**DATA TECHNOLOGIES FOR PLANETARY SCIENCE OF THE NEXT 3 DECADES.** K.-M. Aye[1], A. Muench[2]; [1]Laboratory for Atmosphere and Space Physics, University of Colorado Boulder, 1234 Innovation Drive, Boulder, CO 80303 (michael.aye@lasp.colorado.edu), [2]American Astronomical Society

**Introduction:** With increasing capabilities of remote sensing technologies and improvements in data downlink rates using optical communication, the amount of data available for planetary science will drastically increase, once the required investments in new and more ground stations required for optical communication have been performed.

But already today planetary scientists have in principle access to data from many different planetary missions. We say 'in principle' because in reality there are many hurdles for inter-mission inter-instrumental data analyses to overcome. It starts with identifying the existing data, continues with reading in data formats of vastly different kinds, created within decades of exploration, then combining data taken at different resolutions in time and space.

This also creates a problem for scientific reproducibility. Publishers do not yet have facitilites to store one-click data archives for all data used for a research paper. This problem will only get worse when the amount of available data increases and the frequency of inter-instrumental data analyses increases. Our future vision tries to address several of these obstacles by identifying technologies that exist today, but require to be connected with each other to maximize their benefit to the scientific community.

**Data identification and retrieval:** The Planetary Data System (PDS) and its European pendent Planetary Science Archive (PSA) are currently the most future-proof data storage locations for planetary science data. But data retrieveal from different missions is still hard. In the best case, some meta-data have been combined into databases across all instruments of a mission or even across missions. However, using web-based data search engines is highly time-consuming, automatic search and retrieve interfaces to existing analysis environments like IDL, Matlab and Python are mostly non-existing. Additionally, advanced users that want to combine data from different nodes of the PDS will have to suffer from non-uniform interfaces, requiring relearning each time, and a subsequent combination of data outputs with different structure and formats.

However, we believe the technologies to improve this situation exist today, and are beginning to spread soon. The PDS "Ring-Moon Systems Node" has recently implemented a meta-database that covers a much higher number of science-constraining parameters than other PDS nodes offer. Additionally, this node also offers an easy to exploit application programmer interface (API) for searching and downloading data by creating URL strings. An example of how this can be implemented in Python can be seen at [1]. Using technologies like these, users of all analysis environments that support systematic string creation could create interfaces to planetary science data with the convienience of the analysis environment they are most familiar with. We envision that these kind of easily accessible meta-data interfaces will create vast time savings in future data analyses that encompass multiple instruments on multiple missions.

*Situational awareness (SPICE).* For more efficient identification of data of interest, we envision the use of existing technologies like webGL, directly at browse-time of the PDS and PSA, to be shown a mission and instrument relevant 3D situational overview for any chosen moment in time, similar to what is shown in Figure 1. When a user is browsing data from a mission with orbital travel as complicated as Cassini for example, it would be immensely helpful to have an immediate graphical overview of the current orbital configuration for any time of interest. We only identify lack of funding as a reason to not have these technologies in place today and are hopeful that this instantaneous connection of SPICE data displays to planetary databases will be made at some point in the next decade [2].

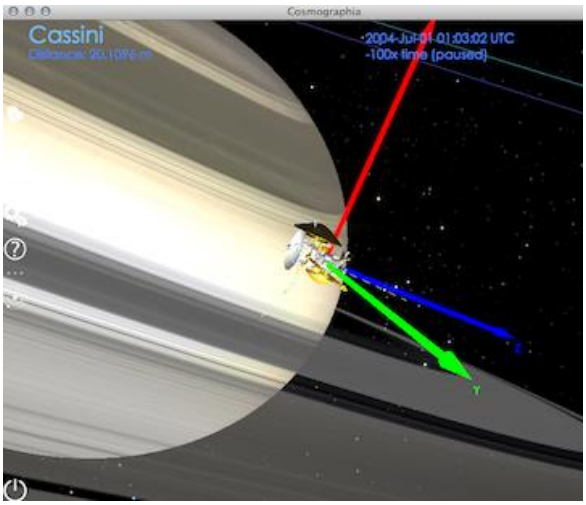


**Figure 1 NAIF SPICE preview using Cosmographia (Source: NASA NAIF Website)**

**Data analyses in parallel:** The amount and size of available data will increase sufficiently that using parallel computing technologies will be absolutely

unavoidable. We have technologies available today, but not yet wide-spread, that make it much easier to work in the parallel computing paradigm than only 5 years ago. Programming and working in parallel requires a quite different mindset from the linear programming techniques that the average planetary scientist applies. But already today, some of these technologies either are using automatic parallelization (Intel numerical libraries), or offering interfaces to parallelization that reduce the learning curve to a minimum.

We envision that the Jupyter notebook technology, funded multiple times by the Sloan foundation, is a key element to provide these technologies. Jupyter notebooks had been developed using Python as the computing kernel, but has since grown to be a computing kernel independent webbrowser based computing system, that supports a multitude of computing languages. Python-based parallel computing libraries enable the average data user to manage dozens to hundreds of cores of large parallel computing clusters, simply by clicking interfaces in their web-browser, but also directly interfaced with their Python functions.

We believe that minor investments in educating planetary scientists in parallel computing and IT departments in deploying these existing technologies will vastly improve the access to more computing power, as is required by the upcoming data challenges of the next decades.

**Reproducability:** We envision much deeper connections between planetary science data, research analysis, and peer reviewed articles. Scholarly publishing, as part of its shifting focus to born digital research results, will begin to incorporate the tools of research directly into the publications of the future.

Tools such as Jupyter notebooks and virtualization containers allow for the packaging and distribution of complete research projects. Notebooks wrap analysis scripts and pipelines in a descriptive narrative that parallel directly the LaTeX formatted papers written today. The difference is that by incorporating these tools into the published article, Journals will bring computational power to the now static text. Virtualization containers such as docker enable the capture of entire workflows for replication and reuse. Future capabilities of planetary science archives will provide deep persistent links between published articles and well documented datasets, while peer reviewed articles will expose these links to ensure reproducibility of results and reuse of data. Connecting these software and data to articles via citation and persistent linking is most important to the broader scholarly commons where a richer set of research objects, e.g., software, are recognized and attributed to individuals and groups.

**References:** [1] pyciss: Python utilities to work with Cassini's ISS camera system. K.-Michael Aye (2016). Zenodo. http://doi.org/10.5281/zenodo.166116 [2] SPICE NAIF system http://naif.jpl.nasa.gov/naif/, [2] Jupyter notebook http://jupyter.org