Rubin Observatory

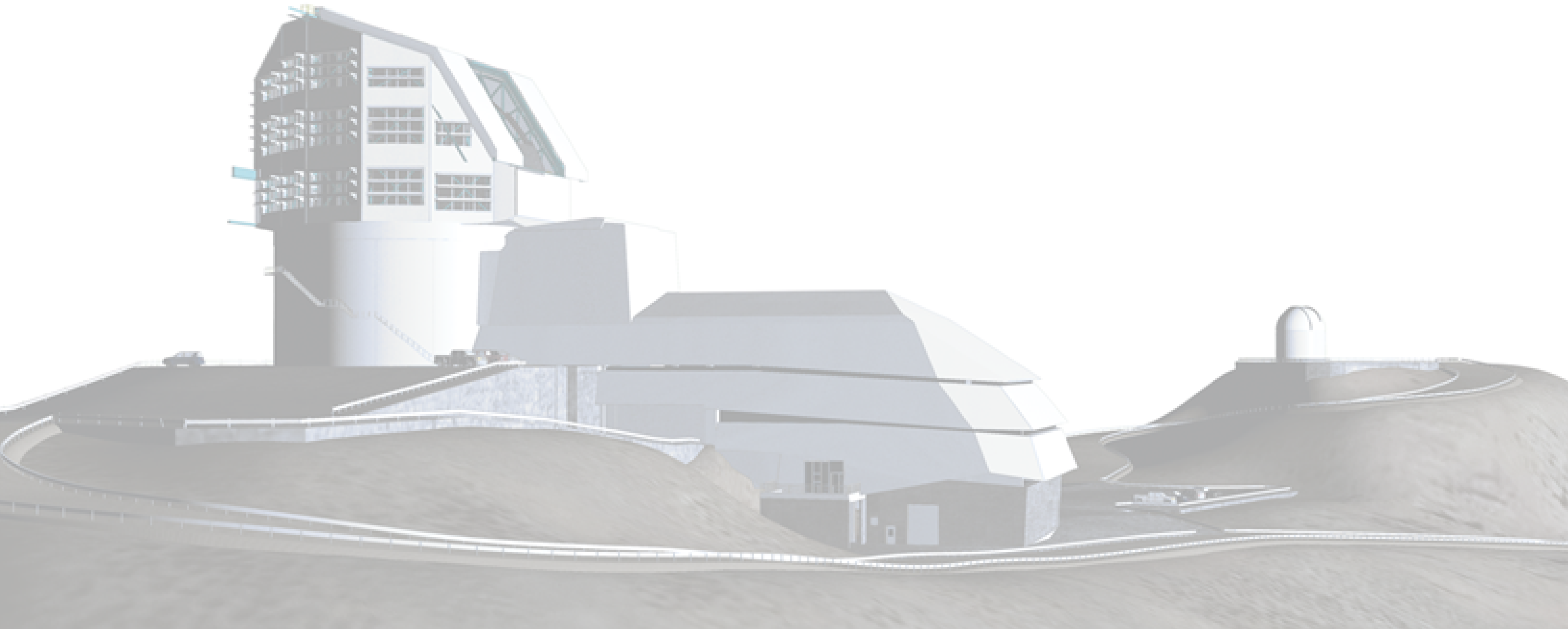
Vera C. Rubin Observatory
Data Management

Proposed Modifications to Solar System Processing and Data Products

Mario Juric, Siegfried Eggl, Joachim Moeyens and Lynne Jones

DMTN-087

Latest Revision: 2020-06-26



Rubin Observatory

Abstract

A high-level presentation of proposed changes to Solar System processing and data products baseline, to bring LSST Solar System processing closer to common asteroid-survey workflows, and help the on-schedule delivery of a scientifically useful set of Solar System data products.

