

Data Management Plan

Data generated

The data generated by this project fall into three major types: the white paper, nine interactive Polynomial Texture Map (PTM) files, and the complete archive of all captured and processed images.

The white paper will be disseminated as Portable Document Format (PDF) with embedded text and images. It will be produced by Hanneken. It will be freely disseminated by web server without restriction. The USC data center is the most robust of the redundant web servers to be used, which will also include St. Mary's and NEH.

The nine PTM files represent the three test objects using three methods (two experimental, one baseline from existing technology). Final assembly and metadata will be performed by Lundberg and Hunt. They will be freely disseminated by web server and the InscriptiFact Digital Image Library.

The complete archive of all captured and processed images will be openly available, for example to scientists seeking improved processing of our captured data. The permanent archive will be stored on magnetic disk in the libraries of St. Mary's and USC. Optical media and the Internet may be used for transmission, but not archiving.

Data management

Intellectual Property: The three test subjects chosen from the collections of affiliates will be unencumbered by restrictions on dissemination. All procedures used and produced will be public. All metadata and file formats will follow open standards. All products will be usable with free software.

Backup and Redundancy During the Grant Period: Hanneken will manage backup and redundancy across media and geographic location. He will maintain the master data collection on a RAID array. Backups will be stored on removable media and remotely. Checksums will be used to verify data integrity.

Webservers: The white paper and PTM files will be available by web server, indexed by search engines, and unrestricted. Mirrors will be available at St. Mary's and NEH servers, but the anchor for reliability and permanence is the USC data center. The white paper and PTM files will be available by web server by the end of the project and, consistent with USC's policies, will be maintained in perpetuity.

File formats: The following digital file formats will be used for the project data:

- TIFF (uncompressed) for captured and processed images
- DNG for registered images (Registered images are like a stack of flat images in an image cube, such that the third dimension represents the same pixel on the object captured or processed in a different way.)
- PDF for documentation (white paper)
- PTM (Polynomial Texture Map) for RTI images
- HTML Some documentation or guides may be stored in HTML to facilitate navigation through a web browser.
- JPG Compressed thumbnails will be used only for navigation to the uncompressed TIFF files.

All formats are open standards viewable and convertible using a variety of free software. The only specialized format is PTM, for which viewers are freely available from InscriptiFact (inscriptifact.com) and HP Labs (www.hpl.hp.com/research/ptm/downloads/download.html).

Metadata standards: Metadata associated with images is recorded in the headers of individual image files and collectively within companion files and database records. Entry of metadata occurs during several stages of the project:

- metadata associated with imaging conditions is recorded before and during image capture, through manual user entry and via automated routines in PhotoShoot software;
- further metadata is recorded during routine processing of acquired images;
- metadata describing image content and specialized processing routines is automatically generated and recorded during processing.

The Qualified Dublin Core standard will be used for meta-data (dublincore.org/documents/2000/07/11/dcmes-qualifiers/).

If the chosen test objects contain text it will be transcribed according to the standards of the Text Encoding Initiative (TEI).

Archival standards: Hanneken will compile the final archive of all project data. Hunt will ensure its suitability for archiving. We expect to use SATA magnetic disks (solid state may become a viable alternative) with USB interfaces, formatted with the FAT 32 file system. Exact duplicates will be kept and made available at the USC and St. Mary's libraries. This agreement is not conditional on continued employment.