

## **Data Management Plan**

### **I. Types of data**

The proposed data management plan for this project satisfies our need to manage large amounts of data, estimated at approximately 50 terabytes (TB). The research data for the proposed project will consist of experimental results (PIV flow measurements, high-speed holographic and standard videos, and time-resolved brightfield images), software, digital data, and digital results. We will use the University of Texas at Austin's 'UTbox' site and for sharing the data among the team members. With these systems each user receives a secure login and password for accessing the project site.

### **II. Data and Metadata Standards**

Substantive metadata will be provided in compliance with the most relevant standard for biological science and will include culture conditions and species along with the digital imagery. New metadata standards have been developed in prior collaborations for data such as 3D holographic recordings where raw images are provided as well as links to reconstruction software. We will continue to be develop standards for modeling results and PIV images.

The software produced in this research will be written using a standard language such as C++ or Python, and will make use of either a commercial software APIs (e.g., Apple's iOS developer tools) or, preferably, open-source APIs.

### **III. Policies for access and sharing, and provisions for appropriate protection and privacy**

At the end of the project, we will make this data available on the project website. We will allow re-use, but not re- distribution, of the software and experiment data created through this project. Any re-use should carry an appropriate acknowledgement or attribution of the investigators, PI institution, and NSF.

The data we collect will not be copyrighted and will have no licenses pertaining to it. There will be no charge for accessing this data. However, to limit the load on our server, we may place data rate or time of access restrictions. We retain the right to use the data before opening it up to wider use but once we publish a paper we will release its corresponding data.

All the students will be required to submit their theses electronically. We will use the worldwide Theses and Dissertations (ETDs) initiative. ETDs are available by browsing authors and departments in the *ETD database* and from the Networked Digital Library of Theses and Dissertations (NDLTD). The NDLTD encourages and supports the efforts of institutes of higher education and their communities to develop electronic publishing and digital libraries (including repositories), thus enabling them to share knowledge more effectively in order to unlock the potential benefits worldwide.

(<http://www.ndltd.org/serviceproviders/scirus-etd-search>).

### **IV. Policies and provisions for re-use, re-distribution**

Responsibility – Data Manager, Brad Gemmell (UTMSI), has overall responsibility for data management over the course of the research project and will monitor compliance with the plan in cooperation with each PI. The data manager working with IT technicians at UTMSI oversees the deposit, collection, quality control and data submission, and dissemination and access control of these data. The data to depository is selected by coPIs.

There will be no permission restriction placed on the data and the code and design files. Other researchers working on privacy in pervasive computing systems are the most likely consumers of the released data. The intended or foreseeable users of the data, code and design files would be those seeking to improve privacy mechanisms. And there are no reasons not to share or re-use data.

### **V. Plans for archiving and Preservation of access**

UTMSI will be the primary site for data storage and archiving. To host large quantities of metadata and results, The Buskey lab has a 96 TB Dell blade sever data storage system maintained by the UTMSI IT department and backed up on central servers at the main campus in Austin, TX. Both databases are cloud-based, enabling data sharing with collaborators and, eventually, anyone with web access.

Subsets of data will also be stored via long-term archival servers using the Biological and Chemical Oceanography Data Management Office (BCO-DMO). The long-term strategy for maintaining, curating and archiving the data is via regular backup of data sets into suitable long term storage media such as optical or magnetic media stored in a secure location or cloud storage, separate from the database server. There are no transformations necessary to prepare the data for preservation and data sharing. As far as metadata goes, documentation will be preserved alongside the data in order to make the data is interpretable by other users.