Rubin Observatory

Change Record

Version	Date	Description	Owner name
1	2018-06-22	2018 Review Release.	Mario Juric
2	2019-08-06	Serving as RFC/LCR summary.	Mario Juric
2.1	2020-06-26	Editorial corrections in preparation for LCR	Mario Juric

Document curator: Mario Juric

Rubin Observatory

Contents

Rubin Observatory

Proposed Modifications to Solar System Processing and Data Products

1 Introduction

Taking the inventory of the Solar System is one of the four main science themes that LSST is designed to enable. As Ivezić et al. (2019) discuss:

> The small bodies of the Solar System, such as main belt asteroids, the Trojan populations of the giant planets and the Kuiper Belt objects, offer a unique insight into its early stages because they provide samples of the original solid materials of the solar nebula. Understanding these populations, both physically and in their number and size distribution, is a key element in testing various theories of Solar System formation and evolution.
>
> The baseline LSST cadence will result in orbital parameters for several million objects; these will be dominated by main-belt asteroids, with light curves and multi-color photometry for a substantial fraction of detected objects. The LSST sample of asteroids with accurate orbits and multi-color light curves will be 10 to 100 larger than currently available sample. LSST will make a significant contribution to the Congressional target completeness of 90% for PHAs larger than 140m, and will detect over 30,000 TNOs brighter than r $\sim$ 24.5 using its baseline cadence. LSST will be capable of detecting objects like Sedna to beyond 100 AU, thus enabling in situ exploration far beyond the edge of the Kuiper belt at $\sim$50 AU. Because most of these objects will be observed several hundred times, accurate orbital elements, colors, and variability information will also be available.

Consistent with the overall data products strategy derived from the SRD, LSST DM has identified Solar System-oriented products and processing that would be i) *broadly useful* and ii) that *LSST is uniquely position to deliver*[1].

[1] i.e., where community-led creation of such products would be infeasible due to technical constraints (such as bandwidth or computing required), or where the creation of products would require a high-degree of LSST-specific expert knowledge unlikely to be readily available in the community (such as understanding of instrumental artefacts or the details of LSST software).

Applied to the Solar System science theme, this overall guidance led us to articulate three specific data products desiderata:

1. Enable real-time (~minutes) identification of fast-moving solar system object candidates. The goal is to facilitate discovery and rapid follow-up of objects in close proximity to Earth.
2. Enable real-time attribution of new observations of known Solar System objects. This facilitates identification of uncommon (i.e., cometary) activity on otherwise known objects, but also reduces the chance an asteroid may be mistaken for a transient.
3. Identify previously unknown solar system objects (those moving on orbits around the Sun), and associate (group) together their repeated observations. This facilitates the census of the Solar System, studies of physical and dynamical properties of individual bodies, as well as studies object populations.

2 Present Baseline

The LSST data products baseline [DPDD] has been constructed to respond to the desiderata of all LSST science themes, including the Solar System desiderata enumerated in the previous section. This baseline is described in detail in the DPDD; we only give a brief summary here, concentrating on Solar System science.

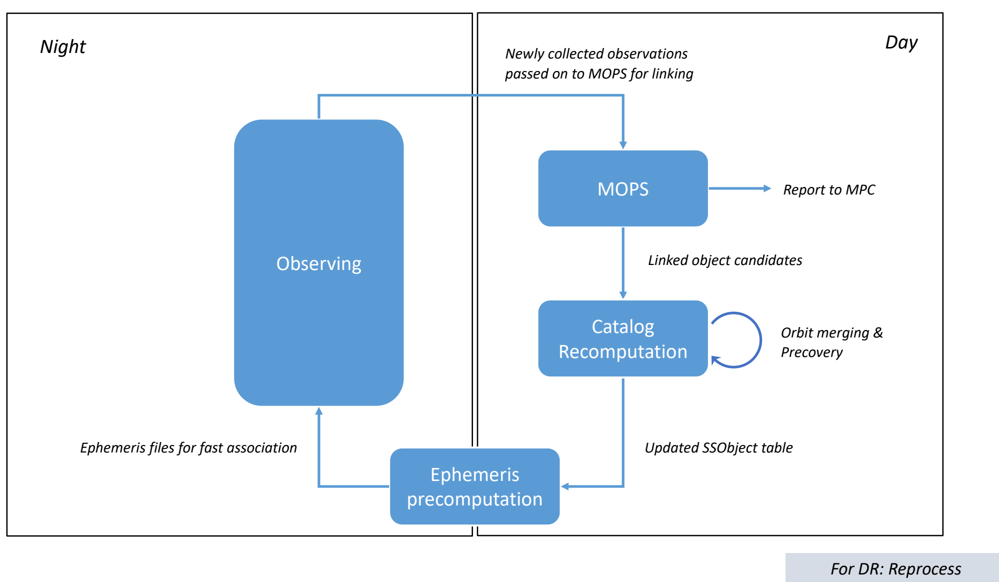
1. To satisfy the real-time identification of fast-moving candidate Solar System objects, we have added the fitting of Trailed Source Model to every DIASource detection. The resulting measurements will be transmitted in the event alert stream.
2. To satisfy the need for real-time attribution, we compute predicted positions (ephemerides) of known objects at time of each observation, and associate observations (DIASources) detected at those positions. We use an internally maintained orbit catalog (see next item), as well as an external catalog (MPCORB). The associations are also transmitted in the real-time event alert stream.
3. To identify previously unknown objects, we are developing an LSST-specific version of the MOPS pipeline [LDM-156]. Orbits are computed for all newly discovered objects, and are added to the LSST orbit catalog (the "SSObject" table). This catalog's primary

