LARGE SYNOPTIC SURVEY TELESCOPE

## Large Synoptic Survey Telescope (LSST)
## LSST Document

# Proposed Policy for Independent Data Access Centers

**William O'Mullane, Beth Willman, Melissa Graham, Leanne Guy, Robert Blum**

**LPM-251**

**Latest Revision: 2019-03-20**

## Abstract

This document describes the proposed policies for groups that are independent from the LSST Project and Operations (i.e. LSST Data Facility) and would like to stand up an independent Data Access Center (iDAC; existing data centers that could serve LSST data products are considered iDACs for purposes of this document). Some iDACs may want to serve only a subset of the LSST data products: this document proposes three portion sizes, from full releases to a "light" catalog without posteriors. Guidelines and requirements for iDACs in terms of data storage, computational resources, dedicated personnel, and user authentication are described, as well as a preliminary assessment of the cost impacts. Some institutions, even those inside the US and Chile, may serve LSST data products locally to their research community. Requirements and responsibilities for such institutional bulk data transfers are also described here. **The purpose of this draft document is to serve as a** ***preliminary*** **resource for partner institutions wishing to assess the feasibility of hosting an iDAC in the era of the Open Data Framework, and is subject to change.**

# Change Record

| Version | Date | Description | Owner name |
|---|---|---|---|
| | 2018-03-24 | Initial version. | WOM |
| | 2019-02-25 | Added site topology schematic | LPG |

*Document source location:* `https://github.com/lsst/LPM-251`

# 1 Introduction

In 2019 LSST has proposed, and NSF and DOE are considering moving to an Open Data Framework (ODF). This means LSST data products will be publicly available to the world science community immediately. LSST data products will be made available to the world community as as well through the LSST Data Access Centers (DACs) in the US and Chile. An ODF has significant advantages for science including extended reach, impact, and legacy of LSST. The ODF vastly simplifies access, publishing, and collaboration for LSST scientists. Finally, an ODF has the advantage that the NSF and DOE can control what data are released to who and when at any time. This is not the case with the current framework which is based on signing agreements with international contributors who pay for data in exchange for rights and access. Getting data to as many scientist as possible without restrictions will foster more collaborations and should give maximum science return. Gaia's [7] open data policy has certainly shown this to be the case with around one thousand publications since the first data release.

The current model which ODF would supplant requires a restricted data rights policy. This policy requires control over access, publication, and sharing of data which any iDACs would have to comply with just as the US and Chile DACs would. In ODF, not such restrictions exist and control and oversight would be significantly easier.

Access to LSST data products for any users will be possible through a Data Access Center (DAC). The United States's DAC will be hosted at the National Center for Supercomputing Applications (NCSA), where registered LSST users will perform scientific queries. Most users will have access to a default set of resources at the DAC sufficient for basic queries and analysis. Users who require more resources will be able to apply for them, and those granted additional resources will be allowed (for example) to perform analysis on the full data releases using the LSST Science Platform (LSP). The LSP is documented with the vision given in LSE-319, with more formal requirements in LDM-554 and the design in LDM-542. The Chilean DAC will be equivalent in functionality to the US DAC, but scaled-down in terms of the computational resources available for query and analysis given the smaller Chilean community [LDM-572][1]. Within the ODF data products may be accessed from a host of other potential locations where one would normally expect to find astronomical data. This is discussed further in Section 2

[1]Most recently the Chilean Government has started the data observatory initiative `https://www.economia.gob.cl/data-observatory` which could see this DAC or other access points moved to the cloud under ODF. This is TBD and outside scope for LSST

