

Software and big data from astronomy

William O'Mullane, AURA/Rubin Observatory
DM Project Manager

March 2, 2020



William O'Mullane

Software and big data





A little about myself

- 1985ish started with BASIC on a Commodore
- 1993 MSc BSc Computer Science, University College Cork, Ireland
- 1993 - 1997 Spacecraft Control Systems (C++), ESA ESOC Germany
- 1997 - 2001 Hipparcos, Integral, Planck, Gaia, Bepi-Sax (C,Java,Oracle, HTM, HEALPix), ESA ESTEC Netherlands
- 2001-2003 Commercial programming - some Astronomy (Java)
- 2003-2005 The Johns Hopkins, SDSS and National Virtual Observatory (C,C#,Java,Sqlserver)
- 2005-2014 Gaia Astrometric Solution and Science Operations (Java, Oracle, Intersystems Cache)
- 2012 PhD on Implementing the Gaia Astrometric Solution, Barcelona University
- 2014-2017 ESA SCI-OD division head - all science ground segments in development
- 2017- LSST Data Management Project Manager (Python,C++), Deputy Project Manager for Software (control systems and IT)



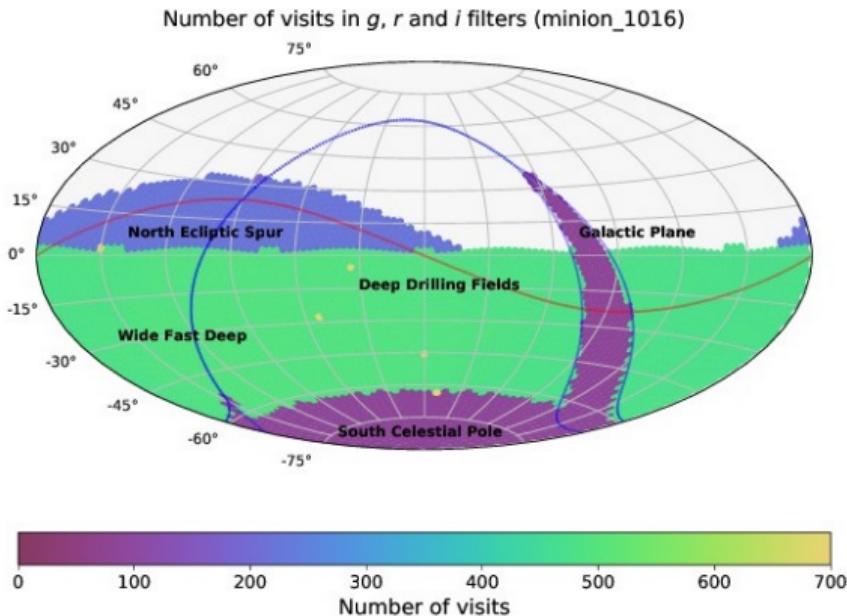
LSST:uniform sky survey

Rubin
Observatory

An optical/near-IR survey of half the sky in ugrizy bands to r 27.5 (36 nJy) based on 825 visits over a 10-year period:
deep wide fast.

- 90% of time spent on uniform survey: every 3-4 nights, the whole observable sky scanned twice per night
- 100 PB of data: about a billion 16 Mpix images, enabling measurements
for 40 billion objects!

see also <http://www.lsst.org> and Ivezic et al. (2019)-arXiv:0805.2366



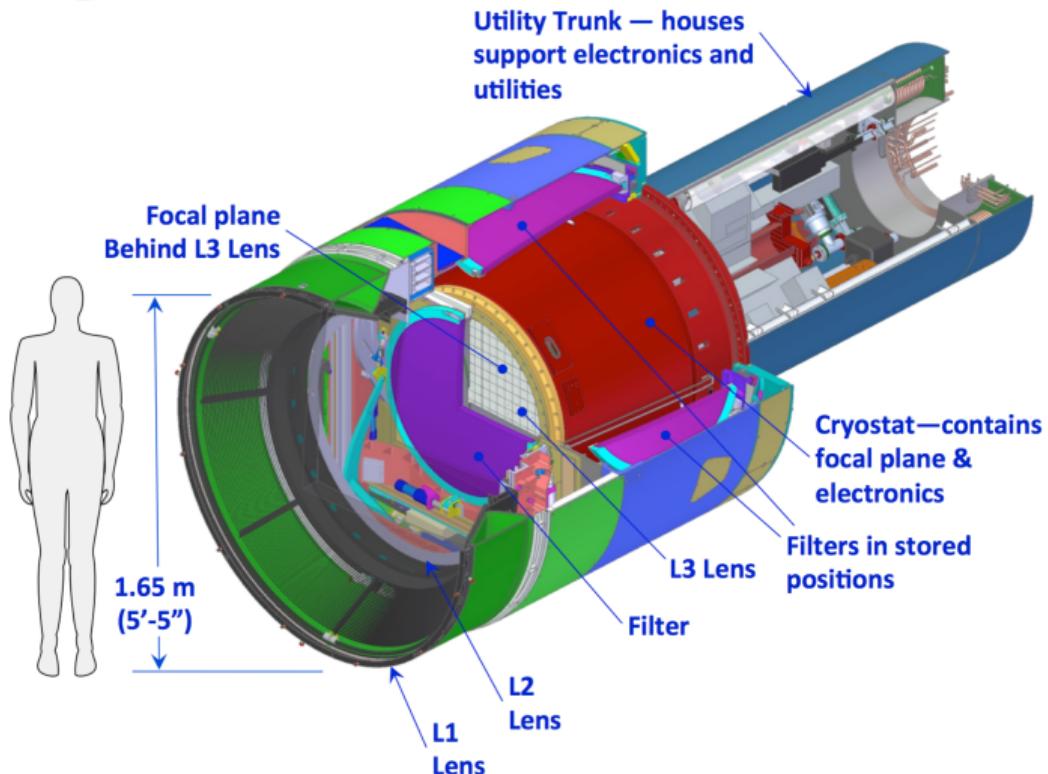
**10-year simulation of LSST survey:
number of visits in u,g,r band (Aitoff
projection of eq. coordinates)**





LSST Camera

Rubin
Observatory



The largest astronomical camera:

- 2800 kg
- 3.2 Gpix



Site as imagined and in March 2019

Rubin
Observatory





DM build and deploy - already challenging

DM must build everything to get LSST products (see <http://ls.st/dpdd>) to the users.

- large data sets (20TB/night)
- complex analysis
- aiming for small systematics
- Science Alerts in under 2 minutes .. (aiming for 1 minute)

About $\frac{1}{2}$ million lines of code (C++/python) all open source on github

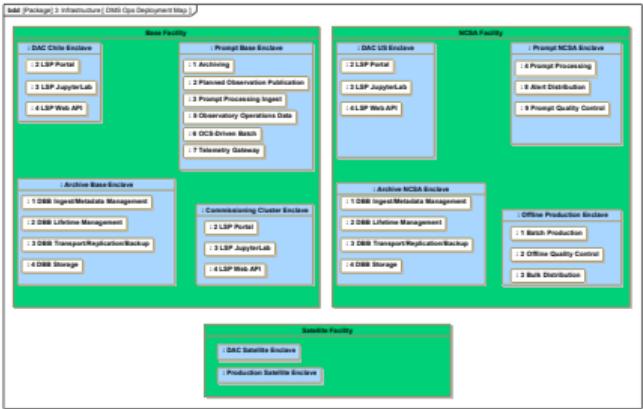
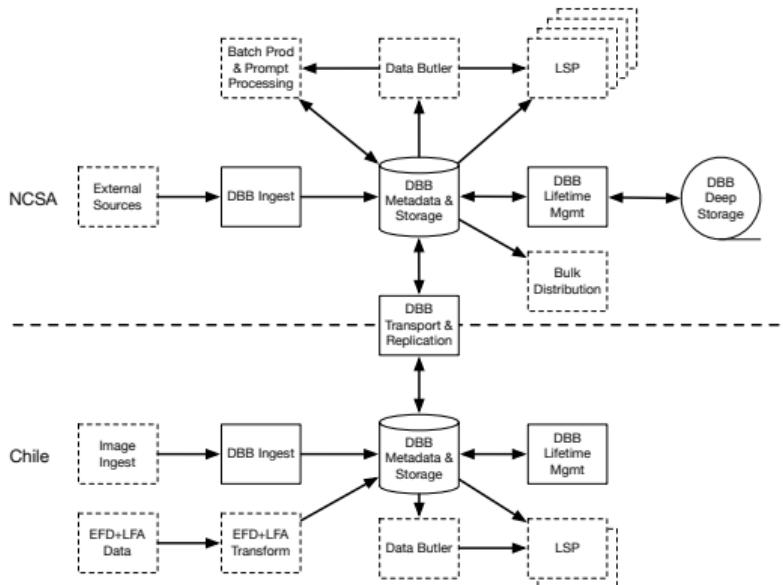


