Solar System Science Collaboration Roadmap: A starting point

The goal of our science is to understand planetary formation and evolution, including placing the Solar System in context with other planetary systems. This requires an understanding of the current status of our Solar System -- the orbital and size distributions of small bodies (e.g. asteroids, comets, planetary satellites, TransNeptunian Objects, Scattered Disk Objects, and other intermediary populations) and their physical properties (e.g. chemical composition, physical shape, masses, density and porosity). LSST will report detections of moving objects in various filters (ugrizy) between approximately 16 and 24.5 magnitudes (in r band) over its observing footprint, link these detections into orbits, and provide metadata on observing conditions, including a generalized completeness estimate for each population. However, it will be up to us to apply a wide variety of methods to translate this information into the data we need to accomplish our scientific goals.

What will LSST provide?

This list isn't intended to be exhaustive, but should capture the main points of what LSST will provide so that we can identify what additional analysis or capabilities we need. The LSST Data Products Document (http://ls.st/dpdd) is the authoritative source; this just an interpretation of that document. LSST will:

1. image the sky with a cadence generally appropriate for detecting small moving objects in the Solar System[1] in filters *ugrizy* over the southern hemisphere, with a northern extension to cover the ecliptic plane + 10 degrees (this 'northern ecliptic spur' may only be imaged in *griz*). More about the observing footprint and plan can be found at http://www.lsst.org/lsst/opsim and simulations of the observing history can be found at https://www.lsstcorp.org/?q=opsim/home. The observing cadence is not 'set in stone' and is open to feedback.
2. create "difference images" where the images in each visit[2] are subtracted from a template image[3] to detect moving objects (plus transient and variable sources) -- these difference image detected sources (diaSources) must appear at 5-sigma above the sky background. In addition, the two exposures within a visit will be subtracted from

[1] This cadence has to be a compromise with other science requirements, but is fairly flexible. The current plan is to observe each field two times (2 visits) per night, separated by 30-90 minutes, then return to each field several times during a month. There should be enough flexiblity to observe each field 3 or 4 times per night, then perhaps have fewer observations of the field in later months or years if this proves necessary or optimal, and appropriately balanced with other science programs.

[2] The 'visit' is LSST's basic block of observing: current plan is 1 visit = 2 x 15-second images, back to back (2-second readout time between exposures, no dither between exposures in the same visit). Total visit time is then 34 seconds; 15s exposure + 1s shutter + 2sreadout + 15s exposure + 1s shutter (and the second readout occurs during the slew to the next field).

[3] The current plan is that the template image is created per filter, using images with various airmass and seeing ranges, and using images from the past year.

each other to look for extremely rapidly varying objects.

3. report diaSources via an Alert stream (like VOEvents) within 60s of taking each image. These alerts will include transients and variables and will not include linking between different visits, but known moving objects (as well as known variables) will be identified with coincident detections in the alerts. The alert information will include measurements of trailing (to identify very fast moving NEOs, even if unknown) and identification of non-stellar PSFs (to identify outbursts or cometary activity).
4. link diaSources from each visit within a night into tracklets, then between nights into tracks, and from there determine orbits -- only diaSources which can be fit to orbits with reasonable residuals are reported as movingObjects[4]. More information on this linking procedure is available in the document at bit.ly/MOPSbaselinePDF.
    a. (two diasources per night is the current requirement to make a tracklet)
5. link additional diaSources into previously known or newly discovered movingObjects to extend the orbital arc for each object (as new images are taken or as new objects are discovered).
    a. (three tracklets within 15 nights is the current requirement to make a track .. goal is to extend this window to 30 nights)
6. evaluate the completeness of the survey for moving objects by running the detection pipeline on artificial sources from a predefined source (such as the S3M, Grav et al 2011).
7. provide metadata on observing conditions, such as the telescope pointing history, seeing history, cloud conditions, and estimated 5-sigma limiting magnitudes in the difference images for each visit.
8. provide access to both the movingObject (i.e. orbits) and observations associated with that movingObject (i.e. the observations themselves, called Sources when measured images or diaSources when measured in difference images) in an online searchable database; this information will be available in Data Release[5] (DR) database versions as well as an up-to-date database (i.e. kept up to date with daily moving object processing). The up-to-date db may suffer discontinuities as DR processing is added back into up-to-date db or as observations are added or removed from linkages. More information on the quantities to be measured and stored in catalogs for each diaSource (bit.ly/MOPSdiaSource), Source (bit.ly/MOPSsource) and movingObject (bit.ly/MOPSmovingObject) can be found at the indicated links[6].
9. provide measurements of each Source with absolute astrometry accurate to better than 0.05” and relative astrometry precise to 0.01” over spatial scales of a few tens of arcminutes after Data Release processing; absolute photometry will be accurate to 10 mmag for bright (r<20) sources and relative photometry precise to 5 mmag for

[4] This is based on the PS MOPS pipeline, although there are some slight differences in how tracks are created. Tracklets assume linear motion, while tracks allow higher order motion.

