LSST Data Management Brief Status Update

Mario Juric
University of Washington
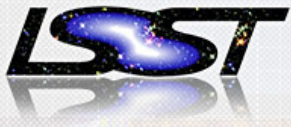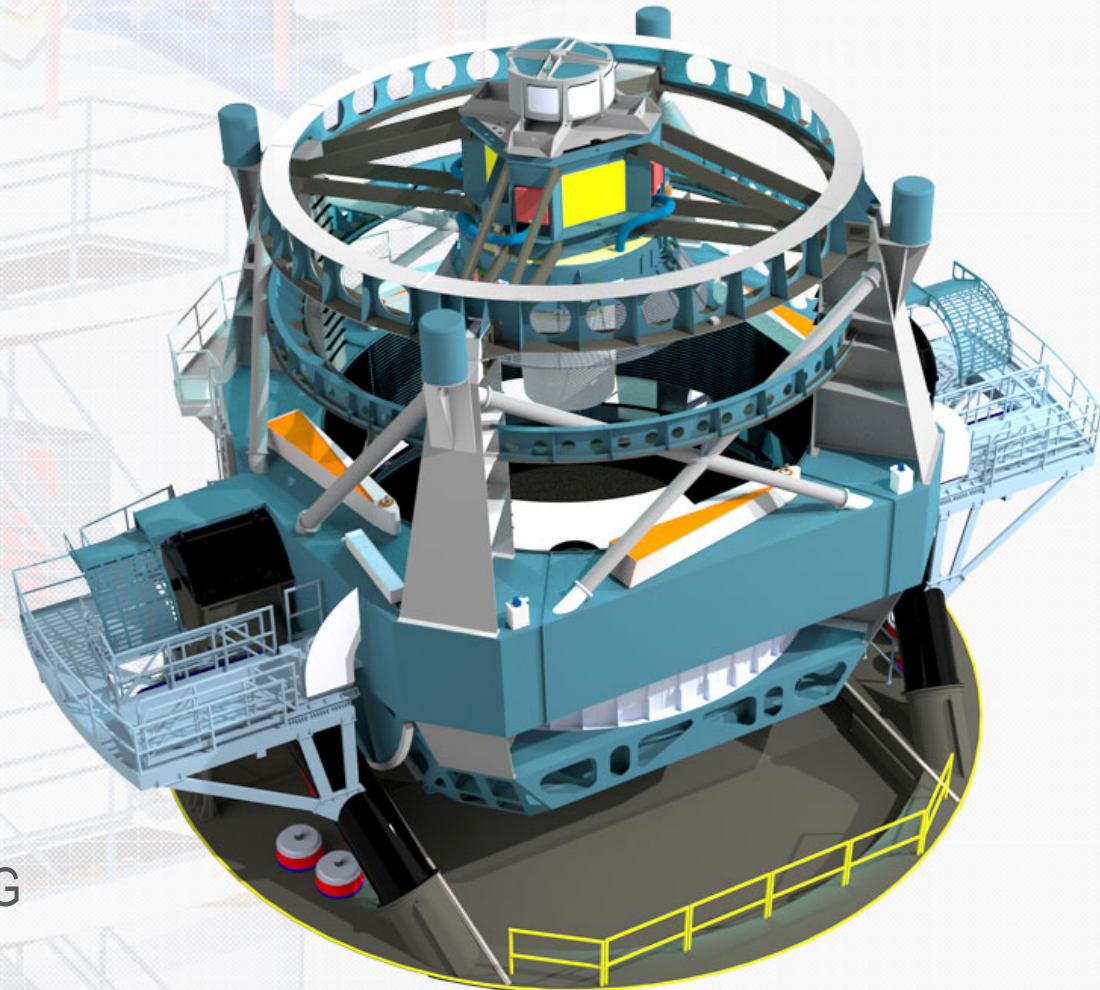*LSST Data Mgmt. Subsystem Lead*

Robert Lupton
Princeton University
LSST DM Science Pipelines Lead

with John Swinbank, Simon Krughoff,
Jacek Becla, K-T Lim, Jim Bosch
and the LSST DM Team

LSST DESC MEETING
March 6, 2016

Status @ Argonne (J. Bosch)

- Jim Bosch gave an excellent summary http://ls.st/4o4; this talk is an update with **some** highlights since last October.

