DATA ORGANIZATION AND ACCESSIBILITY FOR SMALL SOLAR SYSTEM BODIES IN THE ERA OF LARGE SURVEYS. Henry. H. Hsieh[1], Dennis Bodewits[2], Larry Denneau[3], Michael S. Kelley[2], Matthew M. Knight[2], Nicholas A. Moskovitz[4], Cristina A. Thomas[1], [1]Planetary Science Institute (1700 E. Ft. Lowell Rd., Suite 106, Tucson, AZ 85719, USA; hhsieh@psi.edu) , [2]University of Maryland, [3]University of Hawaii, [4]Lowell Observatory.

Background: In the field of small solar system body science, we are currently in an era where we have access to large data sets, and are looking forward to even larger data sets in the future. Many of these large data sets (e.g., catalogs of albedos, diameters, taxonomic classifications, spin vectors, and so on) are available for community use via the Small Body Node of NASA's Planetary Data System (PDS), although other data sets, including those from the Catalina Sky Survey and Pan-STARRS1 survey, are either archived elsewhere or not (yet) publicly available. Even those data available through the PDS are distributed across multiple standalone data sets with only limited functionality for querying multiple data sets simultaneously. This lack of coordinated access to multiple disparate data sets means that substantial additional effort is required on the part of researchers to take full advantage of the broad range of efforts that our community has engaged in over the years to better understand small body populations in our solar system.

Current and Future Large Surveys: The Large Synoptic Survey Telescope (LSST) is expected to revolutionize small body science, increasing the known populations of minor planets and comets by an order of magnitude or more, and acquiring tens of millions of observations of both new and currently known objects. While the exact details of the LSST's Moving Object Processing System (LSST-MOPS) are not yet finalized, its main product will likely be a catalog of individual object detections with calibrated astrometry, photometry, and morphological parameters (e.g., point-spread function widths), and basic metadata such as object designations, orbital elements, observation dates and times, and image quality parameters. This will be accompanied by an alert system for notifying members of the community about observations of objects meeting specified criteria.

While the LSST moving object detection catalog will undoubtedly become an extraordinary resource for a wide range of small body science, it will also be insufficient for actually performing much of that science without significant additional effort. For example, some scientists may be interested in retrieving only those data acquired when an object was within a certain heliocentric distance range or at a particular orbital position or meets other geometric, photometric, or morphological criteria. Other scientists may be interested in selecting data for all asteroids that have certain orbital characteristics and/or physical properties, rather than having a list of specified targets.

Meanwhile, multiple other surveys are currently in operation or are being planned or proposed, including the Catalina Sky Survey, Pan-STARRS, Gaia, the Asteroid Terrestrial-Impact Last Alert System (ATLAS), the Dark Energy Survey, the Zwicky Transient Factory (ZTF, successor to the Palomar Transient Factory survey), and the Near-Earth Object Camera (NEOCam) mission, among many others. Efforts are also underway to extract the vast amount of asteroid data that have been serendipitously obtained over decades of conventional telescope observations of non-solar system targets, much of which currently sits untapped in public archives. Maximizing the amount of science that can be achieved will rely heavily on our future ability to draw on all of these disparate data sources and also connect them with the appropriate metadata (e.g., heliocentric and geocentric distances, phase angles, true anomalies, etc.) needed to properly interpret them. Effective management and synthesis of current and future data streams will also help to maximize the reconnaissance value of these data for future scientific and commercial missions to small bodies,

Proposed Work: With so many large and disparate data sets currently available that are relevant to small body science, and many more to come — the largest being the LSST moving object catalog — we foresee an urgent need for far greater sophistication in the way that we organize and access these data. We are currently in the early stages of an effort to develop tools for producing higher-level SSSB-specific data products from LSST data than will be produced by the baseline LSST and LSST-MOPS pipelines. We also seek to design a database infrastructure and user interface to organize and provide access to those data products and relevant metadata to facilitate scientific usage of LSST data by the broader solar system community. This effort additionally includes plans for cross-linking LSST detection data to external or mirrored data sources such as albedo or taxonomy catalogs, asteroid family lists, or lightcurve databases to provide added physical context. The overall system would ultimately be aimed at allowing users to retrieve

data for objects that simultaneously meet a multitude of observational, physical, and dynamical conditions, thus streamlining the process by which focused scientific investigations of particular types of objects observed under particular conditions can be conducted. We expect that, ultimately, this system could then be expanded to incorporate data and derived products from many of the other current, planned, or proposed survey efforts mentioned above to increase their collective value and impact even more.

Until now, the solar system community has been reasonably successful conducting scientific investigations of survey data on an ad-hoc, individual basis. However, the flood of data that awaits us in the future will require a substantially different approach on our part as a community if we hope to make the most of our current and future investments in survey efforts in pursuit of our scientific, exploratory, and hazard mitigation goals. Given the imminent arrival of this new era in data volume, we argue that the time for careful consideration of how to manage and leverage current and future data streams, and development of the tools to do so, is now.