**Research Data Management Case: In the Right Light - A Mobile and Compact Optical Mammography Instrument Research**

A graduate student in a biomedical engineering lab is working under a PI on a NIH-funded project. The lab is conducting research on designing and building an optical mammography device. The graduate student seeks to create an instrument with broad applicability, allowing for its use at the hospital bedside or in doctors’ offices. Currently, optical mammography instruments, such as one developed in one of her professors’ biomedical engineering lab, operate by scanning an illumination optical fiber over a minimally compressed breast while a collinear detection fiber collects transmitted light through the breast. However, these devices are quite large, and they are composed of two six foot tall units, one serving as the scanning platform and the other housing the optical, electronic, and data processing components.

The graduate student seeks to improve upon this design by engineering and testing a compact, mobile optical mammography device. In order to test the device on human subjects she had to work with her university’s Institutional Review Board (IRB) to get the proper permissions to conduct research involving patients, and for obtaining the consent from these patients to participate in her study. In order to test her device using patients in a clinical setting, she had sought to collaborate with two physicians and nurses at the university’s hospital.

There is literature to support that methods of breast cancer diagnosis using ionizing x-ray radiation often leads to false positive results and unnecessary invasive needle biopsies. There are emerging research findings that support a potentially new method of breast cancer diagnosis, one that uses near-infrared spectroscopy (NIRS) to detect breast tumors. Using NIRS, diffuse optical imaging systems could identify breast tumors on the basis of the measured concentrations of oxygenated and deoxygenated hemoglobin (HbO2 and Hb), and water and lipid content in the breast tissue. NIRS uses the near-infrared region of the electromagnetic spectrum (from about 800 nm to 2500 nm). Each substance absorbs a different amount of near-infrared (NIR) light, and measuring and graphing this absorption creates a unique signature for that substance (% of transmittance, or the ratio of the light energy falling on a body to that transmitted through it, per nanometer of wavelength). Thus physicians could potentially identify a tumor if they could obtain and image the tumors’ NIRS signatures, based on measuring the tumor’s Hb, HbO2, H2O, and lipids NIR absorption.

The graduate student’s hypothesis is that optical mammography can develop into a stand-alone, compact, and portable imaging modality for detection of breast cancer and monitoring of individual response to neoadjuvant chemotherapy. Neoadjuvant therapy is a treatment given before primary therapy. For example, a woman may receive neoadjuvant chemotherapy for breast cancer to shrink a tumor that is inoperable in its current state, so it can be surgically removed later.

After the graduate student obtained her IRB permissions, she worked with the clinicians at the university’s hospital to identify participants for her study. All patient imaging takes place at the hospital. There are two clinical nurses recruiting and obtaining consent from patients for her study. This collaboration between her biomedical engineering lab and the clinical team (physicians and nurses) has brought important insights to this study.

The process begins when the physicians identify patients with abnormal ionizing x-ray mammogram images, possibly showing early stages of breast cancer. They then have their nurses contact these patients and invite their participation in the NIRS imaging study. The nurses explain to patients that they must consent to returning to the hospital and being scanned once more but this time with the NIRS optical mammography instrument. The patients who agree are then given background information and opportunities to ask questions about the research by the nurses. Upon arrival they sign a participation consent form and a data release form administered by the nurses.

The clinical team maintains hardcopy and digital files of the patients’ mammograms, the paper consent forms, the records of who is enrolled, and they have access to the patients’ electronic health records, which contain personal identifying information and protected health information (PHI). The federal Health Insurance Portability and Accountability Act (HIPAA) restricts the use and sharing of PHI. The graduate student and PI do not have access to their participants’ health records and identities, but they do have access to their mammogram images and their physicians’ diagnostic and descriptive information related to these images.