purpose is to enable attribution of subsequently (and previously) detected DIASources to LSST-discovered Solar System objects. It is also made public, on a daily basis, together with some useful pre-computed physical properties (e.g., absolute magnitudes in LSST bands).

4. Finally, to enable an accurate census of the Solar System, we repeat the MOPS-driven detection procedure and object catalog construction for each LSST Data Release. These catalogs are produced with a single, well-characterized, version of MOPS, and validated more extensively than the daily versions. The Data Release catalogs also benefit from the improved astrometry and photometry available in data releases.

In summary: objects with motions sufficient to cause trailing in a single exposure will be identified and flagged as such when the alerts are broadcast. Those that are not trailed will be identified and linked based on their motion from observation to observation, over a period of a few days. Their orbits as derived by MOPS will be published within 24 hours of identification. Better characterized object catalogs, suitable for population studies, will be computed and published with each Data Release. The daily aspect of this production is illustrated in Figure 1.

FIGURE 1: A simplified view of nightly production of Solar System related products.



3 Issues

The described data product baseline has been reviewed and adopted in 2013. Since then, a number of issues have been identified that caused us to re-examine some of the adopted solutions. We discuss them from the point of view of scientific usability (scope), the development schedule, and the development budget (cost).

3.1 Science

Though the baseline presented in previous section satisfies the desiderata for the Solar System science related products, we have identified elements that can be improved. Specifically:

- The current scheme requires associating every newly observed source (a DIASource) with two independently maintained catalogs: the internal LSST orbit catalog, and an external (Minor Planet Center; MPC) catalog[2]. A consequence is that each object would have both an LSST ID and/or a MPC designation. These designations may sometimes be in conflict: e.g., the LSST catalog may attribute two observations to two different (typically short-arc) objects, while the MPC catalog may attribute them to the same object (given it generally has access to more data than LSST). Though not insurmountable, these types of “bookkeeping” issues will make it more difficult to work with the LSST catalog.
- Presently there are no plans to cross-reference the LSST catalog to the MPC catalog. As a consequence, the end-users will have to do this type of orbit-to-orbit matching themselves, which may be rather non-trivial (especially in the case of short-arc orbits).
- Finally, tracklets that are not linked by LSST MOPS are never reported to the Minor Planet Center. Therefore, objects that could be discovered by linking LSST’s tracklet to one from another survey will be lost.

3.2 Budget

The present baseline for the development and commissioning of MOPS has budgeted for 3 FTE-years of graduate student effort, 1 FTE-yr of postdoctoral fellow effort, and 0.75 FTE-

[2] The association with an external catalog is necessary to maximize completeness, especially early on in the survey.

years of supervisory faculty effort. Having recently re-examined the the state of the software prototypes inherited from R&D, and with better modeling of development team productivity, we believe is not possible to deliver the products as presently envisioned with the resources available.

