



DataLab-ORD-1.00

# Operational Requirements Document

For the

## NOAO Data Lab Project

Revised: March 3, 2015



## Revision History

<b>Date</b>	<b>Author</b>	<b>Changes / Comments</b>	<b>Version</b>
Sep 04, 2014	M. Fitzpatrick	First Draft	0.1
Nov 13, 2014	M. Fitzpatrick	Updates	0.2
Nov 20, 2014	M. Fitzpatrick	Restructure / updates	0.3
Nov 24, 2014	M. Fitzpatrick	Full first draft	0.4
Mar 03, 2015	K. Mighell	Edits	0.90
Mar 03, 2015	M. Fitzpatrick	Minor Edits - Final Draft	1.00

## Table of Contents

<b>1 Document Overview</b> .....	<b>4</b>
<b>1.1 Purpose</b> .....	<b>4</b>
<b>1.2 Scope</b> .....	<b>4</b>
<b>1.3 Audience</b> .....	<b>4</b>
<b>1.4 Document Format</b> .....	<b>4</b>
<b>1.5 Referenced Documents</b> .....	<b>4</b>
<b>2 General Project Requirements</b> .....	<b>5</b>
<b>2.1 Project Success Metrics</b> .....	<b>5</b>
<b>2.2 System Documentation</b> .....	<b>5</b>
2.2.1 User Documentation .....	5
2.2.2 Developer Documentation .....	5
2.2.3 Operator Documentation .....	6
2.2.4 Document Repository.....	6
<b>2.3 Source Repository</b> .....	<b>6</b>
<b>2.4 Operations and Monitoring</b> .....	<b>7</b>
<b>2.5 Help Desk and Community Engagement</b> .....	<b>7</b>
<b>3 Science Capabilities</b> .....	<b>7</b>
<b>3.1 Catalog Query and Analysis Tools</b> .....	<b>7</b>
<b>3.2 Image Query and Analysis Tools</b> .....	<b>8</b>
<b>3.3 Time Series Tools</b> .....	<b>9</b>
<b>3.4 Visualization Tools</b> .....	<b>9</b>
3.4.1 2-D/3-D Plotting .....	10
3.4.2 Image Display.....	10
<b>3.5 User-Defined Analysis</b> .....	<b>10</b>
<b>4 System Infrastructure</b> .....	<b>11</b>
<b>4.1 Large Catalog Support</b> .....	<b>11</b>
<b>4.2 Virtual Storage</b> .....	<b>11</b>
<b>4.3 Data Publication</b> .....	<b>13</b>
4.3.1 General Requirements .....	13
4.3.2 Collection Metadata.....	13
4.3.3 Data Ingestion.....	14
4.3.4 Metadata Harvesting.....	14
<b>4.4 Processing Services</b> .....	<b>15</b>
<b>4.5 Machine and Process Virtualization</b> .....	<b>15</b>
<b>4.6 User/Group Authentication</b> .....	<b>16</b>
<b>5 System Interfaces</b> .....	<b>16</b>
<b>5.1 Programmatic Interfaces</b> .....	<b>16</b>
<b>Appendix I: Vocabulary / Acronyms Used</b> .....	Error! Bookmark not defined.

# 1 Document Overview

## 1.1 Purpose

This *Operations Requirement Document* is intended to list the operational and functional requirements for all components of the Data Lab. Each requirement will be uniquely numbered for identification. Where possible, this requirement will be traceable back to the originating Science Requirement(s) or Operational Concept(s).

## 1.2 Scope

The scope of this document is the entire Data Lab Project. This document may evolve over time as requirements and designs are finalized.

## 1.3 Audience

This document is intended primarily for (1) the software developers who will be designing and implementing the Data Lab components, and (2) reviewers of the Data Lab project.

## 1.4 Document Format

This document consists of all requirement statements for each component of the Data Lab. Sections are meant to parallel the structure of the *Operations Concept Document* for consistency; each section will contain a table of the form:

<i>Req. Number</i>	<i>Requirement Statement</i>	<i>OCD Ref #</i>
DL-ORD-00000	The task shall run correctly.	DL-OCD-0000
:	:	:

The requirement number will always be of the form *DL-ORD-1234*, where 'DL' refers to the Data Lab, ORD is this document, and the number is unique. The requirement statement will contain a single requirement, the last column will contain the concept ID as given in the *Operations Concept Document*.

## 1.5 Referenced Documents

This document may reference additional documentation identified below.

- |                                   |       |
|-----------------------------------|-------|
| [1] Science Use Cases             | (SUC) |
| [2] Science Requirements Document | (SRD) |
| [3] Operational Concepts Document | (OCD) |
| [4] System Architecture Design    | (SAD) |
| [5] Project Execution Plan        | (PEP) |

## 2 General Project Requirements

This section specifies the requirements associated with the Data Lab project concepts described in the OCD.

