

LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST) Interface Requirements between Data Management and EPO

Suzanne Jacoby, Brian Selvy, Ben Emmons, Tim Jenness

LSE-131

Latest Revision: July 11, 2018

DRAFT

This LSST document has been approved as a Content-Controlled Document. Its contents are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval. If this document is changed or superseded, the new document will retain the Handle designation shown above. The control is on the most recent digital document with this Handle in the LSST digital archive and not printed versions.



Change Record

Version	Date	Description	Owner name
1	2011-01-02	Initial Version	Suzanne Jacoby
2	2012-05-22	Elevated to controlled document handle	Gregory Dubois- Felsmann
3	2013-09-23	Substantial update	S. Jacoby & Brian Selvy
3.1	2013-10-01	Updated "Annual DM Transfer to EPO" requirement to specify maximum data products transferred in terms of terabytes instead of percentages. Updated the Discussion blocks for the Annual and Nightly transfer requirements.	S. Jacoby, T. Axelrod, K-T Lim, & Brian Selvy
3.2	2013-10-03	Per the recommendations of the CCB held on 10/2/13, the following changes were made: moved EPO-REQ-0009 to DM-EPO ICD (LSE-131) as EP-DM-CON-ICD-0014; moved EPO-REQ-0013 to EM-EPO ICD (LSE-131) as EP-DM-CON-ICD- 0013). Per the email exchange conducted via the LSST-Change Control Board listserve, the following additional change has been made: Moved EPO-REQ-0100 to LSE-131 as EP-DM-CON- ICD-0019.	Brian Selvy
	2013-10-04	Implementation of LCR-147 & above (10/3) changes	Robert McKercher
3.3	2016-07-22	Clarify requirement 2.2 to add a statement specifying there shall be a separate user for each distinct EPO service requiring query access; add requirement 2.3.	K-T Lim
4	2017-07-24	Implementation of LCR-795. Reformatting due to software transition.	K. Wesson, B. Emmons, T. Jenness
5	2018-05-03	Phase 3 updates from EPO JTM 2018 and Citizen	S. Kahn, W. O'Mullane, A.

DRAFT Not Yet Approved - The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval.



		Science Workshop 2018	Bauer, B.
			Emmons, M.
			Butler, C. Banek,
			K-T Lim, S.
			Krughoff, F.
			Mueller, K. Lo, G.
			Dubois-Felsmann,
			E. Bellm, T.
			Rector, C. Lintott
5.1	2018-06-29	Edits based on CCB feedback	B. Emmons, G.
			Dubois-Felsmann
5.2	2018-07-10	Minor ISO date formatting	B Emmons
5.2	2010 07 10		D. Lininons
5.3	2018-07-11	CCB feedback	G. Dubois-Felsmann, K-T Lim

DRAFT Not Yet Approved - The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval.



Table of Contents

Char	nge Record	.2
1	DM Transfer of Catalog Data to EPO	.5
2	DM Generation of a Color Hierarchical Progressive Survey for EPO	.6
3	Sample Collection of UGRIZY+Panchromatic Co-Add FITS Images	.6
4	Catalog Format	.7
5	Citizen Science Data	.7
6	DM Services	.7
7	Data Rights Protection	.8
8	DM to EPO Data Transfer Cadence	.8
9	EPO is an Authorized Science User	.8
10	EPO Quota Management	.8
11	EPO World Public Data Subset	.8
12	No Regulatory Issues from EPO	.9
13	EPO Compute Cluster in the US DAC	.9



LSE-131: Interface Requirements Between Data Management and EPO

Purpose

This ICD is to be used for requirements that support development of final designs of hardware, software, and interface protocols on both sides of the Data Management and EPO subsystem interface. Further updates to this document expected prior to the finalizing of component designs, with all changes subject to the LSST change control and approval processes.

Introduction and Scope

The EPO system provides a public subset of LSST data (notionally referred to as "10%") to nonspecialists, including: the general public, educators/students, citizen scientists, and the informal science center community. Non-targeted audiences, such as amateur astronomers, would also benefit. To build EPO data products for these audiences, the cloud-based EPO Data Center (EDC) receives data products from the U.S. Data Access Center (DAC) at various frequencies, such as a "trickle feed" of DM products as they're processed throughout the year, periodic batches of vetted citizen science subject sets, and on-demand Science Platform queries. The data flow is one-way (DAC \rightarrow EDC) via EPO as an authorized science user of the DAC or initiated by a Citizen Science Principal Investigator (PI) with LSST data rights.

