University of Massachusetts Medical School

eScholarship@UMMS

University of Massachusetts and New England Area Librarian e-Science Symposium

2012 e-Science Symposium

Apr 4th, 12:00 AM - 12:15 PM

Science Panel Discussion presentation: "Data Management in Biomedical Engineering: Needs and Implementation"

Glenn R. Gaudette
Worcester Polytechnic Institute

Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

 Part of the Biomedical Engineering and Bioengineering Commons, and the Library and Information Science Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

Repository Citation

Gaudette, G. R. (2012). Science Panel Discussion presentation: "Data Management in Biomedical Engineering: Needs and Implementation". *University of Massachusetts and New England Area Librarian e-Science Symposium.* https://doi.org/10.13028/w4d1-et88. Retrieved from https://escholarship.umassmed.edu/escience_symposium/2012/program/9

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.
This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Data Management in Biomedical Engineering: Needs and Implementation

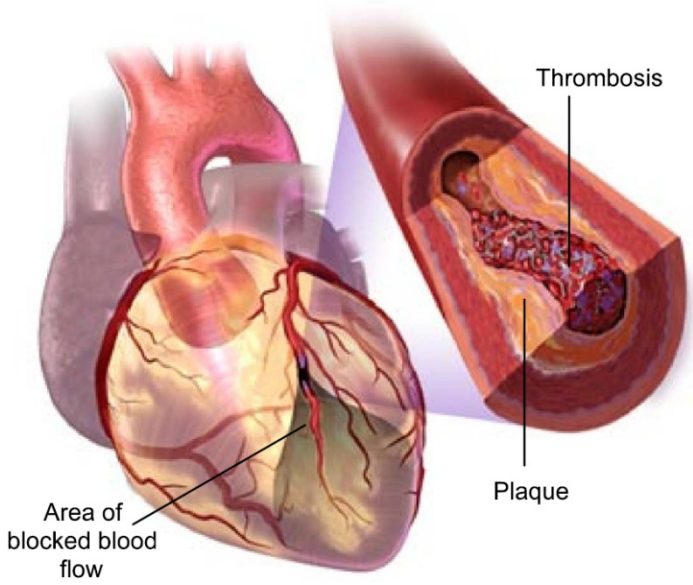
Glenn R. Gaudette, PhD
Worcester Polytechnic Institute

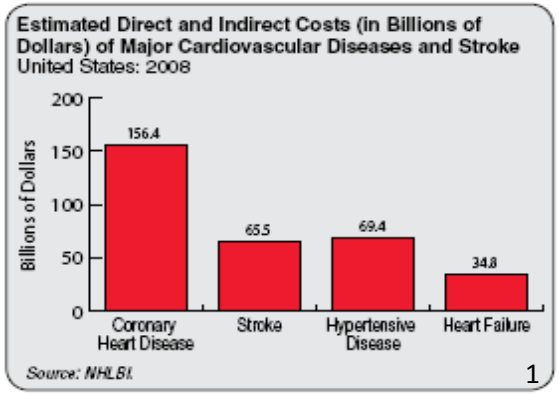
Biomedical Engineering

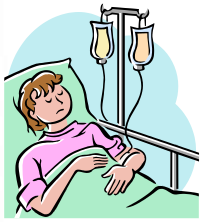
- **Discipline that improves human health through cross-disciplinary activities that integrate principles from:**
  - **Engineering**
  - **Mathematics**
  - **material science**
  - **computational sciences**
  - **life sciences**
  - **clinical practice**
- **Fastest growing engineering field**
- **3rd fastest growing occupation (US Labor Dept)**

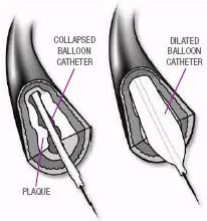
Heart Disease

- Coronary heart disease can lead to ischemia
  - Affects ~ 16 million Americans
- Heart attack (myocardial infarction) results in muscle tissue death
  - Affects ~ 8 million Americans
- Heart does not repair itself
- Cell therapy may improve cardiac function
- Methods to deliver cell therapy are inefficient






