

The NOAO Data Lab: Science-Driven Development

Michael J. Fitzpatrick^a, Matthew J. Graham^{a,b}, Kenneth J. Mighell^a, Knut Olsen^a, Patrick Norris^a, Stephen T. Ridgway^a, Elizabeth B. Stobie^a, Adam S. Bolton^a, Abhijit Saha^a, and Lijuan (Wendy) Huang^a

^aNational Optical Astronomy Observatory, 950 N Cherry Ave, Tucson, AZ 85719, USA

^bCalifornia Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA

ABSTRACT

The NOAO Data Lab aims to provide infrastructure to maximize community use of the high-value survey datasets now being collected with NOAO telescopes and instruments. As a science exploration framework, the Data Lab allow users to access and search databases containing large (i.e. terabyte-scale) catalogs, visualize, analyze, and store the results of these searches, combine search results with data from other archives or facilities, and share these results with collaborators using a shared workspace and/or data publication service. In the process of implementing the needed tools and services, specific science cases are used to guide development of the system framework and tools. The result is a Year-1 capability demonstration that (fully or partially) implements each of the major architecture components in the context of a real-world science use-case. In this paper, we discuss how this model of science-driven development helped us to build a fully functional system capable of executing the chosen science case, and how we plan to scale this system to support general use in the next phase of the project.

Keywords: Science exploration framework, remote execution, data services, application containers

1. INTRODUCTION

The conceptual design of the Data Lab was described earlier,¹ here we review the goals and motivation for the project and provide context for the science demonstration.

1.1 Data Lab Overview

The overall goal of the NOAO Data Lab is to enable efficient exploration and analysis of the large datasets now being generated by instruments on NOAOs wide-field 4-m telescopes. To meet this goal, the Data Lab will support four approaches to science with large datasets:

1. **Catalog Science** – Under this approach, the given science program can be accomplished using only the data from the catalog, with no need for access to the pixels. The Data Lab will support this approach by providing access to large catalog databases, a database query interface, personal database storage space, and basic catalog visualization capabilities.
2. **Data Exploration** – Under this approach, the science program begins with an exploration of the contents of the available datasets. This exploration could involve overlaying select catalog measurements on images, comparison of catalog measurements with external data sources, or prototyping pixel-level analysis using a sample of catalog-selected objects. The Data Lab will support such exploration by providing an efficient image cutout service, a cross-match service, and additional visualization capabilities.
3. **User-defined Custom Workflows** – More complex science cases will require processing or analysis steps in addition to simple catalog queries and image cutouts. The Data Lab will support such workflows by providing a compute service, the ability to interface with users legacy software, and virtual machines and containers to run custom software within the Data Lab environment.

Further author information: (Send correspondence to M.J.F.)

M.J.F.: E-mail: fitz@noao.edu, Telephone: 1 520 318 8387

Software and Cyberinfrastructure for Astronomy IV, edited by Gianluca Chiozzi,
Juan C. Guzman, Proc. of SPIE Vol. 9913, 99130L · © 2016 SPIE
CCC code: 0277-786X/16/\$18 · doi: 10.1117/12.2233791

4. **Collaborative Research** – For teams, the work of exploring and analyzing the large catalogs and associated pixels will often need to be divided among several users, and intermediate steps and results shared between them. The Data Lab will support such work by having virtual storage accessible to lists of specified users and by providing virtual machines to collaborations with a single login.

In order to enable these scientific approaches, Data Lab will rely on the following major design components:

- **Large Catalogs** - Data Lab will serve multiple, TB-scale databases,
- **Pixel Data** - Data Lab will connect users to images and spectra in NOAO Science Archive,
- **Virtual Storage** - Data Lab minimizes data transfer and maximizes data sharing,
- **Visualization** - Data Lab will enable interactive data exploration.
- **Compute Processing** - Data Lab will allow workflows to run close to the data,
- **Additional features** - Data Lab will allow users to published their own datasets and access external data/compute services, export workflows to run at larger-scale facilities, distribute software for use by others.

1.1.1 Architecture Overview

A major feature of the science demonstration is that it relies on nearly all of the components described in the full system architecture. With this core implementation complete to a basic level of functionality, the emphasis of future development will focus on improving reliability and scalability of the system for general use.

The components of the complete Data Lab architecture are shown in Fig. 1 below. New development is centered primarily on the middleware components that provide high-level interfaces to services. Although legacy clients can access storage and data services using the low-level protocols only (primarily Virtual Observatory VOSpace² and TAP (Table Access Protocol³)), we expect that astronomers will prefer the high-level interfaces that hide the complexity of the underlying protocols when creating their own scripted workflows.

