ARCHIVAL DATA AND COMPUTATIONAL POWER IN PLANETARY ASTRONOMY: LESSONS LEARNED 1979–2016 AND A VISION FOR 2020–2050. Mark R. Showalter[1], Matthew S. Tiscareno[1], and Robert S. French[1], [1]SETI Institute (189 Bernardo Ave, Suite 200, Mountain View, CA 94043; mshowalter@seti.org; mtiscareno@seti.org; rfrench@seti.org).

Introduction. When the Voyager spacecraft flew past Jupiter in 1979, the most powerful computer available to most planetary scientists was a PDP-11. With 64K of addressable memory, analyzing even a single 800×800 Voyager image posed serious challenges. Disk storage might hold a few tens of images, but anything beyond that had to be stored off-line on tapes. Retrieving data entailed mailing (not emailing) a request to JPL and waiting for the cross-country shipment of tapes to arrive, perhaps weeks later.

Today, the entire Voyager image archive can be stored on a thumb drive. Scientists routinely analyze thousands of planetary data products on a laptop. Computational advances like these have been fueled by Moore's Law, which predicts that compute power doubles every two years. With our 37-year baseline of experience, we present some ideas about how planetary science can prepare to take advantage of the next decades of growth in compute power. However, we must also acknowledge the perils of making any such predictions, given that so many of today's capabilities were unimaginable in 1979. As a result, our primary focus is on the next 10–15 years. However, we also offer some suggestions for how to prepare for what might come thereafter.

Trends in Scientific Computing: Moore's Law has proceeded uninterrupted for many decades. However, as with any exponential trend, it cannot continue forever [1]. For example, CPU clock rates appear to be maxing out at the level of a few GHz, where heat dissipation becomes a limiting factor. Modern CPUs can execute more instructions per clock cycle, but even this trend cannot continue without limit. However, compensating for these limits is parallelism—the ability to harness very large numbers of CPUs to work on the same problem simultaneously. For example, the display of any laptop is driven by a Graphical Processing Unit (GPU), which is capable of performing many thousands of floating-point operations in parallel. With the proper re-formulation, many computational problems can already take advantage of GPU acceleration.

Data storage capacities and Internet access speeds continue to grow exponentially. However, at the same time, many NASA missions and Earth-based astronomers have demonstrated the capacity to generate exponentially increasing volumes of data. For example, NASA's Large Synoptic Survey Telescope (LSST) will generate 15 TB every night. Although the current Deep Space Network has distinct downlink limits, NASA and other agencies are already experimenting with laser-based optical transmission; this could dramatically increase the data volume from interplanetary spacecraft. Thus, we need to prepare for the possibility that data storage and access times will continue to be limiting factors in scientific research.

Cloud computing will provide a work-around to these limitations. Today, a scientist can construct a "virtual machine" (VM) that contains all of the software and data needed to perform a particular calculation. An arbitrary number of these VMs can execute simultaneously in the cloud. In one recent project, we developed a procedure that processes each image in the Cassini archive. Processing all 400,000 images required 30 days on an eight-core machine. We found that the problem was straightforward to re-formulate for Amazon's EC2 cloud computing platform, where it can now run much faster and for a total cost of $200 (using the lowest-priced computing tier). Depending on the application, other uses of cloud computing could easily run into the thousands of dollars at current prices, but these prices are certain to drop over time. Within the next decade, we can comfortably predict that most scientific computing will be performed in the cloud.

Steps Toward the Future of Cloud-Based Planetary Astronomy: Although cloud computing is our future, it has one marked disadvantage relative to the personal laptop—data analysis cannot be interactive. For this reason, we must explore key and fundamental changes to the way scientists analyze data.

Cloud-based data. Today, the Planetary Data System provides direct on-line access to NASA's planetary data. A typical user searches for data, downloads a small subset to their laptop, and proceeds with the analysis. In the future, it will be much more practical to move the "laptop" to the data. By this, we mean that the PDS should store complete, calibrated data sets in the cloud and make them available through pre-packaged VMs. When a scientist wishes to analyze a data set, s/he will download a VM, install into it the needed software, submit it to the cloud, and let it run there. In many cases, the data will never need to be downloaded.

