DATA MANAGEMENT PLAN

A. Data and Materials:

A.1) Types of Data

A.1.1) Laboratory original data:

The original laboratory data for platelet activation state (PAS) (StonyBrook and Arizona) and fluorescence polarization (FP) (Arizona) are generated as printouts or DAT files. SEM and DEP generated images (Arizona) are saved in JPEG and analyzed by NIH ImageJ. Videos for the platelet flipping and adhesion experiments (StonyBrook) are recorded in Nikon ND2 format, converted to AVI, segmented in JPEG, and analyzed in MATLAB. All laboratory data are recorded in Excel spreadsheets for analysis, and statistical analysis is performed in Excel, SPSS, and SigmaPlot.

A.1.2) Modeling data:

The data consist of platelet model parameters, the computer-generated modeling results, and the algorithm parameters. The platelet model parameters will be in two standard data types: LAMMPS data file (standards in B.1, size scale: ~200MB/file) and standard Protein Structure File (PSF, size scale: ~100MB/file). The former is the inputs for the simulator and the latter for fine-grained structure visualization. The flow system is such a massive particle assemble that is created only at runtime by calling *create_atoms* in LAMMPS. The generated data (positions, velocities and per-atom stresses) at runtime will be the snapshots of platelets and they will be stored every prescribed timesteps, following the data file standard (standards in B.1, size scale: ~500MB/snapshot), and they will be used for visualization and the images generation. The parameters for developed algorithms will be implemented and incorporated directly in the individual packages (details as in A.4).

A.2) Samples:

Platelet samples are discarded after each experiment, as they cannot be stored. Slides on which platelets are mounted and dehydrated for SEM will be stored for several weeks for re-imaging, if necessary.

A.3) Software Suite: The particle-based simulator software is based on modification of LAMMPS software suite and it will encapsulate our mathematical algorithms in the software packages:

(a) The no-slip boundary condition for DPD
(b) The modified Morse potential
(c) The novel multiple time-stepping (MTS) algorithm,
(d) Real-time evaluations of shear stresses on membrane;
(e) Dynamic formation of pseudopodia
(f) The inflow and outflow boundary conditions for complex geometries.

The details of the management of these software packages are as follows:

Packages (a), (b) and (f) will be managed as the *user-contributed package* (standards in B.1). Package (c) will be managed as update of the existing RESPA package. Package (d) is managed as a *standard package* and it will be built based on the existing VIRIAL stress function and VORONOI package. In (d), the data generation rate is several terabytes per second, a rate that require us to analyze the raw data for yielding cumulative shear stress-exposure time values and then discard them for saving disk space. Package (e) is managed as a re-parameterization function in C++ and it is used in the existing NVE integrator to adjust parameters. The filopodia formation formulas and patterns will be implemented and incorporated in the (e). Package (f) is managed as update of the existing *boundary* function, by replacing PBC style. These new packages and codes will implement our developed algorithms and also incorporate latest model parameters. They are designed as individual *patch* files to the copy of original LAMMPS codes and do an incremental update to stay up-to-date with the latest LAMMPS.

A.4) Curriculum and training materials: experimental protocols, manuals for using various codes, software tools and guidelines for adapting them for different uses, manuals for correlating experimental data with simulations data, misc.

A.5) Other: repository of published papers, conference abstracts, references libraries, description of HPC resources-capabilities, accessibility, specific computational tasks requirements, communication tools, operational manuals, etc.

A.6) Scale of Data: The approximate size of the data generated and selectively kept for storage will be (1) Simulations: 850~1000 TB (800~1000 TB simulation results, 50 TB model data and 10 TB source codes) and (2) Experimental data: 4.5-6 TB (2 TB images, 2~3 TB videos, 0.5 TB data files).This storage volume is available to us on SBU Google drive that will be linked to the Harvard Dataverse public data repository (details below)

B. Standards for Data and Metadata:

B.1) Standards for modeling generated data:

Standard for model parameters representation will use the SI base units.
Standard for the model data will follow the LAMMPS data file: http://lammps.sandia.gov/doc/2001/data_format.html.
Standard for the generated data will follow the dump data file standard: http://lammps.sandia.gov/doc/dump.html.
Standard for codes will follow user-contributed package rule: http://lammps.sandia.gov/doc/Section_packages.html.
Standard for the file format of archived data will follow the *tarball* standard: http://en.wikipedia.org/wiki/Tar_(computing).
Standard for the *patch* file: http://en.wikipedia.org/wiki/Patch_%28Unix%29

B.2) Standards for experiments generated data:

Non-industry standards are unavailable for the in vitro experimental data generated in this study. However, widely-accepted file formats (i.e. XLS, DAT, JPEG, AVI) will be used, with data reported using CGS or SI units, as appropriate.

