ASTROMATERIAL DATABASES FROM TWO PERSPECTIVES: USER AND CURATOR | LESSONS LEARNED FROM THE CURRENT STATE OF METEORITE DATA REPORTING AND DATABASE CURATION. A. Ostroverkhova[1], S. M. Morrison[1], J. Mays[2], A. Johansson[2], L. Profeta[2], K. Lehnert[2], [1]Earth and Planets Laboratory, Carnegie Institution for Science, 5241 Broad Branch Road, NW, Washington, DC, USA [2]Lamont-Doherty Earth Observatory, Palisades, NY, USA.

Introduction: The coming era of new space exploration opportunities (samples return missions), expanding and establishing analytical technology, as well as the application of data science and machine learning, provides challenges and opportunities for data curation. Here, we present data curation considerations related to meteorites as they are currently the most available extraterrestrial material, but the findings here will inform data curation of future return missions and other astromaterials.

Meteorite databases and datasets: One of the biggest challenges at this moment is the lack of access and fragmentation of astromaterial data. A few data projects are available for specific types of data, such as meteorites – MetBase [1], the Astromaterials Data System [2], presolar grain database [3-4], ChondriteDB for matrix, chondrules, and clasts [5]. Datasets related to museum collections exist, but are often not publicly available. Numerous personally-collected datasets for projects and PhD theses reside on personal websites or hard drives – not only are they not located in a centralized place, but they are often hard to find or entirely inaccessible. Some of these data overlap, which raises questions about coordination and cooperation for any new or ongoing projects in order to avoid duplication of effort.

Data in publications: Scientific articles are the primary sources of original data to the above databases. Currently, most data uploading is done with human assistance, while machine reading could provide a much more efficient upload rate. Unfortunately, some format features are making machine reading difficult to impossible. For example, converting data tables to graphical formats (JPG, JPEG) dramatically reduces their extractability. For older, non-digitized publications, paper and scan quality can pose challenges to data recovery.

Data reporting and storage: Meteorite names can be reported in slightly different ways over time (e.g., ALH-ALHA, using letters from different languages). Former meteorite names can be correlated to their up-to-date counterparts and thereby signify belonging of meteorites to a particular classification. Data curators and authors can follow The Meteoritical Bulletin Database for the latest naming conventions [6].

Recording of the sample source/name and ID number(s) is particularly important for reusable samples such as thin sections. Ideally, data curators would preserve this information and make it searchable for users. Likewise, the author's practice should be to provide this data in the methods section or directly in tables containing analysis results.

Reporting averages can have a few complications, including: not providing the number of analyses and/or samples; lack of specification of which samples were used (if few different meteorites were analyzed). Keeping detailed records of averages and their components is critical for database curation. For authors, including unaveraged data as supplementary material would help with reproducibility.

Description and classification of analyzed material can dramatically increase data reusability. Descriptions can include texture, classification of sub material (e.g., chondrules and CAI types), and any distinctive features. It is crucial for data curators to preserve as many details as possible (especially in the case of destructive analytical methods). Despite the lack of systematization, authors can still provide functional classification (e.g., texture description for chondrule and Type I/II segregation) to help fill this gap.

The graphical information (e.g., SEM images, elemental maps) is an information-rich part of extraterrestrial samples. These help with classification, interpretation of alteration processes, and understanding the origin of different components. Currently, none of the major databases provide the means to store graphical data. For future planning and information preservation, authors can preserve images as metadata with coordinates (spots for SEM, LA-ICPMS, NanoSIMS analysis), using software like Gigapixel.

Conclusion: Astromaterials data is multifaceted and heterogeneous, and additionally is formatted not only as text and numerical but also as graphical and geographical. Creating and following guidelines for astromaterials data is a complex endeavor, but one that ensure the maximization and preservation of these fundamental and precious data.

References: [1] MetBase. http://www.metbase.org/sites/Metbase_GUI_new. Accessed 9/1/21. [2] Lehnert et al. (2020) *EPSC*, 918. [3] Hynes & Gyngard (2009) *LPSC, 40,* 1198. [4] Stephan et al. (2020) *LPSC, 51,* 2140. [5] Hezel et al. (2018) *Geochem., 78, 1,* 1-14. [6] Piskorski et al. (2013) [6] The Meteoritical Bulletin. https://www.lpi.usra.edu/meteor/ Accessed 9/1/21.