Rubin Observatory

Vera C. Rubin Observatory
Rubin Observatory Document

Proposed Policy for Independent Data Access Centers

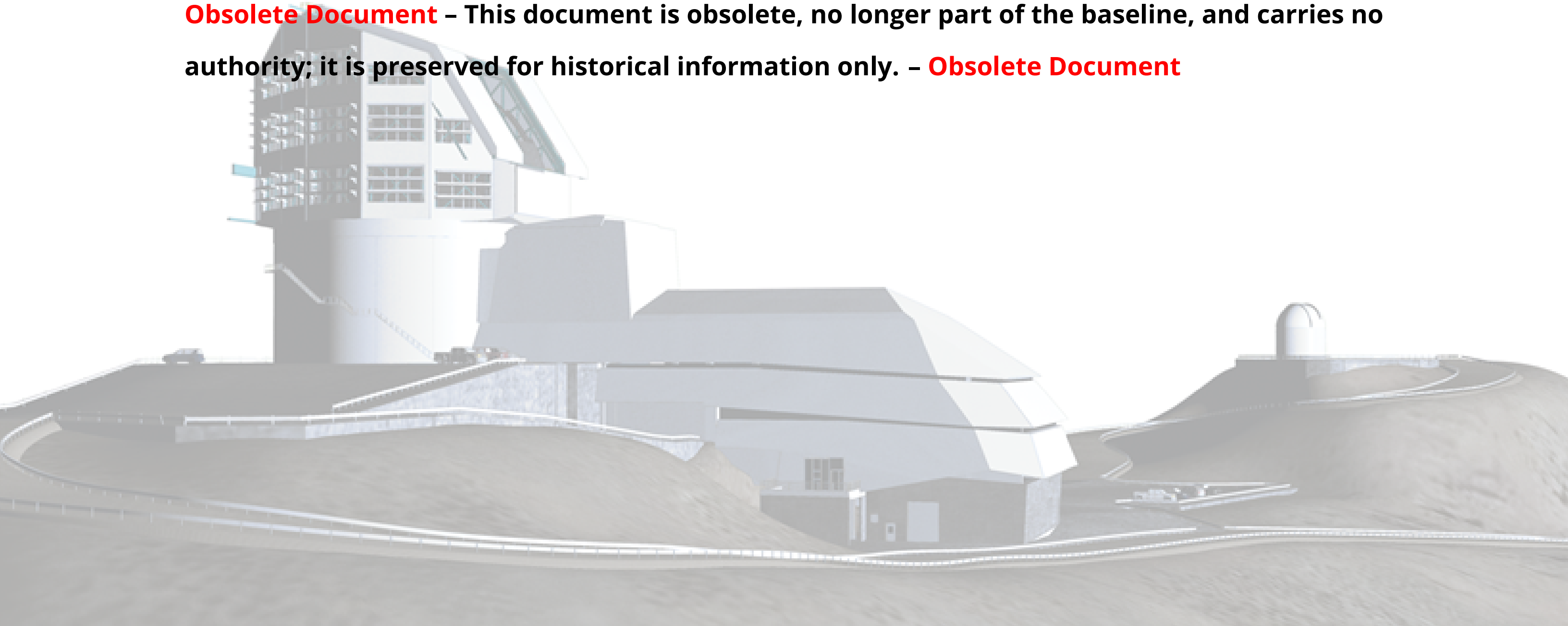
William O'Mullane, Beth Willman, Melissa Graham, Leanne Guy, Robert Blum

LPM-251

Latest Revision: 2020-07-08

Obsolete

Obsolete Document – This document is obsolete, no longer part of the baseline, and carries no authority; it is preserved for historical information only. – Obsolete Document



Abstract

See rather `RTN-003.lsst.io` 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.**

Obsolete

Rubin Observatory

Change Record

Version	Date	Description	Owner name
	2018-03-24	Initial version.	WOM
	2019-02-25	Added site topology schematic	LPG
	2019-06-17	In kind data framework	WOM
	2020-04-01	CEC input	WOM

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

Obsolete

1 Introduction

The current model for LSST is that is provides proprietary data to approved users in Chile and the US. The data access model accommodates this restricted data rights policy. This policy requires control over access, publication, and sharing of proprietary data which any international data access center (IDAC) would have to comply with just as the US and Chile DACs do.

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 US Data Facility[1], 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].

The following sections include the types of data products that could be hosted (Section 2), the requirements and responsibilities that would be expected of an IDAC hosting LSST proprietary data products (Section 3), and a description of the main costs *vs.* their science impacts (Section 5).

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. A simplified checklist is given in Appendix C.

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 LDO-13.