C. Experimental and numerical data and software tools management and dissemination plan:

C.1) Experimental data:

Raw and analyzed images and platelet measurement data will be stored in the appropriate format and shared internally between the research groups in this proposal using a web-based platform (i.e. OneDrive, Google Drive, Dropbox, etc.), and with external groups seeking to validate or use the data through an online data repository: the Harvard Dataverse (http://thedata.harvard.edu/dvn/). The Harvard Dataverse is a public repository, hosted and maintained by Harvard University Information Technology (HUIT). The Harvard Dataverse facilitates data access by providing descriptive and variable/question-level search; topical browsing; data extraction and re-formatting; and on-line analysis. Harvard Dataverse has agreed to establish with us a partnership plan (see details below).

C.1.1) Re-use and re-distribution of experimental data:

The project will create documentation detailing the sources, coding, and editing of all experimental data, in sufficient detail to enable another researcher to replicate them from original sources; and descriptive metadata for each dataset including a title, author, description, descriptive keywords, and file descriptions. The Harvard Dataverse application's "templating" feature will be used for consistency of information across datasets. The Dataverse repository automatically generates persistent identifiers, and Universal Numeric Fingerprints (UNF) for datasets; extracts and indexes variable descriptions, missing-value codes and labels; creates variable-level summary statistics; and facilitates open distribution of metadata with a variety of standard formats (Data Cite, DDI v 2.5, Dublin Core, VO Resource, and ISA-Tab) and protocols (OAI-PMH, SWORD).

C.1.2) Plans for archiving experimental metadata:

The Harvard Dataverse commits to good archival practice, including independent geo-spatially distributed replication, a succession plan for holdings, and regular content migration. Should the archiving entity be unable to perform, transfer agreements with the Data-PASS partnership ensure the continued preservation of the data by partner institutions. All data under this dataset will also be made available for replication by any party under the CC-attribution license, using the LOCKSS protocols – which is fully supported by the Dataverse application.

C.2) Software tools

C.2.1) Re-use and re-distribution of software tools:

The aforementioned developed packages will be distributed as individual *patch files* to the copy of original LAMMPS codes. An incremental update can help re-use some or all of developed algorithms and stay up-to-date with latest LAMMPS codes.

C.2.2) Plans for archiving model data and software tools:

The model data and parameters will be stored together in the same LAMMPS data files and archived together for re-use. The algorithms and associated parameters will be combined in the individual *patch* files that can be archived. These model data files and software tools will be packaged as *source tarball* and archived in Subversion (SVN) (https://subversion.apache.org/) that is a popular archiving system that supports change log, checkout and update to get latest files and models. LAMMPS data files, algorithms, and software tools may also be shared in an online data depository, such as the Harvard Dataverse (http://thedata.harvard.edu/dvn/). Source codes may also be shared on Google Code (https://code.google.com/) or Sourceforge (http://sourceforge.net)

Intra-Project management platform To further facilitate effective real-time communication, a cloud based database was already established (Stony Brook University Google Drive- unlimited storage) which is, and will be accessible to all partners in the project. This database includes an administrative part (e.g., executive reports, progress reports, minutes of meetings, correspondence) and a scientific part (e.g., experimental protocols, research results, code developments, scientific publications, drafts and abstracts, presentation files).

D. Special funds allocation pertaining to the Data Management Plan

Software engineering efforts and data repositories: Starting in the second year a special budget is allocated for software engineering efforts and data repositories. Initially, a budget of $8,375 will be utilized in the 2nd year for maintaining a partnership collaboration between our group and the Harvard Dataverse public repository group at Harvard University Information Technology (HUIT) center for developing effective communication protocols between our storage data bases and their data repositories. The latter will serve as the point of access for both managing and cataloging our data bases and facilitate descriptive and variable/question-level search; topical browsing; data extraction and re-formatting; and on-line analysis. This will be followed by a yearly maintenance budget of $3,875 for: (1) a website to enable user downloads; (2) maintaining the software packages (3) provide basic level support to end users (4) facilitate effective communications between end users and our developers (supported by the project itself). The cost of preparing data and documentation will be borne by the project, and is already reflected in the personnel costs included in the current budget. The incremental cost of permanent archiving activities and maintaining the model repositories will be borne by the project in collaboration with Harvard Dataverse. (see also the budget justification).