Science Validation of the LSST Science Platform

Gregory Dubois-Felsmann
LSST Science Platform Scientist, Caltech/IPAC

Large Synoptic Survey Telescope

Science Validation

Verification: did you build what you said you were going to build? Does it meet the formal requirements?

Validation: did you build the right thing? Does it meet stakeholder needs?

Science Validation of the Science Platform, in essence, means determining whether community users can readily use it to do the science that is enabled by the existence of the LSST datasets.

We feel the best way to do this is to enable testing driven by scientific curiosity. During construction, that means making interesting precursor and simulated datasets available.

Efforts to date:

1) Testing of an early prototype of the Portal Aspect
   - Based on two precursor datasets also used for database testing
   - A formal testing exercise was conducted, leading to report DMTR-22.

2) Exposure of the basic Notebook Aspect environment to users
   - Explored user experience of our version of the JupyterLab environment, its integration with the Portal (Firefly) visualization tools, in the context exposure to the capabilities of the LSST Science Pipelines software stack
   - Carried out at multiple LSST Project and community events, workshops, …

Validation efforts 2017-2018: Portal & APIs

The Prototype Data Access Center (PDAC) effort combined an early version of the Portal Aspect UI with DAX service of two datasets, backed by Qserv.

- Datasets:
  - A 2013 LSST processing of the SDSS Stripe 82 dataset, including measurements on coadds and forced-photometry light curves derived from the sources observed.
  - The AllWISE and NEOWISE-year-1 catalog datasets from the WISE mission
- Portal Aspect: derived from the "IRSA Viewer" Firefly-based application
  - Extensively re-engineered with modern JavaScript frameworks & plot.ly
  - Modified to guide the user to both catalog and image queries for the data
  - Regularly updated with new features (notably HiPS image viewing)
- Data Access services reflected an early non-IVOA-centric concept
  - SQL-based queries, JSON data transport, custom metadata and cutout services
- The datasets presented a useful early challenge for Qserv
  - 60+ billion forced photometry records, all-sky coverage

PDAC testing experience

The PDAC became available for testing starting in early 2017

- The system integration itself provided invaluable feedback on the tools and data services
- Substantial informal feedback was received from project staff
- A formal test was conducted in Spring of 2017, reporting in July.
  - Chris Suberlak, a member of Zeljko Ivezic's research group, was asked to do a thorough evaluation of the capabilities and behavior of the system
  - The results of the testing were reported as DMTR-022
- Feedback covered both functional issues and UX issues with the Portal Aspect
  - Most of the functional issues have been addressed
  - An extensive revision of the Portal UX had been planned for late FY19 and for FY20; this work has been paused by the change of Portal plans.
- Feedback was also received on the behavior of Qserv, particularly regarding query language limitations and the propagation of error messages.

- NB: The original PDAC predated the availability of, and did not include a deployment of the Notebook Aspect

2017 testing: DMTR-22 snapshots

Large Synoptic Survey Telescope (LSST)