diagram K.T. Lim



Data Backbone

Rubin
Observatory



One small box on the previous slide was Data Backbone.
That hides several things

- Qserv - the LSST end user database
 - Custom Massive Parallel Processing (MPP)
 - allow queries on ≈ 20 Petabytes of tabular data
 - 4×10^{10} objects, 4×10^{13} sources (observations)
- All the networks : we now have fiber to the mountain and from La Serena to NCSA (two routes)

diagram K.T. Lim



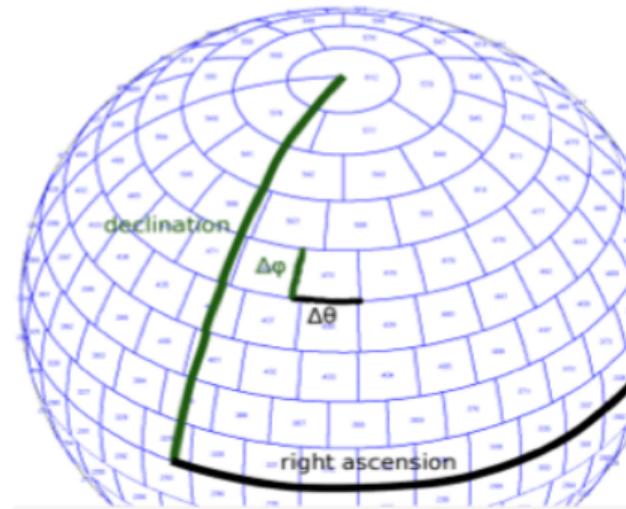


Astronomy catalogs tend to be highly structured, tabular, somewhat predictable access.

- Data, by DR11:
 - 60T rows (mostly ForcedSource)
 - 10PB (mostly Source +
ForcedSource + Object extra)
- Breakdown of most significant
tables (rows x cols, storage):
 - Object: 47B x 330, 100TB
 - Object extra: 1.5T x 7,600,
1.2PB
 - Source: 9T x 50, 5PB
 - ForcedSource: 50T x 6, 2PB
- Get an object or data for small area - <10 sec
- Scan through billions of objects - \approx 1 hour
- Deeper analysis (Object_*) - \approx 8 hours
- Analysis of objects close to other objects - \approx 1 hour, even if full-sky
- Analysis that requires special grouping - \approx 1 hour, even if full sky
- Source, ForcedSource scans - \approx 12 hours
- Cross match & anti-cross match with
external catalogs - \approx 1 hour



- Shared-nothing MPP RDBMS (SQL, throughput, horizontal scaling)
- Spherical partitioning with overlap (near-neighbor self-joins)
- Shared scans (concurrent query load)
- Replicated data (resiliency)
- Fixed-purpose, dedicated hardware (cost, predictability)



Tessellation not O'Mullane et al. (2001)

Design optimized for use case + hardware efficiency **LDM-135**

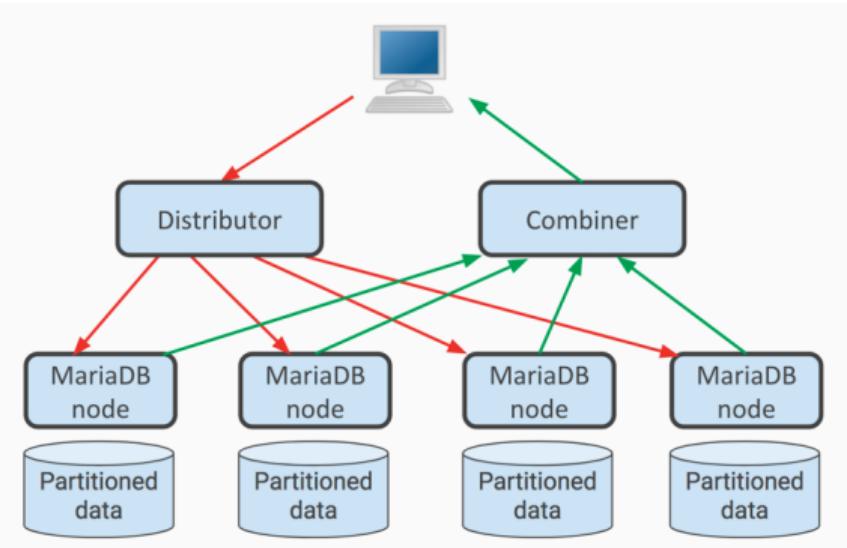
Built on project at SLAC, leverage existing tech within Stanford (MariaDB, MySQL Proxy, XRootD, Google protobuf, Flask)

100% open source <https://github.com/LSST/Qserv>



Shared Nothing Massively Parallel Processing

Rubin
Observatory



Recent scale tests:

<https://dmtr-071.lsst.io>

Perf and BigQuery: Document-31100

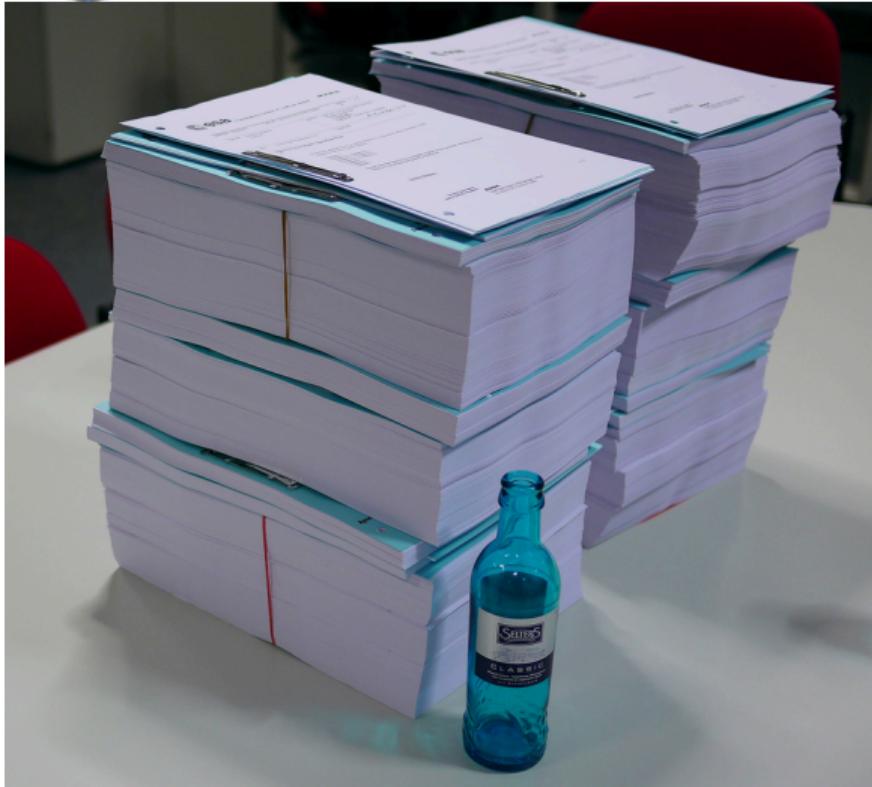
- Ultimate target platform 300 nodes in 2 international data-centers
- Development cluster (CC-IN2P3):
 - 400 cores, 800 GB memory, 500 TB storage
 - 100 TB synthetic dataset on 2 x 25 nodes
- Prototype Data Access Center (NCSA):
 - 500 cores, 4 TB memory, 700 TB storage
 - 100 TB science dataset (SDSS Stripe 82 + WISE) on 30 nodes
 - + HSC reprocessing + GAIA DR2 coming up





Flight Operations Procedures

Rubin
Observatory



The FOP is followed by the spacecraft operators - the paper copy is just in case the computers fail - could be useful!