### 2.1 Project Success Metrics

<b>DL-ORD-2100</b>	Access and usage logs shall be kept for all data services in the Data Lab. Individual components may implement additional logging specific to their functionality.	DL-OCD-2200
<b>DL-ORD-2101</b>	Usage reports shall be generated from usage logs to demonstrate use of the Data Lab to management and to track the most commonly used systems.	
<b>DL-ORD-2102</b>	Publications acknowledging or demonstrating use of Data Lab shall be tracked on the project website as an indication of community uptake.	DL-OCD-2205

### 2.2 System Documentation

#### 2.2.1 User Documentation

<b>DL-ORD-22100</b>	Data Lab shall provide appropriate end-user documentation for each task developed.	DL-OCD-5805
<b>DL-ORD-22100</b>	Data Lab applications shall provide inline documentation (e.g. mouse-over tooltips or a command-line option to print a help summary).	DL-OCD-5815 DL-OCD-5820

#### 2.2.2 Developer Documentation

<b>DL-ORD-22200</b>	Source code for all public programmatic interfaces shall be annotated using language-appropriate tools that permit automatic generation of API documentation.	DL-OCD-5835
<b>DL-ORD-22205</b>	Data Lab shall provide guidelines on required information documenting all user-facing compute or data services.	DL-OCD-5840
<b>DL-ORD-22210</b>	Data Lab shall provide documentation describing the interaction of Data Lab components with common desktop tools (e.g. Topcat, Aladin or DS9) as part of the user documentation.	DL-OCD-5845

### 2.2.3 Operator Documentation

<b>DL-ORD-22300</b>	Data Lab shall maintain a current list of services deployed on the project TWiki site.	DL-OCD-5850 DL-OCD-5851
<b>DL-ORD-22305</b>	Every service deployed to Data Lab hardware shall include a document describing how that service is configured, installed and deployed for use.	DL-OCD-5852
<b>DL-ORD-22310</b>	Data Lab shall provide documentation describing how to administer user accounts and groups.	DL-OCD-5853
<b>DL-ORD-22315</b>	Data Lab shall maintain a current list of software and system dependencies for each deployed service.	DL-OCD-5854
<b>DL-ORD-22320</b>	Data Lab shall provide documentation describing the procedures for monitoring the health and correctness of hardware systems.	DL-OCD-5855
<b>DL-ORD-22325</b>	Data Lab shall provide documentation describing the procedures for monitoring the health and correctness of data services.	DL-OCD-5855
<b>DL-ORD-22330</b>	Data Lab shall provide documentation describing the procedures for monitoring the health and correctness of compute services.	DL-OCD-5855

### 2.2.4 Document Repository

<b>DL-ORD-22400</b>	All system, user and developer documentation shall be maintained in a central documentation repository.	DL-OCD-5860
<b>DL-ORD-22405</b>	Contents of the document repository shall be visible from the project website, allowing access to documents without authentication.	DL-OCD-5861
<b>DL-ORD-22410</b>	The project TWiki site shall be used as a document repository for documents in development or which contain sensitive information.	DL-OCD-5862
<b>DL-ORD-22415</b>	Registered Data Lab users shall be able to upload documents (e.g. cookbooks or tutorials) to the document repository.	DL-OCD-5863
<b>DL-ORD-22420</b>	Data Lab shall store and distribute community-written documentation from the project Document Repository.	DL-OCD-5810

### 2.3 Source Repository

<b>DL-ORD-23000</b>	Data Lab developers shall use a public version control system for software management.	DL-OCD-5700 DL-OCD-5705
<b>DL-ORD-23005</b>	Data Lab shall maintain an internal repository for development purposes.	DL-OCD-5710 DL-OCD-5715
<b>DL-ORD-23010</b>	Data Lab developers shall use a private version control system for managing utility code, configuration files and sensitive user information.	

<b>DL-ORD-23015</b>	Data Lab shall provide long-term storage for operational data, e.g. backups, disk images, etc.	DL-OCD-5720
---------------------	--	-------------

## 2.4 Operations and Monitoring

<b>DL-ORD-23000</b>	Data Lab shall monitor the health of hardware systems.	DL-OCD-5900
<b>DL-ORD-23005</b>	Data Lab shall alert operators to unusual system loads or failures.	
<b>DL-ORD-23010</b>	Data Lab shall periodically validate that all deployed services are available and responding.	DL-OCD-5905
<b>DL-ORD-23020</b>	Data Lab services shall all implement logging to track usage and errors.	DL-OCD-5710
<b>DL-ORD-23025</b>	Usage/Error logs shall be available from a web-based status page.	

## 2.5 Help Desk and Community Engagement

<b>DL-ORD-25000</b>	Data Lab shall provide a support contact email address on the project web site.	DL-OCD-6500
<b>DL-ORD-25001</b>	Data Lab shall maintain a community support forum (details TBD).	
<b>DL-ORD-25002</b>	Data Lab shall maintain a public issue tracking and reporting system.	
<b>DL-ORD-25005</b>	The Data Lab Wiki shall be visible to all users as a documentation resource.	DL-OCD-6505
<b>DL-ORD-25010</b>	The Data Lab Wiki shall enforce write-permission restrictions to include only registered users as a means to avoid spam.	
<b>DL-ORD-25015</b>	Data Lab shall maintain a publically visible project website to serve as a source of information about the project and as the main entry point for registered users.	DL-OCD-6510