Year-2 of the project will include implementation of the operations components of the system (e.g. logging and monitoring tools), stability issues (e.g. failure recovery of services and fixing memory leaks), scalability of tools and services (e.g. deployment across a planned expansion of hardware to support data and compute resources), as well as documentation and packaging work required to support a general release of client- and server-side software. The recent addition of three Data Scientists to the project will allow focused development of new science capabilities and characterization of data holdings to document the suitability of specific datasets for given science cases.

1.2 NOAO Data Services

Like many observatories, NOAO has seen an explosion in its volume of data in recent years, driven by bigger instruments, more ambitious survey programs, and the desire of programs at all scales to make available their derived data products. In our case, the deployment of the Dark Energy Camera (DECam) on the CTIO 4m for use in the Dark Energy Survey (DES⁴) in 2013 greatly increased volume of image data available from our archive in a very short time (See Fig. 2). The result is effectively a growing crowd-sourced all-sky survey, albeit one with a heterogeneous set of observations (the DES survey can be seen in the lower right of Fig. 2, the depth of coverage can be seen compared to SDSS in Fig. 3).

Aside from a growth in image data, major survey programs such as DES* and DECaLS[†] will deliver large uniform catalogs (on order of tens of Terabytes in size and hundreds of millions of objects) derived from the pixel data. Similarly, the wide-field coverage of DECam makes it an excellent survey instrument and various survey programs will deliver significant catalogs as well. While delivery of the largest catalogs is still pending,

*<http://www.darkenergysurvey.org/>

†<http://legacysurvey.org/decamls/>

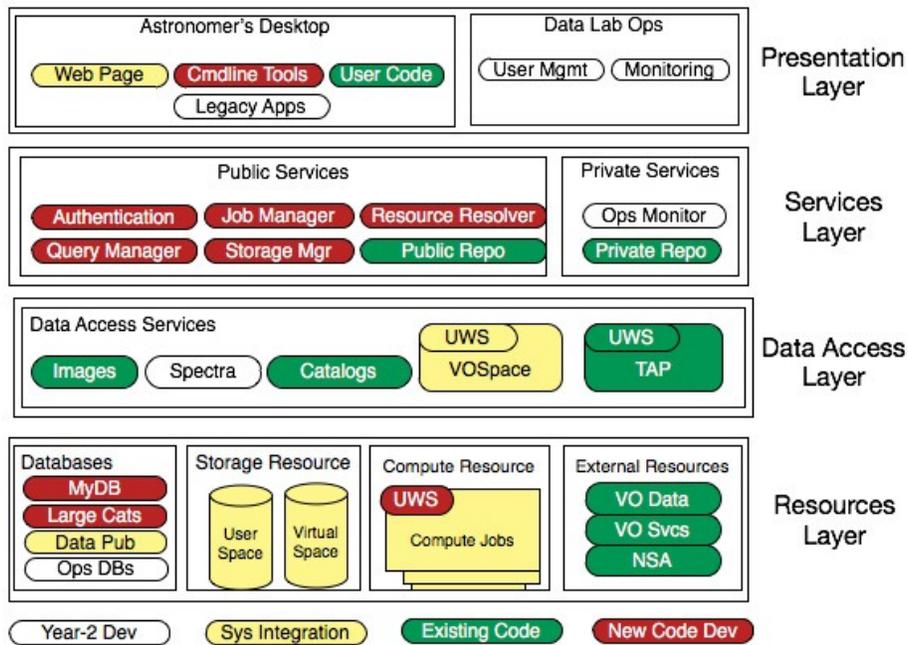


Figure 1. Layer-cake diagram of the Data Lab Year-1 architecture. Uncolored boxes are components not needed until public release (mid-2017). Blue boxes represent existing software, yellow boxes are existing code that requires system integration, the remainder are the only components we develop from scratch.

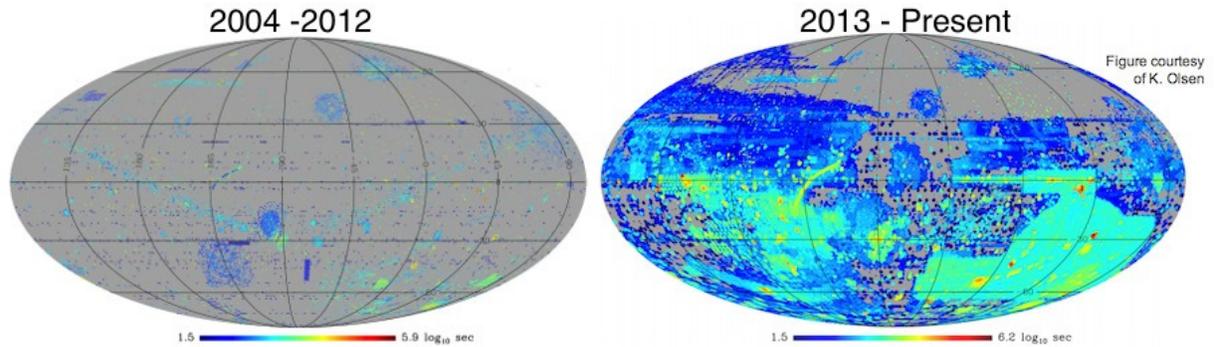


Figure 2. Coverage maps showing the image data available from NOAO Science Archive 2004-2012 (left image) and coverage 2013-present after the deployment of DECam (right image).

we are already starting to see initial data releases from the survey programs that are being used as development datasets. In particular, The SMASH survey,⁵ is not only providing the data for the satellite-finding Data Lab demo, but is beginning to use the Data Lab infrastructure as part of the survey team's ongoing analysis.