Prototype Data Access Center: User

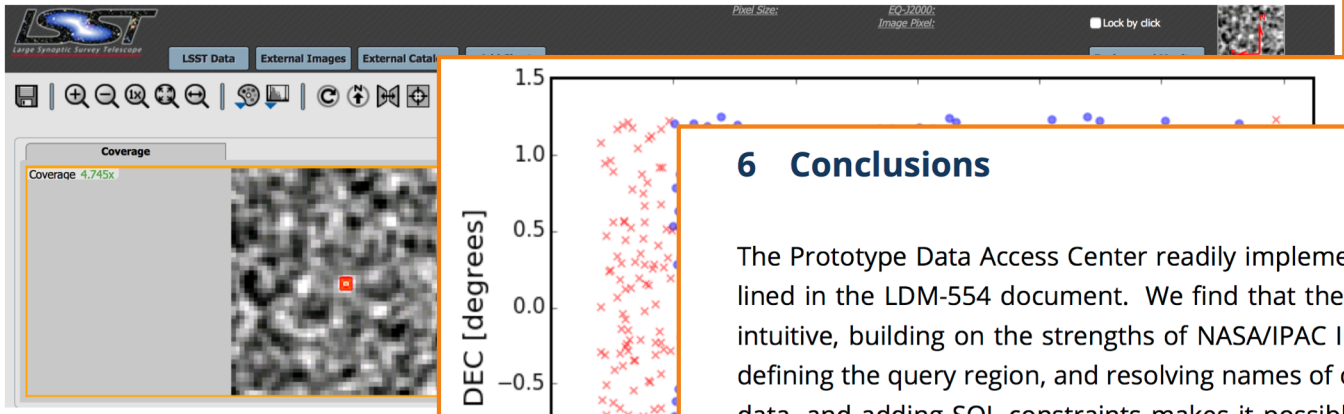


FIGURE 16: A result of cone q
data), with (ra,dec,radius = (
Series' button that links to t

Interface and Tools (SUIT). We employ both in-de
a statistical study of an ensemble of objects. W
current interface, and make recommendations f

FIGURE 37: Results o
RA, Dec. Blue dots a
sec, and red crosses
not alter this result.

6 Conclusions

The Prototype Data Access Center readily implements many requirements for SUIT, as outlined in the LDM-554 document. We find that the overall structure of the interface is very intuitive, building on the strengths of NASA/IPAC IRSA and SDSS CasJobs. Multiple ways of defining the query region, and resolving names of objects with NED helps quickly access the data, and adding SQL constraints makes it possible to reduce the amount of displayed information. An innovative Time Series View, relying on the robust back-end of the NASA Exoplanet Archive Periodogram Tool, has a great potential to aid rapid analysis of time series data. Easy access to external image catalogs allows overlaying information across a range of wavelengths, including a very well designed RGB-making toolbox. The Qserv responds to complicated SQL queries, and scisql functions help to execute very quick joins between catalogs to obtain needed information.

There is a number of areas with possible improvements, including aspects of the User Interface, as well as the SQL query engine. First, certain information, such as calibrated fluxes or magnitudes, is only available via Time Series view or a direct SQL query, while a user may be

Validation Efforts to date: Notebook Aspect

Over the past year, there have been opportunities for a substantial amount of informal user testing of the Notebook Aspect.

- An NCSA-hosted instance of the Notebook Aspect has been available to all DM staff and to selected other users from the Project and community
  - This instance has evolved into the one now known as lsst-lsp-stable
- This instance, as well as pop-up cloud deployments, have been used in various interactions with the LSST SCs, as well as a few activities in the wider community.
- These activities have shared the purposes of:
  - Obtaining feedback on the Notebook Aspect environment itself, including the LSST-specific features of our use of JupyterLab
  - Exposing the community to the Science Pipelines software, taking advantage of the convenience of the provision of pre-installed builds in the Notebook Aspect
  - Exercising the integration of the Portal (Firefly) visualizations in that environment

Thus far no formal testing of the Notebook Aspect has been performed.

Notebook Aspect testing experience

We have obtained useful feedback.

- Thus far the reaction has been enthusiastic.
  - The notebook model has proven very useful for pedagogical purposes and provides a straightforward vehicle for introducing the analysis tools available in the LSST Python environment.
  - Experience to date has focused somewhat more on image analysis than we would anticipate in the era of Data Releases, but is very relevant to the era of commissioning.
- Actionable feedback has been received regarding issues such as:
  - Session lifetime
  - Capabilities for sharing files and other forms of collaborative work
  - The need for remote file access to support users' favorite editors

Future plans – preparation for the next test round

The more complete integration of the three Aspects that has been achieved in early 2019 has paved the way for a new round of formal testing.

- By Fall 2019 we will complete our initial move to an IVOA-centric set of data interfaces, and deploy the data model tooling discussed earlier.
- We then expect to load an LSST processing of publicly released HSC survey data this year and make it available in the integrated LSP
  - A formal "2nd PDAC" test activity will then be carried out early in 2020.
  - The HSC dataset includes coverage of several coordinated community fields and is of substantial scientific value. Our processing will be much more similar to that expected for LSST DRs than the 2013 Stripe 82 processing, and will incorporate much more advanced Science Pipelines capabilities.
  - We will also deploy the Gaia DR2 data, as well as a refreshed ingest of the WISE data including additional years of NEOWISE data.
  - These datasets should meet the criterion of being sufficiently interesting to support "testing driven by scientific curiosity".

