

7. Data management plan

EXPECTED DATA

The data produced for *Digital Historic Skies* will consist of: a) high-resolution digital images of historic celestial charts, and associated metadata; b) software code and documentation for a prototype user interface to interact with these charts; c) user-generated annotations collected through the use of this interface; and d) software code for an alpha prototype of a mobile application.

PERIOD OF DATA RETENTION

Digital images and metadata will be retained for long-term storage in the Adler's digital asset management system.¹ Images and metadata will be accessible through the online public access catalog as they become available.

The software code for user interface, user-generated annotations, and mobile application prototype code will be retained for long-term storage in the Adler's server system, backed up on site at the Adler. Nightly data snapshots are made and stored on the Amazon S3 storage platform. User-generated annotations will be retained for eighteen months for Adler project development, and then released. Software code for both the user interface and mobile application prototype will be made accessible.

DATA FORMATS AND DISSEMINATION

Digital images and metadata. Master images will be created as uncompressed, lossless, color TIFF images. At minimum, technical parameters will be:

- Original 8x10" or smaller: 4000 pixels on long side, excluding mount and borders, between ~400 and ~800 ppi. Bit depth 48-bit RGB. Dimensions at least 100% of original.
- Original 8x10 to 11x17": 6000 pixels on long side, excluding mount and borders, between ~450 and ~600 ppi. Bit depth 48-bit RGB. Dimensions at least 100% of original.
- Original 11x17" or larger: 8000 pixels on long side, excluding mount and borders, ~600 ppi. Bit depth 48-bit RGB. Dimensions as close to 100% of original as possible.

Web- and mobile-ready appropriate derivatives of master files will be created to the specifications of the citizen science project and of the mobile application, to be determined in the course of the project.²

All images will include descriptive, administrative, and technical metadata, utilizing Simple Dublin Core, controlled vocabularies from Library of Congress and Chenhall nomenclature, and the Preservation Metadata Implementation Strategies (PREMIS) data model. Images captured will either have no rights reserved under the attributes of Creative Commons Public Domain (CC0) or be licensed as Creative Commons Attribution-Sharealike (CC BY-SA). Images will be available to the public, both before and after the project is completed, via the Adler's online public access catalog to launch in September 2014. All image metadata will be released as CC0.

¹ The Minisipis Trusted Digital Repository (TDR) is currently in development. The Adler will implement the trusted digital repository or another digital asset management system in winter 2013.

² The Adler's digitization workflow is developed from the Federal Agencies Digitization Guidelines Initiative's *Digitization Activities: Project Planning and Management Outline*. Specifications use the National Archives and Records Administration's *Technical Guidelines for Digitizing Archival Materials for Electronic Access* and the Federal Agencies Digitization Guidelines Initiative's *Technical Guidelines for Digitizing Cultural Heritage Materials: Creation of Raster Image Master Files* as guidelines for digitization standards.

Software code for user interface. Software will be written in JavaScript, Ruby, HTML, and CSS. As with all software developed by the Zooniverse web team, code written will adhere to best practice web standards. W3C compliance will be observed and given the potential broad public audience web tools will be authored with web-accessibility in mind although given the relatively complex nature of the user-tasks the tools are unlikely to be appropriate for all web audiences. As a standard policy, the Zooniverse team publishes their code as Open Source projects on GitHub (<https://github.com/zooniverse/>). Software products and the associated documentation will be released with a liberal open source license on GitHub (Apache 2.0 or MIT).

User-generated annotations. User annotations collected during this program will be stored in the shared infrastructure of the Zooniverse web platform (zooniverse.org) as JSON documents in a MongoDB database. The Zooniverse platform currently supports more than twenty different crowdsourcing projects and over the last five years has collected more than 400 million user annotations. The programmatic interface to this system is a web-based JSON/REST API built upon the Amazon Web Services cloud compute platform.

Mobile application prototype code. Software will be written in Objective C. Adler created code for the mobile prototype code will be published as Open Source projects on GitHub, except where such code uses proprietary elements of for-profit partners (such as existing code and connections to existing mobile applications). Software products and the associated documentation will be released with a liberal open source license on GitHub (Apache 2.0 or MIT).

DATA STORAGE AND PRESERVATION OF ACCESS

Digital images and metadata. The Adler is dedicated to maintaining the digital elements of the collection and includes costs related to its preservation and longevity in its annual budget. The information technology infrastructure in place provides reliable and redundant long term, stable storage for digital assets. Digital assets related to the Adler's collection are stored on a dedicated four-terabyte RAID volume, with the ability for significant expansion. Access to this volume is controlled via user login. Currently, one backup copy is stored on an external RAID volume, and two additional backup copies are stored offsite. After the digitization phase of the project is complete, a second media type will be introduced. Backups of the drives will occur on a regular schedule determined by the Adler's information technology contractors to avoid catastrophic data loss.

The Adler recently hired a full-time digital collections manager, who is responsible for collections-related digital assets, including their long-term preservation. The digital collections manager will codify data standards, workflow, and specific processes related to digitization as well as technical documentation in order to maintain its integrity over time and aims to reach and maintain at least Level 3 of the National Digital Stewardship Alliance's Levels of Digital Preservation for digital collections assets. This includes quality control, instructions for ingestion into the digital asset management system, and fixity checks and checksums to confirm that master files are not altered or corrupted. Training materials and processes will help staff identify corruption as well as anticipate hardware, software, or file format obsolescence and have a plan in place for migration or emulation once those steps become necessary.

Software code for user interface, user-generated annotations, and mobile application prototype code. Zooniverse code, data products, user-generated annotations, and mobile application prototype code will be hosted on the GitHub platform, backed up on site at the Adler and nightly data snapshots are made and stored on the Amazon S3 storage platform. The Director of Citizen Science at the Adler will be responsible for ensuring that these processes are followed for this project.