EXPERIMENTS IN TRANSFERRING, VALIDATING, AND RELEASING MARS 2020 MISSION ARCHIVAL MULTI-MEDIA & IMAGERY DATA DELIVERIES IN THE CLOUD. Rishi Verma[1], Kevin M. Grimes[1], Zachary M. Taylor[1], Michael McAuley[1], Minh Le[1], Rafael Alanis[1]. [1]Jet Propulsion Laboratory, California Institute of Technology.

Introduction: The Planetary Data System (PDS) Imaging Node (IMG) has a rich history of evolving mission delivery and public release mechanisms to keep pace with emerging technologies in the consumer world. During the 1970s, for example, the NASA Viking 1 & 2 missions delivered images of the surface of Mars via magnetic tape media. During the late 1980s and early 1990s, the NASA Galileo and Mars Pathfinder missions pioneered compact-discs (CDs) as a medium to deliver imagery of distant planetary bodies to IMG. Following these missions, IMG quickly began accepting mission archival data deliveries through electronic transfer over the Internet. Marking the beginning of the 2020s, an emerging paradigm is to no longer receive and host mission archival data deliveries to IMG on proprietary, on-premises computational infrastructure, but to both receive and host deliveries on commercial cloud vendors. In addition to this, the computationally expensive tasks of validating, ingesting and archiving these delivered data into IMG's services are performed in the cloud as well.

Motivation: The Mars 2020 (M20) mission is the first NASA Mars mission to operate its entire science data processing system on cloud infrastructure. This is saving costs for the mission but has left archivers like IMG in a challenging position: does IMG continue to operate its archival data processing and release pipeline within on-premises infrastructure and manually scale to match increasing M20 data volumes; or, does IMG make the investments necessary to deploy its infrastructure onto the cloud, potentially reducing long-term costs?

One of the key appeals of leveraging cloud infrastructure is the "pay-by-use" model. Rather than spend upfront costs to procure physical hardware for archival storage and processing, which represent one-time purchases that cannot be easily returned if left unused, the cloud offers a more granular charging model, allowing users to pay by the hour (and, in some cases, by the minute or second). One of the conclusions discussed is this model affords IMG long-term savings in the storage, the validation, the ingest, and the public release of mission data archives that significantly increase in volume over time – like M20.

The paradigm IMG is spearheading for accepting, validating, and releasing mission archival data deliveries in the cloud is not without hurdles. IMG has been performing experiments to ascertain the viability, cost-effectiveness, and long-term stability of shifting the data acceptance, validation, ingest, and release workflow process to a cloud vendor (Amazon Web Services – or AWS – in this case). Experiments are being conducted with the Mars 2020 mission in mind given its historically unprecedented data volume projections that require innovative solutions to address the challenges posed.

Process: There are three important phases of the mission archival data delivery process that IMG has investigated: acceptance, validation, and ingest / release.

The acceptance process covers the mechanisms responsible for allowing a data provider to digitally transfer the contents of a particular mission archival data delivery to IMG quickly, securely, and cost-effectively. We discuss the surprising complications of this process – including technicalities within AWS that limit overall cost-savings.

The validation process ensures mission archival data deliveries (e.g. M20) are validated according to PDS standards both at a wholistic (bundle) level and an individual, per-file (product) level. Here as well, hurdles are crossed to ensure pre-cloud era tools such as the PDS Validate Tool [1] are utilized correctly and efficiently within a cloud environment. Moreover, IMG has experimented with optimization techniques to ensure the computationally expensive validation process occurs in a horizontally scalable fashion as the volume of M20 mission archival deliveries increases over time.

Finally, experiments with the process of parsing, ingesting, and releasing M20 mission archival data deliveries to the public are discussed, including cost savings and expenses. Strategies in crawling data deliveries efficiently are discussed, as well as the subsequent steps of extracting metadata and making it searchable. Moreover, ensuring that the overall cloud architecture supporting this public release is compatible with IMG user-facing software such as the Atlas is also evaluated.

Conclusion: IMG has thoroughly conducted an investigation into a fully end-to-end cloud native mission archival data delivery process to help itself (and the wider PDS) strategically plan for current and future NASA planetary missions. The experiments and conclusions discussed here will be of importance to the community for moving forward with this latest paradigm in NASA mission archival data delivery,

validation, and release of multi-media / imagery to the public and science community at large.

Acknowledgments: This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. © 2021. California Institute of Technology. Government sponsorship acknowledged.

The authors wish to acknowledge members of the M20 mission data delivery team for their support.

References: [1] "PDS Validate Tool." Validate Tool, 2.0.2, GitHub, 7 Apr. 2021, nasa-pds.github.io/validate/.