TABLE 1: Size summary based on LDM-141

| Table | Bytes/row | Rows (DR1 -> DR11) | DR1 (TB) | $\times$ Growth | DR10 (PB) |
|---|---:|---|---:|---:|---:|
| Object_Lite | 1840 | $2.26^{10} - > 4.74^{10}$ | 42 | 2.1 | 0.08 |
| Object_Extra | 20393 | $2.26^{10} - > 4.74^{10}$ | 461 | 2.1 | 0.9 |
| Source | 453 | $4.51^{11} - > 9.01^{12}$ | 204 | 20.0 | 4.0 |
| ForcedSrc | 41 | $1.20^{12} - > 5.01^{13}$ | 49 | 42 | 2.0 |
| DiaObject | 1405 | $7.94^{08} - > 1.54^{10}$ | 1.1 | 19.4 | 0.002 |
| DiaSource | 417 | $2.26^{09} - > 4.52^{10}$ | 0.9 | 20 | 0.002 |
| DiaForcedSource | 49 | $1.50^{10} - > 3.01^{11}$ | 0.7 | 20 | 0.001 |
| Year 1 raw images:$3PB$, tables:$\sim 1PB$, half for Object_Extra,$0.2PB$ Sources | | | | | |
| Year 10 raw images:$30PB$, tables:$\sim 7PB$,$4PB$ Sources,$2.0PB$ Forced ,$1PB$ Object_Extra | | | | | |

The following sections include the types of data products that could be hosted (Section 3), the requirements and responsibilities that would be expected of an iDAC hosting LSST proprietary data products (Section 4), and a description of the main costs *vs.* their science impacts (Section 6).

The contents of this draft document are meant to provide a preliminary resource for partner institutions who may be assessing the feasibility of hosting an iDAC. The specific mechanisms and processes by which future iDACs will negotiate the bulk transfer of data, the installation of software, etc. is considered beyond the scope of this document.

To better understand the sizes of LSST data products, Table 1 gives an overview.

All access to, and use of the LSST data and data products is subject to the policies described in LPM-261.

## 2 Public Data - the Open Data Framework

Within the ODF LSST must support a large number number of users. Compared to the current model, perhaps 30% more professional astronomers. The total is probably comfortably bounder by a factor of two in any case. Alex Szalay (ref) noted there were over a million unique IP addresses which hit the SDSS archive over a one year period. Gaia saw 2K users accounts per hour on the catalogue release. These numbers suggest significant public interest beyond professionals, and this load must be supported by the DAC as well. There are several possible ways to handle the load. [2],

Nationally, we could partner with Google or Amazon who will notionally host public data sets for free - so we could more selectively add users to the DACs and try to put most casual users to the public interface. It is not clear if this would then be the EPO interface, but that is worth considering. EPO would no longer have to select 10% of data, but it would have a bigger job to deal with a full data release. Still, combining DM and EPO seamlessly with respect to data would be favorable. The public data set would be some lite version for the catalog Section 3.3 and the HiPS type color images. No raw data files and potentially no advanced notebook type access. This may not even need a logon for the fast queries like "show me M31" with LSST sources plotted. EPO queries would not include the source catalogue which is large ($10^{13}$ rows). Potential national partners could also host the object catalog e.g.

- MAST at STScI
- DataLab at NOAO
- SAO
- US Naval Observatory
- NED at IPAC
- HEASARC
- CADC - Canadian Astronomical Data Centre - we work with them already in Data Management.

## 2.1 International

Internationally we could partner with a network of sites - it should be a network to allow peer to peer sharing of catalogs. Here we could provide the HIPS color images and again some version of the OBJECT catalogue. We would have to consider if the source catalogue should also be distributed to a smaller subset of centres who could cope with it. Potential International Partners might be:

[2] Gaia reports modest discontent due to user load https://blogs.scientificamerican.com/observations/the-price-of-open-science/

- In Europe we have a few centres that Astronomers expect to find sources at:
  - CDS - Aladin and Vizier - this is a minimum for Europe
  - ESAC - ESASky - European Space Agency
  - ASDC - Italian Data Center
  - GAVO MPA - German Astronomical Virtual Observatory - Max Plank Astrophysics
  - IoA - Cambridge
  - Edinburgh
- In Asia
  - NAOJ Astronomical Data Center, Tokyo
  - Chinese Academy of Sciences - perhaps National Space Science Center

Having the lite catalog hosted at multiple locations where, and in formats which, astronomers would expect to find catalog information would reduce load on the US and Chilean DAC. It will also put the LSST data more readily in the hands of the astronomers and should accelerate science at least in the cases where the catalog is the prime source of information, for example, Galactic dynamics and other large statistical studies.