# 3 Science Capabilities

## 3.1 Catalog Query and Analysis Tools

<b>DL-ORD-31100</b>	Users shall be able to query and access catalog data without requiring login credentials.	N/A
<b>DL-ORD-31105</b>	When presented with login credentials, a service shall provide additional functionality using the resources tied to that user's id.	N/A

<b>DL-ORD-31110</b>	Users shall be able to query and access catalog data using an SQL syntax from a command-line tool.	<b>DL-OCD-2500</b>
<b>DL-ORD-31115</b>	Users shall be able to query and access catalog data using an SQL syntax from a web-based interface.	
<b>DL-ORD-31120</b>	Users shall be able to query and access catalog data using an SQL syntax from a programmatic interface.	
<b>DL-ORD-31125</b>	Users shall be able to query and access catalog data using VO protocols from a command-line tool.	<b>DL-OCD-2502</b>
<b>DL-ORD-31130</b>	Users shall be able to query and access catalog data using VO protocols from a web-based interface.	
<b>DL-ORD-31135</b>	Users shall be able to query and access catalog data using VO protocols from a programmatic interface.	
<b>DL-ORD-31140</b>	Users shall be able to do a positional cross-match of objects in a personal list with catalogs available in the Data Lab.	<b>DL-OCD-2504</b>
<b>DL-ORD-31145</b>	Users shall be able to do a positional cross-match of objects in a personal list with catalogs external to the Data Lab.	
<b>DL-ORD-31150</b>	Users shall be able to upload a local list of objects to the Data Lab cross-match service to be used in a positional cross-match.	
<b>DL-ORD-31155</b>	Users shall be able to specify to URI of a list of objects (e.g. a file in virtual storage) to the Data Lab cross-match service to be used in a positional cross-match.	
<b>DL-ORD-31160</b>	Users shall be able to retrieve a list of non-matches from a cross-match service.	
<b>DL-ORD-31165</b>	Users shall be able to save a query result to a personal database for further analysis.	<b>DL-OCD-2506</b>
<b>DL-ORD-31170</b>	Users shall be able to download a query result for further analysis.	<b>DL-OCD-2508</b>
<b>DL-ORD-31175</b>	Users shall be able to download a query result in a variety of formats suitable for analysis with common desktop tools and systems (e.g. VOTable, CSV, FITS BINTABLE and ASCII).	
<b>DL-ORD-31180</b>	Users shall be able to save a query result to virtual storage.	
<b>DL-ORD-31185</b>	Users shall be able to send a query result to a desktop application via SAMP messaging.	
<b>DL-ORD-31190</b>	Users shall be able to determine statistical properties of columns in a database beyond basic functions available in an SQL query.	<b>DL-OCD-2510</b>

### 3.2 Image Query and Analysis Tools

<b>DL-ORD-32010</b>	Users shall be able to query and access image data using VO protocols from a command-line tool.	<b>DL-OCD-2520</b>
<b>DL-ORD-32015</b>	Users shall be able to query and access image data using VO protocols from a web-based interface.	
<b>DL-ORD-32020</b>	Users shall be able to query and access image data using VO	



	protocols from a programmatic interface.	
<b>DL-ORD-32025</b>	Users shall be able to retrieve DES image data (raw and reduced) that covers a specific sky position from the NOAO Science Archive (NSA).	DL-OCD-2521
<b>DL-ORD-32030</b>	Users shall be able to query and retrieve images from VO image services external to the Data Lab.	DL-OCD-2522
<b>DL-ORD-32035</b>	Users shall be able to save images retrieved from any VO image service to their virtual storage.	
<b>DL-ORD-32040</b>	Users shall be able to download images retrieved from any VO image service to their desktop.	
<b>DL-ORD-32045</b>	Users shall be able to use Data Lab processing services (e.g. cutouts) on image data accessible from the Data Lab.	DL-OCD-2524 DL-OCD-2526 DL-OCD-2528 DL-OCD-2530

### 3.3 Time Series Tools

<b>DL-ORD-33000</b>	Users shall be able to generate light curves from catalogs containing multi-epoch data.	DL-OCD-2540
<b>DL-ORD-33005</b>	Users shall be able to use a Data Lab compute service to estimate the period of an object with variable flux.	DL-OCD-2541
<b>DL-ORD-33010</b>	Data Lab shall provide tools to determine if the flux of an object varies in time for a given statistical significance.	DL-OCD-2542
<b>DL-ORD-33015</b>	Data Lab will provide tools to determine the nature of flux variability (i.e. periodic, aperiodic, random or transient).	
<b>DL-ORD-33020</b>	Users shall be able to apply their own classification algorithms to light curves created in the Data Lab.	DL-OCD-2543

### 3.4 Visualization Tools

<b>DL-ORD-34000</b>	Users shall be able to display images accessible from Data Lab services using common desktop tools (e.g. Aladin or DS9).	DL-OCD-2550 DL-OCD-4270
<b>DL-ORD-34005</b>	Users shall be able to display images accessible from Data Lab services using a web browser.	
<b>DL-ORD-34010</b>	Users shall be able to display images accessible from virtual storage using common desktop tools (e.g. Aladin or DS9).	
<b>DL-ORD-34015</b>	Users shall be able to display images accessible from virtual storage using a web browser.	DL-OCD-2551
<b>DL-ORD-34020</b>	Users shall be able to create arbitrary 2-D (or 3-D) plots of tabular data available from Data Lab services.	
<b>DL-ORD-34025</b>	Users shall be able to create arbitrary 2-D (or 3-D) plots of tabular data available from virtual storage.	

