Images of the heart;
 - Left ventricular pressure data;
 - ◆ Saved in same folder as data set.
 - Mechanical function;
 - ~ 1000 images per data set
 - ~10 data sets per experiment
 - General anatomy;
 - ◆ To document location of ischemic zone;
 - ◆ To document suture location prior to histological sectioning.

A/D (analog to digital) inputs (measure left ventricular pressures) that correspond to the images.

- Histologic sections (sections of heart tissue on slides);
 - ~3 per slide
 - ~500 per experiment

Analyzed Data

The analyzed data sets resulting from one experiment include the following:

- Displacement data between consecutive images over an entire heart cycle;
 - Summary sheet of displacements for each heartbeat;
 - Regional displacement analysis.
- Pressure analysis (heart rate, developed pressure);
- Histology (minute structure of the cardiac tissue as discerned by microscope);
 - Hematoxylin and eosin; Masson's trichrome staining;
 - Immunohistochemistry;
 - ◆ Actinin, troponin, actin, calcium channels, Q dots;
 - ◆ Images;
 - * 5x, 10x, 40x;
 - * Panorama;
 - * Confocal.

Coordinating Efforts in Data Management

Multiple members of the research team analyze data from the same heart. One person will examine the mechanical function of the heart, one will be doing trichrome staining, another will do the actinin staining, and another will do the imaging. All the data sets are linked in an Excel spreadsheet so that these disparate data sets can be integrated for analysis. One graduate student, (generally a more senior PhD student), is charged with the upkeep of the spreadsheet. The spreadsheet includes the following metadata: information about the date of the experiment, strain of animal used, data acquired and analyzed, material and cells implanted, animal weight and comments about the experiment.

Conventions for Naming Files

All the laboratory personnel (e.g. graduate students, research assistants, lab managers) are instructed to use uniform file names. In the case of the image files for an experiment, the file name includes the experiment number, the data set number, and the image number and is formatted in the following manner:

Exp#_DS#_frame#_file format (Experiment Number_Dataset Number_Frame number_File format)

An example of such a file is: 0231_ds2_0008.tif.

In this example, the experiment number is 0231, it is the second data set, and the eighth image in that data set. Most data sets contain 1,000 images. The tif represents the digital image file format. The file can also be stored as an .avi movie file (e.g. 0231_ds2.avi).

The left ventricular pressure data is stored with the following naming convention:

Exp#_DS#_file format (Experiment Number_Dataset Number_File format)

An example of such a file is: 0231_ds2.pdt.

In this example, the experiment number is 0231 and it is the second data set. The data contained within this file includes a list of the image numbers and the left ventricular pressure at the time of image acquisition. For data acquired at 125 frames/second, this file would contain the image number, e.g. 0010, and the time at which that was acquired, e.g. 80 ms.

Naming histology data files

Prior to cutting histological sections, the sample is bisected through the center of the infarcted region, providing two pieces of heart tissue: the apical side, closest to the

apex of the heart (A) and the basal side, closest to the base of the heart (B). The lab's convention for naming histology files is Experiment #, A or B, and Section #. When the sample is stained, an abbreviation for the type of stain used is added to the slide. An example of a histology file name is: 0231A_216_act.

When each participant in the research project names data files in this manner, the process of finding and retrieving data files is easy and no one is left guessing what is represented by an obscure file name. It also helps one glean specific details about the data set at a glance. For example, if someone were to search for data on cardiac tissue stained with actinin, one would be able to identify the slides stained with actinin through the file name act. This convention also reduces the need for a student to use a look-up table to determine what sample a specific image corresponds with.

Data Storage

All digital data sets are stored on institutional servers at WPI. These are maintained by WPI's IT division which performs backup, preservation, and archival services. They do not need further cleaning for archiving and preservation. All reports, presentations, and research papers will be deposited along with the data sets. The data sets will be accessible for at least three years as required by federal funding agencies.

Data Management Plan

Dr. Gaudette and his co-PIs developed a data-sharing plan for this project that is consistent with NIH guidelines. In this plan, they address these key points:

- The raw data generated, the analyzed results, and the simulations are stored on WPI maintained storage arrays and servers with an institutional backup and archiving strategy;

- Public release of data will be made within three years after completion of project;
- There is no embargo period for this data;
- The researcher retains the right to use the data before opening it to a wider audience (e.g. to write manuscript or file patent applications);
- Data sets are not covered by copyright and do not contain any personal data;
- Data sets cannot be used for commercial applications or changed and resubmitted without the PI's permission, and are subject to WPI's intellectual property policies;
- Data will be stored as conventional files that will be easily accessible for other researchers.