But we should avoid *write only* documents.



Guidelines, tools, conventions

Rubin
Observatory

- Its great having extensive guidelines - it was also something super on Gaia
- developer.lsst.io is a full developer guide - everything from git commit messages to style guides for Python and C+.
- pipelines.lsst.io documents the main software release(s)
 - worst of both worlds - it is a monolithic release - but made up of > 120 git repos
 - Still using *in house* tools like EUPS <https://github.com/RobertLuptonTheGood/eups>
 - moving toward conda-forge
 - All open source (GPL) on github.com
- Language (spoken/written) and conventions are also super important
 - Single language projects (like US or UK) fall more easily in the trap of *believing* they speak about the same topics because they speak language X.
 - - on RubinObs we lack something foundational like BAS-003





SDSS image

Rubin
Observatory



Nice colors Lupton et al. (2004)
 $\approx 3.5'$

Image Robert Lupton



Hyper Suprime Cam (HSC) on Subaru

Rubin
Observatory



HSC image (COSMOS) from g,r(1.5 hrs) ,i(3 hrs)
PSF matched co-add (≈ 27.5)
Challenges:

- Unknown statistical distributions,
Truncated, censored and missing data,
Unreliable quantities (e.g. unknown
systematics and random errors)
- PSF - short exposure - atmosphere
dominated ?
 - cosmic shear signal from weak lensing
- Photometry challenging - will Gaia help ..
- Everything is blended!!

Processed with LSST Stack <https://pipelines.lsst.io/>



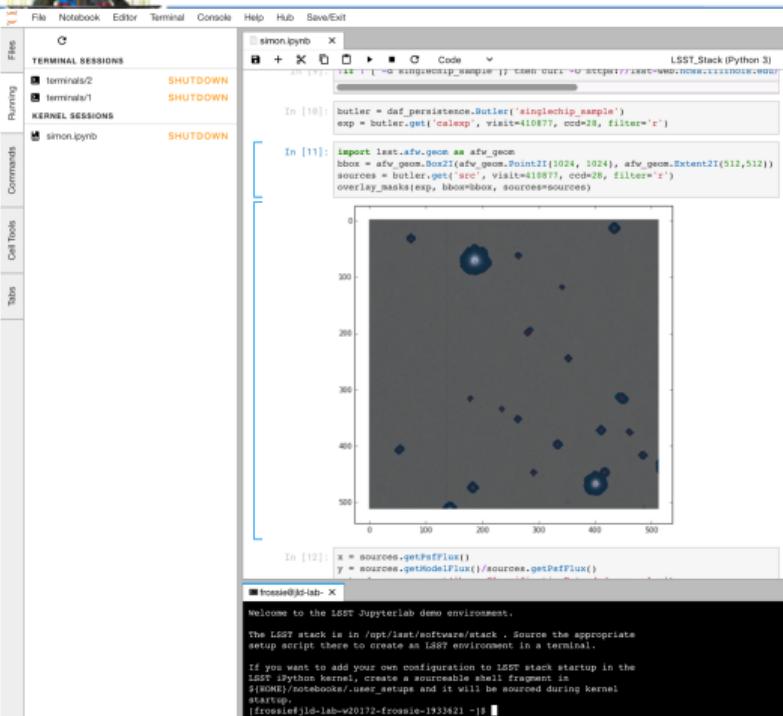
William O'Mullane

Image HSC collaboration, Robert Lupton

14

U.S. DEPARTMENT OF
ENERGY | Office of
Science

Catalog extraction



The screenshot shows a JupyterLab interface with several panels:

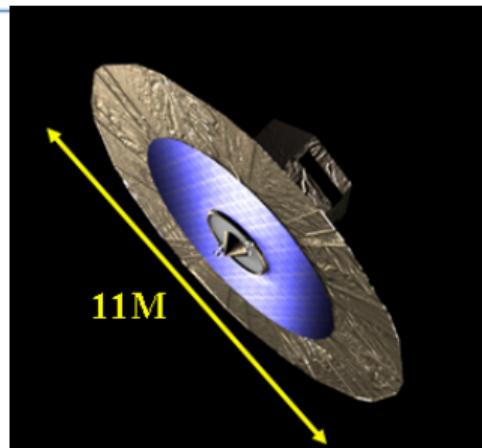
- File Notebooks Editor Terminal Console Help Hub Save/Exit**
- Files** panel: Shows terminal sessions (two shutdown), kernel sessions (one shutdown), and a file named simon.ipynb.
- Running** panel: Shows no active sessions.
- Commands** panel: Shows no active sessions.
- Cell Tools** panel: Shows no active sessions.
- Tools** panel: Shows no active sessions.
- TERMINAL SESSIONS**: Two terminals labeled "SHUTDOWN".
- KERNEL SESSIONS**: One kernel labeled "SHUTDOWN".
- simon.ipynb** notebook:
 - In [10]:
butler = daf_persistences.Butler('singlegchip_sample')
exp = butler.get('calibex', visit=41877, ccd=18, filter='r')
 - In [11]:
import lsst.afw.geom as afw_geom
bbox = afw_geom.Box2I(afw_geom.Point2I(1024, 1024), afw_geom.Extent2I(512, 512))
sources = butler.get('src', visit=41877, ccd=18, filter='z')
overlay_masks = exp.bbox(bbox, sources=sources)
 - A scatter plot showing several sources, with one prominent central source and several smaller ones scattered around it.
 - In [12]:
x = sources.getDefFlux()
y = sources.getModelFlux()
 - In [13]:
Welcome to the LSST Jupyterlab demo environment.
The LSST stack is in /opt/lsst/software/stack . Source the appropriate setup script there to create an lsst7 environment in a terminal.
If you want to add your own configuration to LSST stack startup in the LSST Python kernel, create a sourceable shell fragment in \$HOME/.notebooks/user_setups and it will be sourced during kernel startup.
[frossie@ls-lab-w20172-frossie-1933621 -]\$

- Identifying sources and disentangling them becomes more difficult as we have deeper images.
- Left our typical Jupyter setup runs
 - Instrument signature removal
 - calibration
 - source extraction
 - overlay extracted information on cleaned image



Satellite overview

- Mission:
 - ESA Corner Stone 6
 - ESA provided the hardware and launch
 - Mass: 2120 kg (payload 743 kg)
 - Power: 1631 W (payload 815 W)
 - Launched December 19th 2013
 - Stereoscopic Census of Galaxy over 5 years
 - Extended 2 yrs - request for five
 - μarcsec Astrometry $G < 20$ (10^9 sources)
 - Radial Velocities $G < 16$
 - Photometry millimag $G < 20$