The 2020 testing

- We plan to recruit volunteers for this science validation activity primarily through the Science Collaborations.
  - Volunteers will be asked to describe scientific investigations they wish to perform, and to provide both immediate feedback and a written report of their experience at the end of a designated testing period.
  - Because of the limited hardware and user-support resources available in 2020, we will have to limit this testing exercise to a relatively small number of users.
  - We hope to select them – and their proposed investigations – in a way which explores a wide range of interactions with the LSP, from catalog-centric to image-oriented, and from analysis of targeted sky regions to wide-area statistical analyses.
- Reports will be studied, the issues collated and prioritized, and used to refine the final stages of DM construction activity on the LSP and its support systems.
  - The Portal Aspect maintenance effort will be able to address high-priority bug fixes, but any other Portal Aspect issues will be set aside to be addressed when active development is resumed.

Additional validation opportunities – DESC DC2

- We are working with the Dark Energy Science Collaboration and their "DC2" simulation project to support their use of the data from this project with components of the LSP.
  - Efforts are underway to load DC2 data into Qserv databases in non-Project instances.
  - All the components of the LSP are open-source and meant to be usable by others, and we will encourage their use in working with the DC2 data.
  - We can provide some support for this to the extent that it yields useful test results and does not impede higher-priority activities.
  - The Project itself does not currently plan to ingest this dataset or use it for testing.

- A technical operations rehearsal that involves end-to-end processing, ingest, and service of at least a small simulated dataset would be very useful for verifying that LSST-style data is properly handled throughout the chain.

Additional validation opportunities – ComCam+

- Once ComCam commissioning begins, we plan to make the raw and processed data available in Project LSP instances, and expect extensive informal feedback.
  - The ability to engage the Portal and API Aspects in this process may be limited at first by the absence of organized pipeline processing and output data ingest
  - It will be highly desirable to make at least the observation metadata and raw images available through the API Aspect data services, and thus through the Portal
  - We expect this activity to illuminate the usefulness of the LSP's data-sharing tools
- As commissioning proceeds and the LSST reaches focus and a useful basic automated processing system can be set up, it would be very useful to ingest its outputs in some way visible through all Aspects of the LSP.
- Detailed planning for the role of the LSP in this era has not been completed.

The LSP in Project-level Science Validation

The final phase of the Commissioning era of the project is the Science Validation of the LSST as a whole.

- The LSP will be part of in this in two overlapping senses:
  - The LSP will be the primary tool for the Project's science staff (and any community participants) to access the Commissioning data and perform the analysis tasks required to validate the data quality obtained from the LSST.
  - The LSP, as part of the Project, will itself be the subject of the validation effort.
- We expect that the effort to use the LSP for the larger purposes of the Science Validation phase will inherently drive a validation of many elements of the LSP.
  - In order not to take this for granted, we will prepare a validation plan for the LSP to be used during this phase, with an emphasis on encouraging the formal collection of explicit feedback from the participants in the larger validation effort.
- A further *dedicated* validation of the LSP, beyond the 2020 activities, is still under consideration.