Intravenous



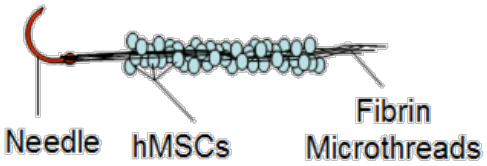
Intra-coronary

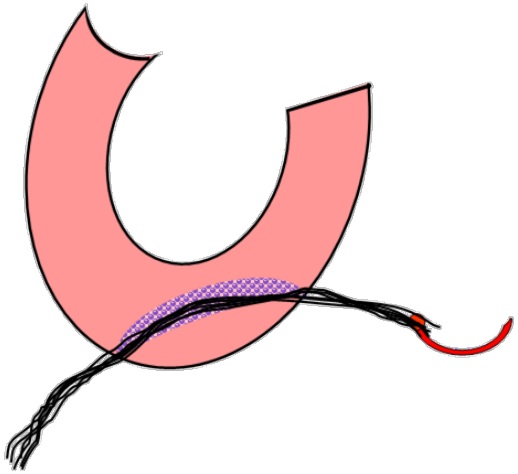

Intra-myocardial

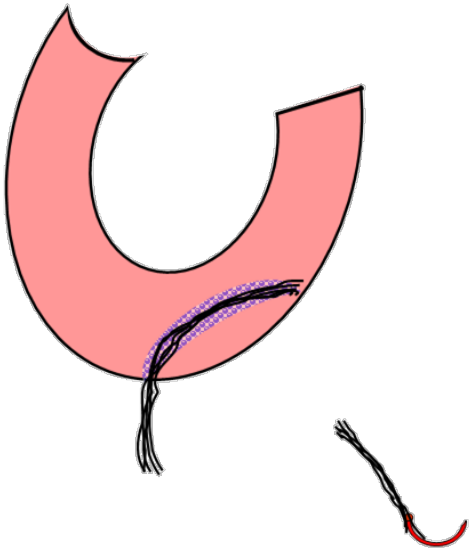

Scaffold

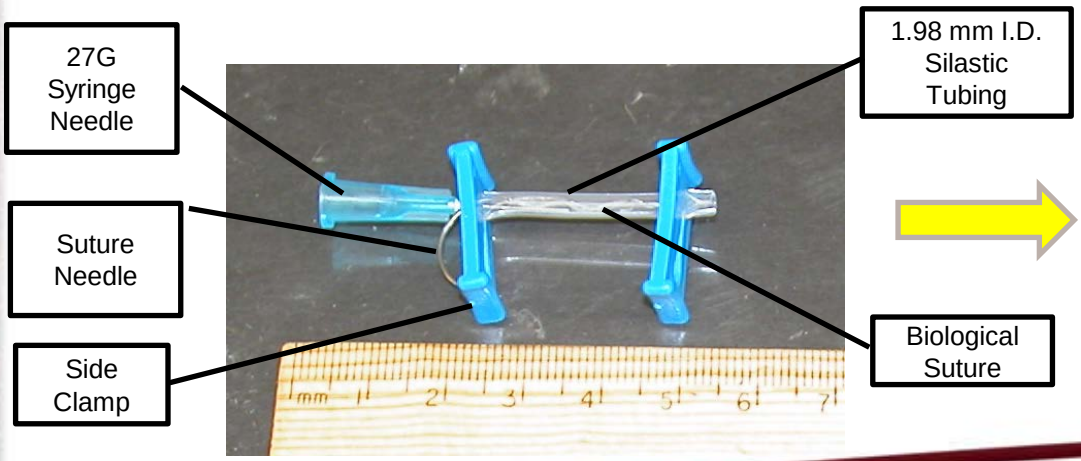
1. American Heart Association. Heart Disease and Stroke Statistics. 2008.