<b>DL-ORD-34030</b>	Users shall be able to generate phase-folded light curves for a given period from tabular data.	DL-OCD-2552
<b>DL-ORD-34035</b>	Users shall be able to create animations from time-ordered stacks of image data.	DL-OCD-2553 DL-OCD-4285
<b>DL-ORD-34040</b>	Users shall be able to easily create Color-Magnitude diagrams from catalog data available from the Data Lab services (inc. virtual storage).	DL-OCD-2554
<b>DL-ORD-34045</b>	Users shall be able to easily create Hess diagrams (with optional contour overlay) from catalog data available from the Data Lab services (inc. virtual storage).	
<b>DL-ORD-34050</b>	Users shall be able to query another data access service using parameters derived from a graphical display.	DL-OCD-2555
<b>DL-ORD-34055</b>	Users shall be able to query another data access service using parameters derived from a tabular display.	
<b>DL-ORD-34060</b>	Users will be able to upload local data for use with Data Lab visualization tools.	DL-OCD-2556

### 3.4.1 2-D/3-D Plotting

<b>DL-ORD-34100</b>	Data Lab plotting tools shall support a variety of plotting styles (e.g. scatter plots, histograms, contours, fill area colors).	DL-OCD-4205 DL-OCD-4220
<b>DL-ORD-34105</b>	Data Lab plotting tools shall support use of error information when available.	DL-OCD-4210
<b>DL-ORD-34110</b>	Users shall be able to control attributes of a plot (window limits, axis labels, annotation, etc).	DL-OCD-4215
<b>DL-ORD-34115</b>	Users shall be able to create plots of the sky when appropriate.	DL-OCD-4230
<b>DL-ORD-34120</b>	Data Lab plotting tools shall support plotting of 3-D data when appropriate.	DL-OCD-4235
<b>DL-ORD-34125</b>	Plotting tools shall have the ability to save to a hardcopy disk format.	DL-OCD-4240

### 3.4.2 Image Display

<b>DL-ORD-34200</b>	Users shall be able to display multiple images (e.g. cutouts) in a single web page display.	DL-OCD-4280
<b>DL-ORD-34205</b>	A web page containing URLs to image data shall be able to load a local desktop tool's display with the image via messaging.	DL-OCD-4275

## 3.5 User-Defined Analysis

<b>DL-ORD-31500</b>	Users shall be able to execute their own analysis codes in the Data Lab.	DL-OCD-2570
---------------------	--	-------------

DL-ORD-31500	Data Lab shall provide programmatic interfaces to all Data Lab components for use by legacy software systems.	DL-OCD-2571
DL-ORD-31500	Data Lab shall provide methods by which legacy software can be deployed as a compute service within the Data Lab framework.	DL-OCD-2572
DL-ORD-31500	Users shall be able to publish reference data for personal use within the Data Lab having the same query/access interfaces as all other data holdings.	DL-OCD-2573

## 4 System Infrastructure

### 4.1 Large Catalog Support

DL-ORD-41000	Users shall be able to save results of a read-only database query to a personal database for further analysis.	DL-OCD-3105
DL-ORD-41005	Users shall be able to save results of a read-only database query to a disk file in virtual storage.	DL-OCD-3110
DL-ORD-41010	Users shall be able to save results of a read-only database query to a disk file in virtual storage.	DL-OCD-3111
DL-ORD-41015	Tuning and optimization of the DES catalog database shall be done prior to deployment of the public service.	DL-OCD-3115 DL-OCD-3120
DL-ORD-41020	Users shall be able to submit DES catalog queries for synchronous execution.	DL-OCD-3125
DL-ORD-41025	Users shall be able to submit DES catalog queries for asynchronous execution.	DL-OCD-3130
DL-ORD-41030	Client applications shall be able to query the status of an asynchronous job.	DL-OCD-3135
DL-ORD-41035	Users shall receive notification (method TBD) if any asynchronous job exceeds its execution time limit.	DL-OCD-3140

### 4.2 Virtual Storage

This section specifies the requirements for the Data Lab Virtual Storage component.

<b>Basic Functionality:</b> <i>Users must be able to upload, see, and move data to and from Data Lab Virtual Storage.</i>		
DL-ORD-4200	A user must have a simple means to copy data files from a local disk to the storage with a given name.	
	A user must be able see or list data files that have been transferred into the storage by name.	

	A user must be able to discover the size of files in the storage.	
	A user must be able to copy data files from the storage to a local disk.	
	A user must be able to organize data into different folders or directories in his space.	
	Virtual storage shall not corrupt data.	