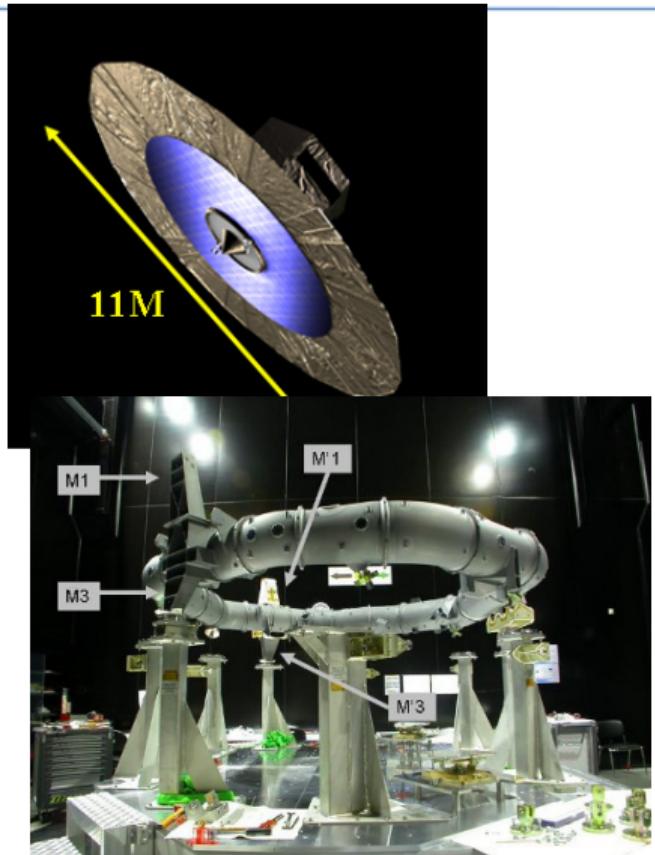


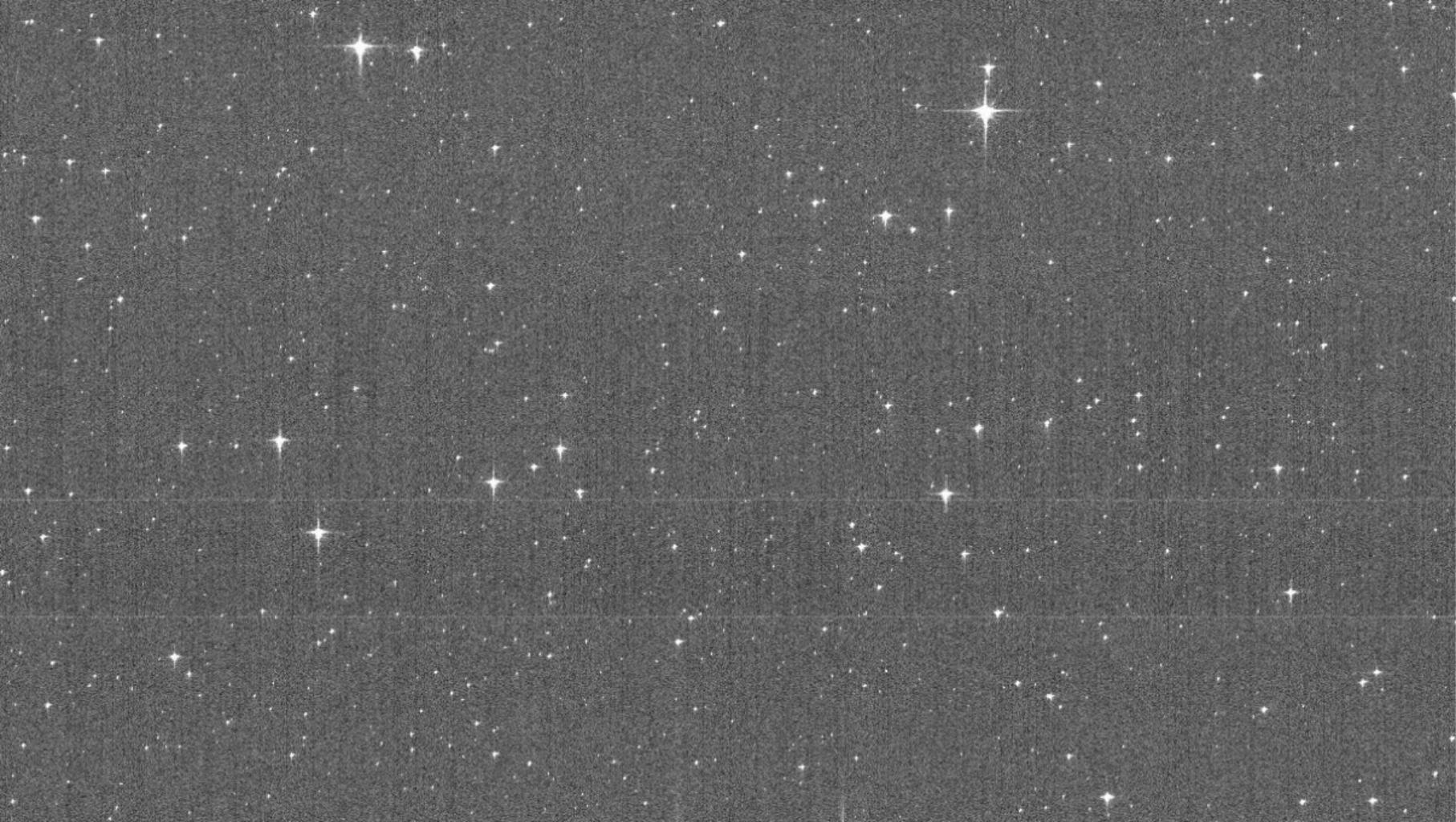


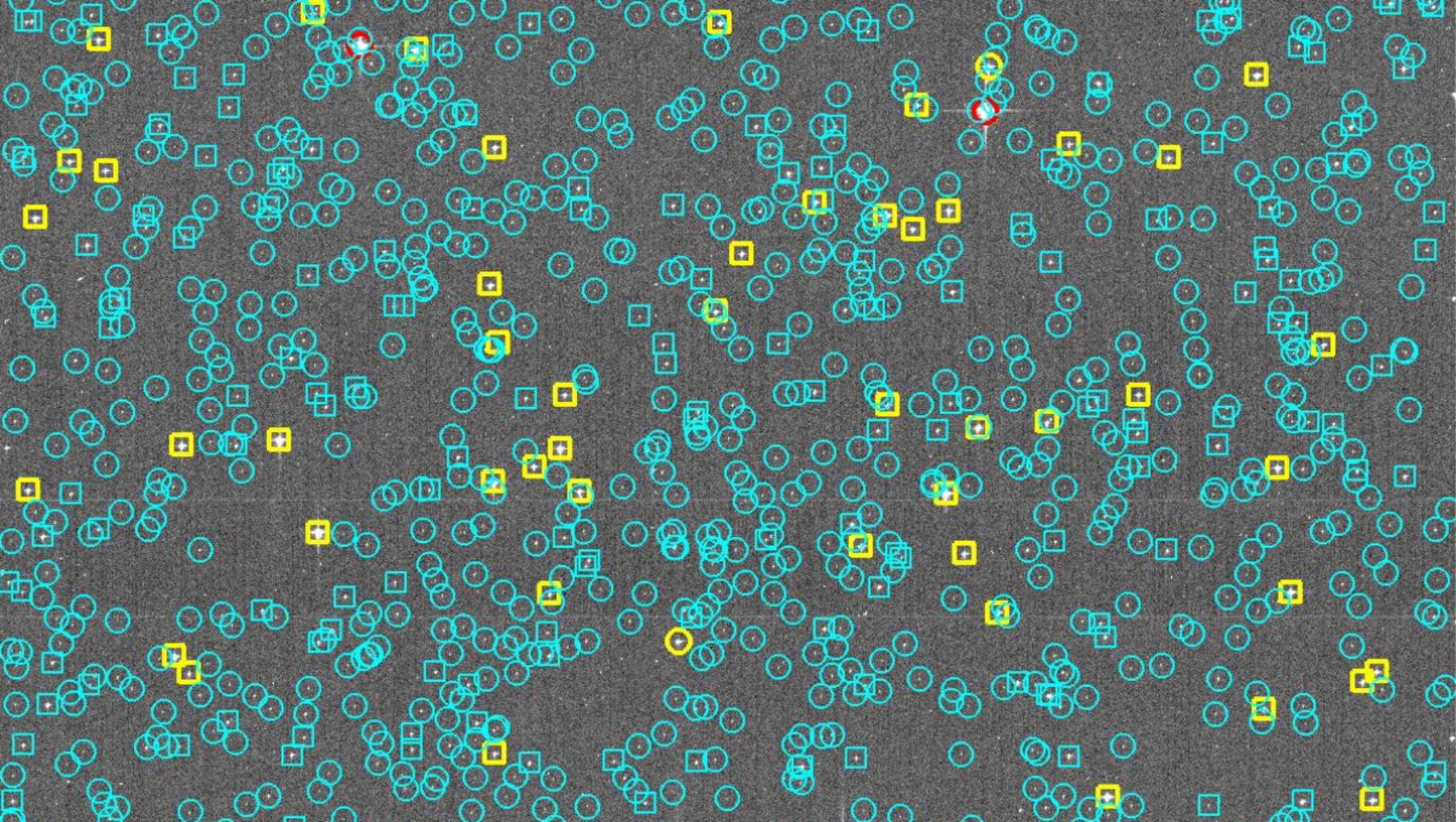
Satellite overview

Rubin
Observatory

- Mission:
 - ESA Corner Stone 6
 - ESA provided the hardware and launch
 - Mass: 2120 kg (payload 743 kg)
 - Power: 1631 W (payload 815 W)
 - Launched December 19th 2013
 - Stereoscopic Census of Galaxy over 5 years
 - Extended 2 yrs - request for five
 - μarcsec Astrometry $G < 20$ (10^9 sources)
 - Radial Velocities $G < 16$
 - Photometry millimag $G < 20$
- Final catalogue $\approx 202?$









- Astrometric centroid of the CCD image to be determined to an accuracy of 1% of the pixel size!
 - There will be 10^{12} images $\approx 100\text{TB}$ downlink need to handle $\approx 1\text{PB}$
 - At 1 millisecond each that is ≈ 30 years
- Reconstructed attitude is required to order $10 \mu\text{arcsec}$
 - Path of light through instrument needed to nanometre level
 - System must be extremely stable
 - Must consider relativistic light bending from solar system objects.



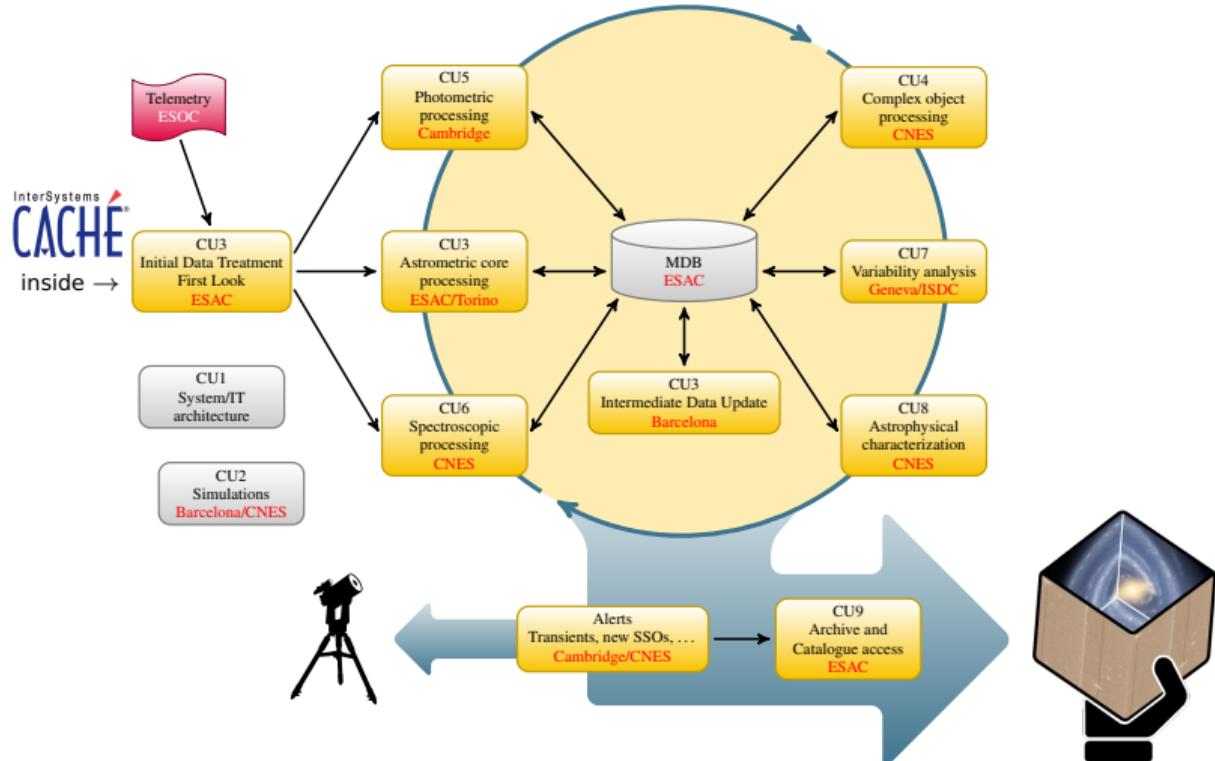
- Astrometric centroid of the CCD image to be determined to an accuracy of 1% of the pixel size!
 - There will be 10^{12} images $\approx 100\text{TB}$ downlink need to handle $\approx 1\text{PB}$
 - At 1 millisecond each that is ≈ 30 years
- Reconstructed attitude is required to order $10 \mu\text{arcsec}$
 - Path of light through instrument needed to nanometre level
 - System must be extremely stable