Table 1 summarizes the various datasets currently being used in developing the science cases and backend services. Part of the shift towards operational readiness will involve refactoring these databases for improvements in efficiency. Although the current database implementation is imperfect, the range in size and scope of these tables has proved invaluable in developing an understanding of the level of curation required for user-provided datasets, and in developing prototype tools needed to ingest a heterogeneous set of tables.

Table 1. Summary of data tables available from the Data Lab.

Dataset	No. Tables	No. Objects (Millions)	No. Meas. (Millions)	Size (GB) data/index
Bulge Survey, Field B5	4	3.55	699.3	111.8 / 60.6
DES SVA1	24	25.2	100.8	117.3 / 2.39
Dwarfs	6	1.45	7.74	3.3 / 0.57
DECaLS DR2	10	288.3	1,729.8	705.2 / 205.5
SMASH	9	411.4	3,583	495. / 539.7
DES Y1 S/N Fields	5	0.274	14.03	1.74 / 0.33
USNO A2 + B1	2	1,571	3,142	28.9 / 98.6
NSA Image Service	6	1.23 (exposures)	64.5 (ccds)	98.26 / 26.8

2. SCIENCE DEMONSTRATIONS

The first major milestone for the Data Lab project was a presentation of the complete science demo at the 228th AAS meeting held in San Diego in June 2016. This demonstration was intended to showcase the prototype system and to show how users can proceed to:

1. Discover a dataset appropriate for a particular science case,
2. Interact with a subset of the data (e.g. explore a small region of sky) to develop an intuition about the analysis procedure to follow based on the data available,
3. Use the intuition gained to develop a scripted workflow to operate automatically upon the full dataset.

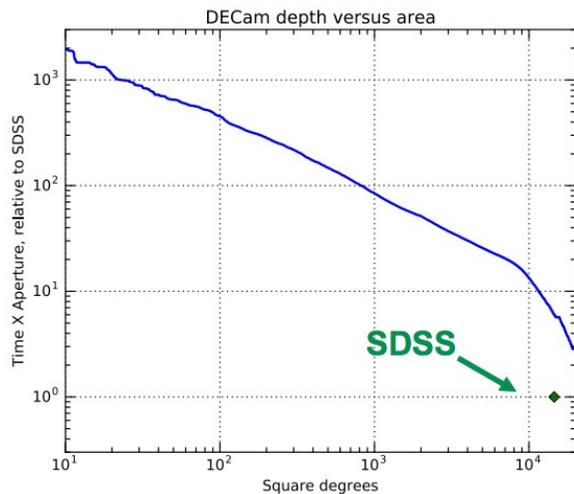


Figure 3. A plot of the spatial and depth coverage of images currently available from the NOAO Science Archive.

would also satisfy our goal of exercising the full system. The final list of science-cases we detailed during the design phase of the project is shown in Table 2.

In the end, we decided on the satellite-galaxy demo because we had a concrete science case with the discovery of the Hydra II dwarf by the SMASH team, participation by that survey team in the ongoing development of

While these steps apply generally to *any* science case, we use them in our chosen use-case of discovering faint Milky Way satellite galaxies as our primary science demonstration, using Data Lab infrastructure to show how the search might be conducted across increasingly large or diverse datasets. The bulk of the demo highlights the use of multiple tools that allow for efficient analysis on existing large catalogs (and their associated pixel data). An extended version of the demo shows how catalogs can be generated from PI-level image datasets to carry out a satellite search in areas of the sky not covered by survey catalogs or as part of a dedicated observing program.

During the project design phase, numerous science cases covering solar system, galactic and extra-galactic programs were explored. These were used to define the boundaries of data volume, analysis workflows, and data services that the Data Lab is expected to support (they map well to observing programs proposed to NOAO facility instruments) and several of these

discovery portal is intended to be a jumping off point for constructing a detailed query to the data services, but does not serve as the primary interface to those services.

