

IMAGE DIGITIZATION OF EARTH ANALOG STUDIES FOR PLANETARY DATA SYSTEM ARCHIVING. D. M. Nelson, A. J. Olson, and D. A. Williams. School of Earth and Space Exploration, Arizona State University, Box 871404, Tempe, AZ 85287 (davidmnelson@asu.edu).

Introduction: The NASA Regional Planetary Information Facilities (RPIFs) are a network of nine US and seven international centers that archive images, maps, literature, and education outreach materials from past and current planetary missions. While each RPIF houses a variety of similar products, each one possesses unique datasets that were developed over time.

The Ronald Greeley Center for Planetary Studies (RGCPS), the RPIF located at Arizona State University (ASU), in Tempe, AZ, hosts a collection of images from Earth analog sites, photographs and films from aeolian and lab experiments, and historic NASA documents, that are currently only available as hardcopy photographs and paper documents. We are digitizing a portion of this unique archive to be added to the Planetary Data System (PDS) for public access.

The RGCPS: The RPIFs were originally founded in the 1970s to serve two primary functions: 1) to archive photographs and literature from active and completed planetary missions, and 2) to provide researchers and the public access to the archive for scientific research, future mission planning, and education/public outreach (EP/O) [1]. During its 40 years of operation, the Center has accumulated over a quarter-million products in its archive, of which over 30,000 are unique (Table 1).

Archive Digitization: In 2014, we were awarded funding to digitize part of our unique collection through the Planetary Data Archiving, Restoration, and Tools (PDART) program. The goal is to distribute these planetary resources so that they can be better utilized by the whole planetary community for future research. To facilitate this distribution, we are making the datasets compliant to NASA's PDS. For this project, we have outlined four tasks:

Task 1) Digitizing the Aerial Photographic Archive. Scan 9" x 9" black and white (b/w) aerial photographs at 600 dots/inch (dpi), and collect metadata from image frame. These 8,249 images were collected by Dr. Ronald Greeley as a result of NASA-funded research in the 1960s-1980s, and are primarily of volcanic features and span the globe.

Task 2) Digitizing the Field Photographs. Scan hand-held camera photographs, b/w and color, at 600 dpi for images 5" x 7" or larger and 1200 dpi for smaller. These 6,557 images were acquired by Dr. Greeley and his colleagues, taken on the ground and from a fixed-wing aircraft, during field studies of aeolian and volcanic features. Sites are primarily in the US, but are also from Iceland, Bolivia, Peru, and Italy.

Task 3) Digitize photographs and films of laboratory experiments. These archives consist of 2,117 still photographs and 459 16-mm movies of experiments conducted at the NASA Ames Vertical Gun Facility (AVGF) and the NASA Planetary Aeolian Laboratory (PAL, which includes wind tunnels at ASU and Ames). This archive, like those of Task 2, contain b/w and color photographs and will be scanned at 600 or 1200 dpi depending on size. The films, which are high-speed movies of PAL experiments, will be scanned by an outside company at high resolution (1920x1080 pixels at 24 frames/second).

Task 4) Scan NASA special publications. There are seven publications in which Dr. Greeley was a lead or co-author that were produced from research focused on terrestrial geological features that serve as analogs to observed planetary features. Each page of the documents will be digitized at 400 dpi to produce PDF versions. The document "Aeolian Features of Southern California" is currently available on our site at ASU (<https://rpif.asu.edu/index.php/downloadables/>).

Digitization Workflow: The workflow for the digitization of the archive is currently conducted by our staff archive librarian, A. Olson.

Each photograph set undergoes an initial assessment to review the physical condition and order of photos. Damaged photos, or those deemed irrelevant to the study, might be scanned but not included to the PDS archive. A naming convention, based on location and feature, is established for the scanning procedure, but remain flexible until all scanning is completed.

The photos are lightly cleaned then scanned on an EPSON flatbed device, without image correction, and are saved as TIFF files. The complete dataset for a single project is scanned before images are straightened and/or cropped using Adobe Photoshop [2].

File Naming Convention: Each image set has a variable for either field or aerial type photo, and one for location (others could be identified later). Thus, file names begin with either the letter "f" for field or "a" for aerial. The location codes are compiled from the Library of Congress's MACHINE Readable Cataloging (MARC) Code List for Countries; this metadata standard encompasses both states and countries [3].