Related Documents:

EPO Subsystem Requirements Document: LSE-89

1 DM Transfer of Catalog Data to EPO

ID: EP-DM-CON-ICD-0004

Specification: As it becomes available, Data Management shall transfer to EPO a subset of catalog data as defined in the table below.

Discussion: Definition of the queries to satisfy the subset requirement will be defined by EPO once ComCam data is available. The historical notion of a single bulk data transfer from DM to EPO just prior to each data release had a number of disadvantages and risks. DM and EPO were both in favor of changing this to a "trickle feed" where data was sent in small amounts over a longer period of time as part of the *solar system object processing* and *data release processing* pipelines. This approach reduces bandwidth spikes, allows for errors/problems to be identified sooner, and gives DM and EPO greater flexibility. DM is permitted to apply additional restrictions in order to stay within the mandatory bounds, but should notify EPO in advance. The values specified in this requirement represent the maximum size of the data release products to be transferred from DM to EPO during the last data transfer that will occur during the ten year survey (DR11), representing the bounding case.

DRAFT Not Yet Approved - The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval.



Product	Table	Columns	Not to Exceed (compressed)
Prompt	SSObject	all	33 gigabytes
Data Release	Object	 objectId bdFluxB bdFluxD bdEllip bdReB bdReD psChi2 psCov psFlux psLnL psNdata psRadec psRadecTai 	10 terabytes
Data Release	ForcedSource	all	10 terabytes

2 DM Generation of a Color Hierarchical Progressive Survey for EPO

ID: EP-DM-CON-ICD-0021

Specification: As part of their co-add image processing pipeline, DM shall create a Hierarchical Progressive Survey (HiPS) for EPO in the form of color JPEG HEALPix tiles limited to 1 arcsecond resolution.

Discussion: These tiles will be used in the EPO Portal Skyviewer (powered by a sky atlas tool, such as Aladin Lite) and our JupyterLab-based educational investigations (powered by a compatible astronomy image viewer extension, such as pyaladin). The EPO Skyviewer will have the same sky coverage as the full survey (southern hemisphere), but the maximum zoom level may be different for different regions of the sky. For some deep drilling fields, we may have additional zoom levels to see extra detail. Definition of EPO's color scheme, definition of the varying depth coverage, and the method for transferring the tiles to the EDC will be defined by EPO once ComCam data is available. EPO may later choose PNG if the JPEG user experience is not satisfactory. Fees related to this deliverable will be paid by EPO but we hope to leverage cost efficiencies by inserting our specific output as part of existing NCSA data processing workflows. Note: these images do not count toward the EPO world public data subset quota because the scientific data is scrubbed and the image format is not FITS-like.

DRAFT Not Yet Approved - The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval.



3 Catalog Format

ID: EP-DM-CON-ICD-0009

Specification: The Data Management System shall deliver catalog data to EPO preferably in Apache Parquet format.

Discussion: Apache Parquet is a promising, emerging data format. More testing will be required to verify it meets all of EPO's needs. As a fallback option, EPO could accept catalog data in Apache Avro or the sub-optimal but universally-supported CSV format.

4 Citizen Science Data

ID: EP-DM-CON-ICD-0034

Specification: EPO shall lead the development of DM stack community modules that provide data processing capabilities needed for citizen science projects as well as a data transfer mechanism and data rights review workflow.

Discussion: Community modules would include: "Color Mixer", "Metadata Scrubber", and "FITS to TIFF/PNG/JPEG Converter". The data transfer mechanism (notionally referred to as a "data funnel") will transfer data from the DAC to a protected S3 API-compliant object storage bucket. A data rights panel will be established to verify proper protocol is followed. EPO will partner with Zooniverse in this development effort. More detail can be found here: https://confluence.lsstcorp.org/display/EPO/Citizen+Science

5 DM Software & Services

ID: EP-DM-CON-ICD-####

Specification: DM shall provide the following software, which is expected to be used in development of the EPO systems: Butler, Supertask.