2.2 Key Science Demonstration Overview

The demo begins by reproducing the discovery of the Hydra II dwarf galaxy in the SMASH (Survey of the MAgellanic Stellar History) dataset. Using both interactive and scripted approaches, the workflow here uses a mix of legacy and custom tools and relies on various parts of the Data Lab infrastructure to provide needed back-end services:

- Catalog access through TAP (Table Access Protocol) and custom Data Lab *query* services,
- Retrieval of associated image data from the NOAO Science Archive using image-cutout services to reduce the amount of data to process,
- Use of Virtual Storage services to make the process efficient by saving intermediate data to storage resources close to the data,
- Custom time-series analysis tools to help confirm candidate objects.

The first step in the demo is to give the user an interactive tour of how to discover a dwarf galaxy in a photometric catalog. We then show how Data Lab services link catalog and image data, give efficient access to time series data, and how the workflow results in a matched filter appropriate for automated use on a wider dataset.

2.3 Finding Milky Way Satellites

The search for new dwarfs involves two related techniques: 1) a catalog-based search for over-densities of point sources with selected colors and magnitudes, and 2) a search for peaks of unresolved light on spatial scales expected for Local Group dwarfs coupled with a measurable over-density of stellar sources. For the second approach, we assume that the users begins with a catalog-based search for stellar over-densities, and then searches for peaks of unresolved light in image cutouts. The alternate approach of searching for candidate dwarfs starting with the images and then proceeding to confirm them by looking for associated sources in the catalog is also valid, but much more resource intensive as it requires processing of the imaging dataset.

2.3.1 Overview of demo procedure

Using a photometric catalog as a starting point, the basic procedure is to identify spatial over-densities in the catalog (relative to surrounding areas) with characteristics of a given stellar population. Various filters can then be applied to the data to remove background galaxies (e.g. using shape information in the catalog) and/or apply color and magnitude cuts to isolate objects representative of the age and distance of the population in question (e.g. by identifying a stellar locus in a color-color diagram, creating notched filters in a color-magnitude diagram, etc). When querying a data service (such as those from Data Lab) these filters may be applied directly to reduce the amount of initial data to be searched; this approach works well in an automated search provided the filters aren't described by some complex shape in color-magnitude space (e.g. multiple or irregular polygons).

Once the data have been sub-selected, candidate over-densities can be identified by further isolating main sequences in a color-magnitude diagram that are also co-located spatially. More detailed analysis to identify RR Lyrae variables that can be used to determine a distance modulus, a search for peaks of unresolved light using image cutouts from the parent dataset, reprocessed photometry of stacked images in the region, or cross-matching against a list of known objects (e.g. globular clusters) can all be used to determine whether the candidate region is of further interest.

2.3.2 Interactive exploration and discovery of Hydra II dwarf

The interactive portion of our demo roughly follows the procedure outlined above. We use the TOPCAT[¶] table visualizer as our main interface because it has the ability to use the Data Lab TAP service to query for data as well as built-in plotting tools which can be used to interactively explore the data in a search for candidate fields. As the Data Lab evolves we expect additional custom clients will be developed, or that users will be able to use the Data Lab interfaces to load data into specialized legacy tools service a similar purpose.

The first step in the process is to use the TOPCAT TAP query panel to select catalog data for a particular field of the SMASH survey (one in which the known Hydra II dwarf exists). The query used here is similar to that which might be used from a scripted environment and provides us with an opportunity to discuss how users can explore the data tables to develop more sophisticated queries. Figure 5 shows an example of the plots used to graphically apply filters to the data when starting from a full catalog of the field. Users are led through the iterations to demonstrate how a matched filter can be created to highlight the candidate over-density (shown in the middle of the spatial plots on the left).

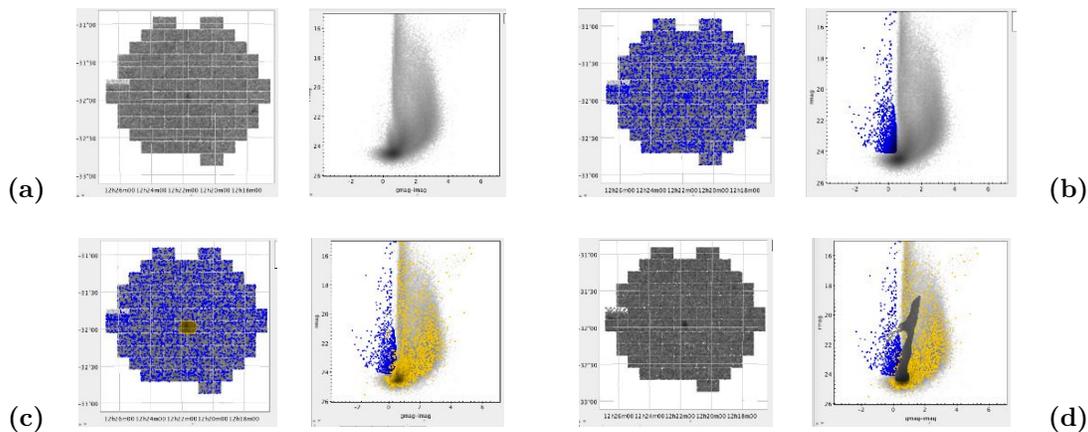


Figure 5. Effects of data filters on the detectability of over-densities. We see the full dataset (a), a cut made to select blue objects ($g-r < 1.0$) brighter than 23 mag to exclude galaxies (b), a selection on the spatial plot to further refine the CMD (c), and finally a hand-drawn main-sequence filter to isolate the candidate region.