Loading Cells on Biological Sutures

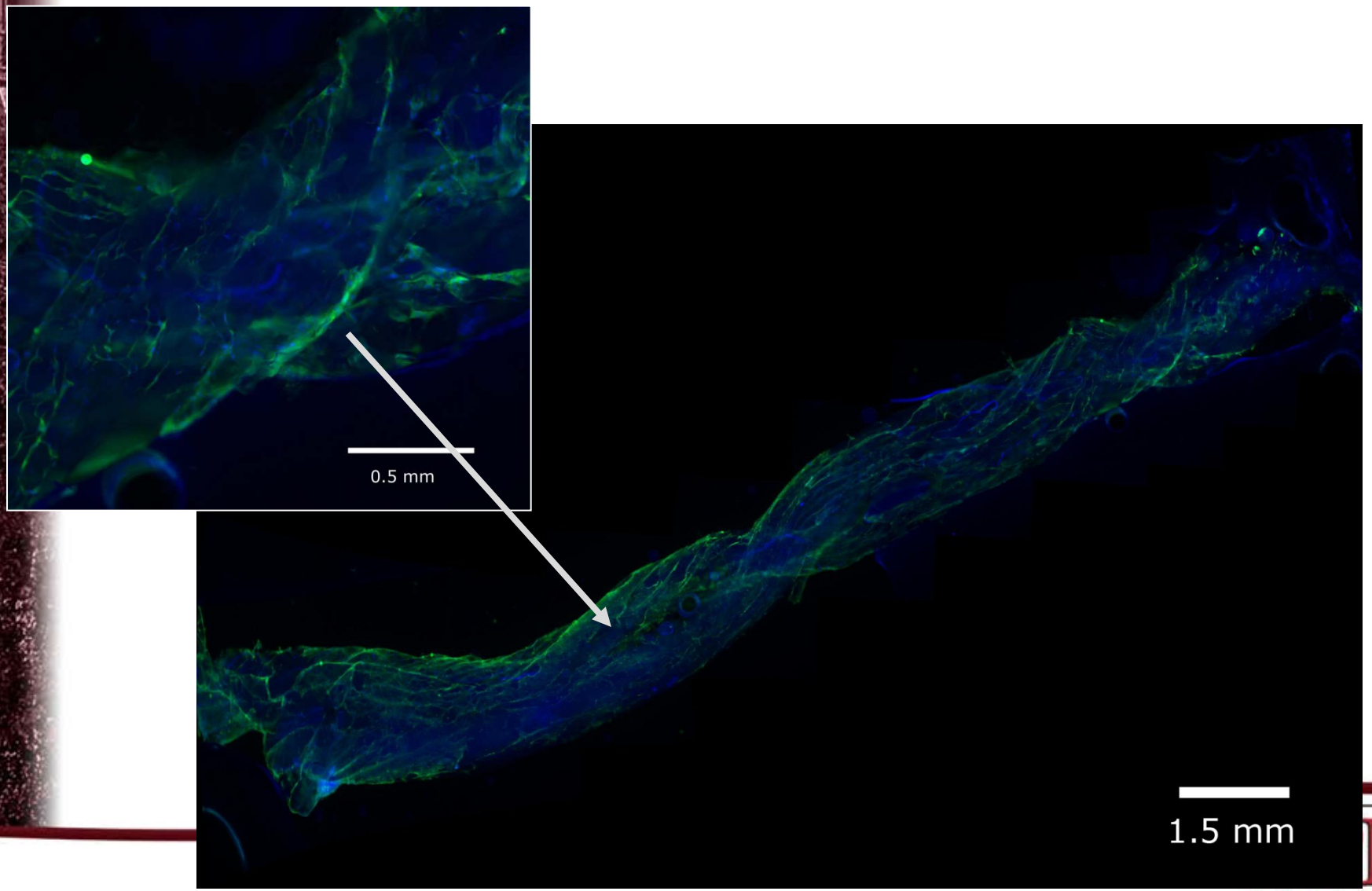




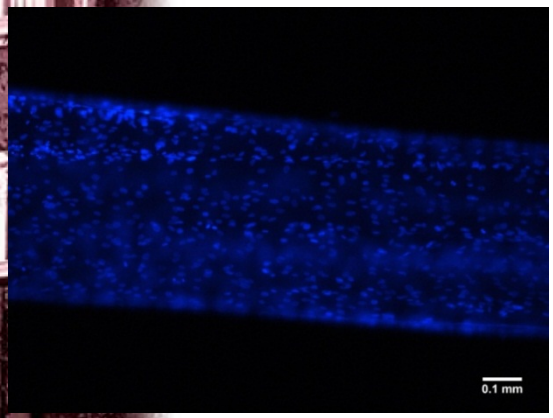


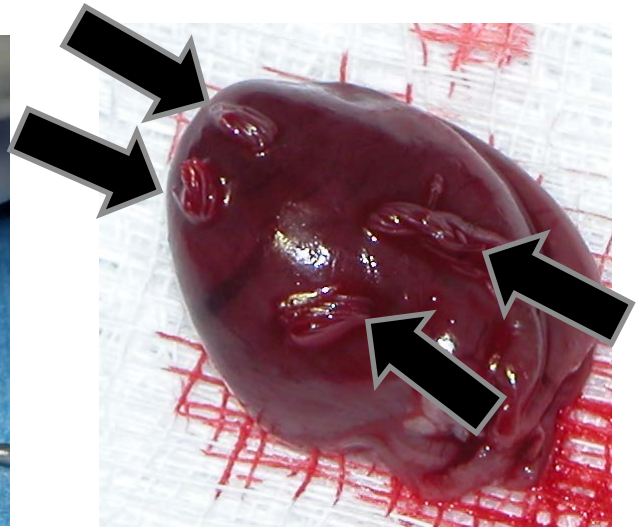


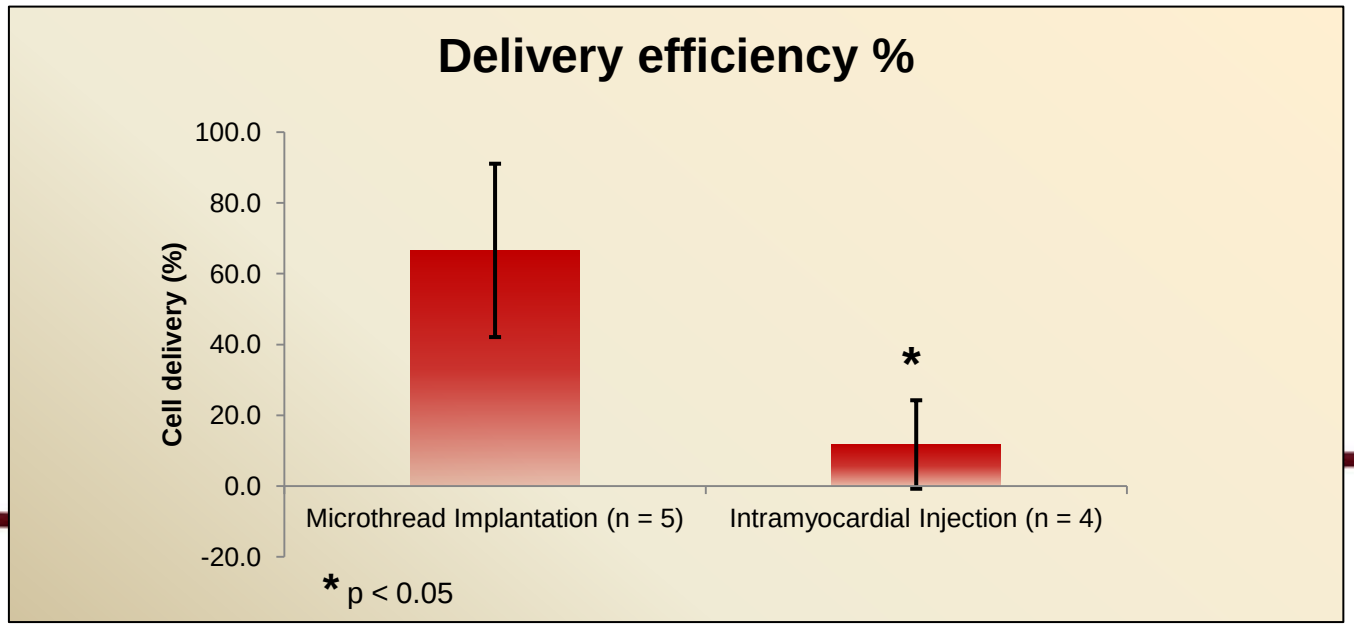
Stem Cells on Microthreads