 - Must consider relativistic light bending from solar system objects.
- Attitude and Geometric calibration can only be done using Gaia's own observational data. (AGIS) (O'Mullane et al., 2011a; Lindegren et al., 2012)
 - this requires a significant portion of the data to be processed iteratively



Simplified processing overview

Rubin
Observatory

Upstream -----> Downstream





Parameters and data models

Rubin
Observatory

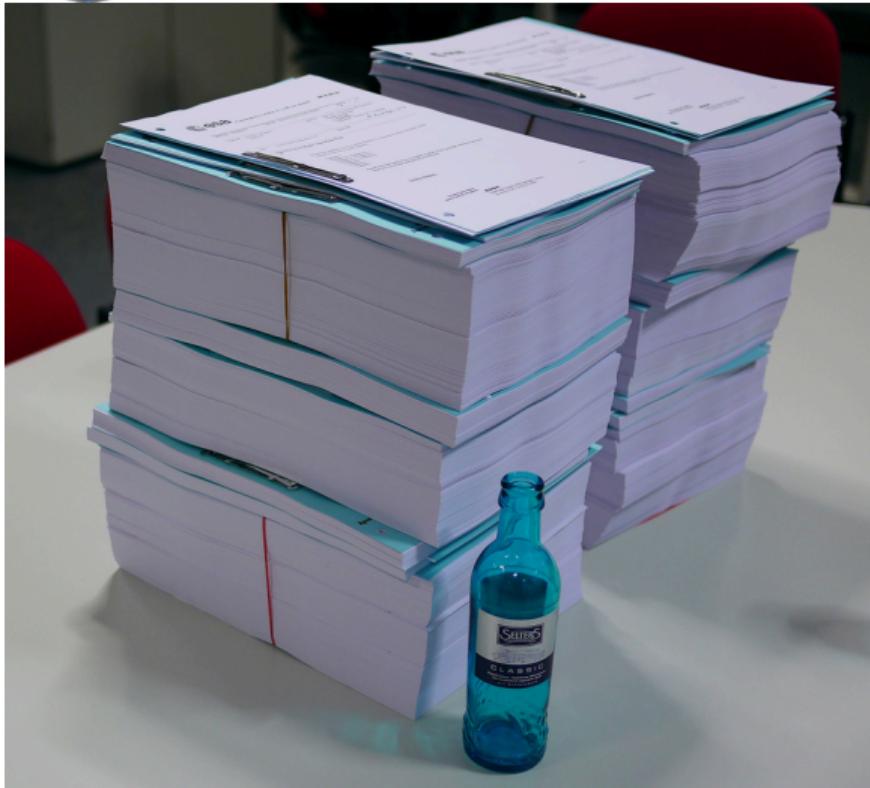
- Avoid different values of constants in peoples code ...
- The Gaia Parameter Database was set up early on for this (de Bruijne et al., 2005)
 - all constants in one place; web searchable configuration controlled (Only updated by Jos De Bruijne)
 - published as constants for Java (can also do C, Fortran...) so you may refer to a particular version
- then the actual data model — what exactly is an AstroElementary?
 - entire data model defined in multi-user dictionary tool; includes Units on each field.
 - good for astronomers — computer people find it harder to handle
 - from it we generate data instance classes and schemas for storage.
 - **ONLY data model not processing** — all objects are dumb (had discussion with KT and Mario on this)
 - This was in UML (Rose) in the 90s but it was impractical to continue..
 - Rubin Obs also had Rational Rose and still have Magic Draw .. but the Data model is in python now more or less.
- **These are logical extensions of having agreed conventions...**