The data subsets created at each iteration can be selectively removed from the plot to reveal only the points associated with the final filtered dataset. At this stage in the demo we use the TOPCAT ability to broadcast messages to other desktop tools (i.e. the *Simple Application Messaging Protocol*, SAMP (ref)) to examine points in the CMD plot by broadcasting the associated table row to a custom client that queries the Data Lab for, and displays, the multi-filter light curves of that object. In the case of Hydra II, a single RR Lyrae associated with the dwarf is found in the instability strip which was used by Vivas⁶ to confirm the discovery of a new dwarf. This light-curve tool is discussed in more detail below.

2.3.3 Automated discovery of Hydra II dwarf

A scripted application of this technique is an efficient way to search large areas of sky (e.g. the current SMASH data holdings comprise 184 individual fields) for candidate over-densities, but is unlikely to lead to truly automated discovery without an expert astronomer to confirm the result. In this approach, color and magnitude cuts of the data will be applied in the database query for efficiency (returning a 10-20x reduction in the size of the query result). Various density-finding algorithms can then be applied to the data to locate a candidate region within a field; and in the course of development, several algorithms were tried with varying levels of success. The differential convolution method of Koposov⁷ has proven to be most effective at allowing us to define search parameters consistent with the expected half-light radius of a dwarf at a given distance (as opposed to more general techniques such as *Kernel Density Estimation*, KDE).

[¶]<http://www.star.bris.ac.uk/~mbt/topcat/>

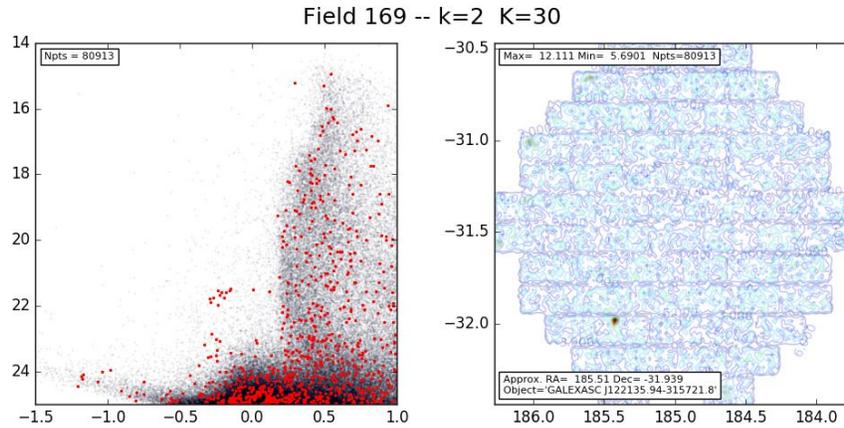


Figure 6. Plots generated automatically using the differential convolution technique, applied to the original Hydra II discovery field. Points within 0.1 deg of the peak over-density found in the clipped histogram are overplotted on the main CMD to highlight the CMD of the local region.

In the differential convolution method, a 2-D histogram consisting of 1-arcmin bins is created from the points in the field. This histogram is convolved with a *small* and a *big* kernel and the difference computed. The result is then further filtered to select only those bins containing values that are some set limit above the mean bin density. Using this approach, an image of the final clipped histogram quite clearly shows any over-density (matching the chosen kernel sizes) present in the field. Figure 6 shows the plots automatically created using this technique on the original Hydra II discovery field. The main sequence isolated interactively above is visible in the left CMD plot, and the over-density corresponding to Hydra II is clearly present in the lower-left of the right histogram plot.

In this example, convolution kernel sizes were chosen based on the known parameters of Hydra II, however it serves to validate the technique. A script to find similar objects in the remaining SMASH fields is trivial, as is running the script over a range of kernel sizes or clipping limits to detect other objects of different size. The Data Lab compute services can be used to execute the script close to the data to minimize transfers, results can be saved to virtual storage (or indeed the script can be triggered to run automatically on query results saved to virtual storage using the capability mechanism⁸ and the summary plot for each field reviewed quickly for candidate over-densities requiring further analysis at some later time. Likewise, implementation of alternative density-finding algorithms can be applied and similar summary plots generated with a new script.