[1] currently National Center for Supercomputing Applications (NCSA)

OBSOLETE – This document is preserved for historical information only. – OBSOLETE

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					

TABLE 2: Potential/estimated usage of proucts in Table 1 and images, this table came from the AMCL originally.

Data Product	Cardinality	Volumei [PB]	Usuage Frequency	Discovery Potential	Replicas
Object_Lite	40M	0.1	95%	20%	0.08
Object_Extra	40M	0.9	4%	24%	0.9
Source	9T	4.0	0.9%	50%	4.0
ForcedSrc	50T	2.0	0.1%	3%	2.0
Image coadds	55K	0.3	0.01%	2%	0.002
Image raw	5.5M	30.0	0.001%	1%	0.002

In addition to the sizes shown in Table 1 it is interesting to consider how much access and potentially how much science there is per table. This is discussed in detail in PSTN-003. The AMCL made an interesting table concerning this topic which is reproduced here in Table 2. Feedback on the correctness of this table has been sought from PST.

Obsolete

2 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.

2.1 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.

2.2 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 ≈ 20 kilobytes per row, is estimated to contain about 40×10^9 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 ~ 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[2].

2.3 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

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

Obsolete

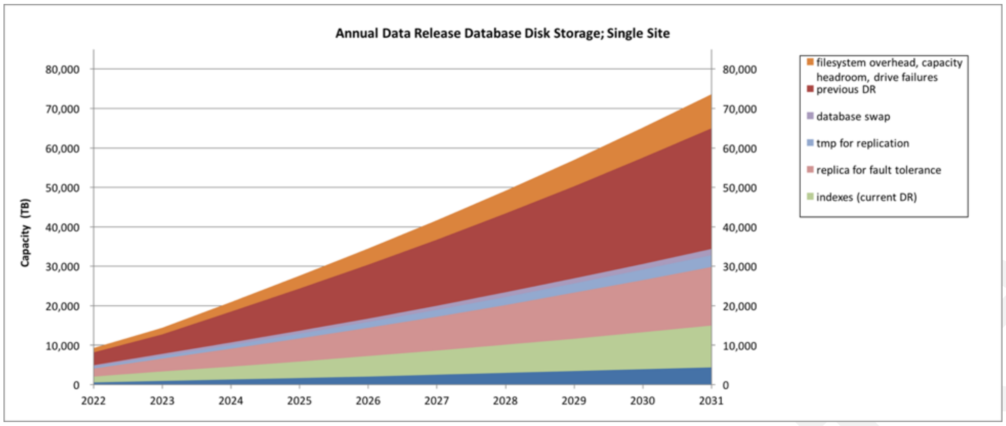


FIGURE 1: Catalog volume over time from LDM-144.

bytes per row for the 40×10^9 objects, giving a size of $\approx 7.4 \times 10^{13}$ bytes (~ 74 terabytes). Even smaller, science-specific versions of `Object Lite` could be envisioned with even less columns 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. 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.

3 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, which are described below. The actual costs of IDAC support and infrastructure development are considered separately in Section 5.2.1.

Rubin Observatory

3.1 LSST site topology

Figure 2 shows a topology for a set of interconnected IDACs. US scientists will have direct access to the LSST Data Access Facility at NCSA. Hosting on the cloud is shown, as described in