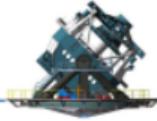
Flight Operations Procedures

Rubin
Observatory



The FOP is followed by the spacecraft operators - the paper copy is just in case the computers fail - could be useful!

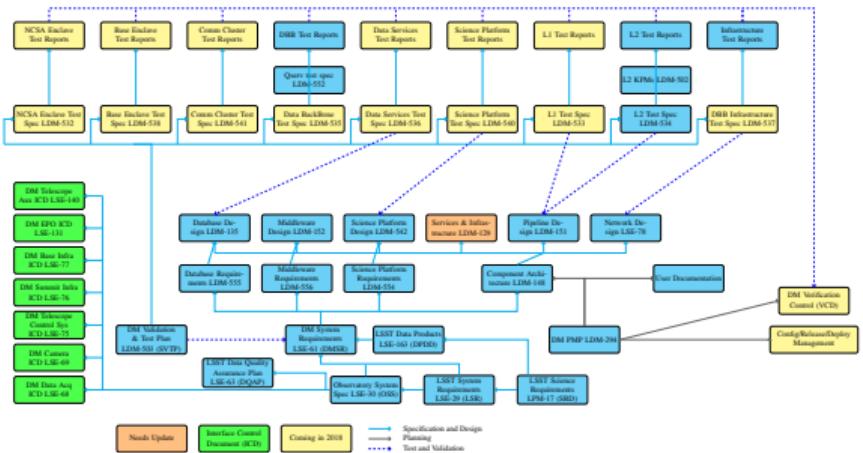
But we should avoid *write only* documents.



Have a standard: DPAC follows ECSS, Rubin MBSE

European Cooperation for Space Standardization

Rubin
Observatory



Rubin Obs DM Doctree (LDM-294)

- ECSS tailored as in figure
 - LaTeX Templates/examples provided (by SOC)
 - Documents are iterated — All of this is done for all products.
 - It is very good to have a standard set of documents augmented by technical notes and streamlined ECSS
 - Some still found it too heavy — other reports requested beyond the standard ones.
- DPAC had sufficient QA people (~ 1/CU) from the start to help with this



Development tools

- All DPAC code and docs in Subversion, hosted at ESAC
 - Access control according to Group membership in the LDAP
- Mantis for centralised issue tracking (includes risks and actions)
 - ALL DPAC issues in one system hosted at ESTEC
 - Jira would probably be better

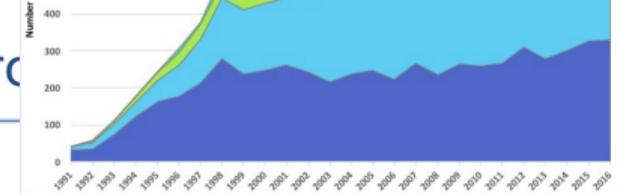


Development tools

- All DPAC code and docs in Subversion, hosted at ESAC
 - Access control according to Group membership in the LDAP
- Mantis for centralised issue tracking (includes risks and actions)
 - ALL DPAC issues in one system hosted at ESTEC
 - Jira would probably be better
- Having one language is good (O'Mullane et al., 2011b) agreed 2006
JH-001 — only one verification part is NOT in Java.
 - Can have a library of standard routines GaiaTools (Relativity, Field Angle Calculator, Ephemeris handling...)
 - The set of routines were not defined hence GaiaTools is a bit of hodgepodge mess...
 - Counter argument for common tools is (unnecessary) interdependence...
 - all libs in Nexus - builds against these no full builds
 - builds with Ant; Maven might be the thing to use now
 - automated builds with Hudson/Jenkins (previously cruisecontrol)
 - virtual machines make some reasons for Java invalid - LSST not ?



The Era Of Surveys and Archival Research



[https://archive.stsci.edu/hst/bibliography/
pubstat.html](https://archive.stsci.edu/hst/bibliography/pubstat.html)

... indicates archival research
probably play an important role
in the scientific success of
XMM-Newton Ness et al. (2014)

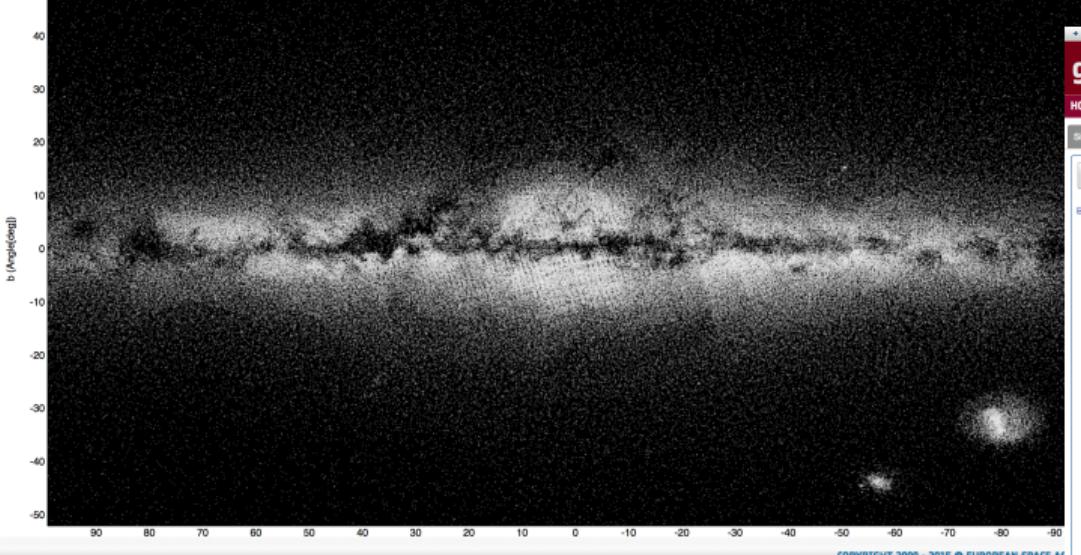


Gaia Archive

Rubin
Observatory

EUROPEAN SPACE AGENCY ABOUT ESAC

gaia archive visualization



+ EUROPEAN SPACE AGENCY ABOUT ESAC
gaia archive

HOME SEARCH STATISTICS VISUALIZATION HELP DOCUMENTATION VOSPACE SHARE

Simple Form ADQL Form Query Results

Job name: Query examples

Job name: Job ID: Type a query to be executed as a job.