## 3 Types of Data Products for iDACs

The three categories of LSST data products, *Prompt*, *Data Release*, and *User Generated* are defined in the LSST Data Product Categories document LPM-231. Both the *Prompt*, *Data Release* data products are produced by LSST and include images, both raw and processed, and catalogs of both `Objects` and `Sources`. The *User Generated* data products are produced the community using the resources of the LSST Science Platform LSE-319. These data products are described in detail in the Data Products Definitions Document LSE-163.

Below, three potential realizations of the the LSST *Data Release* data products that iDACs might consider hosting are described: the full *Data Release* including images, the *Data Release* catalogs only, and a low-volume ("`lite`") subset of the *Data Release* catalogs.

## 3.1 Full Data Release(s)

In this case the iDAC would be hosting all of the raw and processed images, and catalogs, as described in [LSE-163]. Hosting the raw image data at an iDAC requires roughly $6$ petabytes per year of storage, so this represents a significant augmentation of resources in terms of both hardware and personnel. The processed data and associated calibrations bring the total data volume to $0.5$ exabytes for a single data release. Some data volume could be saved by taking only a single calibrated image per band, but the total would still be $60$ petabytes (with compression it may be possible to reduce this even further). Any iDAC considering hosting the full *Data Release* should also deploy the full LSST Science Platform LSE-319 in order to maximize science productivity and their return on investment in hosting an iDAC.

## 3.2 Catalog Server

Alternatively, an iDAC may find that hosting only the *Data Release* catalogs, and not the images, is sufficient for the scientific needs of its community. This will probably require the specific LSST database server [LDM-135] and specific machines, and the deployment of the database system and the associated subset of data access services (DAX; e.g., web APIs, Qserv, LDM-152). The full `Object` catalog, which contains one row per object with a volume of $\approx 20$ kilobytes per row, is estimated to contain about $40 \times 10^9$ objects (even in the first full-sky data release). Adding to this the full `Source` and `Forced Source` catalogs, which contain one row per measurement in each of the $\sim 80$ visit images obtained per year, brings the total storage volume required up into the petabytes range, and will require a serious commitment of resources at the proposed iDAC. The evolution of catalog sizes over the 10-year LSST survey is depicted in Figure 1, from which it is evident that the catalog size for the final release is order $15$ petabytes. For more details on the row counts see the Key Numbers Page[3].

## 3.3 An "`Object Lite`" Catalog

Many – perhaps most – astronomers' science goals will be adequately served by a low-volume subset of the `Object` catalog's columns that do not include, for example, the full posteriors for the bulge+disk likelihood parameters. This `Object Lite` catalog would nominally contain $1840$ bytes per row for the $40 \times 10^9$ objects, giving a size of $\approx 7.4 \times 10^{13}$ bytes ($\sim 74$ terabytes). Even smaller, science-specific versions of `Object Lite` could be envisioned with even less columns

[3] https://confluence.lsstcorp.org/display/LKB/LSST+Key+Numbers

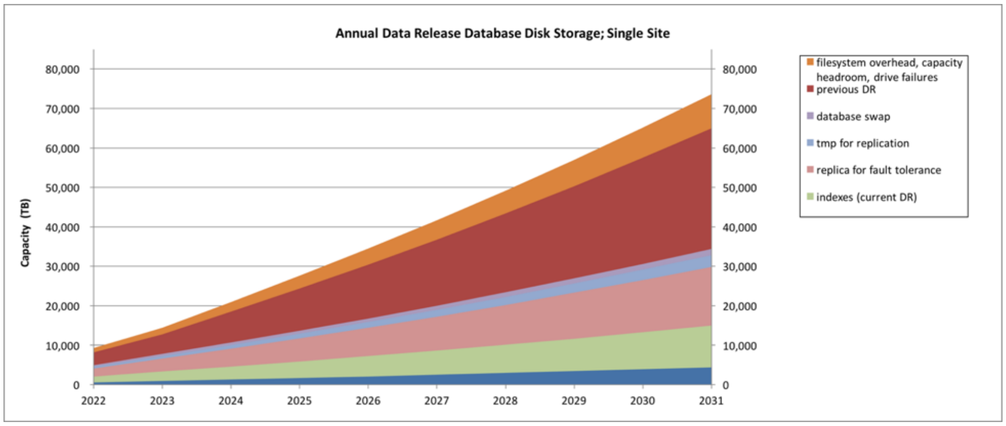


FIGURE 1: Catalog volume over time from LDM-144.

and/or separate star and galaxy catalogs. The Solar System community for example will be primarily be interested in the contents of just the `SSObjects` table.