2.3.4 User-generated catalogs as a starting point

The science demo until this point has used existing catalogs as the starting point for the workflow, however large catalogs such as those from DES and the final catalog from the DECaLS and other Survey programs are still in preparation. As seen in Figure 2, the area of sky covered by reduced image data is considerably larger than what is currently available in the form of source catalogs. The heterogeneous nature of the image data make processing this data into an all-sky catalog (even if resources were available to do so) impractical, however the creation of source-detection catalogs for the purpose of identifying candidate over-densities is more reasonable.

A SExtractor-based pipeline tool (called *NIKE*) was developed for the purpose of creating exploratory catalogs from archive DECam image data. While these catalogs generally don't go as deep as those produced by more detailed pipelines, they were adequate to recover all known dwarfs for which image data is available (see Figure 7) using the convolution techniques described above. The major advantage of the NIKE pipeline is its speed: catalogs of full DECam exposures can be generated in a few minutes and individual CCDs can be processed fast enough to support interactive web services.

To provide a training set for experimenting with other detection techniques, NIKE catalogs were ingested into the Data Lab TAP service to provide a catalog-based starting point. Jupyter notebooks are used to further

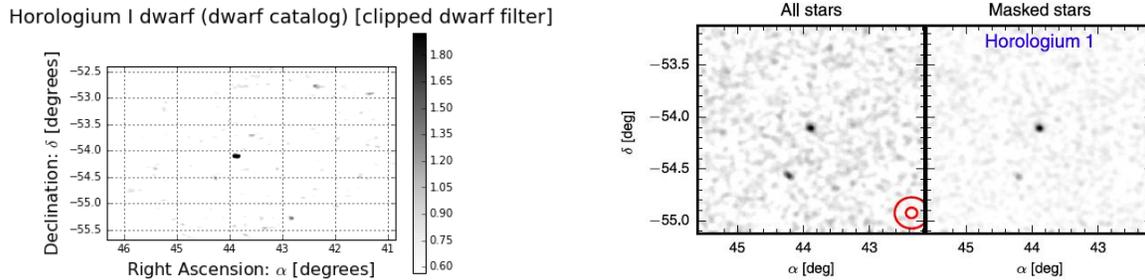


Figure 7. Detection plots for the Horologium I dwarf using catalogs generated by the NIKE pipeline (left) compared with the figure from the discovery paper by Koposov et al (2015) (right).

explain the convolution technique using real data as well as to execute the pipeline on data selected from the image archive. The combination of notebooks then serves as the basis for a scripted application that could be used to search large areas of the sky not already covered by catalog data.

2.4 Rediscovering the Pal-5 Tidal Tails

The dwarf-finding workflow can be applied to detecting other objects in catalog data as well, requiring only a few minor parameter changes to be modified to create a filter matched to a different type of stellar population. As an example, spatial analysis of SDSS photometry by Odenkirchen et al⁹ discovered substantial tidal tails in the sparse globular cluster Pal-5 extending symmetrically from the main cluster body. Subsequent analysis further refined the Pal-5 population using a variety of techniques. For our satellite-finding demo, we developed a tool that combines elements of the interactive search technique that allows for interaction between the spatial convolution display (left plot of Figure 8) with zoomed displays of the CMD and the catalog area under the cursor. By selecting larger convolution kernel sizes than are used in the dwarf search (small=40 and big=60 arcsec in this case) we can remove the effects of the nearby Messier 5 globular cluster. Similarly, by selecting a color and magnitude cut from the CMD representative of the Pal-5 population we can isolate stars in the cluster, and by again filtering on points near the mean of the density histogram (as opposed to extrema when dwarf-finding) we can highlight the sparse population of the tidal tail itself (see Fig'8). In all, the tail can be revealed with this type of exploratory analysis by the user in under a minute, considerably less time than was required for the original discovery techniques.

The data shown here are from the DECaLS DR2 catalog which only has multi-band information for the northern tail, future releases of this survey would permit examination of the southern branch of the tail using this same technique. Targeted observations along the extent of the Pal-5 tails exist in the NOAO Science Archive that would allow user-generated catalogs to be created to permit other types of analysis, and this technique could be similarly applied in a search for other low-density populations in the various catalog and image datasets.

2.5 Time Series Processing Using Data Lab

The 'Deep Synoptic Survey of the Galactic Bulge' (PI A Saha, NOAO) is a DECam program to derive the stellar population mix of the Bulge using RR Lyrae variables as probes in constructing reddening maps. The data consist of 5 fields observed in *ugriz* filters over roughly 40 epochs; the resulting catalog will contain information on approximately 30 million objects and is expected to contain some 300,000 variable objects of all types (only about 500 RR Lyrae are predicted). Although the object table is not large, the measurements table will be more than a billion rows and the time-series aspect of the project means the database will be used in ways that differ from other survey products. Specifically, these data will primarily be queried to get time series of individual objects rather than for all objects of a certain class or by sky position.