The study has enrolled 99 patients. The clinical team collects and stores the consent and release forms and assigns each patient a unique identification number (ID), thus the graduate student does not know the patients’ identities. The clinical team prepares a folder for each participant labeled with this id number that contains 2-4 scanned ionizing x-ray mammogram image .jpeg files, and the NIRS mammogram .jpeg file; these image files range from 500KB – 3.500KB. There are also several .doc and .xls(x), .doc(x), and .pdf files containing the physician’s descriptive and diagnostic information regarding the images. The clinical team organizes the files in these folders by labeling them with the participant ID and type of image (x-ray or NIRS).

In addition to building the NIRS optical instrument, the graduate student’s lab has custom coded optical mammography instrumentation software that collects data on light intensity measured through the breast tissue during the scan using MatLab™. This software records metadata related to the intensity of light at different wavelengths at different points of time during the breast scan (e.g., codes such as DC for continuous wave data and TRUN for truncated data). The lab’s instrumentation software, .mat code files, is proprietary intellectual property and will possibly be patented, and thus it cannot be shared; the .mat analysis code, which the student uses to analyze her data and which a person would need to verify her results, cannot be shared until the research is complete. She also collects NIRS and image data in LabView™ software and exports .txt text files into MatLab™ (.mat) for analysis. The patients’ image variables have been documented in a data dictionary (this is a .txt file created by a previous graduate student who had worked on the protocol).

The graduate student also records data in paper lab notebooks. These are kept on desks in a lab; they are not locked up; anyone in lab has access to them. The notebooks have a variety of data: pasted graphs and images of experiments run with the optical mammography instrument, and related notes and calculations.

In some instances the graduate student will get data from the clinical team via email, but all PHI and identifiers have been stripped. She keeps her data on a password-protected PC hard drive at the hospital, on her password-protected laptop, on USB drives, and on the university’s server. She has created a password-protected website for sharing her NIRS images and her analyzed data with her team members. She feels that when her research is complete she would like to make these data files openly accessible. However, the diagnostic instrument and technology, and any related intellectual property such as specs, documents, and software programs have proprietary and commercial potential, and they will not be shared.

Points to consider for writing a data management plan:

1. Types of data

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

b. What would be needed in a data management plan to describe use of proprietary equipment or software?

c. What analytical methods and mechanisms will be applied to the researcher’s data either prior to or post integration?

d. What type of outcome data will be generated?

2. Contextual details

a. What file formats and naming conventions will be used for the separate data sources and for the integrated file used for analysis?

b. What impact would the naming conventions have on later data access?

c. What other contextual details would the researcher specifically need to document to make her data meaningful to others?

d. In what form will the researcher capture these details?

3. Data Storage, Backup, Security

a. Where and on what media will the data from each data source be stored?

b. How, how often and where will the data from each source be backed up?

c. How long following the completion of her study will she need to store the data?

d. What impact would the proprietary software, and software updates have on storage and access?

4. Data protection/privacy

a. Who will own any copyright or intellectual property rights to the data from each source?

b. How will the datasets be licensed if rights exist?

c. How can the researcher make the participants’ mammograms openly available?

d. How is she addressing any ethical or privacy issues?

e. What mechanism is the team using to identify individual patients?

5. Policies for reuse of data

a. Will the data be restricted to be re-used only for certain purposes or by specific researchers?

b. Are there any reasons not to share or re-use data?

c. How will she create a de-identified copy of the data?

d. Will a new patient consent be required for subsequent re-use or publication of data?

6. Policies for access and sharing

a. Will some kind of contribution or fee be charged for subsequent access to this data?

b. What process should be followed to gain future access to the researcher’s study data?

7. Archiving and preservation

a. What is the long-term strategy for maintaining, curating and archiving the data?

b. What data will be included in an archive?

c. Where and how will it be archived?

d. What other contextual data or other related data will be included in the archive?

e. How long will the data be kept beyond the life of the project?

[Creative Commons License](http://creativecommons.org/licenses/by-nc-sa/3.0/us/)

The research case, *In the Right Light - A Mobile and Compact Optical Mammography Instrument Research: A Biomedical Engineering*, is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 United States License.

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
* the title of the work: New England Collaborative Data Management Curriculum
* the URL where the original work can be found: <http://library.umassmed.edu/necdmc>

Revised June 12, 2015