Data Management Status

Jim Bosch
Princeton University

- Status in ~October 2015:
  - We were porting over science pipelines changes from HSC
  - We were improving the long-term development plan
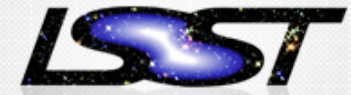 - We were in the midst hiring

A year of growth

- Staff has quadrupled since start of MREFC, will be at **~80 people** by ~October this year (including contractors and fractional staff)
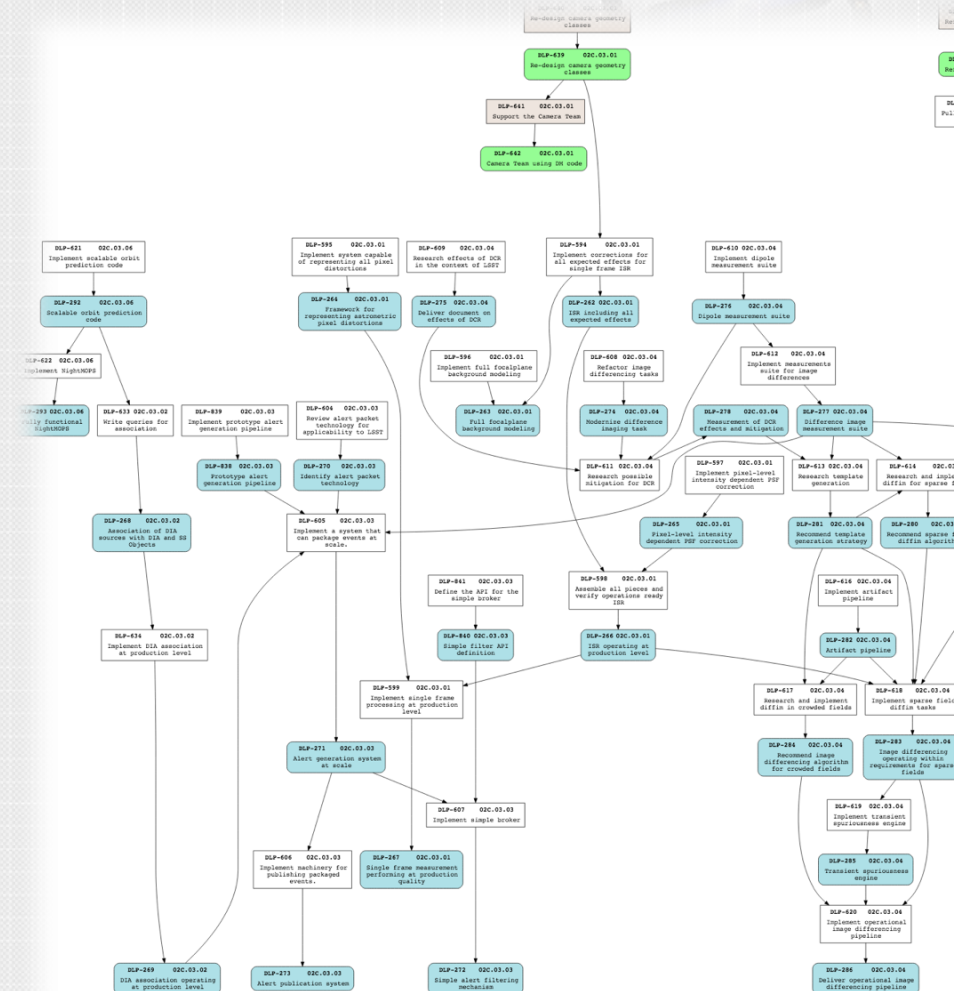
- Some are fractional; **~54 FTE hired/assigned**

+more

Status of the development plan

- https://jira.lsstcorp.org/projects/DLP

- A long-term plan for science pipeline development and related development.

- A significant update since construction start, but more work still remains

- Next refresh will come later this year

Plan visualizations: http://ls.st/zmf, http://ls.st/ms1

DM Leadership Changes

- We've streamlined the leadership and structure of the DM subsystem.

