LROC PDS TOOLCHAIN REDEVELOPMENT AND FASTER RELEASE SCHEDULE. N. M. Estes, K. S. Bowley, J. Barnett, and the LROC team. School of Earth and Space Exploration, Arizona State University, nme@ser.asu.edu

Introduction: The Lunar Reconnaissance Orbiter Camera (LROC) Science Operations Center (SOC) processes and releases to the Planetary Data System (PDS) all LROC observations. These data include Narrow Angle Camera (NAC) and Wide Angle Camera (WAC) Experiment Data Records (EDRs), Calibrated Data Records (CDRs), and Reduced Data Records (RDRs). Up to 450 Gib/day are downlinked from LROC, and this data volume grows to approximately 7 times its original size after processing into EDR and CDR products, resulting in a nominal quarterly release volume containing >300,000 files at a size of ~20 TiB (Table 1). Producing an LROC PDS volume requires preparing and populating the filesystem directory tree as well as updating multiple databases (Figure 1). The original collection of scripts and programs (the toolchain) worked well for the first eight years of operations; however, there were weaknesses in the toolchain that lead the LROC SOC team to redevelop the entire process.

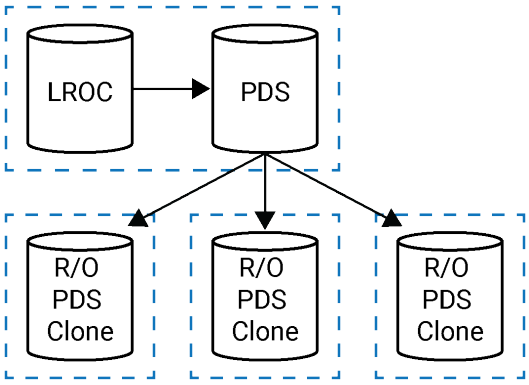

Figure 1: The LROC database holds all operations information including regions of interest, targets, observation records, file catalog, and all processed product data. This database is the source of PDS data which gets inserted into the PDS database when publishing. The master PDS database is then replicated to three read-only clones of the public PDS data to support all public web services using this dataset. The original process handled each database individually and took several hours from filesystem publishing to the completion of the database publishing. The new process prepares the data in all databases simultaneously and is made live all together at the same time as publishing in the file system (<1 second lag time).

Improvements made: The original process handled the filesystem and database components of the release separately with a delay of several hours between filesystem publication and adding/updating that data in all affected databases. Combining those two operations resulted in a more reliable release process, and is more easily recovered in the event of a problem. With the original process, re-releasing reprocessed data required extensive manual effort; the new process completely automates the release, regardless of the state of data (published/unchanged, unpublished, or published/reprocessed). While reliable, if any problem occurred during processing, the original process often required manual intervention to recover the filesystem and database to a consistent state. The new toolchain avoids undefined intermediate states in the event of any problem by approaching the release process in a different way.

Approach: To solve these weaknesses in the original system, the new LROC PDS toolchain combined all filesystem and database operations into a single process. This process was split into two parts: preparation and publishing. In an ideal world, the filesystem and multiple databases would all be published as a single atomic operation (in programming, 'atomic operation' refers to an action that either entirely completes or not with no possible intermediate state). The new process gets remarkably close to this ideal using prepared transactions in the databases and staging files on the PDS data node

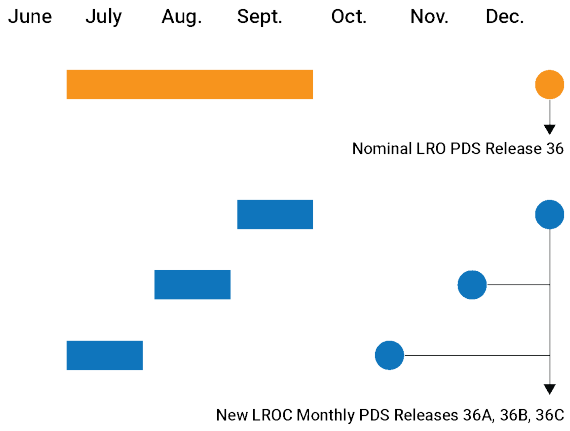

Figure 2: Figure showing how the nominal volume 36 release includes data between three and six months old; however, by splitting into monthly releases, LROC monthly release data is only three to four months old in addition to only being a third the size.