These would not be small enough to handle on a laptop, but might be served by a small departmental cluster (a mini-iDAC). Searching even a small `Object Lite` catalog would require some form of database, but many institutes would already have a system which may be capable of loading this data. In this case, LSST might only ship files with documentation and not provide administrative support for the system, but this would allow the `Object Lite` catalog to be widely available to all partner institution iDACs. Distribution options such as peer-to-peer networking to avoid download bandwidth limitations might be possible to implement in this case. See also Section 2.



# 4 Requirements and Guidelines for iDACs

Since creating, delivering, and supporting the implementation of LSST data products via iDACs creates some cost to the LSST Project, iDACs will be expected to follow some basic requirements and guidelines that are described below. The actual costs of iDAC support and infrastructure development are considered separately in Section 6.

## 4.1 LSST site topology

Leanne to describe the flow in the topology diagram.

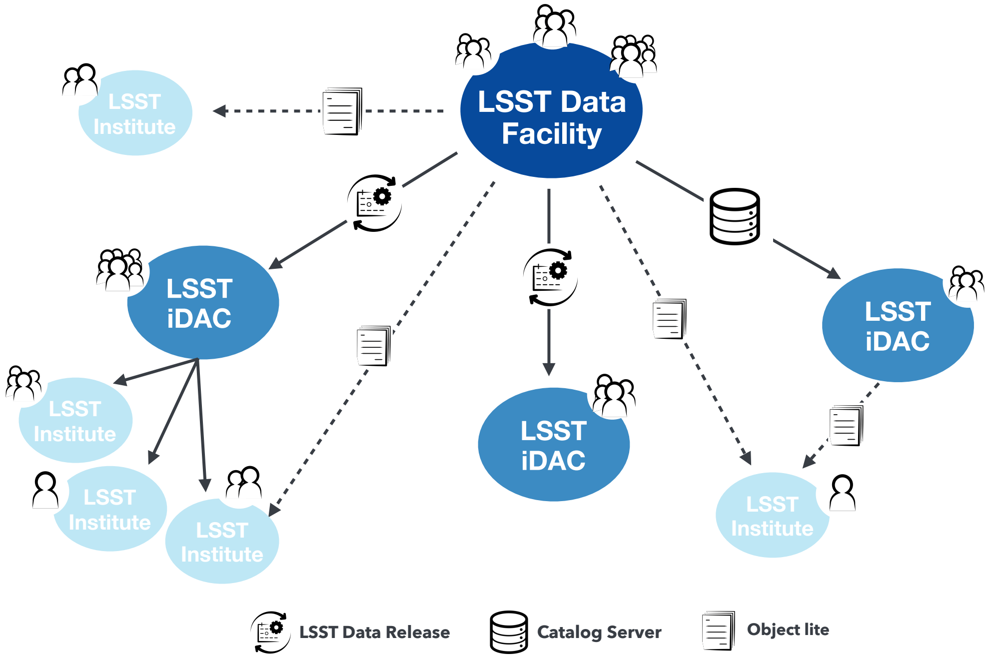


FIGURE 2: LSST Data Facility and Independent DAC network topology.

## 4.2 Required Resources

Institutions or organizations wishing to set up independent data access centres will be expected to have sufficient resources and commitments before we discuss data transfers and support. See also Section 6.3 for a discussion on compute vs storage.

### 4.2.1 Data Storage

Any institution considering setting up an iDAC will need to show commitment on purchasing sufficient storage and CPU power to hold and serve the data. Sufficient storage ranges from $0.5$ exabytes for the full data release(s) down to $100$ terabytes for a catalog server, and potentially further down to $70$ terabytes if the `Object Lite` option is offered. For the full catalog , of order 100 nodes are required to serve it up. To serve images, a DAC would need some additional servers; depending on load this may be order 10 additional nodes.