Efficient Cell Delivery with Biological Sutures

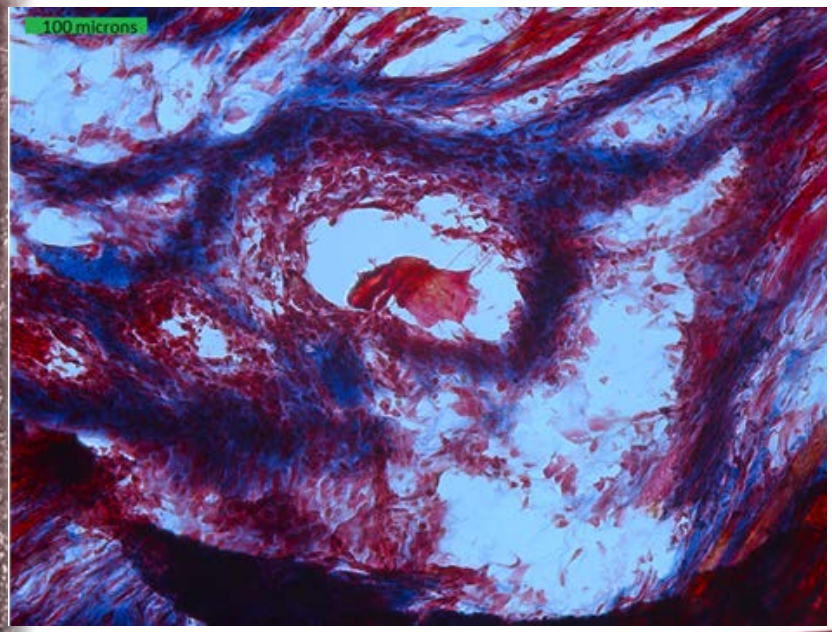


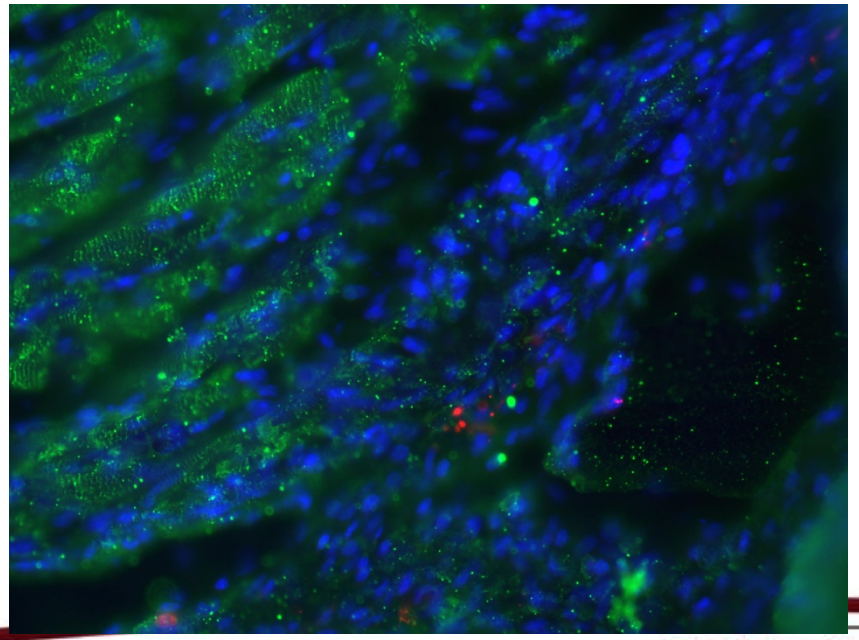




Biological Sutures Delivery Cells to the Beating Heart

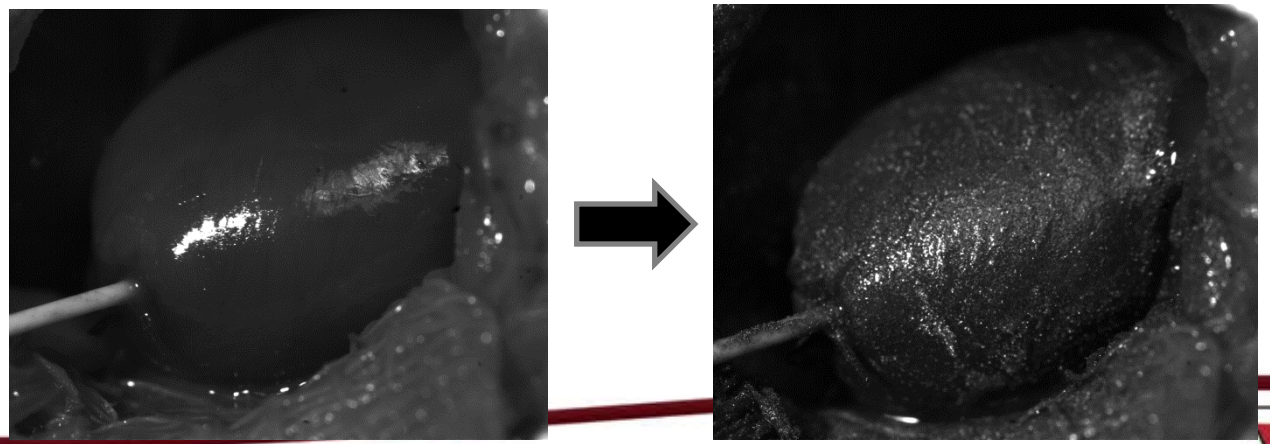
- Histology images (400x)



