IMWEL Engineering Research Center for Integrated Mechanobiology for Women's Health

DATA MANAGEMENT PLAN

Significant barriers confront researchers interested in entering the field of women's health, and particularly reproductive health as accessible models of the female reproductive system do not exist. The limited availability of data requires investigators to have access to a major medical system and willing clinical collaborators. Researchers are often limited to retrospective imaging and mechanical data of already diseased or injured patients, making it challenging to define normal anatomy and biomechanical function. Moreover little information on the mechanical relevance of most animal models to humans exists, while the gold standard, non-human primate, presents ethical, infrastructural, and financial challenges.

IMWEL will overcome these distinct challenges through a collaborative computational, experimental, and clinical effort spanning institutes with large and diverse clinical volumes. IMWEL will create a free, open source, and online database of female reproductive, cardiovascular, and musculoskeletal anatomical data, tissue material properties, computational models, and corresponding clinical attributes, which conform to FAIR principles (Findability, Accessibility, Interoperability, and Reusability). IMWEL will generate foundational large sets of experimental and simulation data from studies of natural, model synthetic and model computational tissues including cells, cell-extracellular matrix, and cell-hydrogel constructs. IMWEL will also generate an unprecedented raw medical image database with corresponding de-identified clinical information (provenance) to spur engineering discovery by lowering the barrier of entry to the women's health field. IMWEL will disseminate these data through both traditional publication routes, where information is synthesized and analyzed, and through a dedicated project website, linking datasets and open-source computational models from publicly accessible digital repositories.

1. Products of Research

IMWEL will produce *convergent research* data in the form of image files (e.g., jpeg, tiff, DICOM), videos files (e.g., mp4), raw data sets (text or binary), theoretical models, medical images with corresponding clinical info (provenance), discretized segmented anatomy using open source frameworks, in silico biomechanical models (and all model by-products), data analysis, methods, and results. The investigators will compile metadata to ensure that IMWEL documents information pertaining to data format, contents, conditions of data generation, and software compatibility including descriptions of experimental conditions and information on electronic files such as author, file type, date file was created, and type of contents. IMWEL will charge all team members with managing and labeling their data in a systematic way for ease of use by the project team as needed. IMWEL will keep all meta-data electronically with the data files.

All data related to specimens from human tissue samples will respect protected health information (PHI) as the samples will be de-identified. The project leadership team from all four institutions will have primary responsibility for all data, and in particular for ensuring that the research records maintain no PHI, related to the technical aspects of this project at each site and overall. Investigators will discard physical samples of cells and tissues in Biohazard tissue disposal bags and send for incineration following any biomechanical testing, imaging studies, or other experiments.

In addition to the IMWEL convergent research data, as part of the project's *broader impacts* plan, IMWEL will collect separate data streams including curriculum materials and evaluation data. Evaluation data generated could include the following: (1) students' answers to survey questions, discussions and exit surveys; (2) student metrics (e.g., gender, ethnicity, major, GPA) provided by the student or university institutional offices (e.g., departments, Office of Institutional Research, Graduate Division); and (3) overall trend results from the analysis of these data. IMWEL will report and present only aggregate data and will

release no identifiable data. IMWEL will archive and make public these datasets through the same approaches as for the convergent research data.

2. Data Formats and Standards

IMWEL will standardized filenames for all electronic files and, as needed, will convert files to accessible, operating system-independent and open source formats (e.g., ASCII, text files, jpeg, tiff, mp4, pdf) for file sharing with team members or other researchers. All data will be backed up using an external hard drive, a server or an additional computer, utilizing each campus's own research data storage protocols and facilities. All four partner institutions will cross-post data from each site to enable four copies of all data.

3. Dissemination, Access and Sharing of Data

IMWEL will make available all de-identified datasets, computational models (and their by-products) and results of analyses on Academic Commons, a publicly accessible digital repository maintained by Columbia University and SimTK, a free federally-supported project-hosting platform for the biomedical computation community. Whenever possible, IMWEL will also publish datasets and results in supplementary materials or in a summarized format in peer-reviewed journal articles, theses, conference proceedings, book chapters and other print or electronic publishing formats, plus manuscripts archived in Academic Commons and/or PubMed Central. IMWEL will also make software and open-source analysis codes and computational models developed in the course of a project publicly available through SimTK.

4. Re-Use, Re-Distribution and Production of Derivatives:

Individual faculty members will determine the terms of re-use and re-distribution of data for general access or sharing. However, IMWEL will encourage, whenever possible, re-use of data, and without restrictions aside from appropriate attribution, utilizing CC-BY licensure. The project website and in other repositories under a CC-BY license for reuse will also make public the educational and outreach materials.

5. Archiving of Data

Lead institution Columbia University will ensure data archiving through Academic Commons, is a publicly accessible digital repository that provides open, persistent access to research produced at Columbia University. The Columbia University Libraries manages Academic Commons as part of the Libraries' long-term digital storage system, which ensures that files are replicated and stored in at least two distinct locations. The repository follows FAIR data principles and uses unique, persistent identifiers and rich metadata to enhance the accessibility of each work.

<u>Project website</u>
IMWEL will create and maintain a project website using WordPress as a content management system, which will connect users to all publicly available resources archived on Academic Commons and SimTK. The project website will host announcements including significant milestones of the project (publications, presentations at national and international meetings, student achievements) and broadcast availability of data and resources from IMWEL as they become available.