<b>DL-ORD-42002</b>	Data Lab shall be able to allocate exclusive access to finite amounts of storage to individual users. Access shall be given only to authorized users.	
<b>DL-ORD-42004</b>	Users shall be able to access to their allocation after having authenticated through a login portal.	DL-OCD-3205
<b>DL-ORD-42006</b>	Users shall have a means to grant/deny other users access to files stored in their space.	
<b>DL-ORD-42008</b>	Users shall be able to request an alternate format of some types of files stored in the Data Lab. Specific formats are TBD.	DL-OCD-3210
<b>DL-ORD-42010</b>	Virtual storage containers shall have the ability to allow processing of data moved to the space.	DL-OCD-3215
<b>DL-ORD-42012</b>	Users shall have the ability to control which capability is enabled on a specific container	DI-OCD-3220
<b>DL-ORD-42014</b>	Users shall have the ability to enable/disable capabilities on containers.	
<b>DL-ORD-42016</b>	Users shall be able to execute a user-defined application automatically on data placed in virtual storage containers.	DL-OCD-3225
<b>DL-ORD-42018</b>	Users shall have the option to make data in virtual storage available through a standard access protocol appropriate for the type of data.	DL-OCD-3230
<b>DL-ORD-42020</b>	Users shall be able to make the data access service on a container a public data service requiring no special authorization.	
<b>DL-ORD-42022</b>	Users shall be able to access and manipulate contents of their virtual storage from their desktop.	
<b>DL-ORD-42024</b>	Users shall be able to access and manipulate contents of their virtual storage from a web browser.	DL-OCD-3235
<b>DL-ORD-42026</b>	Users shall be able to access and manipulate contents of their virtual storage from a programmatic interface.	
<b>DL-ORD-42028</b>	Users shall be able to access multiple instances of a Data Lab virtual storage service.	DL-OCD-3240
<b>DL-ORD-42030</b>	Users shall be able to move data between multiple instances of a Data Lab virtual storage service.	DL-OCD-3250
<b>DL-ORD-42032</b>	Users shall be able to download and install a Data Lab virtual storage service for use on local data.	DL-OCD-3245 DL-OCD-5205 DL-OCD-5210
<b>DL-ORD-42034</b>	Virtual storage space shall be subject to allocations and quotas	DL-OCD-3255

	based on the intended use of the space.	DL-OCD-3260
<b>DL-ORD-42036</b>	Data Lab operators shall have the ability to modify space allocations.	DL-OCD-3265 DL-OCD-5215 DL-OCD-5225
<b>DL-ORD-42038</b>	Virtual storage shall use a normal Unix filesystem as the backend storage to allow direct access and manipulation of data.	DL-OCD-5200 DL-OCD-5220 DL-OCD-5240
<b>DL-ORD-42040</b>	Virtual storage shall be expandable by adding new physical storage to the system.	DL-OCD-5220

### 4.3 Data Publication

#### 4.3.1 General Requirements

<b>DL-ORD-43000</b>	Users shall be responsible for converting data file into formats that can be ingested for publication by the Data Lab.	DL-OCD-3300 DL-OCD-5350
<b>DL-ORD-43005</b>	Users shall be responsible for providing all data collection metadata required for a public data service.	DL-OCD-5351
<b>DL-ORD-43010</b>	Users shall be able to publish catalog, image and spectral data through standard access protocols in the Data Lab.	DL-OCD-3305
<b>DL-ORD-43015</b>	Users shall be able to publish complex catalogs (i.e. databases with multiple tables) in a way that permits a query using the entire database schema.	DL-OCD-3310 DL-OCD-2500 DL-OCD-2502
<b>DL-ORD-43020</b>	Users shall be able to publish catalog data using only a parameterized query of fundamental metadata (e.g. spatial, temporal and bandpass information).	DL-OCD-3315 DL-OCD-2502
<b>DL-ORD-43025</b>	Users shall be able to connect higher-level data products (e.g. catalogs) to the original data that created them.	N/A
<b>DL-ORD-43030</b>	Users shall be able to create private data services with the same functionality as public data services.	DL-OCD-3325 DL-OCD-3230
<b>DL-ORD-43035</b>	Users shall be able to easily make private data services public.	DL-OCD-3330

#### 4.3.2 Collection Metadata

<b>DL-ORD-43050</b>	Registration (with a VO Registry for public discovery) shall only be required for public data services.	DL-OCD-5300
<b>DL-ORD-43055</b>	Data services shall be usable without being registered.	DL-OCD-5305
<b>DL-ORD-43060</b>	Data Lab shall use an existing public VO Registry of services during the initial development phase of the project.	DL-OCD-5325
<b>DL-ORD-43065</b>	Metadata extraction tools shall be extensible to produce the proper XML records required to register a data collection according to the latest standard.	DL-OCD-5330