Instead, in Section 4, we will present potential modifications to the products and processing strategy. These retain the scientific value while reducing the resources needed to construct the processing software.

3.3 Schedule

The present development plan assumed full engagement on MOPS construction starting in 2017. As hiring in this area is difficult, the actual work lagged relative to the plan.

It's unlikely we could restore this schedule by adding significantly more people at this phase of the project[3]. A better strategy may be to re-examine the work to be done, with a goal of simplifying the design while retaining most (or all) of the scope (while continuing to make well-suited hires, if possible). We discuss such a proposal in Section 4.

3.4 Operational Considerations

The presently defined scope calls for daily production of an updated orbit catalog. This process is largely automated, but it has been known to occasionally require manual intervention by a domain-expert (curation). This is especially true when it comes to incorporation into the orbit database of numerous newly discovered short-arc objects. The Minor Planet Center dedicates approximately ∼2 FTE to maintenance of a comparable catalog.

At present, plans for LSST operations do not explicitly include staff with Solar System / orbital dynamics expertise to fulfill this type of role. This carries a significant operational risk, should the orbit catalog maintenance code not be sufficiently autonomous.

[3]This is commonly known as Brooks' Law: "adding human resources to a late software project makes it later" (Brooks, 1982)

4 Overview of Proposed Changes

The issues described in the previous section point to serious obstacles to an on-schedule delivery of the LSST Solar System Data Products, as presently envisioned. We have therefore holistically re-examined the deliverables in this area, looking for redesign opportunities to allow for deliver close to desired schedule, with a minimum impact to the budget.

4.1 Processing Workflow

Before looking at LSST, it is instructive to sketch out how analogous present-day asteroid surveys operate. We take the example of a typical night in PanSTARRS: i) observations are taken, ii) known objects are flagged by comparing detections to external catalogs (MPCORB), iii) tracklets are constructed by PS1-MOPS, iv) likely asteroids are submitted to the Minor Planet Center (MPC), v) the Minor Planet Center confirms the detections and (where possible) computes and publishes orbits in time for the next night of observing. Comparing this workflow to Figure 1, we see that this closely follows the model adopted by LSST. The major difference is that all the steps in the LSST processing are performed in-house, whereas PS1 relies on the Minor Planet Center.

The Minor Planet Center[4] (MPC) operates at the Smithsonian Astrophysical Observatory (SAO), under the auspices of Division F of the International Astronomical Union (IAU). The MPC is responsible for the designation of minor bodies in the solar system: minor planets; comets; and natural satellites. They are also responsible for the efficient collection, computation, checking and dissemination of astrometric observations and orbits for minor planets and comets, via its various journals. The MPC was founded at the University of Cincinnati in 1947, and has been operating at the SAO since 1978. Reporting small body discoveries to the MPC, who construct and curate the resulting orbit catalogs, has been an established community practice at least since the ~1970s.

The rationale for LSST not utilizing MPC services are largely historical. When LSST was originally envisioned, the Minor Planet Center infrastructure was not going to scale to the sheer volume of data LSST was predicted to produce[5]. Furthermore, keeping full control over all aspects of the processing minimized operational risk.

[4] `https://www.minorplanetcenter.net/iau/mpc.html`
[5] LSST was originally designed to begin operating in early 2010s.

The LSST start date has shifted by a ~decade since these original designs. In this time, the predicted data volumes for LSST have stayed the same, while the generally available computing capabilities have increased. The MPC has demonstrated the ability to scale with the growth of the community (e.g., it successfully handles PanSTARRS PS1 data). Today, the MPC routinely manages a database of $\sim 750,000$ orbits, only a factor of ~ 7 relative to what is expected for end-of-mission LSST.