[5] Data Releases occur on an annual schedule and involve reprocessing *all* data received prior to the Data Release cutoff date. This allows processing with improved photometric and astrometric pipelines, as well as improved template images.

[6] These quantities subject to revision.

observations over spatial scales[7] small relative to the visible sky in *griz* (7 mmag in *uy*) after Data Release processing.
10. provide access to postage stamps of each diaSource and/or Source for a particular movingObject as requested, as well as supply postage stamps LSST images which match a user-defined orbit or user-defined RA/Dec/time regions.
11. provide the capability to do forced photometry (or low-threshhold object detection) along a user-supplied orbit.
12. provide catalog and image query capabilities through a Science User Interface (SUI), where users can select catalog entries or images of interest and download the results.
13. provide computer resources through Data Centers for extended analysis of catalogs or images (such as coadding visits for detection of fainter objects, running clustering analysis on MBAs, etc.), where the analysis routines are written by users and interface with APIs provided by LSST. Available resources are expected to be on the order of 10% of the LSST Data Release processing capabilities.

These basic capabilities mean we will have orbital distributions for large populations of Solar System objects, which can be de-biased using either the nominal completeness function or a user-supplied 'truth' population (through the metadata on observing history), with accurate photometry in multiple bandpasses and accurate astrometry. We will have color information, although objects will be measured in different filters at different times (with a variety of times between measurements that could vary from a few minutes to many days or months). We will have lightcurve information, including upper limit data points where LSST could not detect a source, although these lightcurves will have variable time sampling. The catalogs will include measurements of the PSF of each source (and a measurement of the deviation from the stellar PSF) and we will be able to retrieve cutout images of each source to search for outbursts or cometary activity; markers of activity will also be reported in nightly alerts. We will be able to ask for forced photometry or low-SNR object detection along a user-defined orbit, in order to look for detections of objects which may have fallen below the 5-sigma diaSource threshhold. We will be able to ask for subsets of the data (images or catalogs) for further analysis outside the LSST system, and we will be able to ask for computer resources to carry out larger catalog or image analysis on machines colocated with the data.

As an estimate of how big this data will be: in the LSST Science Book we predicted sample sizes on the order of 100K NEAs, 5.5M MBAs, 280K Jupiter Trojans, and 20-40K TNOs and SDOs (Table 5.2), with between 200-350 observations per object (for bright objects) (Figure 5.4).

[7] These are approximate (my best guess) spatial scales for distinguishing where 'relative' and 'absolute' calibration will cross over.

Collaboration activities to be spelled out in the roadmap: There are some basic questions related to how LSST is producing data that we should weigh in on but there are also longer term 'roadmap' questions related to how we need to prepare to analyze the data, some of which are hinted at the paragraphs above or relate to what additional information may be needed to accomplish our science goals. Some of these tasks may be most appropriate to address as individual investigators or independent teams, while other tasks - especially if they are crucial to our overall goals - may be best served by organizing as a community to be sure they are addressed. These are tasks we need to identify and plan for.

Where do we need to specify/evaluate/add requirements on the data products provided by LSST?

1. We need to evaluate the observing cadence for Solar System studies, and determine our requirements (both basic, minimum requirements and optimal, desired requirements). By giving the project enough feedback to balance our requirements with other groups, we are most likely to get what we need in terms of observing cadence.
2. We need to evaluate what level of completeness is required for various Solar System populations, based on science goals, as this ties in with both the observing cadence requirements and completeness requirements for MOPS linkages. A draft with preliminary completeness requirements is at https://dev.lsstcorp.org/cgit/contrib/solar_system_requirements.git/plain/ssoreq.pdf; are these requirements and completeness levels appropriate?
3. We need to specify how the values defined for movingObject (bit.ly/MOPSmovingObject) will be calculated (such as albedo and pole locations) and define what the final revision of this table, as well as the diaSource and Source tables, should look like.
4. We should help define how we want or need the up-to-date database to be reconciled with the Data Release database, and what kind of consistency we need in the up-to-date db (i.e. Data Release db will be stable over time by definition, but the up-to-date db may change as objects are refined and previous observations added or removed from linkages).
5. We need to identify problems with or gaps in the data produced by or analysis capabilities provided by LSST vs what we need for doing science.

