On the Use of Planetary Science Data for Studying Extrasolar Planets: Enabling cross-SMD Archive Interoperability

D. J. Crichton, T. J. Lazio, J. S. Hughes, J. Jewell, E. Law, G. M. Roudier, J. Padams

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109

Corresponding Author: Daniel. J.Crichton@jpl.nasa.gov

Summary

The characterization of extrasolar planets and the search for life beyond Earth is a topic that is naturally cross-disciplinary between Planetary Science and Astrophysics, and may touch upon Heliophysics and Earth Science. However, the relevant data are spread over multiple NASA data archives (including the PDS). NASA could accelerate and enable new approaches to studying both extrasolar planets and bodies within the solar system by investing in developing analytical data services infrastructures, post-archive, that provides the ability to extract, integrate, and stage data across archives that is structured to support interdisciplinary science with appropriate data-driven approaches [1]. This is well articulated in SMD's Strategy for Data and Computing in 2019 [2]. Observations within the solar system provide a spatial and spectral resolution unimaginable for extrasolar planets, but those extrasolar planets greatly outnumber the relatively small number of targets available nearby, providing a more complete picture of the wide diversity of planetary systems and allow for a generic, less parochial understanding of how planets form and evolve. It is only through comparative and joint analysis of the solar system and extrasolar planetary systems that we can piece together a more complete understanding of the formation of planetary systems, leading to insight into the frequency of occurrence of habitable worlds within those planetary systems.

1. Science needs and opportunities across data archives

An example of the type of analysis that could be enhanced by cross-disciplinary access between the PDS and Astrophysics Data Archives is given by extrasolar planet phase curve modeling validation using *Cassini* ISS images of Jupiter. Phase curve analysis of extra-solar systems (spectroscopic observations of the unresolved stellar and orbiting planet during a full orbital revolution), in combination with transit spectroscopy, probes the content, structure and dynamics of exoplanets atmospheres. The next generation of datasets (JWST, ARIEL) will not only push the on-going theoretical effort into more detailed and complex modeling but will also unveil a regime of extrasolar planetary atmospheric physics that has been poorly explored, due to the natural observational bias towards high temperature atmospheres of gas giants. In the near-infrared (NIR), the contribution of the orbiting body to the total (host star + planet) signal is essentially thermal emission for a vast majority of characterized exoplanets with HST and Spitzer, state-of-the-art datasets in the field.

As the community interests are trending towards colder targets such as super-Earths and warm Neptunes with the incoming generation of datasets from JWST and ARIEL, the relative contribution (dominance) of thermal emission from the planet compared to the stellar light bouncing through the extrasolar planet's atmosphere will change, requiring the modeling to consider not only absorbers, but also scattering and reflecting components such as aerosols and clouds, in particular in the context of phase curve science.

An extreme case of this regime of radiative transfer modeling is the modeling, in the ultraviolet (UV) and NIR, of Jupiter---a cold gas giant compared to most known extrasolar planets---for which scattered and reflected light dominates the signal. To validate the augmentation of the current extrasolar planet transit modeling to phase curve analysis, we can stress-test both the forward model and the retrieval facility against *Cassini* ISS images (extremely accurate compared to extrasolar planetary data), challenging our modeling effort on extrasolar planet aerosols and clouds into an unexplored regime where the scattering and reflective contribution create the majority of the measured signal.

Other opportunities exist to model solar system objects as "extrasolar planetary analogs." As the Juno mission travelled from Earth to Jupiter, ground-based facilities performed complementary observations of Jupiter. In particular, NASA's InfraRed Telescope Facility (IRTF) undertook an extensive observing campaign covering a range of infrared wavelengths over more than 1000 epochs. If the data are easily accessible and discoverable from both the PDS and the Astrophysics Data Archive, they would provide an opportunity to 1) to quantify variations of Jupiter's flux as it rotates, particular at longer wavelengths (5 microns) where thermal emission is stronger than scattered light, 2) to retrieve a 2-D

surface map of Jupiter as a function of wavelength, and 3) to use the 2-D model as a template for simulated extrasolar planet observations, for which the aim is to image directly an extrasolar planet's rotational modulation.

2. Data and Computing Architecture and Infrastructure:

Data captured in NASA archives is often optimized for archiving. It is not structured in a way that enables it to be quickly extracted, combined with other data, and analyzed. As science moves to become more data intensive, staging data in such environments is going to open opportunities to apply more data-driven methods. As it is, groups that want to apply different data mining approaches often need to restage and build their own data environments to be able to begin applying different techniques (e.g, machine learning, deep learning, etc). Further, building well-labelled data is important to improving the efficacy of different approaches, particularly deep learning which requires massive amounts of labeled data. The NASA Planetary Data Ecosystem (PDE) needs a strategy for extract, staging, and constructing such environments. This will give way to more interdisciplinary capabilities by laying the necessary foundation to not only bring data together but to do it in a systematic manner using modern data science approaches.

As a result, the PDE should recognize analytical capabilities beyond basic archives. The PDS is highly critical to supporting both missions and the research community. However, architecturally, there needs to be a distinction between archival data and data prepared to support analysis. This is particularly important for leveraging modern data science approaches where data must be staged and structured to support analysis that may span multiple disciplines, missions, instruments, and observations. As a result, the PDE should explicitly extend the science data lifecycle to recognize analytical capabilities beyond archives.

Building a more rigorous platform to enable data science beyond archives that is constructed through systematic approaches such as on-demand computational workflows can yield significant improvements towards repeatability. Today, much of this is ad hoc and mechanisms to repeat and attempt to reproduce science results often cannot be accomplished. Archives themselves generally are capturing raw and processed observational data, but the shift to integrate data across archives coupled with processing requires a more rigorous approach to delivering computational capabilities and services in a disciplined manner. Constructing more rigorous pipelines that can also integrate data science methodologies (e.g., machine learning, deep learning, applied statistics, etc.) can be used to provide more routine extraction and preparation of data to enable analysis. This requires an explicit data science strategy from PDS and the NASA planetary science division along with investments in analytical data infrastructures.

3. Examples: The JPL Interdisciplinary Environment for Exoplanet Research

JPL has developed a prototype infrastructure linking planetary and astrophysics archive together towards this vision of an interdisciplinary analytical capability. This included integrating data from across the PDS with astrophysics data into an online data portal based on the PDS4 model to support search and usability of the data. The virtual portal pulls data from the Ring Moon Systems Node of the PDS combing it with other data for exoplanet research. The results will enhance the analysis and interpretation of data from many NASA missions, for example Cassini and Juno, as well as improve our interpretation of exoplanet observations from HST now and JWST, and WFIRST in the future. Moreover, this will enable new data products from ground-based observations (such as the Infrared Telescope Facility, Keck, and others) archived and/or available to the PDS to be available for comparative studies of our Solar System and exoplanetary systems across discipline archives.

4. References

[1] Crichton, D., Hughes, J. S., Roudier, G., West, R., Jewell, J., Bryden, G., … Lazio, T. J. W. (2019). On the Use of Planetary Science Data for Studying Extrasolar Planets. Bulletin of the AAS. Retrieved from https://baas.aas.org/pub/2020n3i130

[2] Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science, 2019. https://science.nasa.gov/science-pink/s3fs-public/atoms/files/SDMWG%20Strategy_Final.pdf