An RGCPS-created geographical code is assigned to each set and incorporated into the naming convention. These 2- to 3-letter codes are unique according to the geographical purpose/subject of each image set.

The file name ends with a 3-number suffix, assigned from the order of images in our physical collec-

tion. These numbers begin at 001 for each set and end with the corresponding last image number in the set.

For example, the name of the first field image in the Amboy Crater, California collection is:

fcau_ac_001.tif, where

f=field; cau=CA, US; ac=Amboy Crater;

001=number in series; .tif= file format extension

Metadata Compilation: Information for the images is sparse and we included feature interpretation, physical and digital maps, field notes, print materials on geomorphology, and publications by Dr. Greeley and his colleagues (where available) as part of the metadata. This info is included in a Microsoft Excel document as part of the digital archive. Each collection will have its own metadata spreadsheet.

Field types include: Product ID, Date Taken, Location, Description, Named Feature, Feature Type, Purpose Photo Taken, Scale, Source/Photographer, Instrument, and Related Publications.

PDS Archive Procedure: Once the data have been collected and reviewed, each project is compiled into a directory structure and supplemented with supporting XML documents, as described by the conventions outlined in the PDS4 Data Provider's Handbook [7]. Each Task project (with the exception of the Task 4 gray literature documents), is compiled as a single Bundle, e.g.: aerial photos; field photos; lab photos and films. This Bundle includes all image files, Excel spreadsheets, and supporting documents. Within each Bundle, individual projects form a Collection (e.g., Field photos of Amboy Crater, CA). Collections contain the individual Products, which in this case are image files.

Progress and Continuing Work: The implementation of this digitization project required the development of a cataloging procedure that would standardize the sometimes-sparse document information, while providing users with enough information to identify specific image archive products.

To develop an initial methodology that would meet these requirements, it was necessary to begin recording metadata from the photograph sets without a specific procedure. This allowed us to identify a variety of problems, from missing or incomplete information to out-of-order collections, which made it difficult to establish a naming convention and directory structure.

Approximately two months was needed for this trial-and-error approach to identify a standard that could remain consistent throughout the duration of the project. Metadata and cataloging standards, set by the Society of American Archivists and the Library of Congress, provided insight on how to create the final naming convention and continue to guide the ongoing process of dataset archiving.

To date, one large photograph Collection is currently under review by PDS: the pilot study of field work done in Amboy Crater, California. This Collection consists of 266 images and corresponding metadata and other supporting documents.

In addition, several other projects are in progress: Task 1) Aerial photographs: approximately 260 images completed; Task 2) Field photographs from Alaska, Australia, and Bolivia (several hundred images) are being completed with the finalized cataloging procedure; Task 3) Lab photographs: 234 images have been scanned. Lab movies: 10 have been digitized, 100 are in process; Task 4) NASA gray literature documents: all have been scanned and are currently under review. While the initial development of the procedure for this project took longer than anticipated, we are currently making considerable progress towards our goal. We plan to submit more collections of digitized documents to PDS this year.

References: [1] <http://www.lpi.usra.edu/library/RPIF/>
 [2] Adobe: <http://www.adobe.com>
 [3] <http://www.loc.gov/marc/>
 [4] https://pds.jpl.nasa.gov/pds4/doc/dph/current/PDS4_DataProvidersHandbook_1.9.0.pdf (2017)

Data Product Type	Quantity	Comments
Photographs of Field Work	6,557	Combination ground & aerial images, sizes $\leq 8 \times 10''$
Photographs of Laboratory Work	2,117	PAL and Vertical Gun, sizes $\leq 8 \times 10''$
Aerial Photographs	8,249	Standard: $9.5 \times 9.5''$; also sizes: $8 \times 10''$, $5 \times 7''$, $4 \times 5''$
Site Studies	2,224	Radar, aerial. Overlays, negatives, maps (oversized)
35 mm slides	11,038	Field and experiment
Film	459	PAL wind tunnel experiments
VHS tapes	120	Mission animations, NASA mission press conferences
Publications and reports	30	Not listed on WorldCat, avg. 110 pages/pub.
Total Products	30,794	

Table 1. List of the unique photograph and film archives of the ASU RPIF.