PLANETARY DATA: HIDDEN BARRIERS TO DIVERSITY IN THE PLANETARY SCIENCE COMMUNITY. L. Keszthelyi[1], M. Milazzo[2], M. Miller[3], E. Lakdawalla[4] [1]U.S, Geological Survey, Astrogeology Science Center, Flagstaff, AZ (laz@usgs.gov), [2]Other Orb LLC, [3]Jacobs/NASA Johnson Space Center, [4]The Lakdawalla Group LLC.

Introduction: NASA has placed importance on building diversity in the planetary science community [1]. A variety of NASA programs (e.g., MUREP) seek to increase the participation of people from groups that have historically been under-represented in planetary science. Here we highlight the role that planetary data plays as an unacknowledged systemic barrier to entry into the planetary science community.

As with many systemic barriers to inclusion, there is no obvious intentional effort to exclude specific groups. Instead, steps taken to serve the existing community have often been made without explicit consideration of the challenges being created to those outside of the community. This inadvertently builds barriers that are essentially invisible to those within the community.

There will always be high barriers to working in planetary science. The field is “rocket science” and requires highly specialized knowledge. The existence of these significant unavoidable barriers makes it all the more important to remove *unnecessary* barriers to entry into the planetary science community. Here we narrowly focus on issues related to planetary data, i.e., the topic of this workshop.

While the discussion below is informed by the experiences of all the authors, it draws specifics from an FY20 project to determine the challenges to investigating the photometry of planetary surfaces. This study was conducted from the perspective of a first-year graduate student starting new research into lunar photometry. This “persona” sought to find and gather appropriate data products, process them to extract photometric information, compare the observations to various photometric models and, ultimately, share the results. During this exercise, it became clear that the outcomes would be very different for a student at an institution with strong existing ties to lunar science versus one that was largely outside of the planetary science community.

We describe our observations through the lens of the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [2] which we found useful for framing the issues we encountered.

Findable: Mission data are meticulously archived in the NASA Planetary Data System (PDS) and similar archives maintained by other space agencies. The PDS has a mandate to focus on *archival* of data which drives an emphasis on raw and lower-level products [3]. Expert users who intend to produce their own higher-level products are the intended customers of this system. While the PDS is not perfect, it is generally observed that expert users are well-served by it [4].

The same cannot be said for a novice user at an institution without established links to the planetary science community. Such persons lack ready access to advisors or mentors who can point them to start their search for data from a specific instrument on a specific mission (in this case LRO LROC-WAC images on the PDS Imaging Node). Novice users without community connections may struggle to even find the PDS, much less discover which instruments on which missions obtained data relevant to their study. For example, nothing on the first page of a Google search on “lunar photometry” is helpful (i.e., points to the LROC WAC images or publications from them). The problem is not that key information is hidden but rather that it is obfuscated by vast quantities of information that all appear relevant. Even after deciding to use the LROC WAC data, the challenge in selecting a subset of the data to analyze is similar. The PDS provides seven different search tools with no clear guidance for which to use in a particular case. Furthermore, some of the most powerful tools, such as the JMoon and MoonTrek are not listed. While our persona highlighted concerns with the LROC-WAC data, this is just one specific example of issues endemic to searching planetary science data.

The challenges are far greater when it comes to finding higher-level data products. Only a small fraction of such products are hosted in the PDS. Of the rest, only some are uploaded to a myriad of data repositories (typically associated with the authors’ institution). Search is sometimes possible within a given repository, but there is no central portal that is aware of the contents of the many dozens of separate repositories that may contain relevant data. In practice, the most reliable way to find these products is by word of mouth within the community. This obviously excludes anyone who is not already a member of that science network. Lacking such access, one usually must search for the relevant journal publications and contact the authors.

Accessible: Once data have been located, barriers to access lower-level data are minimal. Mission data are all provided without cost. Transfer of large volumes of data is always a challenge but relatively easy-to-use free tools exist. However, higher-level products are often less accessible. Some are provided as supplemental materials to journal articles that are behind a paywall.