Conclusions

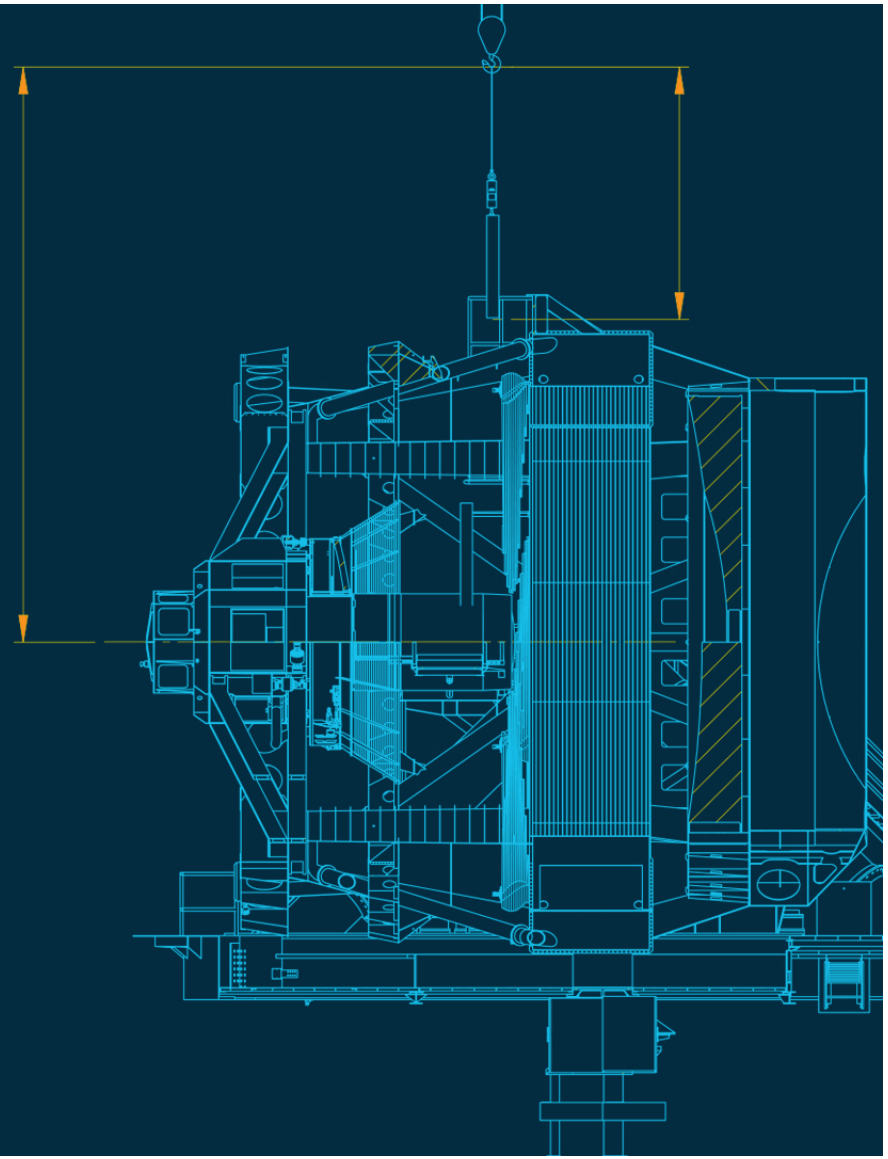
Science Validation of the LSP is relevant to these charge points:

4. Are the verification, validation and software quality assurance plans adequate?
5. Does the performance of the current system and its development status inspire confidence that both the interim and operations-era functionality can be delivered?

Summary

- We have performed a formal scientific validation of a prototype of the Portal Aspect and certain API Aspect services, supported by Qserv. The feedback was used in development.
- We plan a second formal validation exercise, covering the full integrated LSP, based on several scientifically interesting datasets (HSC, Gaia, WISE, …), in early 2020.
- Extensive informal user testing has occurred and will continue.
- The LSP will be used throughout commissioning, with user experience monitored.
- Formal feedback on the LSP itself will be collected as part of the final Project-level Science Validation effort.

Additional details



Large Synoptic Survey Telescope

Details of the PDAC datasets: Stripe 82 & WISE

The two datasets provided in the PDAC were:

- A 2013 LSST processing of the SDSS Stripe 82 dataset
  - Coadds were generated in each band and sources detected on them
  - *i*-band coadded sources were used to seed forced photometry in all epochs
  - These catalogs and calibrated coadded and single-epoch images were made available.
  - Xxx coadded sources; Yyy forced-photometry data points
  - NB: The data model is different from the DPDD-based data model planned for LSST.

- The principal catalogs and image metadata tables from the "AllWISE" processing of the original WISE mission data, as well as from the first year of the NEOWISE "reactivation mission"
  - Xxx coadded sources; Yyy forced-photometry data points; Zzz single-epoch sources
  - Xxx coadded image tiles; Yyy single-epoch images