filesystem in a private location where the whole volume can be moved into the public location with a single filesystem hard link. Prepared database transactions wrap any number of database updates into a single unit and perform all necessary steps to guarantee the data can be committed, preserving that state indefinitely until the prepared transaction is either committed or rolled back. Using these two strategies, the final publish step of the new toolchain is composed of a set of "COMMIT PREPARED" statements in the affected databases and a single filesystem move command. All of these operations are individually atomic and are nearly 100% reliable, which minimizes the potential for failure at the final publishing step. Where the old process involved an hour-long gap between filesystem and database changes, the new approach allows both portions to be complete in <1 second. Each step through the new toolchain performs checks to ensure the resulting published volume is consistent between the PDS index files, PDS database clones, and LROC database. If for any reason a volume should not be published, or if a problem occurs while preparing a volume, the isolated staging area, in combination with the use of prepared transactions, allows a volume in progress to revert back to the original state with 100% reliability, regardless of what state it was in when the process failed or was aborted.

Results: Over the eight years of LROC operations, there have been a small number of events resulting in a portion of old data being reprocessed. Due to the manual nature of reprocessing old volumes to incorporate this small number of updated products, these products had not been re-released. Starting in March 2018, the new toolchain was used to start reprocessing those volumes to make the latest processed data available to the entire community. This effort was completed in September 2018, at which time the LROC SOC switched from quarterly releases to monthly releases. The reduced size of the monthly volumes combined with the efficiencies gained with the new toolchain have resulted in an easier to manage process that delivers data to the planetary community more frequently (Figure 2).

Volume	PDS Product Count	Total File Count	PDS Data Size	Total Data Size
1	230,518	655,715	39 TB	44 TB
2	102,412	307,254	17 TB	19 TB
3	136,924	410,790	24 TB	27 TB
4	129,750	389,268	22 TB	25 TB
5	139,080	417,258	28 TB	32 TB
6	133,027	399,099	27 TB	30 TB
7	153,690	461,088	29 TB	33 TB
8	166,136	498,426	29 TB	32 TB
9	164,514	493,560	29 TB	33 TB
10	146,722	440,184	27 TB	30 TB
11	155,444	466,350	28 TB	31 TB
12	152,262	456,804	27 TB	30 TB
13	150,556	451,686	27 TB	30 TB
14	141,638	424,932	26 TB	29 TB
15	139,628	418,902	27 TB	30 TB
16	138,344	415,050	27 TB	30 TB
17	128,731	386,211	24 TB	27 TB
18	139,162	417,504	25 TB	28 TB
19	131,976	395,946	25 TB	28 TB
20	141,526	424,596	26 TB	29 TB
21	135,076	405,246	26 TB	29 TB
22	135,912	407,754	25 TB	28 TB
23	137,200	411,618	27 TB	30 TB
24	135,696	407,106	26 TB	30 TB
25	126,748	380,262	25 TB	28 TB
26	128,680	386,058	25 TB	27 TB
27	125,284	375,870	25 TB	28 TB
28	119,682	359,064	22 TB	25 TB
29	119,626	358,896	23 TB	26 TB
30	109,476	328,446	21 TB	24 TB
31	119,356	358,086	23 TB	26 TB
32	118,802	356,424	23 TB	26 TB
33	117,836	353,664	23 TB	26 TB
34	109,860	329,598	22 TB	25 TB
35	121,130	363,408	24 TB	27 TB
36A	41,014	123,060	8 TB	8 TB
36B	38,158	114,492	8 TB	8 TB
36C	41,224	123,690	8 TB	9 TB
Total	**4,902,800**	**14,673,365**	**916 TB**	**1,027 TB**

Table 1: Volume file counts and sizes for all released LROC EDR and CDR PDS volumes through the end of 2018. The reduced size of the new monthly volumes helps make the large volumes more manageable and makes data available to the community more frequently.