#### 4.2.2 User Computational

If the full set of data release products including images and catalogs are desired, it is highly recommended that the iDAC deploy the LSST Science Platform (LSP). The LSP serves as a portal to the data, and provides a user interface of web services and Jupyter notebooks for scientific queries and analysis, an open software framework for astronomy and parallel processing, and the associated computational, storage, and communications infrastructure needed to enable science. The LSP is described in full in LSE-319 and LDM-554. Depending on the assumed load, the LSP is relatively modest as it requires only $\sim 2$ servers to set up, and it is recommended to have 2 CPUs per simultaneous user (e.g., if the iDAC's desired capability is to serve 200 users, but only expect 50 to be active at a time, then 100 CPUs would be sufficient). From that starting point, the amount of next-to-the-data computational resources can be as large as the data center wishes to provide, and may make use of connecting to e.g., local super computer resources.

#### 4.2.3 Dedicated Personnel

The significant hardware required by an iDAC is above the normal level for most astronomy departments, and would require dedicated technical personnel to set it up and keep it running. For an `Object Lite` catalog running on existing hardware, this might not be a significant increase in person power if the hardware is already serving on order $50$–$100$ terabytes. Still, it is recommended to assume $\gtrsim 0.25$ full-time equivalent (FTE) personnel hours for `Object Lite`, and perhaps closer to $\sim 2$ FTE for the full catalogs, which includes setting up and maintaining the service, and installing new data releases and software updates every year. For iDACs wishing to host the full data releases' images and catalogs and deploy the LSST Science Platform, it becomes necessary to employ $1$–$2$ storage engineers to mange the large amount of data, and possibly one more FTE to keep the Kubernetes (or equivalent) system updated with the latest software deploys. If the iDAC intends to support the science of many local users, support will become a specific issue which may not be covered by the usual institutional funding, and will require further effort. It is therefore recommended that any partner institution wishing to host a full-release iDAC provide a minimum personnel of 5 FTE to be considered viable.

### 4.3 Networking and distribution

There is an assumption than any prospective iDAC will have a high bandwidth connection with demonstrated sustained $40Gb/s$ to enable data transfer and sufficient bandwidth for access by users. In addition all iDACs should be ready to serve the `Object Lite` catalog to any institution worldwide but especially any *local* institutions.

### 4.4 Services

Leanne

Services provided by LSST partner iDACs

#### 4.4.1 The LSST Science Platform

All LSST independent data access centres must run an instance of the LSST Science Platform

#### 4.4.2 User Generated data products

## 5 Responsibilities of the LSST Data Facility

I prefer we not call them LSST iDACs or iLSST DACs. These are independent DACs which comply with basic requirements to support LSST data. By calling them LSST DACs we imply that we have some responsibility for them. I prefer the figure also not use the term "LSST iDACs"

This section describes the services that the LSST Data Facility (LDF) will provide in support of all LSST iDACs.

The LDF will prepare data products for distribution to iLSST DACs along with documentation of hardware and software that will make serving these data consistent with the serving of data from the LSST DACs. LSST will provide (modest) technical support consistent with available resources to assist groups setting up iLSST DACs.

LSST, through the LDF will establish a process for potential iDAC groups to interface with and establish data transfers to their iDACs. It is expected that iDAC groups will propose to LSST what their iDAC would support and then LSST will work with them to establish requirements to receive LSST data. One approved, LSST will provide (modest) technical support consistent with available resources to assist groups setting up their iDACs provided they comply with prerequisites discussed in this document and especially in Section 4.2.

### 5.1 Proprietary Data Access Policies

Defer for now until further guidance received.
=======

### 5.2 Data Distribution

NCSA will have $100Gb/s$ connections on ESNET which has interconnects with Internet2 - this should provide a distribution mechanism for getting data to iDACs, it will however be limited by the fact that much of our bandwidth is already allocated for data transmission to IN2P3 and alert distribution.

A tiered model as used by CERN for high energy physics would seem a desirable way to achieve big transfers. Hence we would have a small selection of tier 2 centres with all data products from which tier 3 centres could copy the subsets they wish to work with. Other alternatives are discuss in Section 6.2.

In HEP experiments such as BABAR various physics analysis groups (science collaborations in LSST ) were assigned to specific international centers as their primary computing and analysis facility, thereby distributing the computing load around the ”network.” Users naturally tend to use the facility with available resources and cycles.