Data Management Instruction

At the beginning of each semester Dr. Gaudette provides detailed instruction to his students on keeping the paper lab notebook and file naming conventions for his lab's experiments. Unfortunately laboratory notebooks are not currently digitized. The laboratory notebooks are kept in the lab at all times. This requires students to be in the lab to access data. Future improvements in data management in Dr. Gaudette's lab may include digitized laboratory notebooks. Providing training to students new to the lab maintains consistent data management practices and ensures that any unclear issues regarding data management are addressed prior to the students engaging in the research project. While consistent data management instruction has been a priority in Dr. Gaudette's lab, data management instruction varies greatly from one research environment to another.

Until recently, little attention has been given to the need for formal data management in-

struction in science and engineering curricula. Typically graduate students, as front-line workers in the research lab handle experimental data and oversee the handling of it by other students and lab assistants. Most students learn data management haphazardly from fellow graduate students, who in turn had learned from their predecessors in the lab. This inconsistent teaching of data management is often sketchy and fails to cover all the practices that ensure discoverability, access, storage, and long-term preservation of data. Moreover, graduate students come and go frequently during the course of a research project; working for a year or two in the lab before graduating and leaving the university.

Recognizing the need for teaching data management instruction to students in the sciences, both coauthors of this paper have collaborated on an IMLS planning grant between the University of Massachusetts Medical School and Worcester Polytechnic Institute, to plan and develop a data management curriculum that can be easily integrated into science, health sciences, and engineering courses at the undergraduate and graduate levels (University of Massachusetts Medical School and Worcester Polytechnic Institute, 2012). Implementing instruction on the value of data sharing and data management best practices into undergraduate and graduate courses will not only ease data management in the individual research lab, it will serve to foster open sharing of data both within a lab and with external labs.

Data management should be incorporated into every graduate program that relies on the generation or analysis of data. The Biomedical Engineering program at Worcester Polytechnic Institute is considering adding these modules from the data management curriculum to its seminar program, a required course for all graduate students. Specific examples of data management should be provided to all students; and, potentially as part of a homework assignment, students

should be asked to design their own data management plan.

Conclusions

The specific example of data management in biomedical engineering presented here illustrates an organized method by which laboratory members can store and retrieve files. A universally used naming convention quickly provides lab members with an understanding of what each file contains. Detailed metadata is contained within a master spreadsheet. Digitizing laboratory notebooks might provide an inclusive data set that could be assessed remotely. This case study illuminates the need for teaching data management to students involved in data heavy research.

References

- American Heart Association. 2012. "About heart attacks." Accessed Sept. 11, 2012 at http://www.heart.org/HEARTORG/Conditions/HeartAttack/AboutHeartAttacks/About-Heart-Attacks_UCM_002038_Article.jsp.
- Bureau of Labor Statistics, U.S. Department of Labor, "Biomedical engineers." *Occupational Outlook Handbook, 2012-13 Edition*, Accessed Sept. 17, 2012 at <http://www.bls.gov/ooh/architecture-and-engineering/biomedical-engineers.htm>.
- Gaudette, G. 2011. Keeping a lab notebook in the Gaudette lab. Unpublished.
- Imperial College London, Department of Engineering. N.D. "Definition of biomedical engineering." Accessed Nov. 19, 2012 at <http://www3.imperial.ac.uk/pls/portallive/docs/1/51182.PDF>.
- National Institute of Health. 2003. "Final NIH statement on sharing research data." Accessed Jan.23, 2013 at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.

Rosen et al. "Finding fluorescent needles in the cardiac haystack: tracking human mesenchymal stem cells labeled with quantum dots for quantitative in vivo three-dimensional fluorescence analysis." *Stem Cells* 25, no. 8 (2007): 2129-38. Accessed Oct. 25, 2012, <http://dx.doi.org/10.1634/stemcells.2006-0722>

University of Massachusetts Medical School and Worcester Polytechnic Institute. 2012. "Frameworks for a Data Management Curriculum." Accessed Oct. 25, 2012 at http://library.umassmed.edu/data_management_frameworks.pdf.

Acknowledgments

Image acquisition and analysis was performed by John Favreau. This work was supported by National Institutes of Health R21HL093639, 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.

Disclosure: The authors report no conflicts of interest.

All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a Creative Commons Attribution-Noncommercial-Share Alike License <http://creativecommons.org/licenses/by-nc-sa/3.0/>

ISSN 2161-3974