Software preservation. Each NASA mission's data pipelines are written for its current era. As we all know, old source code is extremely difficult to maintain because hardware, operating systems and programming languages continue to evolve. The calibration pipelines of some NASA missions are already nearly impossible to run, because they were developed for architectures that no longer exists. For the sake of each mission's long-term legacy, it is critically important to begin preserving these pipelines. VMs provide a partial solution, because they also capture the OS and the environment needed to run software. Although a VM may not continue to run once the associated hardware is obsolete, hardware emulators can, in principle, preserve the functionality indefinitely. The broader computer science community has recognized the importance of this problem; NASA should endorse and support these efforts.

Calibration. Many mission data sets are still delivered to the PDS in raw form. In the future, it will be increasingly impractical to expect users to calibrate their own data products (except perhaps under specialized circumstances). Although preserving software pipelines via VMs, as discussed above, could be part of the solution, the ideal is for calibrated data always to be preserved as a part of the permanent PDS archive.

Precision navigation. One of the key steps in analyzing almost any planetary data product is navigation—aligning the data with a geometric description of the field of view. In the case of images, navigation can be based on the locations of fiducial features such as stars, moons, rings or specific craters. Instrument pointing is imperfect, so until the product is navigated, the predicted and observed locations of features in the field of view will disagree. In the past, navigation was almost always done by hand, product by product; such manual navigation is obviously impractical when TB of data are being generated each day. We have been developing procedures for the automated navigation of Cassini Saturn images, and are achieving a high rate of success [2,3]. Such procedures must be developed and perfected for all NASA data sets. Alternatively, spacecraft systems should be developed such that pointing and telemetry are sufficiently accurate that navigation becomes unnecessary.

Metadata and backplanes. A complete geometric description of each product's field of view serves two purposes. First, it potentially simplifies the analysis by automatically associating the geometry of each pixel with its geometric content. Second, it provides the robust information that might be needed for a user to determine, before checking, whether the geometric content of a particular product meets certain scientific requirements such as target body, resolution and lighting geometry. As one example, such information is critical to our ability to track new discoveries back through old data sets. Accurate metadata is contingent upon already having well-navigated products. VMs that generate specialized metadata and backplanes could also have a place in the archive.

Pattern recognition and neural networks. Many pocket cameras and phones now contain sophisticated algorithms for smile detection and blink detection. Additionally, popular photo software is capable of classifying images according to categories like "dog" and "birthday cake." One can only imagine the power of similar algorithms when applied to our planetary archives, where features like "cloud", "dust devil", "new crater" or "impact event" could potentially be identified automatically. The technology behind these algorithms must be harnessed for planetary data analysis and discovery.

Longer-term trends. The future of computing is driven by for-profit companies investing in tools to reach large markets. Beyond the next decade or so, it is impossible to predict where the big breakthroughs will occur. Planetary science will never be a large market, so we cannot necessarily expect profit-driven corporations to produce scientifically useful tools. However, we can and should piggy-back off the newest technologies as they emerge. We note that the greatest benefit from our tools can only emerge if we share them and agree to build upon them; this requires that NASA and the planetary science community reaffirm our commitment to open source.

Conclusions: Any survey of the literature from the 1980s and 1990s will confirm that it was a vibrant time in planetary astronomy, even though we were working with hardware that might today be compared to stone knives and bearskins. This illustrates the point that computers alone are insufficient to solve our scientific problems; the fundamental first steps will always originate with humans, via their ideas and innovations. Computers can, however, eliminate much of the hands-on drudgery that went into the pixel-by-pixel data analysis of earlier decades, streamlining the path from a germ of an idea to a refereed publication. With Moore's Law still in force, or nearly so, we will also be able to revisit old data and perform analyses that were once computationally impossible. Thus, computational advances will not just support future missions; they will, with the proper groundwork, make it possible to uncover fundamental new discoveries in NASA's existing archives.

References: [1] Waldrop, M. M. (2016), *Nature* **530**, 145–147. [2] French, R. S., M. R. Showalter, and M. K. Gordon (2014), DPS meeting #46, 422.01. [3] French, R. S., M. R. Showalter, and M. K. Gordon (2014), DPS meeting #48, Abstract 121.14.