

Data management plan**Summary**

This project represents a collaboration of researchers in Art History and Electrical Engineering. Both groups will generate research results that are subject to the Data Management requirements of the National Endowment for the Humanities.

Scope

“Data” is defined as materials generated or collected during the course of conducting research. In the context of this project, data will consist of software/algorithms, digital data sets, and reports. Research will be conducted on a variety of images from a variety of sources. In some cases, we will have the rights to perform research on an image, but not to republish that image.

In accordance with guidance from the Office of Management and Budget, this plan does not cover 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.

Expected 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.

Data Retention

We expect to retain all data indefinitely in our own files. Processed data necessary to support our findings and advance the state of the fields of study will be retained and disseminated as discussed below.

Data Storage and Preservation of Access

The Bourns College of Engineering (BCOE) at the University of California, Riverside, in partnership with the UCR Libraries and the California Digital Library, has established a new, custom data management system designed to make results of our research readily and reliably accessible. This system combines two services of the California Digital Library for the first time: an eScholarship series, a curated site where all BCOE investigators can store and disseminate their data, and EZID permanent identifiers for long-term identification and management of data resources.

All principal investigators in BCOE are responsible for complying with sponsor policies regarding data management and accessibility. All have sufficient computing resources to store and process the data generated in their research. 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, the eScholarship series (<http://escholarship.org/uc/search?entity=bcoe>) provides a ready-made resource for dissemination.

Researchers can submit materials to the eScholarship series site, where it is automatically given an EZID persistent identifier. All material on the site is indexed on Google, Google Scholar, and Melvyl/WorldCat. The California Digital Library also provides data backup services to avoid accidental loss of data.

This partnership, which has resulted in the development of a new service that CDL can provide to other research organizations, makes our data widely available and searchable under the taxonomies, identifiers, and standards of universal data sharing systems envisioned for the near future (see *Changing the Conduct of Science in the Information Age*, NSF, June 2011).

Materials will be searchable through a number of databases and search engines, including Google and Google Scholar.

Stored materials are downloadable but not editable, and they are backed up to assure no loss of data. All uploads are approved by a curator before they are published.

Data Formats and Dissemination

We are not aware of any standards or conventions for dissemination of data in this field. The formatting of data will be at the discretion of the principal investigators. The eScholarship site can accommodate portable document format (PDF) files as well as spreadsheets, database files, images, and sound/video clips.