

STANDARDS FOR TRACEABILITY AND NON-DESTRUCTIVE CONSTRUCTION IN PLANETARY SCIENCE DATA SETS: AN EXAMPLE FROM THE CTX GLOBAL MOSAIC. J. L. Dickson¹ and B. L. Ehlmann^{1,2}, ¹Division of Geological and Planetary Science, California Institute of Technology, 1200 E California Blvd, MC 170-25. Pasadena, CA, 91125. (jdickson@caltech.edu), ²Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109.

Introduction: Recent attention has been given to defining *foundational* data sets within a Planetary Spatial Data Infrastructure (PSDI) [1]. These standards pertain to the knowledge of geodetic accuracy of a product (be it coordinate systems, elevation or orthoimagery [2]), such that foundational data are of engineering utility as a reference product for other data sets. Less work has been done to develop standards that maximize the scientific utility and traceability of planetary data sets. Presently, a product can be termed *foundational* without its own foundation being communicated to the end user. There are not currently requirements for (1) independent assessment (peer-review), (2) publication of detailed methods and code used to create the product, or (3) ready traceability between the delivered product and the raw data used in its creation. We believe that *foundational* science data should include its traceability. The value of these products will be amplified if they include sufficiently rigorous documentation for scientific analysis.

In this contribution we augment the existing engineering-focused definition of *foundational*, or develop a separate definition apart from *foundational*, with scientific standards of traceability: transparency and reproducibility. We broadly discuss the range of traceability achievable as standards must accommodate a diversity of data types and processing techniques. As Mars provides a range of global data sufficient for demonstration, we use Martian data products for illustrative purposes and examples from our considerations during recent construction of a global CTX mosaic [3].

Range of Traceability: The level of traceability that a *foundational* product can attain at the time of dissemination can vary as a function of (1) the type of data (orthoimagery vs. topography) and (2) the methods used in its construction (destructive vs. non-destructive – is information preserved throughout processing?). Pixel-for-pixel vectorized mapping of source data (Fig. 1), which we discuss below (*Stage 4*), may not be possible for some data products. Thus, one blanket requirement for maximum traceability is not appropriate and we suggest that traceability be a ladder of successive stages that data providers attempt to climb and communicate that level of traceability to the end user. This affords some discretion to the data provider to assess how users will interface with their product and provide a commensurate level of traceability. For instance, the global

THEMIS IR daytime mosaic [4] provides a quantitatively controlled reference framework for other data sets, but end users are unlikely to need to trace pixels back to raw THEMIS IR data, thus pixel-for-pixel mapping of the product is not necessary. For the global CTX mosaic [3], which is uncontrolled but provides the highest resolution imagery of 95% of Mars, we predicted that users will want to access pre-mosaicked versions of the data, so we prioritized seam mapping in its construction. Our beta testing program with > 30 Mars scientists revealed that this was of value to the science community, and was essential for efficient quality control during its creation.

Stages of traceability: Here we propose a ladder of traceability, from presentation of metadata to pixel-for-pixel mapping of products (Table 1). Stages 0-2 are largely subjective (the creator has latitude in terms of how much information they provide) while stages 3-4 are objective. Situations may arise where more direct, objective stages of traceability are achieved before less advanced stages (i.e., code and seam maps are disseminated before a peer-reviewed manuscript is published).

Stage 0. Basic bookkeeping of metadata associated with the final product (resolution, projection, list of source instruments, list of images/DEMs included, date of creation, etc.), but no information about the process of constructing the product.

Stage 1. Unvetted (not peer-reviewed) narrative of the data processing pipeline, software packages utilized, and known deficiencies/limitations of the product.

Stage 2. Vetted (peer-reviewed) manuscript that covers in detail the processes used in the construction of the data and its limitations/uncertainties.

Stage 3. Scripts used in the construction of the product with instructions for proper execution.

Stage 4. Attribute-laden vectorized feature maps of input data that trace every pixel of the rendered product to EDR and RDR versions of the original data (Fig. 1), and contain all original metadata for each pixel.

Discussion: We propose that all *foundational* data (coordinate systems, topography, orthoimagery [2]) be traceable through, at minimum, *Stage 2* (Table 1), such that the product is accompanied by a peer-reviewed publication. This publication should provide the user with a clear articulation (interpretable to scientists as well as technicians) of what has been done to the data and what its limitations and uncertainties are.