Gaia Data Release 1

- gaia1r1.allwise_best_neighbour
- gaia1r1.allwise_neighbourhood
- gaia1r1.allwise_original_valid
- gaia1r1.aux_gao_lsqr2_match
- gaia1r1.cepheid
- gaia1r1.ext_phot_zero_point
- gaia1r1.gai_source
- gaia1r1.gsc23_best_neighbour
- gaia1r1.gsc23_neighbourhood
- gaia1r1.gsc23_original_valid
- gaia1r1.phot_variable_time_s
- gaia1r1.phot_variable_time_s
- gaia1r1.ppmal_best_neighbour
- gaia1r1.ppmal_neighbourhood
- gaia1r1.ppmal_original_valid
- gaia1r1.rnyne
- gaia1r1.sdes_d9_best_neigh
- gaia1r1.sdes_d9_neighbour

Ctrl+Space to query autocomplete

Status	Job	Creation date	Num. rows	Size
<input type="checkbox"/>	14977193855390	17-Jun-2017, 19:09:45	0	0 KB
<input checked="" type="checkbox"/>	oulierParax	05-Nov-2016, 17:27:34	313	10 KB
<input checked="" type="checkbox"/>	14783608081700	05-Nov-2016, 16:48:48	0	0 KB
<input checked="" type="checkbox"/>	14783607846200	05-Nov-2016, 16:48:24	0	0 KB
<input checked="" type="checkbox"/>	parallax_dff	04-Nov-2016, 14:47:33	93635	2 MB
<input checked="" type="checkbox"/>	14776461654710	28-Oct-2016, 11:49:25	16285	4 MB
<input checked="" type="checkbox"/>	xmatch_lycha2_sgn	10-Jun-2016, 18:10:54	0	0 KB

Download format:

COPYRIGHT 2017 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED. [v1.1]

All Gaia data is publicly accessible at <https://gea.esac.esa.int/archive/>



William O'Mullane

Software and big data

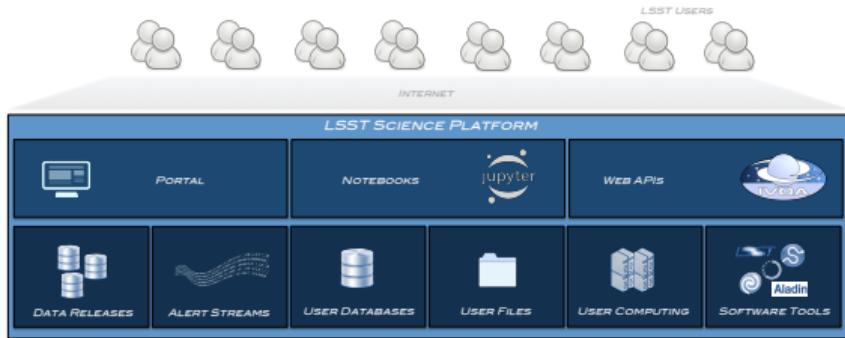
25

U.S. DEPARTMENT OF
ENERGY
Office of Science



LSST Science Platform

Rubin
Observatory



For DR2:

- Computing: 2,400 cores (≈ 18 TFLOPs)
- File storage: $\approx 4\text{PB}$ (VOSpace)
- Database storage: $\approx 3\text{PB}$ (MYDB)

The Science Platform has three user facing aspects: the Portal (novice), the JupyterLab (intermediate), and the Web APIs (expert and remote tools).

Vision: LSE-319 — Design: LDM-542 — Test: DMTR-51

This is the sort of environment users now expect to have - it is no longer novel. We are finally **Bringing code to the data** - almost did with GAVIP Vagg et al. (2016)

Bruno will have more to say on this topic.





General Challenge : Data volume

Rubin
Observatory

See also Bauer et al. (2019)

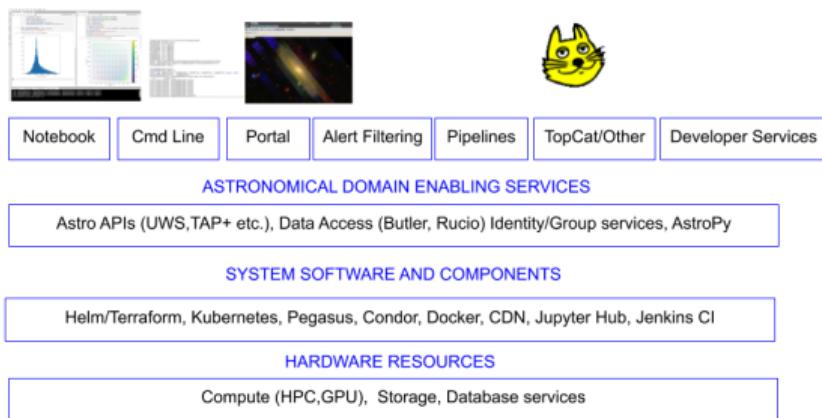
Soon, if not already, Data will be looking for astronomers not vice versa. Looking at Hubble archive 7K datasets have no publications.

- Proprietary data may have had its day .. if we want people to look at data we need to remove barriers.
- Networks and infrastructure are improving but more needs to be done
- Who looks after all the data ?
 - For space science in Europe ESAC preserves data - it is relatively small
 - IPAC, HEASARC and STScI pretty much deal with NASA data - still not one location.
 - Who looks after all the ground based data ?
 - There is no long term preservation plan for LSST or other big telescopes in the USA (Alex wil probably have more to say on this)





We all constantly redesign and rebuild wheels, this will become too expensive as data volume grows ..



- We should agree a component based cyber infrastructure model and work on improving specific components to plug in - right now we are all building TAP, Designing Databases , deploying Jupyter ...
- Data models like CAOM from IVOA are going to be essential going forward so we can inter-operate on data

Filesystems are End Of Life - object stores should not be confused with the object databases many of us struggled with in the 90s and 00s .. Google and Amazon do not run filesystems they run object stores.



Processing patterns

- LSST allows 9 months for a Data Release Processing Cycle - probably about 6 months actual processing.
- The original requirement for Gaia astrometric solution was three months.
- Traditional batch systems and shared nothing architecture may not always work
 - Gaia Astrometric Solution required temporal spatial access and had global matrices
 - There are processes for LSST which will have similar problems e.g. Forward Global Calibration Model (FGCM) <https://github.com/lsst/fgcmcal> (Burke et al., 2018)
- preparing and staging for tasks can take ever longer as our processing becomes more sophisticated
- And then there is machine learning!
 - Automated discovery - need it great !
 - Reproducibility, understandability ..



Software is no longer an afterthought

Rubin
Observatory

- In ESA (personal opinion) Gaia was a game changer - the Software was seen as critical - I had to fight to keep it off the Launch Critical Items List. We still had a meager 10% or so of the budget.
- LSST recognized early on that the processing was as important as the telescope - DM is one of 4 Subsystem with about 20% of the project budget
- We need to consider long term software support and the software eco system
 - Open source and publicly scrutinized algorithms
 - Agree community cyber infrastructure model (previous slide)
 - Though I made my career in astronomy/computing .. its not easy .. and I really had to do the management thing
- Education and outreach - How do we do better?
 - Does everyone need to be a data scientist/programmer ?
 - Educate astronomers/managers on how to *open source*
 - How do we foster inclusion ?





Its all team work on big projects

Rubin
Observatory



The End



Acronyms I

Acronym	Description
AGIS	Astrometric Global Iterative Solution
Archive	The repository for documents required by the NSF to be kept. These include documents related to design and development, construction, integration, test, and operations of the LSST observatory system. The archive is maintained using the enterprise content management system DocuShare, which is accessible through a link on the project website www.project.lsst.org .
CAOM	Common Astronomical Observation Model
CCD	Charge-Coupled Device
CI	Continuous Integration
CU	Coordination Unit
Camera	The LSST subsystem responsible for the 3.2-gigapixel LSST camera, which will take more than 800 panoramic images of the sky every night. SLAC leads a consortium of Department of Energy laboratories to design and build the camera sensors, optics, electronics, cryostat, filters and filter exchange mechanism, and camera control system.
DM	Data Management
DMTR	DM Test Report
Data Backbone	The software that provides for data registration, retrieval, storage, transport, replication, and provenance capabilities that are compatible with the Data Butler. It allows data products to move between Facilities, Enclaves, and DACs by managing caches of files at each endpoint, including persistence to long-term archival storage (e.g. tape).
Data Management	The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson.
Data Release	The approximately annual reprocessing of all LSST data, and the installation of the resulting data products in the LSST Data Access Centers, which marks the start of the two-year proprietary period.



