Data Management

Products of Research

During the course of this project, it is anticipated that the following products will be developed:

1. Physical samples (sensor devices): These are the semiconductor sensor devices. It is important to note that the primary focus of this research is to study the behavior of the polymer film, and many of these films degrade upon exposure to oxygen or UV. Therefore, even with stringent storage conditions, the samples will degrade over time.
2. Physical samples (reagents): These are the polymers synthesized to fabricate the polymer thin films. They will be stored in an inert environment. Many of the monomers and related catalysts, which are subject to oxidative or UV degradation, will be discarded after a year of storage.
3. Material characterization data sets: The XPS, FTIR and ellipsometry data which verifies the attachment of the polymer thin films.
4. Sensing data sets: The data of the sensor detecting the polymer film behavior.
5. Labview code: LabView codes or the computer software which will be developed during the course of this program will be actively used on the testing computer. All codes are maintained on at least two computers in the lab: the testing set-up and a "ghost" computer.
6. Simulation code: (finite element method): The simulations of the optical device and the lipid bilayer transport dynamics. All simulations are performed in COMSOL Multiphysics.
7. MATLAB code: Programs will be developed to accelerate and automate the data analysis.
8. Microscope data: Images of the devices taken with both optical and electron microscopes.
9. Sensor system: The schematics of the sensor system (machine drawings and a bill of materials) will be developed.
10. Curriculum and outreach materials: A series of outreach materials will be developed, including several presentations on nanotechnology and an undergraduate lab manual. All outreach materials are posted on the group website.
11. Additionally, peer-reviewed publications and presentations will be given.

Format of Data

1. Detection Data: All detection data is saved as a .csv file. This format is fairly universal and can be opened in MATLab, Origin, Excel, and Igor.
2. Microscopy Data: Microscopy data will be stored as 16-bit TIFF images; a standard file format that most image processing programs handle natively.
3. MATLAB code: MATLAB programs are stored as .m files. Data generated by running these programs is saved in the standard .mat file format as well as plain text and graphical output (saved as JPEGs or TIFFs).
4. LabView Code: We develop our own LabView code for controlling and integrating instruments. After being developed, these codes are saved as executables (.exe) which can be operated on most computers and do not require the purchase of LabView.
5. FEM Simulation: We use COMSOL Multiphysics for our simulations. Our simulation files are saved accordingly.
6. Sensor System schematics: This will be saved as a single pdf file.

These data formats were chosen as they are the most generic file formats for each application and will enable easy sharing with other research groups.

Data Access and Policy

The results from this proposal will be published in peer-reviewed journals and will be presented at conferences. The papers and supporting information will all be accessible on the appropriate publisher's website in pdf or html format, and when permissible by the publisher, will also be accessible on the personal website of the PI.

The PI routinely shares the original files of her group's published data (simulations,

LabView codes, detection data) with and gives advice on instrumentation design to other research groups. Because COMSOL is rather expensive and not all groups have access to it, as an alternative to sending our code to a collaborator, we have hosted them in our lab.

Depending on the size of the file, we will either email it directly to the requestor or send a DVD with the data within a reasonable timeframe. Before publishing, any necessary IP forms are filed, to enable us to share our results with our colleagues without delay. The majority of the materials developed for outreach efforts, such as slides for high school presentations, is posted on the website and is freely available. Although each investigator controls the content of their group's website, the website server is maintained by the University. Therefore, there should not be problems with website maintenance.

Policies and provisions for re-use, re-distribution, and the production of derivatives

The PI currently provides several images (renderings as well as optical and electron micrographs) of devices on her group website for general public use. She also provides sample data sets and simulations which did not appear in publications but contain similar data. The primary requirement for image use is that our research group receives credit. We plan to continue this practice. We always provide detailed information about sensor sample fabrication and chemical synthesis methods in papers or in supplemental information, and we plan to continue this policy. We also routinely field questions about our methods.