Spatial indexing combined with clustering of the (PostgreSQL) database table physically locates all rows for an object on the disk to provide fast (i.e. sub-second) access times for an object light curve. Additionally, all light curves within a region of the sky are similarly co-located making a spatial query to retrieve multiple curves

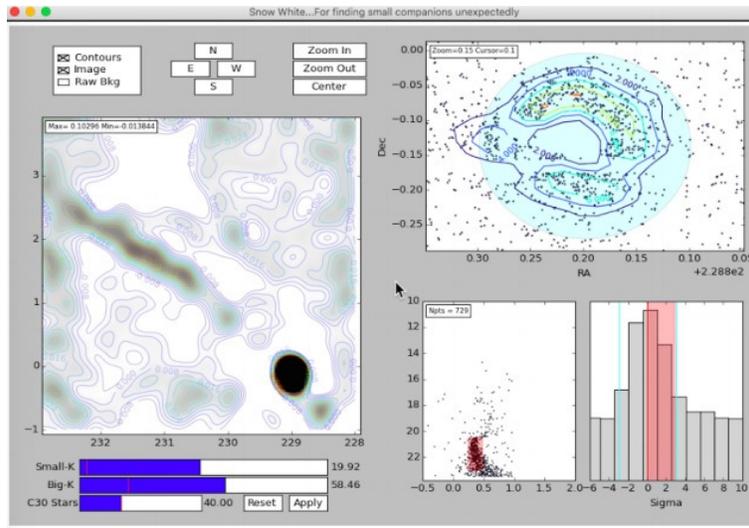


Figure 8. Plots generated automatically using the differential convolution technique, applied to the original Hydra II discovery field. Points within 0.1 deg of the peak over-density found in the clipped histogram are overlotted on the main CMD to highlight the CMD of the local region.

more efficient as well. Figure 9 shows example plots from a light-curve visualization tool we were able to develop to identify variables found in the CMD of candidate dwarf regions. For the Bulge survey dataset, legacy tools developed by the PI were extended to use Data Lab services to provide fast data access to specialized science code running on the user's desktop machine.

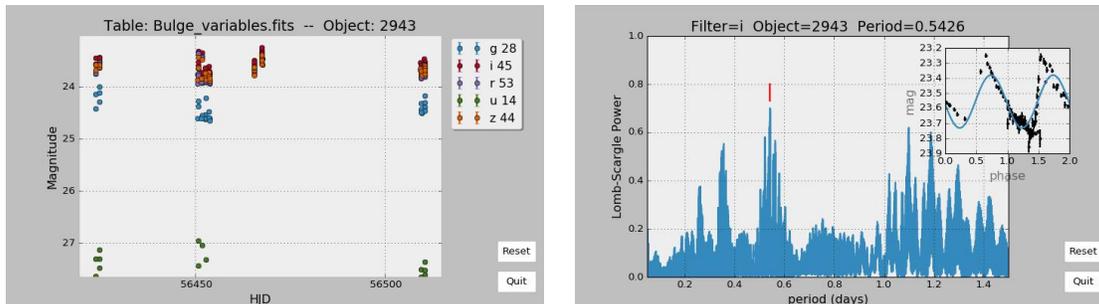


Figure 9. Example output of the light-curve demo tool used to quickly browse the more than 300,000 variable objects in the Bulge survey dataset. Individual object IDs may be broadcast from tools such as TOPCAT when exploring plots of outliers in the data table.

While the spatial indexing of time-series data optimizes the query, the time to build the index and cluster the database can be an issue; in our case, 30 hours was required to create the index initially and we've noticed poorer performance when updating the table with new data as it becomes available.

3. TECHNOLOGY DEMONSTRATIONS

Aside from the main science demonstrations, a number of targeted technology demonstration have also been developed to guide the implementation of needed functionality in the context of real-world use cases.

3.1 Virtual Storage

The Data Lab Virtual Storage system⁸ is critical to making science workflows efficient by providing storage resources close to the data. As a proof of concept we invented a scenario in which a user named Sarah wants to

see whether data exist in the archive at sufficient depth and resolution to allow her to conduct her experiment on a sample of Late Type Galaxies (LTG). She begins by doing some exploratory analysis of data in the Stripe 82 region before deciding whether a full analysis of all DES data is possible.

Using her Virtual Storage, the steps involved in this demo include:

- Mounting the Data Lab Virtual Storage on Sarah's machine as a local filesystem using FUSE.
- Initialize virtual storage working directories, attaching processing capabilities to each as needed.
- Download and filter a catalog containing known LTGs to obtain positions for sample objects.
- Generate image cutouts around each object for analysis using legacy software.
- Analysis of cutouts using legacy software.

In the demonstration, a sample of 585 LTGs are found to lie within the Stripe 82 region. A cross-match against the NOAO Science Archive holding show that an upper limit of more than 48,000 cutouts from approximately 1200 full DECam exposures are possible (this was severely filtered in the demo for time consideration). The traditional method of downloading the exposures and generating the cutouts locally would have required more than 15 hours of network transfer time, however in this demo we show that the use of Virtual Storage allows cutouts to be generated and used by the Sarah within 15 minutes.