- **Used to be three-in-a-box**
  - DM Project Manager: Jeff Kantor
  - DM Project Engineer: K-T Lim
  - DM Project Scientist: Mario Juric

- **Leadership structure reorganized**
  - Subsystem Lead: **Mario Juric** (interim)
    - Project Manager: **Jacek Becla** (interim)
    - Project Scientist: **Zeljko Ivezic** (interim)
    - System Architect: **K-T Lim**
  - **Andy Connolly** named interim Level 1 processing science lead
    - Search underway for permanent lead

- Announced two weeks ago at the LSST Joint Technical Meeting. Transition still in process (not all documents & websites updated to reflect the new structure).
- We will be conducting a world-wide search for the Project Manager manager position

New Working Groups

– Established working groups to make a push to the next level of DM architecture fidelity. Two of special interest to DESC:

- **Science Pipelines Definition Workgroup (scipi-wg)**
  - Co-Chairs: Zeljko Ivezic & Robert Lupton
  - Refine and improve the high-level architecture of the DM Science Pipelines to support further refinement into detailed design, the creation and execution of well planned and resourced construction activities, as well as provide science pipelne related inputs necessary for the work of dps-wg. This includes **enumerating all pipelines required to deliver products in the DPDD, defining overall processing flows, and specifying algorithms (as best understood at this time)**. For the purposes of this WG, science pipelines include SDQA and Calibration.

- **Data Processing System Architecture Workgroup (dps-wg)**
  - Co-Chairs: K-T Lim and Don Petravick
  - **Refine and improve the overall architecture of the data processing system and its coarse-grain concept of operations**, to support further refinement into detailed design, the creation and execution of adequately planned and resourced construction activities, as well as project-wide operations planning.

https://confluence.lsstcorp.org/display/DM/DMLT+Working+Groups

Significant New Science Pipeline Features

- 'Multi-band' coadd processing.
  - Provides consistent outputs across filters.
- 'Safe' clipping in coadd construction.
  - Much more robust and reliable rejection of outliers.
- Prototype 'Brighter-Fatter' correction.
- System for removing false detections around bright objects.
- Integrated external PSFEx package for state-of-the-art PSF estimation.
- Joint astrometry/photometry calibration (impl. by Pierre Astier's grop)

- ctrl_pool task parallelization framework
  - High level solution for MPI-based task distribution across a cluster.
  - Not the long-term solution to LSST's needs, but good enough for now.
- Ability to inject 'fake' sources into processing.

- Rework of ProcessCcdTask

IMPORTANT: All of these are currently on master, and will appear in the the next release (~3 months). Intermediate builds may be available sooner.

More from Robert in just a minute!

Credit: Princeton & UW Science Pipelines Teams

DMTN-006: False Positive Rates in the LSST Image Differencing Pipeline

Colin Slater, Mario Jurić, Željko Ivezić and Lynne Jones

Latest Revision: 2016-03-04

Significant work done on understanding of current image differencing pipeline performance. With DECam data, we're capable of producing a clean sample of difference image detections at roughly the 200-400 per square degree level (early results)

- Raw counts, before real-bogus filtering (only a simple dipole detector applied)
- A significant number of these is likely due to real variability + asteroids
- This is image-to-image (not image-to-template) differencing
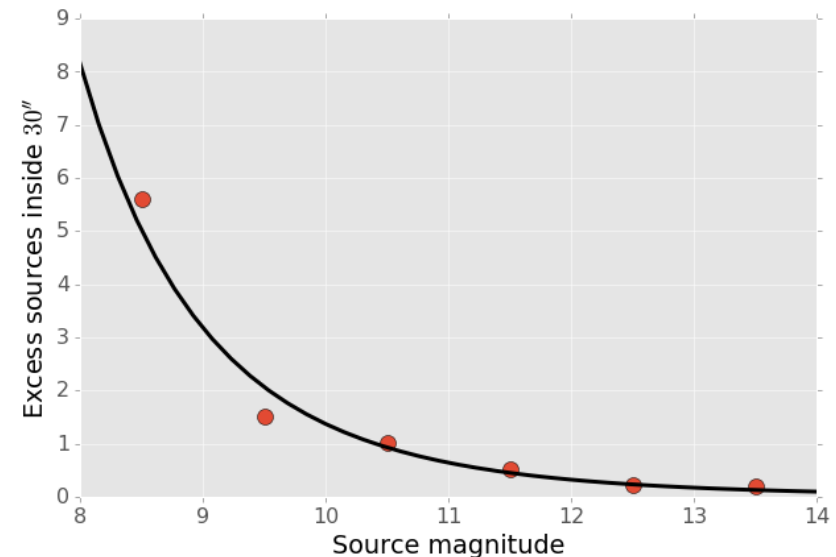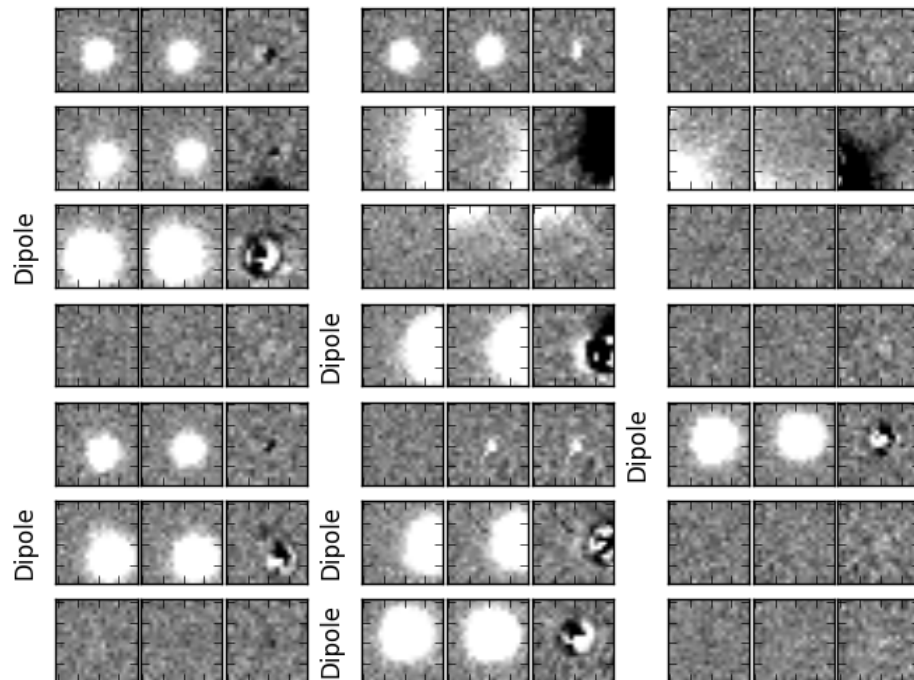
Left: A sample of (false) detections
Top: Detection excess near bright sources