USC's policy is to encourage, wherever appropriate, research data to be shared with the general public through internet access. This public access will be regulated by the university in order to protect privacy and confidentiality concerns, as well to respect any proprietary or intellectual property rights. Administrators will consult with the university's legal office to address any concerns on a case-by-case basis, if necessary. Terms of use will include requirements of attribution along with disclaimers of liability in connection with any use or distribution of the research data, which may be conditioned under some circumstances.

Archiving of data

Data sets and other associated research products will be made available immediately after the publication of a peer-reviewed journal article or a conference proceeding. Final peer-reviewed journal manuscripts, and supplemental information such as data tables for graphical information in manuscript figures and for statistically processed averages, which arise from NSF funds, will be posted in the USC Digital Repository and will be available on this publically available website no later than 12 months after publication. Authors will ensure their publishing agreement allows the paper to be posted to the archive; alternatively the USC website will provide a reference to the journal articles coupled with the supplemental information. These records will be durable, accessible through web protocols, and made safe from tampering. The storage media will be updated as necessary to keep it current.

The USC Digital Repository (USCDR) provides fee-based consulting and services to help USC researchers meet NSF requirements. Services include digitization, cataloging, preservation, archiving, and online access. USCDR is a center that is jointly operated by the USC Libraries, the USC Shoah Foundation Institute (SFI), and USC's Information Technology Services (ITS) division. As such, USCDR is able to offer researchers access to the professional expertise and technological infrastructure of the SFI, Libraries, and ITS.

After consultation with the appropriate NSF program officer to ascertain any exceptions, items will be discarded no sooner than 3 years after the conclusion of the grant or the public release, whichever is later. Research data that support patents will be retained for the entire term of the patent.

In addition to the electronic records and the written records contained in the lab notebooks, we also maintain hard copies of all data which is included in publications or which has significantly contributed to a peer-reviewed publication. This hard-copy volume is generated every year and is hard-bound, printed on archival paper, and is kept in Prof. Armani's office. The schematic of the sensor system will be included in this book.

Collaborative Research: SDCI Net: Policy-driven Large Scale Data Access Framework with Light-weight Performance Monitoring and Estimation

Formatted: Font: Times New Roman, 11 pt

Data Management Plan

Formatted: Font: 11 pt

1. Products of the Research: types of data, samples, physical collections, software, curriculum materials, and other materials to be produced

This project will produce a software framework that we will make available to the community freely under open source software distribution licenses. We will provide source code and may also, as appropriate, make available binary distributions for a set of typical hardware platforms to provide easier deployment for our application partners, such as Earth System Grid. In addition, we will provide documentation including Manual pages, user guides, administrator guides, etc.

During our development and evaluation of the software framework, we will measure how well our transfer estimates and transfer advice correspond to actual measured performance. We will promptly publish the results of our evaluation in peer-reviewed conference papers and journals that contain these measurements in graphs and tables. We will make the raw data for these graphs and tables available on our project web site and will maintain this web site for at least three years after the completion of our project. We will also describe, both in publications and on our project web site, the provenance of these data sets, including information on the experimental conditions under which these measurements were collected and the hardware and software versions that were used in producing the data sets.

Once deployed, the software will collect and generate additional data, including transfer performance measurements, estimates of future performance, advice on transfer parameters, etc. However, these data sets will be the intellectual property of the Virtual Organizations in which our software is deployed, which may prefer to keep these data sets private. We will not collect these data sets ourselves and will not make them available to the community.

2. Standards to be used for data and metadata format and content:

Our open source software will be written in standard programming languages and scripting languages, including C, C++, Java, Python and PERL. Software documentation will use standard tools such as JavaDoc to generate much documentation automatically. Other documentation, including Manual pages, user guides, administrator guides, etc., will be written in English and made available on project web pages and as PDF documents. Graphs and tables will be generated using standard tools such as Microsoft Excel spreadsheets, gnuplot and others. Our publications will be made available in PDF format. Presentations will be provided in Microsoft PowerPoint and PDF versions.