Discussion: The purpose is as follows: Butler for accessing formal DM data products, Supertask for grouping and running related tasks.

ID: EP-DM-CON-ICD-####

Specification: DM shall ensure that the following services are made available to the EPO systems: Image Cutout Service, Metaserv, Mini-Broker, TAP, and ObsTAP.

Discussion: The purpose is as follows: Image Cutout Service for obtaining single-band co-add images, Metaserv for querying the database, Mini-Broker for applying EPO-provided filters to obtain variable star classifications, and TAP/ObsTAP for ad-hoc queries. In addition to the EPO-provided filters, we also hope to leverage the simple filters mentioned in the DPDD: "we will provide a limited number of predefined filters for a small number of object types of common interest. These will answer non-exclusive questions such as 'is the light curve consistent with an RR Lyra?"

DRAFT Not Yet Approved - The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval.



6 Data Rights Protection

ID: EP-DM-CON-ICD-0031

Specification: EPO shall not provide products, interfaces, or services that could allow users without data rights to query, access, or otherwise interact with an LSST Data Access Center (DAC).

Discussion: See https://jira.lsstcorp.org/browse/LIT-97 for further elaboration.

7 DM to EPO Data Transfer Cadence

ID: EP-DM-CON-ICD-0019

Specification: The cloud-based EPO Data Center (EDC) shall receive data products from the U.S. Data Access Center (DAC) at various frequencies.

Discussion: Such as: a "trickle feed" of DM products as they're processed throughout the year, periodic batches of vetted citizen science subject sets, and on-demand Science Platform queries.

8 EPO is an Authorized Science User

ID: EP-DM-CON-ICD-0002

Specification: DM shall provide to EPO a single DAC account with the same permissions as an authorized science user.

Discussion: This account will enable EPO to access the Science Platform, make queries, process data within our Kubernetes cluster, transfer data from the DAC to our EDC, etc. Note: data transfer related to citizen science by a PI will be logged as that individual.

9 EPO Quota Management

ID: EP-DM-CON-ICD-0033

Specification: EPO shall be responsible to ensure EPO data usage falls within the quota terms outlined in this document.

Discussion: DM/NCSA will not need to programmatically restrict data usage by the EPO account accessing the DAC but can at their discretion monitor usage. The onus is on EPO to conform to the quota agreements stated herein. There may be situations where the default scientist account quota is too low for allowed EPO usage and, upon mutual agreement, an exception will need to be implemented by DM/NCSA.

10 EPO World Public Data Subset

ID: EP-DM-CON-ICD-0032

Specification: EPO shall be able to use and distribute its data subset publicly, without access restrictions, data rights control, or tracking required.

DRAFT Not Yet Approved - The contents of this document are subject to configuration control and may not be changed, altered, or their provisions waived without prior approval.



Discussion: In short, all EPO data is world public.

11 No Regulatory Issues from EPO

ID: EP-DM-CON-ICD-0020

Specification: EPO shall ensure that the DM system (particularly the DAC) will never need to be concerned with any regulatory issues coming from EPO or its users.

Discussion: Possible sources of regulation include the Children's Online Privacy Protection Act (COPPA) and the Family Educational Rights and Privacy Act (FERPA), among others. It is expected that EPO will meet this requirement by not passing any identifying information about its users to DM and by not storing any state for identifiable users within the DM system. Any "citizen science" results incorporated into the DM system (e.g. object annotations or classifications) will be the responsibility of the project's Principal Investigator (PI).

12 EPO Compute Cluster

ID: EP-DM-CON-ICD-####

Specification: NCSA shall host a compute cluster for EPO and allow it to transfer approved public subset data to the EPO Data Center (EDC).

Discussion: This will be similar to the Kubernetes cluster they currently host for DM and will be paid for by EPO. This EPO cluster will be used to apply EPO-specific processing (such as converting FITS to TIFF) close to the data as well as to act as a storage buffer while transferring data from the DAC to the EDC. Special VPN or firewall configuration will be required to allow the cluster to push data to the cloud-based EDC.