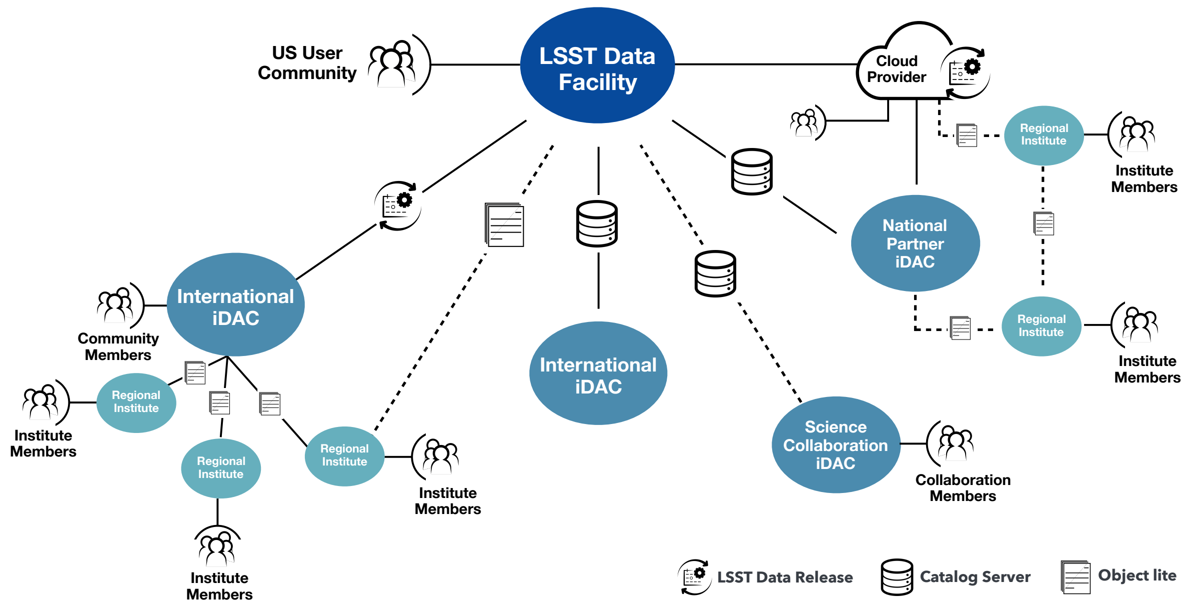


FIGURE 2: LSST Data Facility and Independent DAC network topology.

3.2 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 5.3 for a discussion on compute vs storage.

3.2.1 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.

3.2.2 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 ~ 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.

3.2.3 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 ~ 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.

Obsolete

3.3 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.

3.4 Services

Independent DACs will be expected to provide services analogous to those provided at the LSST Data Facility at NCSA.

3.4.1 The LSST Science Platform

The *LSST Science Platform* LSE-319 is a set of integrated web applications and services deployed at LSST Data Access Centers through which the scientific community will access, visualize, subset and perform next-to-the-data analysis of the data collected by the LSST; it is envisioned to enable science cases that would greatly benefit from the co-location of user processing and LSST data. It will provide users access to the `Data Products` described in 2, such as, resources for image reprocessing, access to the LSST processing framework, and many other services as described in LSE-163. All LSST `Independent Data Access Centers` will be expected to run and support the LSST Science Platform.

3.4.2 User Generated data products

User Generated data products will be created by the community deriving from the *Prompt* and *Data Release* data products, and making full use of the power of the LSST database systems and Science Platform for the storage, access, and analysis of their results. The Science Platform will allow for the creation of *User Generated* data products and will enable science cases that greatly benefit from co-location of user processing and/or data within the LSST Archive Center. Independent DACs will be expected to provide support for the creation of *User Generated* data products and their federation with the LSST Data products.

Obsolete

4 Responsibilities of the LSST Data Facility

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

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

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 3.2.

4.1 Proprietary Data Access Policies

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

4.2 Data Distribution

NCSA will have $100 Gb/s$ 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 5.2.

In HEP experiments such as BaBar various physics analysis groups (science collaborations in

Obsolete

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.

5 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.

5.1 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.

5.2 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.

5.2.1 Transfer cost use case

If we take the final number from the key numbers page [3] 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 55 days[4]. 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 [5] or a Content Delivery Network (CDN)[6] 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[7]. 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 [8] may still be cost effective in the LSST era, which is predicted in the a paper.

5.3 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., aug-

[3] https://confluence.lsstcorp.org/display/LKB/LSST+Key+Numbers
[4] day = 86400
[5] https://cloud.google.com/storage/pricingi#network-pricing
[6] https://cloud.google.com/cdn/pricing
[7] https://newatlas.com/sony-ibm-magnetic-tape-density-record/50743/

Obsolete

menting 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 LDO-13. 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 Science Collaborations with large compute needs.

The full sizing model is in DMTN-135 - any IDAC should have a similar sizing model. They may not need as much compute or as many copies of data as we have but the raw information to make such calls are in the technote. For ball park figures the construction to first year of ops table is copied here as tabreftab:Summary[8].

Table 3: This table from DMTN-135 pulls together a high level summary of costs using conservative price factors.

Year	2020	2021	2022	2023
Compute (2019 pricing)	$690,000	$0	$1,510,000	$2,820,000
Storage (2019 pricing)	$190,702	$126,563	$1,216,867	$7,864,125
Qserv (2019 pricing)			$560,000	$3,240,000
Total (2019 pricing)	**$880,702**	**$126,563**	**$3,286,867**	**$13,924,125**
Compute (applying price factor)	$621,000	$0	$1,057,000	$1,692,000
IN2P3 (50% of compute in ops)				-$846,000
Storage (applying price factor)	$181,167	$113,907	$1,034,337	$6,291,300
Qserv (applying price factor)			$434,000	$2,268,000
Hosting cost NCSA	$110,802	$62,802	$238,012	$536,801
Total budget (using price factors)	**$912,969**	**$176,709**	**$2,763,350**	**$9,942,100**

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] **[LDO-13]**, Blum, R., et al., 2019, *LSST Data Policy*, LDO-13, URL https://ls.st/LDO-13

[8] There is no guarantee of being in sync with DMTN-135 but as an order of magnitude it is good.