Documentation: DM Technical Notes

A system to make it easy to publish, maintain, and cross-reference technical notes (credit: Jonathan Sick and the SQuaRE team).

Current list of tech notes is on community.lsst.org (link below).

Plan to have it indexed by ADS.

Same mechanism may be usable for DESC purposes.

https://community.lsst.org/t/listing-of-available-dm-technotes/496

Communication: community.lsst.org

A forum for both <u>internal</u> and <u>external</u> communication and interaction.
Note: More formal mechanisms for Project-Collaboration interactions are also being put in place.

User and Developer Guides cleaned up, updated

Credit: Jonathan Sick and the SQuaRE team.

A significantly improved software user guide (http://ls.st/ug) and the developer guide (http://developer.lsst.io)

These will be interesting for DESC to mine for standards, best practices, etc.

Upcoming work (next ~6-9 months)

- We're inserting a short ~three-month development cycle
  - The "X16" (for "eXtra") cycle
  - Having releases coincide with project meetings didn't work well
  - Next cycle will be "Fall 2016"
- Expect to do significant work on the overall system architecture (working group process).

- **Upcoming science pipelines work (some highlights):**
  - MultiFit (initial version) and deblender improvements
  - Astrometry
  - QA Pipelines
- **Architectural and middleware work (some highlights):**
  - Looking at better integration with AstroPy
  - Work on data access components (see next slide)
- **More work on documentation and usability**

Tracking X16 planning and work:
https://jira.lsstcorp.org/issues/?filter=14115

Key Data Access Layer Components: Butler

http://ldm-463.lsst.io/en/tickets-dm-5091/#

Usability: universal binary installs with conda

We're working to make it easier to install the LSST stack, and provide binary distributions:

- Internally already building OpenStack images for our Nebula cluster.
- Containers, native packages, CERN-VMFS builds
- **Working on conda packaging of the LSST stack.**

Installing and using

No root privileges are required to install or use this code.

If you're using Anaconda (or Miniconda) Python distribution, you should be able to install the LSST stack by doing the following:

```
conda update conda   # to ensure you're running latest conda

conda config --add channels http://research.majuric.org/conda/stable

conda install lsst-distrib
```

An experimental conda build of Summer 2015 is at http://ls.st/k2i. More recent builds are also available (ask on community.lsst.org).

Summary

- We've staffed up (with a bit more to go)

- Visible progress in a number of areas, especially in science pipelines and documentation. <u>Much more underneath the surface</u>.

- Key challenge for the next year: keeping the development coordinated (both internally and with the community)

- A push to strengthen the architecture and project management

- Continue the conversation on http://community.lsst.org

Questions?