We've therefore examined the elements in Figure 1 with the aim of bringing LSST Solar System processing closer to present day practices that utilize Minor Planet Center services. There are two major sources of Solar System-specific pipelines (in terms of construction and operations effort):

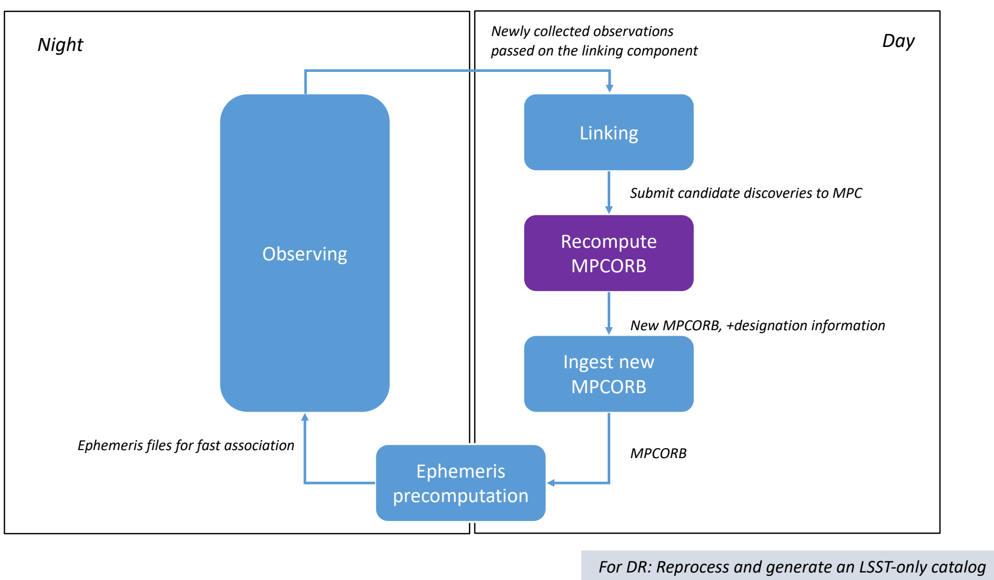
1. Delivering the MOPS pipeline to link tracklets into high-probability candidate tracks
2. Delivering the pipeline to construct, manage, and maintain (in operations) the resulting orbit catalog.

The specific method that LSST adopted for discovery and linking of solar system objects requires a MOPS pipeline with a computationally intensive "link tracklets" step. The novelty of the method, and the sheer computational intensity, argue for continued in-house development of this step.

The second step – construction and maintenance of the orbit catalog – is an example of "bread and butter" operation that the MPC specializes in. If LSST delivered a list of already linked tracklets, the MPC has the software, staff, processes, and capabilities to ingest them, verify the linkage, add them to the orbit database, and perform all the typical curation activities. Assuming MPC services are used for this step, the resulting daily cycle is presented in Figure 2.

By switching to the scheme presented in Figure 2, the newly discovered high-confidence tracks would now be submitted to the MPC for inclusion into the central orbit database they maintain and regularly curate. Following incorporation of nightly data received from LSST and other surveys, the MPC would publish an updated orbit catalog in time for the next night of LSST observing. LSST would download the published catalog, and use it for attribution of known objects in the following night.

FIGURE 2: A simplified view of nightly production of Solar System related products, with orbit catalog generation and curation performed by the Minor Planet Center. A detailed workflow can be found in Figure 6 of LDM-151.