## 6 Cost Impacts

As previously mentioned, standing up and maintaining multiple iDACs comes at a significant cost impact to both the LSST Project and the partner institutions. Minimizing these costs – or at least maximizing the amount of science they enable – should be at the forefront of all considerations concerning partner iDACs, such as the following propositions.

## 6.1 Maximizing Profits with Science-Driven iDACs

There are two main cost impacts of iDACs being set up outside of the US and Chilean DACs: the positive impact is that some computational load may be taken off of these existing DACs, but the negative impact is the level of support required from the LSST Project in order to get them set up and running. This negative impact could be mitigated by ensuring that science productivity is maximized as a result of this extended effort. One way to do this might be to associate specific areas of science to a given iDAC, and encourage users working in that field to use that iDAC. This could create a customer base for the iDAC, bring together like-minded experts, and effectively distribute the computing load across a network of iDACs. This might also enhance internal funding arguments for investment resources by arguing for synergies with local science goals and attracting international users and official endorsement.

## 6.2 Data Transfer

Even with good networks the data transfer will not be trivial, and could be quite expensive. LSST is not currently set up to distribute data to multiple sites, i.e., there is no form of peer-to-peer sharing. The bandwidth at NCSA is adequate for receiving data and delivering `Alerts` to brokers during the night; perhaps some day time bandwidth could be used to transfer data to iDACs. A full data release of images and catalogs does not have to transferred within a given day; if the correct agreements are in place with an iDAC, a full release could be transferred slowly as it is produced, and then made available to the iDACs users in whole on the official release day.

### 6.2.1 Transfer cost use case

If we take the final number from the key numbers page [4] we could consider DR1 as about 6 PB (10% of the final size).

We would have at least two ways to transfer this : via the network, via physical devices.

A network transfer at 10Gbps of 6 PB would take $8 * 6 \times 10^{12} \; 10^{7} = 4.8 \times 10^{6} \; seconds$ or about

[4] https://confluence.lsstcorp.org/display/LKB/LSST+Key+Numbers

55 days[5]. Many institutes have 100 Gbps connections so this should be an upper limit and a transfer should be order one week. If we had a peer to peer network this may go down somewhat and we may be able to support it from NCSA.

Alternatively we could host the data on Amazon or Google and let people download it from there - they would have more capacity. Storage on the cloud for public data would be theoretically free - download (egress) would cost. Transfer to another cloud [6] or a Content Delivery Network (CDN)[7] end up costing about a cent a GB which for an open science project and at our volume should be negotiable. At one cent a transfer would cost $\sim \$0.01 * 6 \times 10^{12}\ 10^{6} = \$60K$.

For physical devices, today apparently we could get a device like Petarack https://www.aberdeeninc.com/petarack/ for $300K. Theoretically we could get this cheaper though this is close to the drive price, Tape may also be a possibility especially if Sony/IBM commercialize high density tape with >300TB per cartridge[8]. A current 6TB cartridge is about $30, so enough tapes for 6PB would cost about 30K. If the density increased this could come down significantly. This could be be a partner data center cost as well as shipping it. Transfer of data on to this would be about the same as the network rate above so 7 days. SneakerNet [? ] may still be cost effective in the LSST era, which is predicted in the a paper.

## 6.3 Compute *vs.* Storage Resources

Data storage is a large cost to iDACs, and could be considered as an overhead relative to the amount of computational resources an iDAC can offer. If an iDAC is set up without a large compute capacity, the facility might be less useful to the science community than e.g., augmenting an existing DAC or iDAC to have more computational resources. It is conceivable that a partner institution may prefer to spend their money increasing the computational quotas available for a given collaboration or set of PIs, and it would be scientifically beneficial if this was possible at all DAC and iDACs. The notion of standard compute quotas and resource allocation committees to adjudicate on large proposals for substantial increases to computational allocations are described in LPM-261. Another way to approach a solution to this issue might be to have a *Cloud*-based iDAC where a user or PI could buy nodes on the provider cloud to access the holdings put there by LSST. Such an option may be particularly useful to

[5] day = 86400
[6] https://cloud.google.com/storage/pricingi#network-pricing
[7] https://cloud.google.com/cdn/pricing
[8] https://newatlas.com/sony-ibm-magnetic-tape-density-record/50743/

Science Collaborations with large compute needs.