3.2 Querying NOAO Catalog Services

The TOPCAT table visualizer is used as a convenient interface to the Data Lab TAP service, however it can be used to access other Data Lab services to query for images and (eventually) spectra, as well as services which support the same protocols at other data centers. Targeted demonstrations of this capability can serve as an introduction to TOPCAT for first-time users.

In addition, the *Jupyter*^{||} notebooks which demonstrate the scripted approach to the main science demo can be used as a basis for exploring the python interface to Data Lab for users interested in incorporating Data Lab services into their existing code base.

4. CONCLUSIONS

The explosive growth in data being produced by NOAO wide-field imagers and the large catalogs being produced by survey programs means that traditional methods of retrieving and analyzing data on the desktop are becoming increasingly impractical for many science programs. The NOAO Data Lab will address this problem by providing data access, storage and compute resources close to the data to allow researchers to efficiently carry out experiments over very large datasets. In designing this system, the use of real-world science cases has helped us to refine development priorities in order to deliver an infrastructure which demonstrates a complete science case and which can be expanded prior to public release to provide robust scientific capabilities which scale to support our expected user community.

ACKNOWLEDGMENTS

We gratefully acknowledge the many useful comments and suggestions offered by review committees and local NOAO staff. The National Optical Astronomy Observatory (NOAO) is operated by the Association of Universities for Research in Astronomy (AURA) under contract to the National Science Foundation (NSF).

^{||}<http://jupyter.org>

REFERENCES

- [1] Fitzpatrick, M., Graham, M., Mighell, K., Olsen, K., Norris, P., Ridgway, S., Stobie, E., Bolton, A., and Huang, L., “The NOAO data laboratory: a conceptual overview,” in [*Observatory Operations: Strategies, Processes, and Systems V*], Peck, A., Benn, C., and Seaman, R., eds., *Proc. SPIE* **9149**, 128 (2014).
- [2] Graham, M., Morris, D., Rixon, G., Dowler, P., Schaaff, A., and Tody, D., “VOspace specification Version 2.0.” IVOA Recommendation 29 March 2013 (Mar. 2013).
- [3] Dowler, P., Rixon, G., and Tody, D., “Table Access Protocol Version 1.0.” IVOA Recommendation 27 March 2010 (Mar. 2010).
- [4] Anonymous, “Description of the Dark Energy Survey for Astronomers.” <https://www.darkenergysurvey.org/survey/des-description.pdf>. (Accessed: 1 May 2012).
- [5] Nidever, D. and Team, S., “Survey of magellanic stellar history - smash,” in [*Fifty Years of Wide Field Studies in the Southern Hemisphere: Resolved Stellar Populations in the Galactic Bulge and Magellanic Clouds*], Points, S.; Kunder, A., ed., *PASP* **491**, 325 (2015).
- [6] Vivas, A. K., Olsen, K., Blum, R., Nidever, D. L., Walker, A. R., Martin, N. F., Besla, G., Gallart, C., van der Marel, R. P., Majewski, S. R., Kaleida, C. C., Muñoz, R. R., Saha, A., Conn, B. C., and Jin, S., “Variable Stars in the Field of the Hydra II Ultra-faint Dwarf Galaxy,” *The Astronomical Journal* **151**, 118 (May 2016).
- [7] Koposov, S., Belokurov, V., Evans, N. W., Hewett, P. C., Irwin, M. J., Gilmore, G., Zucker, D. B., Rix, H.-W., Fellhauer, M., Bell, E. F., and Glushkova, E. V., “The Luminosity Function of the Milky Way Satellites,” *The Astrophysical Journal* **686**, 279–291 (Oct. 2008).
- [8] Graham, M., Fitzpatrick, M., Norris, P., Mighell, K., Olsen, K., Ridgway, S., Stobie, E., Bolton, A., Saha, A., and Huang, L., “The NOAO Data Lab virtual storage system,” in [*Software and Cyberinfrastructure for Astronomy IV*], Chiozzi, G. and Guzman, J. C., eds., *Proc. SPIE* **9913**, 12 (2016).
- [9] Odenkirchen, M., Grebel, E. K., Rockosi, C. M., ter Dehnen, W., Ibata, R., Rix, H.-W., Stolte, A., Wolf, C., John E. Anderson, J., Bahcall, N. A., Brinkmann, J., Csabai, I., Hennessy, G., Hindsley, R. B., Ivezic, L., Lupton, R. H., Munn, J. A., Pier, J. R., Stoughton, C., and York, D. G., “Detection of massive tidal tails around the globular cluster palomar 5 with sloan digital sky survey commissioning data,” *The Astrophysical Journal Letters* **548**(2), L165 (2001).