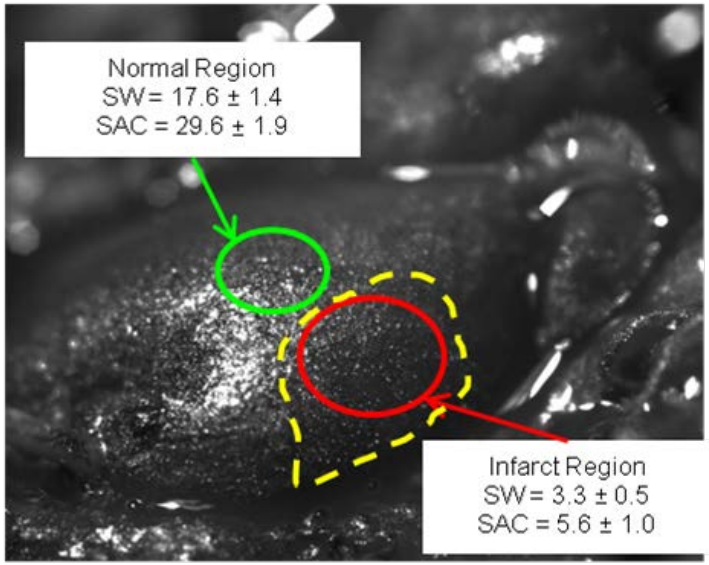


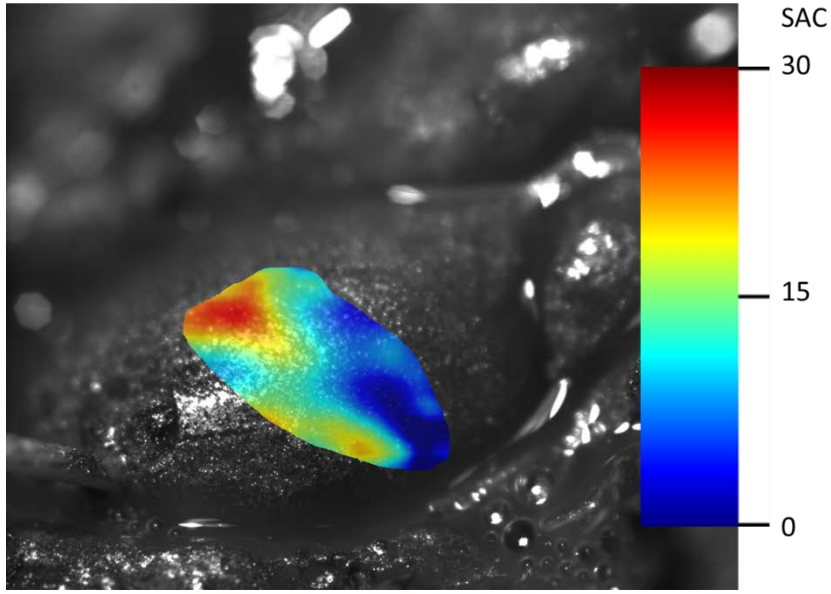
Data Collection

- High frame rate (~250 fps) videos
  - Target: >=30 Frames per cardiac cycle
- 8-bit grayscale
- Speckle application
  - Silicon carbide nanoparticles with retro-reflective beads
  - Increases light intensity distribution



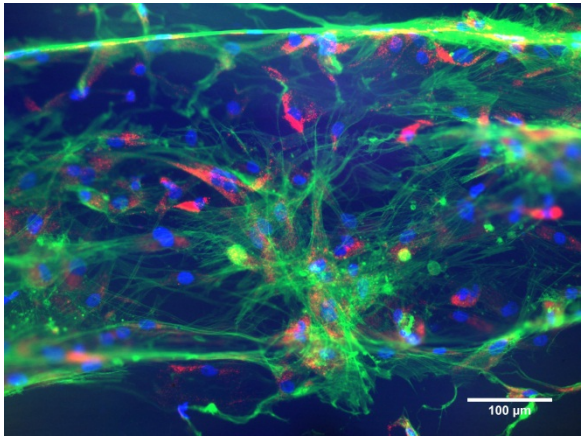
HDM Can Determine Regional Mechanical Function in the Infarcted Rat Heart





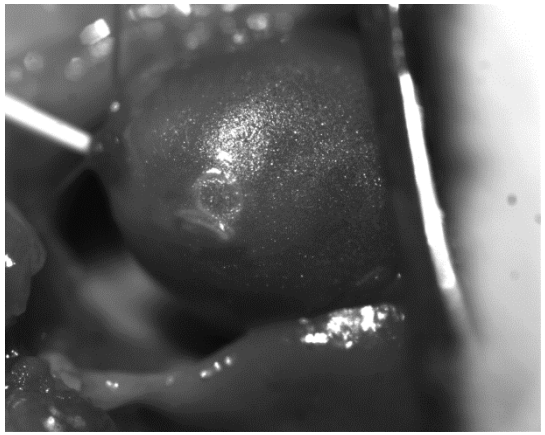
Typical Data Sets for One Experiment – Pre-Experiment

- Pre-operative record
  - Weight, heart rate, oxygen saturation
  - Drug dosage
- Images of cells loaded on suture
- Time of procedure
  - Incisions
  - Cell delivery
- Images of cell delivery



Typical Data Sets for One Experiment – Raw Data

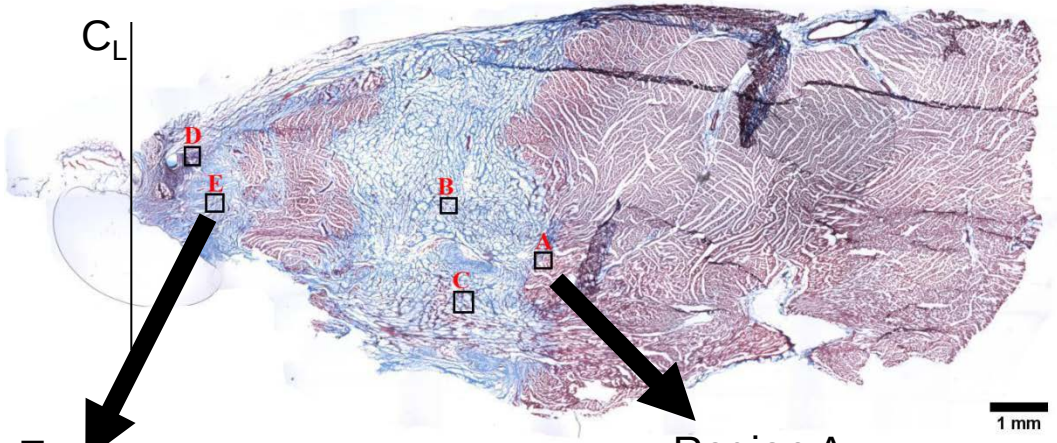
- Images of Heart (mechanical function)
  - 1,000 images/data set
  - ~10 data sets/experiment
- A/D inputs (correspond to images)
- General images
- Histological sections
  - ~3/slide
  - ~500 per experiment