Foundational data will be of most use to the scientific community if they are accompanied by source code used in its construction (*Stage 3*) and dynamic pixel-for-pixel vectorized seam-maps (*Stage 4*). Pixel-for-pixel traceability mapping (*Stage 4*) (e.g., Fig.1) is, we believe, the highest achievable standard for data traceability within a GIS framework. This allows the user to readily compare and quantify the difference between the derived product and the original data used in its construction, and to distinguish processing artifacts from surface features. Geospatial and spacecraft data processing software [5] still predominantly depend upon sequential, destructive (loss of information) frameworks that make seam-mapping of data challenging or computationally impossible for large data products. We incorporated non-linear, non-destructive (information-preserving) software in our otherwise destructive sequential pipeline to generate the CTX mosaic [3], inherited from photography, graphic design, and film, where the demand for rapid iteration over the last two decades has

resulted in efficient software that preserves all information and mitigates the need for inefficient and expensive high-performance computing. While this provides pixel-for-pixel traceability, this is more challenging for products (specifically topographic maps) where quantitative values must be preserved. This leaves a challenge for products that merge data from multiple instruments, like blended global DEMs [6-7], where users would *most* want to determine which instrument provided the data being observed, while the tools that facilitate the product’s construction [5] are destructive and not designed to efficiently preserve this information.

Thus, we consider the incorporation of more non-destructive data integration tools to be of long-term importance as geospatial data processing software continues to evolve.

References: [1] Laura J. et al. (2018) *LPSC*, 49, 1426. [2] Laura, J. et al. (2017) *IJG-I* doi: 10.3390/ijgi6060181. [3] Dickson et al. (2018) *LPSC* 49, 2480. [4] Fergason, R. et al. (2013) *LPSC* 44, 1642. [5] Beyer, R. (2015) *Geomorphology*, 240, 137-145. [6] Fergason, R. et al. (2018) http://bit.ly/HRSC_MOLA_Blend_v0. [7] Barker, M. et al. (2016), *Icarus*, 273, 346-355.

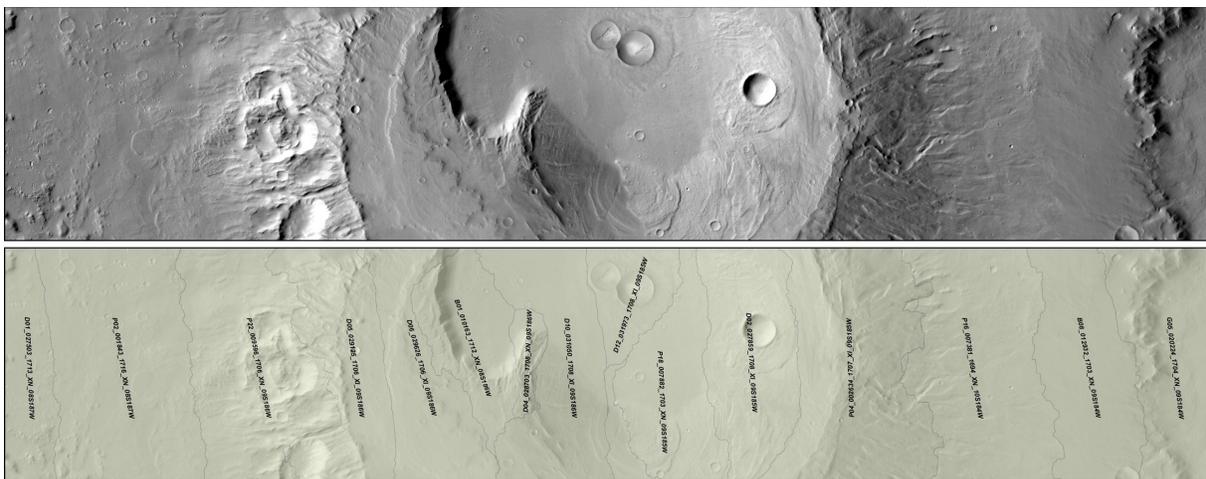


Figure 1. Portion of the global CTX mosaic of Mars [3] (top), with vectorized seam mapping (below) over the caldera of Apollinaris Patera. Scene covers ~200 km (40K px). Each feature in the map contains metadata and links to raw versions of each orbit.

Subjective Amount of information provided is determined by the data creator			Objective Information directly describes data heritage	
Stage 0 Metadata	Stage 1 Unvetted Description	Stage 2 Peer-reviewed Manuscript	Stage 3 Code-provided	Stage 4 Pixel-for-pixel Mapping
Functionality	Methods & Uncertainties		Reproducibility	Traceability

Table 1. Proposed stages of data reproducibility and traceability for *foundational* data products. All five stages may not be possible or valuable for every data product, thus we consider these to be a framework by which data creators communicate the degree of data heritage knowledge to the end user.