

## Data Management Plan

### 1. Expected Data

#### *1.1 Types of Data*

This project is expected to produce three types of data:

- a. Still photographic images (.jpg or similar machine-readable format).
- b. Video images (.mpg or similar machine-viewable format).
- c. Algorithms for comparison of images.

Some or all of the algorithms will be based on prior work by one of the investigators (Roy Chowdhury) on recognition of human faces. Some or all of the algorithms may have already been published or otherwise disseminated.

#### *1.2 Management of Unprocessed or Work-in-Progress Data*

All principal investigators in the Bourns College of Engineering (BCOE) are provided with sufficient computer storage to store and process the data generated in their research. The College of Humanities, Arts and Social Sciences (CHASS) also has central computing and data storage for all departments in the college. Every principal investigator has the ability to create his or her personal web site, and to make data sets available through that web site. Additionally, BCOE has established an archive based on the open-source EPrints system ([www.eprints.org](http://www.eprints.org)) to be a permanent, accessible repository for articles and processed data sets to facilitate data sharing.

Complementing this infrastructure, the UCR Libraries and the California Digital Library/UC Curation Center are available to principal investigators to consult on the formatting of data from project inception through dissemination of finished products. Further, the California Digital Library has established Merritt as a repository for published articles and data sets that can be searched along with other resources of the University of California system, including the highly popular Melvyl catalog. As Merritt develops, the BCOE intends to make data sets available through both the locally housed EPrints archive and Merritt, which will provide two channels for access to the results of our research.

#### *1.3 Limitations and Restrictions*

This Data Management Plan addresses the policy that primary data “commonly accepted in the scientific community as necessary to validate research findings” should be made available to the research community. In accordance with this policy and guidance from the Office of Management and Budget, this plan does not include preliminary analyses (including raw data), drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. Data that must be withheld long enough to enable peer review and publication/dissemination or protection of intellectual property is subject to this plan only after those steps have taken place.

The release of data about students or any other human subjects is subject to policies and restrictions in protocols adopted by the relevant Institutional Research Board (IRB) and the Family Educational Rights and Privacy Act (FERPA) regulations.

It is common for museums to restrict the publication of images of works in their collections. The investigators intend to secure permission (free or paid) to reproduce images for use in published articles, lectures, presentations at professional societies, and classroom instruction. However, depending on the

image, the owner, and the agreement governing each work of art, the rights to use these images might not extend to other researchers or to follow-on or derivative research by others. In light of this, materials will be disseminated only to the extent that the owners of images allow, and reproduction or reuse may be restricted.

#### *1.4 Data Sharing/Dissemination*

All applicable electronic materials produced in BCOE will be made available through the BCOE EPrints server and, subsequently, through Merritt. Data will be made available only after appropriate steps have been taken to protect intellectual property. Confidential material will be handled according to policies and protocols for human subjects, FERPA, and any other applicable regulations and restrictions.

The EPrints server will be read-only; therefore, data will be secure from tampering. Files will be stored on a Net App RAID file server and will be backed up twice per week using an unencrypted LTO5 tape drive system.

### **2. Period of Data Retention**

The BCOE EPrints archive is intended to be a permanent, read-only repository. There will be no schedule for deleting the data from the archive.

If any PI leaves UCR, the products in the BCOE EPrints database will remain stored there. PI is free to reproduce and repost these materials at his/her new institution.

### **3. Data Formats**

It is anticipated that data will be in the form of images (.jpg or other common machine-readable formats) and video images (.mpg or other common machine-viewable formats). Any algorithms will be developed using commercial off-the-shelf software and thus will be usable by the broad research community.