OBSOLETE – This document is preserved for historical information only. – OBSOLETE

[4] **[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

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

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

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

[8] Gray, J., Chong, W., Barclay, T., Szalay, A., vandenBerg, J., 2002, arXiv e-prints, cs/0208011 (arXiv:cs/0208011), ADS Link

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

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

[11] **[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

[12] **[DMTN-135]**, Michelle Butler, W.O., Kian Tat Lim, 2019, *DM sizing model and purchase plan for the remainder of construction.*, DMTN-135, URL http://DMTN-135.lsst.io

[13] **[PSTN-003]**, O'Mullane, W., Mueller, F., 2019, *Discussion of Object vs. Source table queries and data distribution*, PSTN-003, URL http://PSTN-003.lsst.io

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

B Acronyms

Acronym	Description
AMCL	AURA Management Council for LSST
ASDC	ASI Science Data Center (Italy)
CADC	Canadian Astronomy Data Centre

Obsolete

CAOM	Common Archive Observation Model
CDN	Content Delivery Network
CDS	Centre de Donnes astronomiques de Strasbourg
CEC	International in-kind Contribution Evaluation Committee
CERN	European Organization for Nuclear Research
CPU	Central Processing Unit
DAC	Data Access Center
DAX	Data Access Services
DM	Data Management
DMTN	DM Technical Note
DR1	Data Release 1
DR10	Data Release 10
DR11	Data Release 11
EPO	Education and Public Outreach
ESAC	European Space Astronomy Centre
ESNet	Energy Sciences Network
FTE	Full-Time Equivalent
FY22	Financial Year 22
FY34	Financial Year 34
GAVO	German Astronomical Virtual Observatory
GB	Gigabyte
Gb	Gigabit
HEASARC	NASA's Archive of Data on Energetic Phenomena
HEP	High Energy Physics
HIPS	Hierarchical Progressive Survey
IBM	International Business Machines
IDAC	Independent Data Access Center
IN2P3	Institut National de Physique Nucléaire et de Physique des Particules
IP	Internet Protocol
IPAC	No longer an acronym; science and data center at Caltech
IVOA	International Virtual-Observatory Alliance
IoA	Institute of Astronomy (Cambridge; also denoted IOA)
LDF	LSST Data Facility
LDM	LSST Data Management (Document Handle)

Obsolete

LDO	LSST Document Operations (Document Handle)
LPM	LSST Project Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)
LSP	LSST Science Platform
LSST	Legacy Survey of Space and Time (formerly 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)
PB	PetaByte
PI	Principle Investigator
PST	Project Science Team
PSTN	Project Science Technical Note
SAO	Smithsonian Astrophysical Observatory
SDSS	Sloan Digital Sky Survey
TAP	Table Access Protocol
TB	TeraByte
US	United States

C IDAC Proposal Checklist

There is a spectrum of possibilities for an IDAC's scope, from just hosting the Object lite database, to serving full copies of the current data release and prompt products database. Here an attempt is made to have a set of check lists that can be used to look at an IDAC. Section C.3 covers a full capability IDAC, while Section C.2 gives the criteria for a minimum capability IDAC. There will undoubtedly be proposals in between. There are some criteria that any IDAC must meet, in order to comply with Rubin Observatory data policy. These are given in Section **??**.

Obsolete

C.1 Any DAC

- ☐ Authentication/Authorization system inline with Rubin Observatory Access
- ☐ Agreement to make broadly accessible to all Data Rights holders

C.2 Lite IDAC

All criteria in Section C.1, and then, in addition:

- ☐ Database system capable of handling 4^{10} rows.
- ☐ IVOA TAP interface, MyDB and Table Upload, CAOM support.
- ☐ About 500TB of disk for catalogs + MyDBs.
- ☐ Professional support staff (min 0.25 FTE)
- ☐ Sufficient connectivity to support users

C.3 Full DAC

All criteria in Section C.1, and then, in addition:

- ☐ Staff (about 5 FTE) to handle major hardware installation
- ☐ Agreement to stand up standard Science Platform (Puppet/Kubernetes etc.)
- ☐ Database system capable of handling all catalogs (or Qserv) with IVOA interfaces.
- ☐ Understanding of sizing model in DMTN-135 - sizing model and a cost model for the IDAC.
- ☐ ”Commitment to fund the IDAC through the LSST operations period, FY22-FY34 (probably $> \$6M/year$ based on hardware cost model and labor plan).”.
- ☐ Sufficient connectivity to support data transfer in and user access out at least 20Gbps of free bandwidth.

OBSOLETE – This document is preserved for historical information only. – OBSOLETE