Typical Data Sets for One Experiment – Analyzed Data

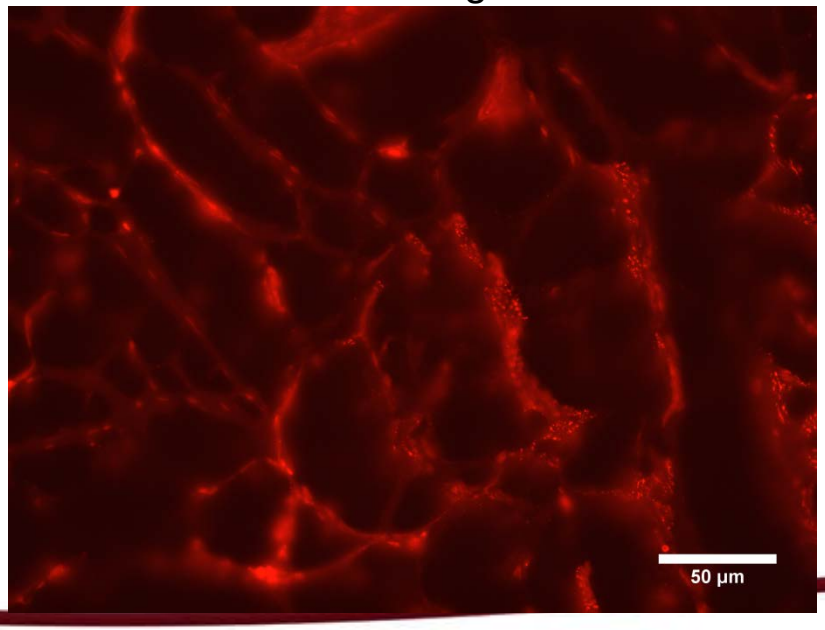
- U and V displacement data between 2 images
  - Summary sheet for each heart beat
  - Regional analysis
- Histology
  - H&E; trichrome; phase contrast
  - Immunohistochemistry
    - Actinin, troponin, actin, calcium channels, Qdots
    - Images
      - 5x, 10x, 40x
      - Panorama
      - confocal

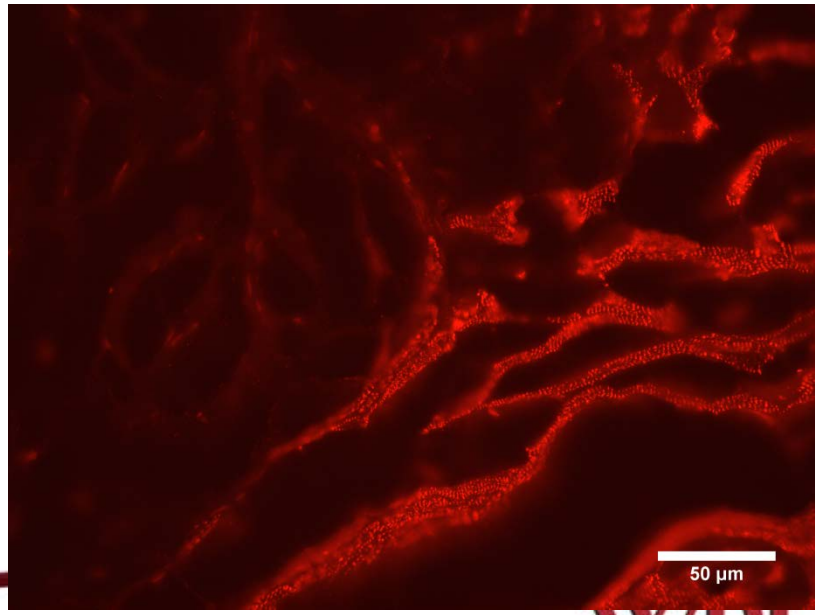
Trichrome Staining
Immunohistochemistry



Region E

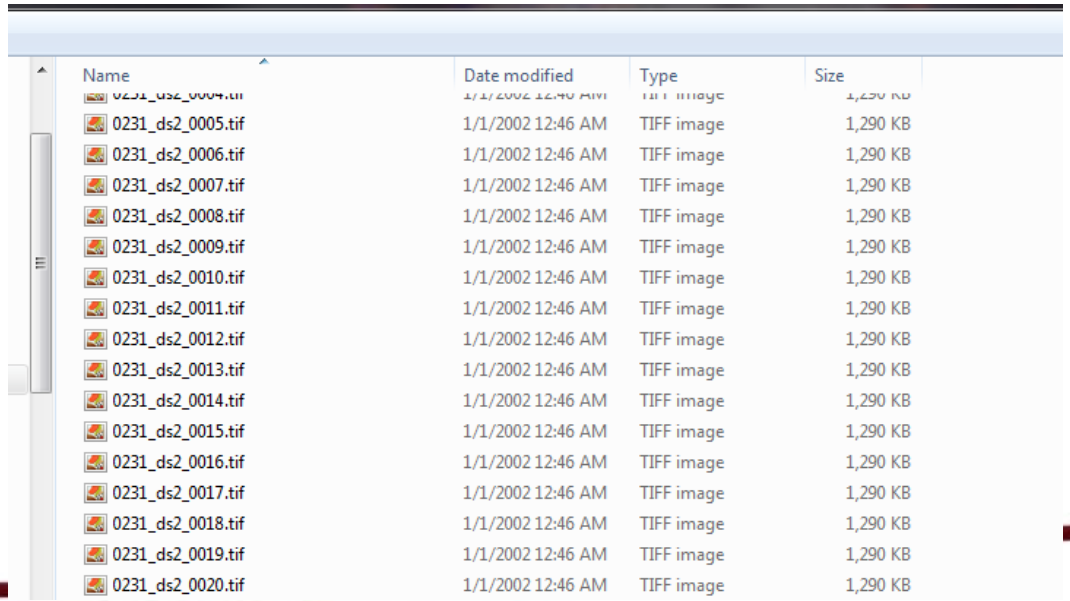
Region A





Data Management: Images for Mechanical Function

- File names must be uniform
  - Exp#_DS#_frame#
- Example:
  - 0231_ds2_0008.tif