<b>DL-ORD-43070</b>	Users wishing to create public data services through the Data Lab shall be responsible for providing all necessary metadata to properly register the data collection.	DL-OCD-3300
<b>DL-ORD-43075</b>	Data Lab shall provide tools and/or methods for collection metadata necessary to properly register a data collection.	DL-OCD-5335
<b>DL-ORD-43080</b>	Data Lab shall create a Naming Authority within an existing VO Registry for the purpose of registering data collections.	DL-OCD-5340
<b>DL-ORD-43085</b>	Data Lab shall be able to resolve a service URI to a service endpoint without relying on the VO Registry.	DL-OCD-5245
<b>DL-ORD-43090</b>	A URI that cannot be resolved within the Data Lab shall be passed to the VO Registry before returning a failure code.	DL-OCD-5246

### 4.3.3 Data Ingestion

<b>DL-ORD-43050</b>	Data Lab shall provide a means for users to transfer their data for publication.	DL-OCD-5353
<b>DL-ORD-43055</b>	Users shall provide catalog data in a format that can be easily loaded into a database (e.g. CSV, FITS BINTABLE, SQL dump).	DL-OCD-5360
<b>DL-ORD-43060</b>	Data Lab shall provide a tool to create the service-specific configuration file from user-supplied metadata.	DL-OCD-5361 DL-OCD-5372 DL-OCD-5382
<b>DL-ORD-43065</b>	Users shall provide image data in a format that can be returned to users directly.	DL-OCD-5370 DL-OCD-5380
<b>DL-ORD-43066</b>	User-supplied image data shall pass basic FITS validation.	
<b>DL-ORD-43070</b>	Data Lab shall provide guidance on required/optional FITS keywords for an image data service.	DL-OCD-5371
<b>DL-ORD-43075</b>	Data Lab shall provide guidance on supported FITS conventions (e.g. WCS, inheritance).	
<b>DL-ORD-43085</b>	Data Lab shall provide guidance on supported spectral data formats (tabular, native FITS, etc).	DL-OCD-5381

### 4.3.4 Metadata Harvesting

<b>DL-ORD-43050</b>	Data Lab shall provide guidelines and tools for users to specify the metadata mapping needed by their data.	DL-OCD-5397
<b>DL-ORD-43055</b>	Metadata tools shall be useful as a means of validating a dataset meets all requirements before publication.	DL-OCD-5398
<b>DL-ORD-43060</b>	Users shall be able to download and run harvesting tools on their local machine to verify their data before transferring to Data Lab for ingestion.	DL-OCD-5399

#### 4.4 Processing Services

<b>DL-ORD-44000</b>	A processing service shall be able access files in Data Lab virtual storage when necessary.	DL-OCD-3400 DL-OCD-5400 DL-OCD-5410
<b>DL-ORD-44005</b>	A processing service shall be able to access data referenced by a URL when necessary.	DL-OCD-5410
<b>DL-ORD-44010</b>	A processing service shall be executable outside the Data Lab framework (e.g. as a command-line tool).	DL-OCD-5405
<b>DL-ORD-44215</b>	Users shall be able to submit a job to a processing service as a synchronous task.	DL-OCD-3410
<b>DL-ORD-44020</b>	Users shall be able to submit a job to a processing service as an asynchronous task.	DL-OCD-3415
<b>DL-ORD-44025</b>	Users shall be able to invoke multiple instances of a processing service in order to parallel process lists of items.	DL-OCD-3420
<b>DL-ORD-44030</b>	Users shall be able to create and deploy processing services for their own use.	DL-OCD-3425
<b>DL-ORD-44035</b>	Users shall be able to access Data Lab processing services from their desktop.	DL-OCD-5415
<b>DL-ORD-44040</b>	Users shall be able to view the status of all submitted jobs.	DL-OCD-3415
<b>DL-ORD-44045</b>	Users shall be able to view the error message of a failed job.	

#### 4.5 Machine and Process Virtualization

<b>DL-ORD-45000</b>	Data Lab shall deploy virtual machines when necessary.	DL-OCD-3500
<b>DL-ORD-45000</b>	Virtual machines shall be transportable between physical machines to allow operational load balancing.	DL-OCD-3505
<b>DL-ORD-45000</b>	Data Lab shall maintain a list of base operating system images from which standard virtual machines can be built.	DL-OCD-3510
<b>DL-ORD-45000</b>	Data Lab shall permit the use of Linux Containers for lightweight task virtualization.	DL-OCD-3525
<b>DL-ORD-45000</b>	Data Lab shall permit multiple instances of a specific container to be run as a means of increasing compute capacity.	DL-OCD-3530
<b>DL-ORD-45000</b>	Users shall be able to download and execute containerized processes on their local desktop.	DL-OCD-3535 DL-OCD-5505
<b>DL-ORD-45000</b>	Users shall have access to interactive shells and persistent storage in containers running interactive tasks.	DL-OCD-3540 DL-OCD-5500 DL-OCD-5515
<b>DL-ORD-45000</b>	Users shall be able to create containerized tasks for use in the Data Lab.	DL-OCD-5505 DL-OCD-5510
<b>DL-ORD-45000</b>	Users shall be able to share containerized applications with other users.	DL-OCD-3425

