FIVE DECADES OF ASTROMATERIALS DATA: CAPTURING THE PAST AND PLANNING FOR THE FUTURE. R. A. Zeigler[1], J. Gross[1,2], F. M. McCubbin[1]; ryan.zeigler@nasa.gov; [1]NASA, Johnson Space Center, Mail Code XI2, Houston, TX, 77058, [2]Rutgers University, 610 Taylor Road, Piscataway 08854.

Overview: The Astromaterials Acquisition and Curation Office at NASA – Johnson Space Center (hereafter the Curation Office) is the past, present, and future home of NASA's astromaterials sample collections [1]. Over the past 50 years, the curation office has accumulated a huge volume of data and enormous variety of data types across 9 extant sample collections: Apollo, Luna, Antarctic Meteorites, Cosmic Dust, Microparticle Impacts, Genesis, Stardust – Cometary, Stardust – Interstellar, and Hayabusa. Additionally, data collection has already begun for future sample return missions such as Hayabusa2, OSIRIS REx, Mars2020, and Artemis. We are also increasingly incorporating modern instrumentation in the curation process, such as 3D laser scanning, high resolution precision photography (HRPP), X-ray computed tomography (XCT), mapping raman, and soon benchtop X-ray fluorescence (XRF).

This abstract will discuss: (1) the types of legacy data we have and efforts to modernize that data and present it to the public (where appropriate); (2) emerging high-volume data sets and the unique challenges they represent; (3) our efforts to work with our colleagues in the community to cross-reference our data with similar data preservation efforts. This abstract will focus largely on the Apollo collection, but much of what is said here is broadly applicable across all of our collections, though each current collection [2], future collection [3], and instrument [4] has its own unique challenges.

Existing Sample Data: As soon as the Apollo samples came back, data acquisition on the samples began. This includes a complete record of the sample processing history of each of the 2196 Apollo samples. These "Data Packs" are hard-copy mixed media documents that contain >50,000 official NASA sample photos, type-written instructions, hand-written notes, sketches, annotated polaroid photos (~40,000), etc. There are over 300 shelf feet of Data Packs. In addition to the record of what has happened to the samples while in NASA possession, there is also a fairly comprehensive record of the sample history after they are allocated, in the form of over 3300 sample requests and 57,000 sample history reports. Beyond the sample data, there is a record of the environment that the samples have been stored and processed in, including: (1) facility and laboratory schematics and documents of the original lunar receiving lab and the current lunar sample facility (stretching back to the mid-1960s); and (2) a limited record of the lab cleanliness (e.g., particle counts) and the cleanliness of the other materials that come into contact with the samples (e.g., N2 gas composition including isotopes; UPW composition; Lab temperature/humidity/pressure, etc.). Recently, the Apollo samples have undergone a retroactive preliminary examination (PE), at least in part. Modern non- or minimally-destructive techniques, especially HRPP and XCT, have been applied to portions of the collection [5,6], and increasingly being used as part of the day-to-day curation efforts [4].

Publicly served Data: Since the earliest Apollo days, efforts to catalog and summarize the information about Apollo samples have been captured in the form of sample catalogs [7], and more recently online compendia (collections of important sample information gleaned from the scientific literature [8]) and online catalogs of sample images and characteristics [9].

External Partners: In addition to data sets that are wholly or primarily the bailiwick of curation to preserve and present to the public, there are also efforts that we try to provide support for, either by providing advice and sample numbers [10] or through access to the underlying data on the public websites [11]. Similarly the curation office is actively working with multiple PDS nodes to make large volumes of image and instrument data available to the public.

Moving Forward: "Old" analog data sets are invaluable and need to be properly captured and preserved for future use. Even data sets that cannot be made publicly available (like sample requests) play an important internal role in better preserving a collection. There is more to preserving these data sets than simply scanning them, however. Without proper indexing of the data through some combination of handwriting recognition, machine learning, and brute force labelling, their usefulness is limited. Similarly, ensuring that the data format used to preserve the data are long lasting is key. Finding ways to make the sample environment data available to the community is an important future step. Redesigning and modernizing our website to provide more and better data to our PIs, as well as make the data more easily accessible through APIs to external users is also important. Finally, staying ahead of the curve on large-volume data sets in terms of both storage and public access is an important goal moving forward.

References: [1] McCubbin F. M. et al., (2016) 47th LPSC, abstract # 2668. [2] Allton J. H. (2021) This volume. [3] Mitchell J. L. et al. (2021) this Volume. [4] Eckley, S. A. et al., (2021) this volume. [5] Astromaterials 3D. [6] Apollo Sample News. [7] Apollo Sample Catalogs. [8] Lunar Sample Compendium. 9] Lunar Sample and Photo Catalog. [10] Astromaterials Data System. [11] Apollo in Real Time.