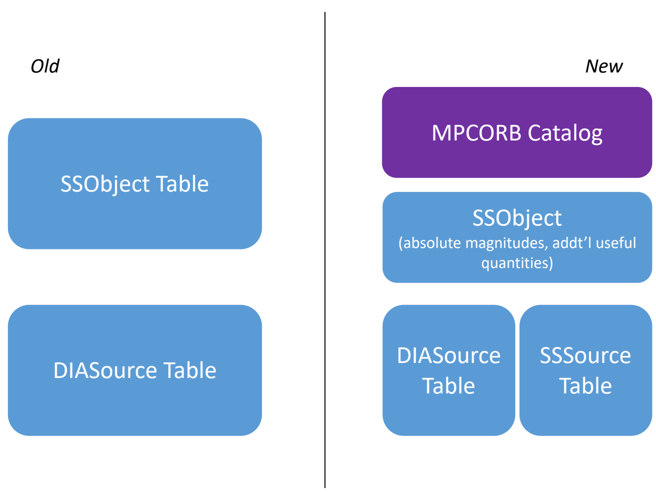


4.2 Data Products

From the data products point of view, the change can be viewed as having an SSObject table divided into two parts: dynamical and physical characteristics. The dynamical characteristics (orbital elements) would now be computed by the Minor Planet Center (the "MPCORB" catalog); LSST would ingest a new copy of the MPCORB every evening. We note that this means that **the orbits in the ingested catalog will be computed using not only LSST observations, but also other observations the MPC has access to**. The biases, selection function, and the quality of orbits in the catalog will therefore be complicated, making this catalog unsuitable for population studies. The DRP catalogs (see below) should be used for this purpose.

Given that information, the physical characteristics such as absolute magnitudes would still be computed by LSST (also on a nightly basis, forming tne updated SSObject table). This is schematically illustrated in Figure 3.

FIGURE 3: Comparison of present data products baseline, and the proposed changes. See text for details.



To facilitate the determination of biases LSST will, as a part of DRP, compute an orbit catalog and the table of physical properties using LSST observations only and utilizing one, well-characterized, version of the Solar System Processing pipeline. This products is designed to enable population studies based on objects discovered by LSST.

4.3 Linking Algorithm

A new tracklet linking algorithm, HelioLINC, has recently been published and demonstrated by Holman et al. (2018). This algorithm brings significant (>10x) computational improvements over the present baseline, as well as being conceptually simpler.

We have analyzed its linking performance. We find that present day implementation (`https://github.com/pytrax/pytrax`) is already able to link $> 90\%$ of non-NEO objects and $> 70\%$ of NEOs in a simulate 3-month LSST survey. Technical analysis of possible improvements indicates that with additional development (budgeted in this change request) this codebase can be updated to meet LSST requirements (95% linking efficiency). The computational performance is approximately 10x over the present LSST MOPS prototypes.

Given the new algorithm's simplicity, increased computational performance, and the ability to share code and know-how with the Minor Planet Center (the original developers), we propose to switch to HelioLINC as the baseline linking algorithm.

5 Impact Assessment

5.1 Science

Our analysis and feedback from the LSST Solar System Science Collaboration indicates this change will have an overall positive impact. It will lead to improved usability, asteroid characterization, and enhanced discovery yields. Specific benefits include:

- As MPCORB includes not only LSST information, but other surveys as well, the orbit catalog used for association is now maximally complete at all times. This also automatically satisfies DMS-REQ-0288 ("Use of External Orbit Catalogs").
- There is now only one catalog now (a opposed to having an LSST-specific orbit catalog). This solves the non-trivial cross-match problem, simplifies bookkeeping, and reduces community confusion.
- For the community that already works with the MPC databases, there will be no need to do anything special to take advantage of LSST data (as it will already be included). This makes the LSST measurements significantly more accessible and useful.

Rubin Observatory

- This change places LSST into a more general (and already largely existing) framework of how asteroid surveys operate world-wide
- The change would open the possibility to submit all tracklets to the MPC, including trails
- The change would enable future cross-survey linking either at MPC or by a third party. For example, LSST's first tracklet may complete a track that some other survey has started nights before. This increases the overall object discovery yields.

On the negative side, not having an LSST-only nightly catalog diminishes its value for precision population studies. However, the data release catalogs are arguably better (and specifically designed) for that use case.

5.2 Budget and Schedule

We estimate that with this change the construction scope of prompt Solar System processing can be executed with the presently hired staff in time for the beginning of operations. We estimate the DRP elements will be possible to construct in the same time-frame *assuming no significant changes are needed to prompt processing elements as a result of feedback from commissioning.*

A secondary positive effect is that MOPS remains the most complex Solar System-specific pipeline to be developed, in addition to the Solar System-specific spatially indexed internal database. The work required on MOPS matches closely the expertise available on-project. Utilization of the next-generation HelioLINC algorithm and its `pytrax` implementation reduces the computational risk further.