Some preliminary science-related questions to address:

1. How sufficient are the colors derived from LSST measurements:
    a. For understanding relationships between populations (Trojans & Centaurs & TNOs & SDOs & Incoming Oort Cloud Objects, for example)?
    b. For determining MBA albedos (from which to determine their sizes) & general taxonomy (and what constraints does this provide on the chemical

composition)?
    c. For determining NEO albedos (and thus sizes)?
    d. What additional tools do we need to calculate colors from measurements obtained at different times? How do we account for (or potentially benefit from) phase curve effects, which can be different in different band passes? How do we account for potential light curve effects? What kind of errors should we expect on colors, and can the range of these errors be characterized for different populations?
2. How sufficient are the lightcurves which could be derived from LSST measurements:
    a. For understanding shape or albedo distributions?
    b. For sparse lightcurve inversion for accurate shapes of specific objects?
    c. For determining pole distributions of MBAs?
    d. What tools do we need for calculating light curves for different populations given varying observing cadences? What kind of errors can we expect? What additional data might be useful?
3. What tools do we need to have to understand the debiased populations?
    a. What models are needed to link the observed (biased) populations with Solar System evolution - models of early Solar System formation are sometimes (often?) terminated before evolving over the lifetime of the Solar System so don't include later gravitational sculpting.
    b. The same (or very similar) code that MOPS uses to generate completeness estimates could be used to generate completeness estimates for user-specified populations, but could be quite time-consuming if the entire Solar System (plus the user-specified population) must be simulated. Is this necessary, or can just the user-specified population be simulated -- what effect does confusion from other linkages create?
4. What additional information (if any) is needed to translate from the observed luminosity and distance/orbit distributions to underlying size distributions?
    a. For MBAs and NEOs we know there is a distribution of albedos, but we also know there is a correlation with color. How good is this correlation and what is the final error in the size distribution (is this error appropriate for science requirements)?
    b. We know collisional families influence the calculation of the size distribution (as we usually need to know the background size distribution or the family size distribution separately). How can we separate families and background populations - what tools do we need, or what calculations must be done, and do existing tools scale to the 5.5M MBAs expected to be discovered with LSST?
    c. For TNOs, we know there is a distribution of albedos; how do we account for this when attempting to interpret the size distribution? What about binaries in the observed population (which influence the measured luminosity function)?
5. How can we identify binaries in the catalogs?
    a. Wide binaries? Will these confuse orbit fitting?

   b. Sometimes-resolved binaries? Can we find these through astrometric residuals, light curves, shape measurements?
   c. How many binaries should we expect to find in LSST?
6. How do we identify outbursts or cometary activity in the moving objects?
   a. Will the measurements provided by LSST DM be sufficient? If not, can we define additional measurements that can be routinely carried out by DM and added to the catalogs?
7. How do we understand relationships between different populations? Colors, binary fractions, volatile activity, physical characteristics can help but we will need new models as well. Are there particular observations or measurements that would help spur on this activity, and/or are there model predictions we should be testing?
8. How can we incorporate multi-wavelength measurements with LSST catalogs to determine albedos for many objects?
   a. How should this information be folded back into determining the underlying population's orbital or size distributions (i.e. if there are many 'dark' objects)?
   b. How will this information tie in with LSST's precision astrometry measurements over long time periods to help measure Yarkovsky and YORP effects?
9. What other datasets might we want to incorporate with LSST data?
   a. Additional lightcurve (or even individual observation) data so that longer term light curves or phase curves or colors can be derived? (e.g. adding PanStarrs, DES and other survey measurements to LSST measurements)
   b. Computed proper orbital elements?
   c. Family identifications?
   d. Spectral information?
      i. A database combining spectra from many MBAs (potentially from many sources) with their observed LSST colors, orbital parameters, and lightcurves (and phase curves) could perhaps identify additional physical characteristics or features common to small groups of MBAs, and would be useful for a variety of scientific investigations.

Some specific examples of things to work on short-term

1. Develop a roadmap (with dates and more detailed work plan)
2. Push on solar system simulations
   a. MAF simulations to specify requirements on cadence
   b. MOPS simulations to specify requirements on MOPS
3. Develop tools for working with large datasets - coordinate with other science collaborations, such as Transients/Variables and AstroInformatics
   a. determining light curves
   b. identifying binaries
   c. identifying family clusters
4. Work with precursor surveys / archival studies
   a. learn to work with large data sets

    b. learn to work with data sets with non-optimal / non-standard cadences
    c. learn more about the solar system - underlying theory advances -> learn more from LSST (and know where we need to learn more) & improve models feeding into LSST simulations (improve orbital models, add lightcurve simulations, color simulations, unusual objects...)
5. Develop real and/or simulated databases
    a. for lightcurve retrieval
    b. joining with data from other sources & other wavelengths
    c. data access patterns - database capabilities
    d. what information should be present for each object? forced photometry?
6. Feedback to LSST Data Management Software
    a. requirements (15s exposures vs visits)
    b. shape measurements for binary determination
    c. Alert requirements
    d. MOPS up-to-date database vs. Data Release database & their integration
    e. requirements on MOPS & linking to non-MOPS detections
7. Alternative detection algorithms to MOPS
    a. Really slow-moving objects? Detections when objects appear stationary?
8. Followup and supplemental observations
    a. observations at other wavelengths
    b. observations with different cadences (lightcurve improvements)
    c. observations of very fast moving objects
9. Train students
10. Citizen science & outreach