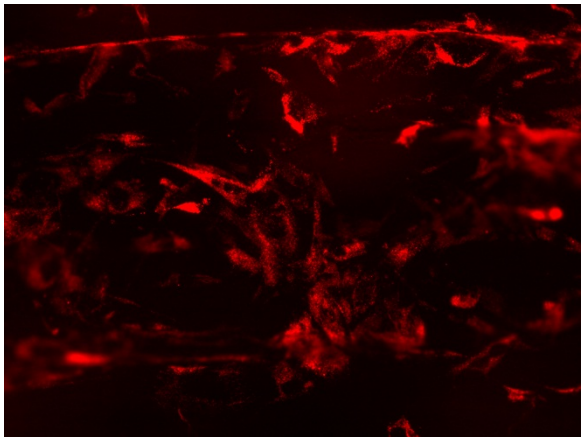
Name	Date modified	Type	Size
0231_ds2_0005.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0006.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0007.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0008.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0009.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0010.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0011.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0012.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0013.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0014.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0015.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0016.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0017.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0018.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0019.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB
0231_ds2_0020.tif	1/1/2002 12:46 AM	TIFF image	1,290 KB

Data Management: Images for Histology

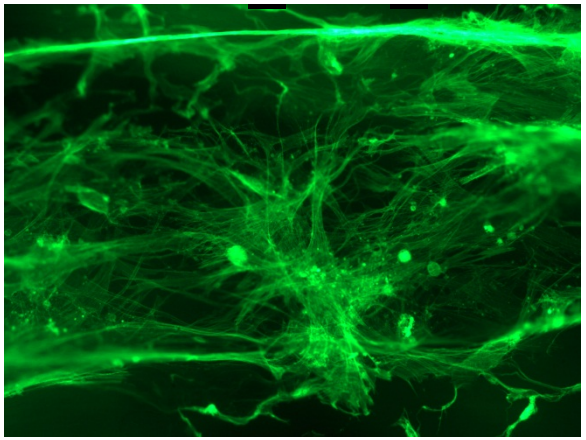
- File names must be uniform
  - Exp#A(or B)_section#_stain
- Example:
  - 0231A_216_act

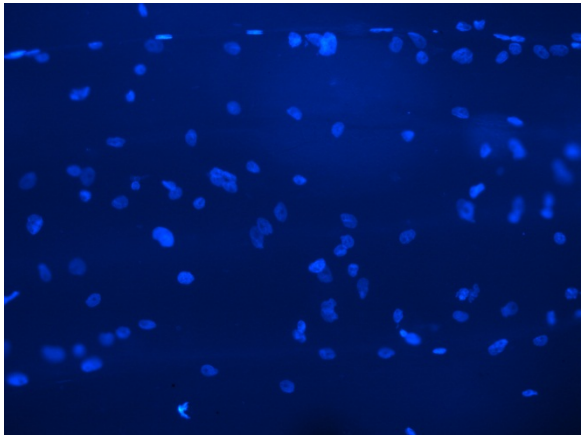
Cells on Suture

0402_001_qd

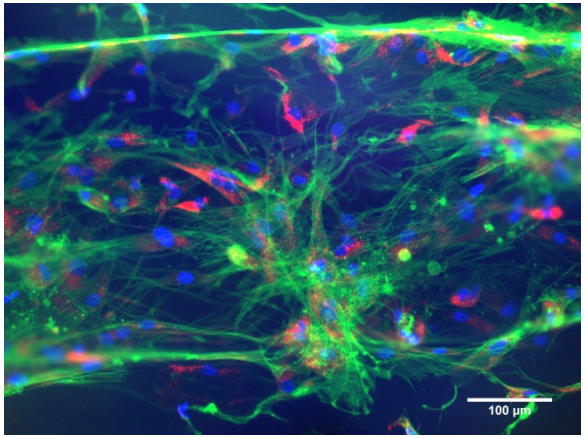


0402_001_act





0402_001_ho



0402_001_merge

How to Educate Students

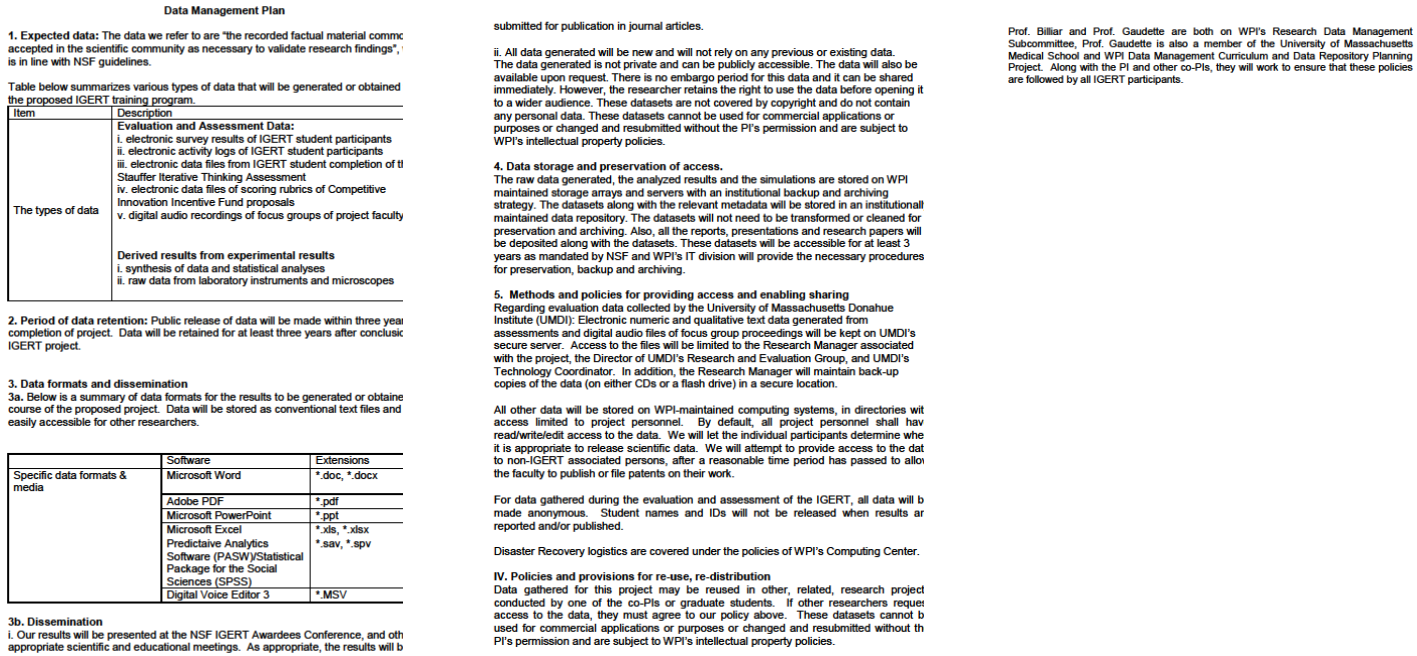
- In lab:
  - Mention it once per semester
  - Files with naming conventions
- In courses:
  - Seminar series (mandatory for grad students)
  - Need “buy in” from colleagues