5.3 Operational Impacts

Operationally, this change significantly reduces the daily operational complexity, as all catalog maintenance now resides at the Minor Planet Center. It brings the remaining tasks (characterization of Data Release catalogs) closer to what has been budgeted for operations. It will still requires some operational expertize in the area of the Solar System, but at a reduced FTE level.

5.4 Risks

The introduction of a partner into daily workflow brings along a degree of third-party operational risk. We've identified two specific (and correlated) risks: i) present management and technology evolution challenges at the MPC, and ii) the MPC scaling up approximately 10x to be able to intake the daily LSST volume.

The first risk is related to MPCs migration from legacy VMS systems and largely manual processes, to UNIX systems with a high degree of process automation. The MPC is also increasing their staff by about $\sim 2x$ and professionalizing internal business processes along the way. This has not been without challenges. The migration has taken longer than expected and was executed earlier than optimal. On June 10th 2019., a disk failure on a machine that handles several observation processing pipelines caused suspension of normal operations for approximately six days, and forced a somewhat premature switch to the new Linux based system. The switch caused a number of subsequent smaller issues, the major of which were corrected within a week, with a long tail of smaller problems persisting over the subsequent three months[6]. An event like this in LSST operation would not be disastrous, but would generate additional LSST staff workload to implement mitigation procedures and monitor for subsequent issues. We've been continuously monitoring the MPC's response and lessons learned, and have offered advice and assistance in areas where the LSST SSP's expertise could help.

The work on scaling up the MPC systems to LSST data rates has been partially delayed due to the extended Linux migration and handling of the disk crash crisis. Nevertheless, there has been progress. For example, the MPC has demonstrated code capable of orbit computations on $O(1M)$ objects with just a few -hours of computational effort, albeit with a still too high rate of cases requiring manual intervention. Furthermore, the database migration (which should allow for larger workloads) is expected to complete by March 2020. Overall, the work will continue through 2020-2021 in order to be ready for LSST in 2022. We plan to monitor the MPC developments closely, and continuously evaluate the level of this risk.

Nevertheless, if the analysis presented in Section 3.3 is correct the schedule risk within the present baseline is already quite serious. The change to incorporate the MPC into our baseline – notwithstanding the issues discussed above – would still reduce it.

[6] https://minorplanetcenter.net/iau/MPCStatus.html

6 New and Changed Documents

6.1 LSE-163: Data Products Definition Document

The DPDD has been updated to:

- Reflect the new workflow where candidate discoveries are sent to the Minor Planet Center, and the orbit catalog is now downloaded from the Minor Planet Center on a daily basis.
- Deprecate the usage of the term "MOPS" (Moving Object Processing System), in favor of the more accurate and less confusing term Solar System Processing. In the community, MOPS is widely understood to be one specific implementation of the asteroid linking algorithm, which we will stop using after this change request is adopted. Secondly, linking (which MOPS refers to) is a subset of the overall processing required for Solar System objects.
- State that the Data Release products will consist of two elements: the reprocessed prompt processing data products, but also a special Solar System object catalog constructed from LSST observations only (to enable population debiasing and studies).
- Update and increase the level of detail provided by the high-level table schemas for Solar System data products. These LaTeX versions are now auto-generated from the human-readable spreadsheets maintained at `http://ls.st/j7f`.

The table schemas in the DPDD have been updated consistent with Figure 3 to explicitly call for three tables related to Solar System objects: the MPCORB table (with orbits received from the MPC), the SSObject table (with physical quantities computed from LSST objservations), and SSSource tables (with per-observation quantities).

6.2 Database Schemas

An easily maintainable human-friendly versions of the new schema is maintained at `http://ls.st/j7f`. Based on this spreadsheet, we have developed the updated database schema in the Felis/YAML format (delivered as attachment to `http://ls.st/rfc-620`), as well as generated the LaTeX tables in the DPDD.

The DPDD representation, followed by the Felis/YAML representation (which contains additional detail) are considered canonical (i.e., defining the baseline).