## 4.6 User/Group Authentication

<b>DL-ORD-46000</b>	Data Lab shall not require login authorization to access public data services.	DL-OCD-3600
<b>DL-ORD-46005</b>	Users shall be able to use an NOAO identity to create a Data Lab identity.	DL-OCD-3605 DL-OCD-5600
<b>DL-ORD-46010</b>	Users shall be able to create a Data Lab identity from a web login portal.	DL-OCD-3610
<b>DL-ORD-46015</b>	An NOAO identity shall not be required to create a Data Lab identity.	DL-OCD-3615
<b>DL-ORD-46020</b>	A login portal shall allow users to view and modify details of their account (e.g. contact information, password, etc).	DL-OCD-3620
<b>DL-ORD-46025</b>	Users shall be able to logically connect their NOAO identity to their Data Lab identity.	DL-OCD-3625
<b>DL-ORD-46030</b>	Users shall be able to create groups of other users with whom they can share data and resources.	DL-OCD-3630
<b>DL-ORD-46035</b>	Users shall be able to independently control the access permissions for their data and resources.	DL-OCD-3635
<b>DL-ORD-46040</b>	Users shall be able to control access permissions using user or group names.	DL-OCD-3640
<b>DL-ORD-46045</b>	Users shall be able to transfer the administrator role of a group to another member of the group.	DL-OCD-3650
<b>DL-ORD-46050</b>	Data Lab operators shall be able to allocate and modify quotas associated with an individual or group resource.	DL-OCD-3640
<b>DL-ORD-46055</b>	Data Lab operators shall be able to create, manage and delete users and groups.	DL-OCD-5605 DL-OCD-5610 DL-OCD-5615
<b>DL-ORD-46060</b>	Data Lab identity certificates shall be compatible with similar services deployed by other data centers (e.g. VOSpace)	N/L
<b>DL-ORD-46070</b>	Data Lab shall enable logging of authorization services for the collection of usage statistics.	DL-OCD-2200
<b>DL-ORD-46075</b>	Data Lab logging shall be sufficient to determine when failures are due to user error (e.g. wrong password) or a service failure.	N/A

## 5 System Interfaces

### 5.1 Programmatic Interfaces

<b>DL-ORD-51000</b>	Data Lab shall implement client-side interfaces only for core components and services. Application developers will use 3 <sup>rd</sup> party libraries for all other standard interfaces (e.g. http,	DL-OCD-3700 DL-OCD-3705 DL-OCD-3710
---------------------	--	---



	database accee	DL-OCD-3715
<b>DL-ORD-51005</b>	Interfaces developed in the Data Lab shall be available for user download.	DL-OCD-3720
<b>DL-ORD-51010</b>	Data Lab shall not require a specific language for implementation of services or applications.	DL-OCD-3725
<b>DL-ORD-51015</b>	Data Lab shall not guarantee support for all programming languages.	DL-OCD-3730
<b>DL-ORD-51020</b>	All programmatic interfaces shall be fully documentated.	DL-OCD-3740

## Appendix I: Vocabulary / Acronyms Used

AAS	(American Astronomical Society)
ADASS	(Astronomical Data Analysis Software and Systems) conference
ADQL	(Astronomical Data Query Language) An SQL-like language which includes astronomical facilities to query a database.
AGN	(Active Galactic Nucleus)
API	(Application Programming Interface) The documentation of the interface to a software library or tool.
ASCII	(American Standard Code for Information Interchange) A character-encoding scheme based on the English alphabet where 128 specific characters are encoded into 7-bit binary integers.
ASV	(ASCII Space Values)
AURA	(Association of Universities for Research in Astronomy)
CADC	(Canadian Astronomy Data Centre)
CDS	Centre de Données astronomiques de Strasbourg
CMD	(Color Magnitude Diagram)
CSV	(Comma Separated Values)
CTIO	(Cerro Tololo Inter-American Observatory)
DAL	(Data Access Layer) The VO protocols that define how VO applications access data resources.
Datalink	VO protocol for associating complex astronomical data
DECaLS	DECam Legacy Survey
DECam	(Dark Energy Camera) A 520 megapixel digital camera on the Blanco 4-m telescope at CTIO.
DES	(Dark Energy Survey) a survey to probe the origin of the accelerating Universe and help uncover the nature of dark energy by measuring the 14 billion-year history of cosmic expansion with high precision over five years beginning in summer 2013.
DESI	(Dark Energy Spectroscopic Instrument) An instrument to measure the effect of dark energy on the expansion of the universe by obtaining optical spectra for tens of millions of galaxies and quasars (beginning 2018).
DESDM	(Dark Energy Survey Data Management) Project that developed and operates the DESDM system at NCSA.
DL	(Data Lab)
DAOPHOT	Package for crowded field stellar photometry.
Docker	An open platform for developers and system administrators to build, ship, and run distributed applications.
DoPHOT	CCD PSF fitting photometry program.
DS9	SAOimage DS9, an astronomical imaging and data visualization application.
DSS	(Digitized Sky Survey)
ESO	(European Southern Observatory)
FITS	(Flexible Image Transport System) An open standard defining a digital file format for storage, transmission, and processing of astronomical (and other scientific) data.
FTP	(File Transfer Protocol) A standard network protocol used to transfer computer files from one host to another host over a TCP-based network.