NASA's new mandate that all publications (and associated data) be stored in a public repository begins to address this issue, at least going forward. Other data are available "on request" from authors. The response to such queries is anecdotally almost always prompt and positive when it comes from a known member of the community. While we do not have quantified metrics on the responses to queries from individuals outside the community, the need to ask for help from a stranger is itself a non-trivial barrier.

Interoperable: Once a researcher has the data, working with it requires specialized software. Data from the PDS are not immediately ready to import using typical data analysis software (e.g., ArcGIS or MATLAB). There is a host of programs and scripts to convert PDS data into formats more amenable to analysis tools. Some of these, like JMARS, even integrate search, download, and analysis into a single package. These tools, and expert users to help explain their use, are part of the basic infrastructure at institutions that are integrated into the planetary science community. Given the availability of in-house training for those within the planetary science community, there has been little perceived need for formal or online training in the installation and use of these types of tools. However, this capability is largely alien to other institutions. Tools are relatively easy to discover, but very little guidance is provided on which is the most appropriate for a given use case.

The depth and breadth of this issue is hard to overstate. As a simple but critical example, one can take standards for longitude. Terrestrial geospatial tools typically present longitude with values between -180° and +180° with positive to the east. But for planetary data, many data sets are provided with longitude values from 0° to 360°. Some data have positive to the east and others are positive to the west. Another example is that some raster data index the first pixel as (0,0) while others define it as (1,1). The position of the corner of the image can refer to the position at the center of the reference pixel or its outer corner. There are different, but parallel, issues translating between standards used by astronomers and planetary scientists.

The PDS4 standard does begin to address some of these problems. The geospatial portion of the PDS4 metadata now follow the FGDC (Federal Geographic Data Committee) standards. And the FGDC standards are being brought in line with international (ISO) standards. However, not all interoperability issues can be solved by planetary science simply adopting existing standards. For example, it is not uncommon for existing standards to have planetary radius hardwired for the Earth, providing no option for different planetary radii. Active and effective engagement with international standards groups outside of planetary science is required.

Reusable: Once a researcher has created a data product, there is no standard process to make it available to others. Given the legal requirements set by many funding agencies to make derived data public, this is no longer a problem that researchers can address on a "best effort" basis.

For a researcher at an institution that is integrated into the planetary science community, placing data onto their institutional data repository is a viable option, especially if followed up with vigorous institution-supported publicity at key meetings (e.g., handouts at an institutional LPSC booth). The planetary science community is small enough that this type of word-of-mouth communication is effective – but also effectively excludes those not already within the community.

For a researcher seeking to enter the community, the options are all suboptimal. Their institutional data repository will not be found by the community. Their ability to publicize data as an individual is limited. The process for entering data into the PDS, designed for archiving streams of data from active spaceflight missions, is inappropriate for many one-off, highly processed products. Many journals have noted this gap and are creating private data repositories that may meet the letter of the legal requirement but do not meet the intent of making data products freely available.

Conclusion: Existing methods of saving and sharing planetary science data favor people already in the community and needlessly exclude those outside it. Integrating FAIR data principles into planetary science would reduce those barriers to entry. This work cannot be performed in isolation; it requires collaboration with many other communities (Earth science, astrophysics, software developers, etc.). It includes adopting existing standards and instigating changes to standards that exclude planetary data. Metadata that follow a common standard open the door to interoperability with widely used tools, data repositories, and search methods.

As a final note, making it easier for everyone to find, use and share our data will also help those already in the community. We make the case that attending to the needs of those outside the community is an important factor when prioritizing improvements to the planetary data ecosystem.

References:

[1] https://science.nasa.gov/science-red/s3fs-public/atoms/files/2020-2024_Science.pdf.
[2] https://www.go-fair.org/fair-principles/.
[3] https://pds.nasa.gov/home/about/PlanetaryDataSystemRMS17-26_20jun17.pdf.
[4] https://pds.nasa.gov/home/NASAPDSUserSatisfaction2020final2.pdf.