6.3 LSE-61: DM Systems Requirements

This change request **does not change the fundamental DMSR-level requirements**, but rather addresses their implementation. All changes to the DMSR are of terminological nature (changing MOPS and related terms to Solar System Processing and analogous terms) or clarifications.

The requirements being modified are:

- DMS-REQ-0273: SSObject Catalog
- DMS-REQ-0323: Calculating SSObject Parameters
- DMS-REQ-0089: Solar System Objects Available Within Specified Time
- DMS-REQ-0185: Archive Center

6.4 LDM-151: Science Pipelines Design

LDM-151 has been updated to:

- Switch to using "Solar System Processing", rather than MOPS
- Reflect the incorporation of the Minor Planet Center into the daily workflow
- Switch to HelioLINC as the baseline linking algorithm
- Define the new pipeline tasks
- Provide the inputs, outputs, and specification of individual pipeline tasks
- Clarify that the DRP Solar System data products involve reprocessing of the prompt products, but also generation of a separate catalog constructed out of LSST observations only.
- Specify failure modes and recovery options for individual pipeline component

- Specify the interface to the Database (PPDB and/or L2DB) will largely be file based, with the SSP maintaining an internal spatially indexed database.

6.5 LDM-156: Moving Object Pipeline System Design

Having adopted HelioLINC as the algorithm, this document is to be taken out of the official baseline and kept for historical documentation. It is superseded by Section 3.5 of LDM-151 and the Holman et al. (2018) paper.

6.6 Annual Milestones

The implementation of this change request will incorporate the milestones enumerated below into LSST's integrated PMCS. The plans will be refined on a cycle-by-cycle basis, consistent with usual LSST-DM practice. A more detailed notional schedule is kept locally in a PMCS tool.

Date	Milestone
Dec 31st, 2019	Moving object linking (based on Heliolinc), tested on LSST simulation.
Dec 31st, 2020	Moving object linking (based on Heliolinc), running at design spec, tested on LSST simulation.
	MPC integration (Submission of discoveries to MPC; Retrieval of updated MPCORB – the Minor Planet Center report submission pipelines).
	Ephemeris cache precomputation pipeline delivered
	Attribution service delivered
Dec 31st, 2021	Daily Data Products Pipeline delivered.
	Prompt Processing pipeline integration
Sep 31st, 2022	DRP-specific pipeline elements delivered
	Regular processing of commissioning data
	Final construction version of Solar System Pipelines

The presented re-baseline has been developed under the following high-level guidance: keep the cost fixed (at approximately 1.5 FTE/yr through the end of construction), keep the schedule for final delivery fixed (end of construction), and attempt to minimally affect the scope. The resulting plan and schedule should therefore be considered success oriented and with no built-in margins. We are particularly exposed to external third party risks (the MPC), but also internal integration risks (e.g., the delivery of the attribution service at the end of 2020 depends on the readiness of the prompt processing system to receive it). We caution that any unforeseen issues and/or delays are likely to result in a slip of DRP-specific functionality into YR1 of operations.

6.7 Supplementary Documentation

A Memorandum of Understanding has been developed by the Minor Planet Center and LSST wherein both parties commit to working to realize the smooth operation of the LSST Solar System Processing as described in Section 4.1.

References

Brooks, F.P., 1982, *The Mythical Man-Month: Essays on Software Engineering*, ADS Link

Holman, M.J., Payne, M.J., Blankley, P., Janssen, R., Kuindersma, S., 2018, AJ, 156, 135, doi:10.3847/1538-3881/aad69a, ADS Link

[LPM-17], Ivezić, Ž., The LSST Science Collaboration, 2018, *LSST Science Requirements Document*, LPM-17, URL `https://ls.st/LPM-17`

Ivezić, Ž., Kahn, S.M., Tyson, J.A., et al., 2019, ApJ, 873, 111 (`arXiv:0805.2366`), doi:10.3847/1538-4357/ab042c, ADS Link

[LSE-163], Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL `https://ls.st/LSE-163`

[LDM-156], Myers, J., Jones, L., Axelrod, T., 2013, *Moving Object Pipeline System Design*, LDM-156, URL `https://ls.st/LDM-156`