Data Sharing/Management Plans

2010

Data Sharing Plan. The investigators agree to participate in a data sharing plan consistent with NIH guidelines. These data would be made available to interested researchers as data become available. Any model organisms developed in the course of these studies would likewise be made available. In later years of the proposal, we have requested funding to present our preliminary data at a national meeting, such as the American Heart Association, Society for Thoracic Surgery or American College of Cardiology, as appropriate. Additionally, a manuscript of the results of this project will be submitted to an appropriate peer reviewed journal.

2010



Data Management Plan

1. Expected data: The data we refer to are "the recorded factual material commc accepted in the scientific community as necessary to validate research findings", is in line with NSF guidelines.

Table below summarizes various types of data that will be generated or obtained the proposed IGERT training program.

Item	Description
The types of data	**Evaluation and Assessment Data:** i. electronic survey results of IGERT student participants ii. electronic activity logs of IGERT student participants iii. electronic data files from IGERT student completion of tl Stauffer Iterative Thinking Assessment iv. electronic data files of scoring rubrics of Competitive Innovation Incentive Fund proposals v. digital audio recordings of focus groups of project faculty **Derived results from experimental results** i. synthesis of data and statistical analyses ii. raw data from laboratory instruments and microscopes

2. Period of data retention: Public release of data will be made within three yeai completion of project. Data will be retained for at least three years after conclusic IGERT project.

3. Data formats and dissemination
3a. Below is a summary of data formats for the results to be generated or obtaine course of the proposed project. Data will be stored as conventional text files and easily accessible for other researchers.

	Software	Extensions
Specific data formats & media	Microsoft Word	*.doc, *.docx
	Adobe PDF	*.pdf
	Microsoft PowerPoint	*.ppt
	Microsoft Excel Predictaive Analytics Software (PASW)/Statistical Package for the Social Sciences (SPSS)	*.xls, *.xlsx *.sav, *.spv
	Digital Voice Editor 3	*.MSV

3b. Dissemination
i. Our results will be presented at the NSF IGERT Awardees Conference, and oth appropriate scientific and educational meetings. As appropriate, the results will b

submitted for publication in journal articles.

ii. All data generated will be new and will not rely on any previous or existing data. The data generated is not private and can be publicly accessible. The data will also be available upon request. There is no embargo period for this data and it can be shared immediately. However, the researcher retains the right to use the data before opening it to a wider audience. These datasets are not covered by copyright and do not contain any personal data. These datasets cannot be used for commercial applications or purposes or changed and resubmitted without the PI's permission and are subject to WPI's intellectual property policies.

4. Data storage and preservation of access.
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. The datasets along with the relevant metadata will be stored in an institutionall maintained data repository. The datasets will not need to be transformed or cleaned for preservation and archiving. Also, all the reports, presentations and research papers will be deposited along with the datasets. These datasets will be accessible for at least 3 years as mandated by NSF and WPI's IT division will provide the necessary procedures for preservation, backup and archiving.

5. Methods and policies for providing access and enabling sharing
Regarding evaluation data collected by the University of Massachusetts Donahue Institute (UMDI): Electronic numeric and qualitative text data generated from assessments and digital audio files of focus group proceedings will be kept on UMDI's secure server. Access to the files will be limited to the Research Manager associated with the project, the Director of UMDI's Research and Evaluation Group, and UMDI's Technology Coordinator. In addition, the Research Manager will maintain back-up copies of the data (on either CDs or a flash drive) in a secure location.

All other data will be stored on WPI-maintained computing systems, in directories wit access limited to project personnel. By default, all project personnel shall hav read/write/edit access to the data. We will let the individual participants determine whe it is appropriate to release scientific data. We will attempt to provide access to the dat to non-IGERT associated persons, after a reasonable time period has passed to allo the faculty to publish or file patents on their work.

For data gathered during the evaluation and assessment of the IGERT, all data will b made anonymous. Student names and IDs will not be released when results ar reported and/or published.

Disaster Recovery logistics are covered under the policies of WPI's Computing Center.

IV. Policies and provisions for re-use, re-distribution
Data gathered for this project may be reused in other, related, research project conducted by one of the co-PIs or graduate students. If other researchers reque access to the data, they must agree to our policy above. These datasets cannot b used for commercial applications or purposes or changed and resubmitted without th PI's permission and are subject to WPI's intellectual property policies.

Prof. Billiar and Prof. Gaudette are both on WPI's Research Data Management Subcommittee, Prof. Gaudette is also a member of the University of Massachusetts Medical School and WPI Data Management Curriculum and Data Repository Planning Project. Along with the PI and other co-PIs, they will work to ensure that these policies are followed by all IGERT participants.

Thank You


Gateway Park
Life Sciences
Bioengineering
WPI
LIFE SCIENCES
BIOENGINEERING