3. Access to data and data sharing practices and policies:

Our software, documentation and publications will be made available on a project web site that will be maintained for at least 3 years after the conclusion of the award. Our software will be released for free under open source licenses. It will be shared with any parties who agree to use the software under the terms of the open source license.

All participants in this proposal will conduct research and publish the results of their work. Papers will be published in a peer-reviewed conference paper, scientific journal or book that publishes in English, or as a peer-reviewed data report. We will make all publications available on our web site, except where prohibited by the copyright of the publisher. We will also provide access to raw data used in our

publications on the web site, and this data will be available for free to any scientists who want to use it for comparison or analysis.

Because of privacy and intellectual property concerns related to the Virtual Organizations that may deploy our software, we will not collect or make available data regarding how they use the software in their environments, such as transfer measurements, transfer estimates, advice on future transfers, or an evaluation of how effectively the software performs in their environments.

4. Policies and provisions for re-use, re-distribution, and the production of derivatives:

Software provided through the project web site may be used as specified under the open source license. Publications provided on the web site may be distributed freely in academic environments and may be cited with appropriate attribution according to standard academic practice. Raw data provided on the web site may be used for analysis and comparison of scientific results.

USC's policy is to encourage, wherever appropriate, research data to be shared with the general public through internet access. This public access will be regulated by the university in order to protect privacy and confidentiality concerns, as well to respect any proprietary or intellectual property rights. Administrators will consult with the university's legal office to address any concerns on a case-by-case basis, if necessary. Terms of use will include requirements of attribution along with disclaimers of liability in connection with any use or distribution of the research data, which may be conditioned under some circumstances.

5. Plans for archiving and for preservation of access:

The proposal team is committed to preserving the project web site for at least three years after the completion of funding using resources at our respective institutions.

We will also use the USC Digital Repository to archive and preserve our results. Final peer-reviewed journal manuscripts, and supplemental information such as data tables for graphical information in manuscript figures and for statistically processed averages, which arise from NSF funds, will be posted in the USC Digital Repository and will be available on this publically available website no later than 12 months after publication. Authors will ensure their publishing agreement allows the paper to be posted to the archive; alternatively the USC website will provide a reference to the journal articles coupled with the supplemental information. These records will be durable, accessible through web protocols, and made safe from tampering or falsification. The storage media will be updated as necessary to keep it current.

After consultation with the appropriate NSF program officer to ascertain any exceptions, items will be discarded no sooner than 3 years after the conclusion of the grant or the public release, whichever is later.

DATA MANAGEMENT PLAN

In the course of testing our traffic generator we will produce a number of models of legitimate traffic, mined from traffic traces. Where we used a public trace to mine this data we will make the link to the trace and the resulting models publicly available via the project Web page that we will host at USC/ISI. If we use a private traffic trace to mine traffic models we will examine the output for privacy leaks before making it publicly available. If no leaks are found, we will post the models at the project Web page along with a description of the traffic trace that they were mined from. The description will be generic enough to hide the specific source of data while aiding researchers in evaluating data utility (e.g., "packet size distribution from an enterprise trace").

The exact format of the models is to be determined in this research, but we expect that it will be a text file containing some required fields such as the fidelity dimensions being modeled, the distribution name for the model and the parameters of that distribution. The models will be released under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. This license allows reuse only for non-commercial purposes and it allows modification of the work if the results are shared.

All the source code developed under this proposal will be released via the project Web page under the GNU General Public License version 3 that allows free use and modification of the code, as long as the result is released under the GNU GPL v 3 License. All data and software placed on our project Web page will be preserved for at least 10 years, and will remain accessible to public during this period. After 10 years we will evaluate if the data/code are still popular and if not, we will archive them on a DVD and remove them from the server. The DVD will remain available at ISI for the foreseeable future and a copy of it will be sent to interested researchers that request it.