Acronyms II

Rubin
Observatory

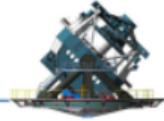
ESA	European Space Agency
ESAC	European Space Astronomy Centre
ESOC	European Space Operations Centre
ESTEC	European Space Technology Engineering Centre
FGCM	Forward Global Calibration Model
GAVIP	Gaia Added Value Interface Platform
GB	Gigabyte
HEALPix	Hierarchical Equal-Area iso-Latitude Pixelisation
HEASARC	NASA's Archive of Data on Energetic Phenomena
HSC	Hyper Suprime-Cam
HTM	Hierarchical Triangular Mesh
IPAC	No longer an acronym; science and data center at Caltech
IR	Infra Red
ISSI	International Space Science Institute
IVOA	International Virtual-Observatory Alliance
LDM	LSST Data Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)
LSST	Large Synoptic Survey Telescope
MPP	Massively Parallel Process
NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputing Applications
Operations	The 10-year period following construction and commissioning during which the LSST Observatory conducts its survey
PB	PetaByte
PSF	Point Spread Function





Acronyms III

Project Manager	The person responsible for exercising leadership and oversight over the entire LSST project; he or she controls schedule, budget, and all contingency funds
Qserv	Proprietary Database built by SLAC for LSST
SDSS	Sloan Digital Sky Survey
Science Platform	A set of integrated web applications and services deployed at the LSST Data Access Centers (DACs) through which the scientific community will access, visualize, and perform next-to-the-data analysis of the LSST data products.
Subsystem	A set of elements comprising a system within the larger LSST system that is responsible for a key technical deliverable of the project.
TAP	Table Access Protocol
TB	TeraByte
XMM	X-ray Multi-mirror Mission (ESA; officially known as XMM-Newton)
arcmin	arcminute minute of arc (unit of angle)
arcsec	arcsecond second of arc (unit of angle)
astrometry	In astronomy, the sub-discipline of astrometry concerns precision measurement of positions (at a reference epoch), and real and apparent motions of astrophysical objects. Real motion means 3-D motions of the object with respect to an inertial reference frame; apparent motions are an artifact of the motion of the Earth. Astrometry per se is sometimes confused with the act of determining a World Coordinate System (WCS), which is a functional characterization of the mapping from pixels in an image or spectrum to world coordinate such as (RA, Dec) or wavelength.
calibration	The process of translating signals produced by a measuring instrument such as a telescope and camera into physical units such as flux, which are used for scientific analysis. Calibration removes most of the contributions to the signal from environmental and instrumental factors, such that only the astronomical component remains.
camera	An imaging device mounted at a telescope focal plane, composed of optics, a shutter, a set of filters, and one or more sensors arranged in a focal plane array.



Acronyms IV

metric	A measurable quantity which may be tracked. A metric has a name, description, unit, references, and tags (which are used for grouping). A metric is a scalar by definition. See also: aggregate metric, model metric, point metric.
--------	---



References

- [BAS-003]**, Bastian, U., 2007, *Reference systems, conventions and notations for Gaia*,
GAIA-CA-SP-ARI-BAS-003, URL <http://www.rssd.esa.int/cs/livelink/open/358698>
- Bauer, A.E., Bellm, E.C., Bolton, A.S., et al., 2019, arXiv e-prints, arXiv:1905.05116 (arXiv:1905.05116), ADS Link
- [LDM-135]**, Becla, J., Wang, D., Monkewitz, S., et al., 2017, *Data Management Database Design*, LDM-135, URL <https://ls.st/LDM-135>
- [DMTR-51]**, Bosch, J., Chiang, H.F., Gower, M., et al., 2017, *LDM-503-02 (HSC Reprocessing) Test Report*, DMTR-51, URL <https://ls.st/DMTR-51>
- Burke, D.L., Rykoff, E.S., Allam, S., et al., 2018, AJ, 155, 41 (arXiv:1706.01542), doi:10.3847/1538-3881/aa9f22, ADS Link
- de Bruijne, J.H.J., Lammers, U., Perryman, M.A.C., 2005, In: C. Turon, K. S. O'Flaherty, & M. A. C. Perryman (ed.) *The Three-Dimensional Universe with Gaia*, vol. 576 of ESA Special Publication, 67–+, ADS Link
- [LDM-542]**, Dubois-Felsmann, G., Lim, K.T., Wu, X., et al., 2017, *LSST Science Platform Design*, LDM-542, URL <https://ls.st/LDM-542>
- [JH-001]**, Hernandez, J., 2014, *Main Database Interface Control Document*,
GAIA-C1-SP-ESAC-JH-001, URL <http://www.rssd.esa.int/cs/livelink/open/2786145>
- Ivezić, Ž., Kahn, S.M., Tyson, J.A., et al., 2019, ApJ, 873, 111 (arXiv:0805.2366), doi:10.3847/1538-4357/ab042c, ADS Link
- [LSE-319]**, Jurić, M., Ciardi, D., Dubois-Felsmann, G., 2017, *LSST Science Platform Vision Document*, LSE-319, URL <https://ls.st/LSE-319>
- Lindgren, L., Lammers, U., Hobbs, D., et al., 2012, A&A, 538, A78 (arXiv:1112.4139), doi:10.1051/0004-6361/201117905, ADS Link
- Lupton, R., Blanton, M.R., Fekete, G., et al., 2004, PASP, 116, 133 (arXiv:astro-ph/0312483), doi:10.1086/382245, ADS Link
- Ness, J.U., Parmar, A.N., Valencic, L.A., et al., 2014, Astronomische Nachrichten, 335, 210 (arXiv:1311.5751), doi:10.1002/asna.201312001, ADS Link
- O'Mullane, W., Banday, A.J., Górski, K.M., Kunszt, P., Szalay, A.S., 2001, In: Banday, A.J., Zaroubi, S., Bartelmann, M. (eds.) *Mining the Sky*, 638,
doi:10.1007/10849171_84, ADS Link
- O'Mullane, W., Lammers, U., Bailer-Jones, C., et al., 2006, ArXiv Astrophysics e-prints (arXiv:astro-ph/0611885), ADS Link
- O'Mullane, W., Hernández, J., Hoar, J., Lammers, U., 2009, In: D. A. Bohlender, D. Durand, & P. Dowler (ed.) *Astronomical Data Analysis Software and Systems XVIII*, vol. 411 of *Astronomical Society of the Pacific Conference Series*, 470, ADS Link
- O'Mullane, W., Lammers, U., Lindegren, L., Hernandez, J., Hobbs, D., 2011a, Experimental Astronomy, 31, 215 (arXiv:1108.2206),
doi:10.1007/s10686-011-9248-z, ADS Link
- O'Mullane, W., Luri, X., Parsons, P., et al., 2011b, ArXiv e-prints (arXiv:1108.0355), ADS Link
- [LDM-294]**, O'Mullane, W., Swinbank, J., Jurić, M., DMLT, 2018, *Data Management Organization and Management*, LDM-294, URL <https://ls.st/LDM-294>
- [Document-31100]**, Thomson, J.R., 2019, *LSST Benchmarking of Qserv and BigQuery*, Document-31100, URL <https://ls.st/Document-31100>
- Vagg, D., O'Callaghan, D., O'Hóigáin, F., et al., 2016, In: Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 9913 of SPIE, 99131V (arXiv:1605.09287), doi:10.1117/12.2233619, ADS Link