FUSE	(FileSystem in User Space) An operating system mechanism that lets non-privileged users to create their own file systems.
GAVO	(German Astrophysical Virtual Observatory)
GMS	(Group Management Services)
GPFS	(General Parallel File System) A high-performance clustered file system developed by IBM
Hess diagram	Plots the relative density of the occurrence of stars at different color-magnitude positions of Hertzsprung-Russell diagram for a given galaxy.
HSB	(High Surface Brightness)
HST	(Hubble Space Telescope)
HTTP	(HyperText Transfer Protocol) An application protocol for distributed, collaborative, hypermedia information systems.
IDL	(Interactive Data Language) A programming language used for data visualization and analysis.
IPAC	(Infrared Processing and Analysis Center)
IRAF	(Image Reduction and Analysis Facility) NOAO image reduction/analysis and visualization software system.
IVOA	(International Virtual Observatory Alliance) The international VO community responsible for developing VO standards.
JIRA	A commercial tool for software teams to plan, build, and track projects.
JPEG	(Joint Photographic Experts Group) Lossy compression for digital images
LDAP	(Lightweight Directory Access Protocol) An industry standard application protocol for accessing and maintaining distributed directory information services over an Internet Protocol (IP) network.
LSB	(Low Surface Brightness)
LMC	(Large Magellanic Clouds)
LSST	(Large Synoptic Survey Telescope)
MAST	(Mikulski Archive for Space Telescopes)
MPC	(Minor Planet Center)
MCs	(Magellanic Clouds)
MySQL	Popular open source database.
MyDB	Simple database wrapper for MySQL.
NASA	(National Aeronautics and Space Administration)
NCSA	(National Center for Supercomputing Applications)
NHPPS	(NOAO High-Performance Pipeline System) An event-driven, multi-process executor system developed to manage pipeline applications in a coarse-grained, distributed processing environment.
NOAO	(National Optical Astronomy Observatory)
NSA	(NOAO Science Archive)
NSSDC	(NOAO System Science and Data Center)
OCD	(Operational Concept Document)
ORD	(Operational Requirements Document)
OS	(Operating System)
PCMS	(Positional Cross-Match Service)
PNG	(Portable Network Graphics) Raster graphics file format that supports lossless data compression.
PSF	(Point Spread Function)
QServ	The LSST database management system.
R	A programming language and software environment for statistical computing and graphics.

RDBMS	(Relational DataBase Management System) A DBMS that represents data using a relational database.
Relational database	A database that stores data in a structure consisting of one or more tables (aka relations) of rows and columns, which may be interconnected.
ReST	(Representational State Transfer) An approach to web services that uses the standard HTTP GET and POST protocols.
SAD	(System Architecture Design) document
SAMP	(Simple Applications Messaging Protocol) A VO protocol for desktop messaging.
SCS	(Simple Cone Search)
SDM	(Science Data Management) group
SDSS	(Sloan Digital Sky Survey)
SED	(Spectral Energy Distribution) Plot of brightness of flux density versus frequency or wavelength.
SEtractor	A program that builds a catalogue of objects from an astronomical image.
SIA/SIAP	(Simple Image Access Protocol) A VO protocol that supports queries for images available in a given data collection near a given position on the sky.
SMASH	(Survey of the MAgellanic Stellar History) PI: Nidever
SMC	(Small Magellanic Cloud)
SN	(Super Nova)
SQL	(Structured Query Language) The standard language used to communicate with RDBMS's.
SQLite	A software library that implement a self-contained, serverless, zero-configuration, transactional SQL database engine.
SRD	(Science Requirements Document)
SSh	Secure Shell
SSA	(Simple Spectral Access) A VO protocol for spectral query/retrieval.
SSO	(Single Sign-On)
SUC	(Science Use Case) document
SVC	An abbreviation for a Web service.
SWIG	(Simplified Wrapper and Interface Generator) An open source software tool used to connect C or C# programs or libraries with scripting languages.
TAP	(Table Access Protocol) A VO protocol for querying remote databases.
TCP	(Transmission Control Protocol) One of the core protocols of the Internet protocol suite, commonly referred to as TCP/IP.
TSV	(Tab-Separated Values) A simple file format often used to move tabular data between computer programs that support the format, e.g., transferring information from a database program to a spreadsheet.
URI	(Uniform Resource Identifier) An address standard for a resource available on the Internet.
URL	(Uniform Resource Locator) The global address of documents and other Resources on the World Wide Web. The address contains 2 parts: specification of the protocol to be used in accessing the resource and its network location.
UWS	(Universal Worker Service) pattern defines how to manage asynchronous execution of jobs on a service.
VAO	(Virtual Astronomical Observatory) The US VO project.
VM	(Virtual Machine)
VO	(Virtual Observatory)

VOSI (VO Support Interfaces) The minimum interface that a SOAP or REST-based web service requires for compatibility with the IVOA.

VOSpace The IVOA interface to distributed storage that specifies how VO agents and applications can use network attached data stores to persist and exchange data in a standard way.

XML (eXtensible Markup Language)

2MASS-PSC (2 Micron All Sky Survey – Point Source Catalog)