# A References


[1] **[LDM-141]**, Becla, J., Lim, K.T., 2013, *Data Management Storage Sizing and I/O Model*, LDM-141, URL https://ls.st/LDM-141

[2] **[LDM-135]**, Becla, J., Wang, D., Monkewitz, S., et al., 2017, *Data Management Database Design*, LDM-135, URL https://ls.st/LDM-135

[3] **[LDM-542]**, Dubois-Felsmann, G., Lim, K.T., Wu, X., et al., 2017, *LSST Science Platform Design*, LDM-542, URL https://ls.st/LDM-542

[4] **[LDM-554]**, Dubois-Felsmann, G., Ciardi, D., Mueller, F., Economou, F., 2018, *Science Platform Requirements*, LDM-554, URL https://ls.st/LDM-554

[5] **[LPM-231]**, Dubois-Felsmann, G.P., Ivezic, Z., Juric, M., 2018, *LSST Data Product Categories*, LPM-231, URL https://ls.st/LPM-231

[6] **[LDM-144]**, Freemon, M., Pietrowicz, S., Alt, J., 2016, *Site Specific Infrastructure Estimation Model*, LDM-144, URL https://ls.st/LDM-144

[7] Gaia Collaboration, Prusti, T., de Bruijne, J.H.J., et al., 2016, A&A, 595, A1 (arXiv:1609.04153), doi:10.1051/0004-6361/201629272, ADS Link

[8] **[LSE-319]**, Jurić, M., Ciardi, D., Dubois-Felsmann, G., 2017, *LSST Science Platform Vision Document*, LSE-319, URL https://ls.st/LSE-319

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

[10] **[LDM-152]**, Lim, K.T., Dubois-Felsmann, G., Johnson, M., Jurić, M., Petravick, D., 2017, *Data Management Middleware Design*, LDM-152, URL https://ls.st/LDM-152

[11] **[LDM-572]**, O'Mullane, W., Petravick, D., 2017, *Chilean Data Access Center*, LDM-572, URL https://ls.st/LDM-572

[12] **[LPM-261]**, Willman, B., Graham, M., O'Mullane, W., Petravick, D., 2018, *Access Policy for LSST Data and Data Access Center*, LPM-261, URL https://ls.st/LPM-261

# B Acronyms

| Acronym | Description |
|---|---|
| AAS | American Astronomical Society |
| ASDC | ASI Science Data Center (Italy) |
| CADC | Canadian Astronomy Data Centre |
| CDN | Content Delivery Network |
| CDS | Centre de Données astronomiques de Strasbourg |
| CPU | Central Processing Unit |
| DAC | Data Access Center |
| DAX | Data access services |
| DM | Data Management |
| DOE | Department Of Energy |
| EPO | Education and Public Outreach |
| ESAC | European Space Astronomy Centre |
| FTE | Full Time Equivalent |
| GAVO | German Astronomical Virtual Observatory |
| GB | GigaByte |
| HEASARC | NASA's Archive of Data on Energetic Phenomena |
| HIPS | Hierarchical Progressive Survey |
| IBM | International Business Machines |
| IP | Internet Protocol |
| IPAC | No longer an acronym |
| IT | Integration Test |
| IoA | Institute of Astronomy (Cambridge UK) |
| LDF | LSST Data Facility |
| LDM | LSST Data Management (Document Handle) |
| LPM | LSST Project Management (Document Handle) |
| LSE | LSST Systems Engineering (Document Handle) |
| LSP | LSST Science Platform |
| LSST | Large Synoptic Survey Telescope |
| MAST | Mikulski Archive for Space Telescopes |
| MPA | Max Planck Institute for Astrophysics |
| NAOJ | National Astronomical Observatory of Japan |
| NCSA | National Center for Supercomputing Applications |

| | |
|---|---|
| NED | NASA/IPAC Extragalactic Database |
| NOAO | National Optical Astronomy Observatories (USA) |
| NSF | National Science Foundation |
| ODF | Open Data Framework |
| PB | PetaByte |
| PI | Principle Investigator |
| Qserv | Proprietary LSST Database system |
| SAO | Smithsonian Astrophysical Observatory |
| SDSS | Sloan Digital Sky Survey |
| TB | TeraByte |
| US | United States |