

Crafting big data systems for astronomy

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

March 6, 2020



William O'Mullane

Thoughts on crafting big data systems



what did and did not work ..

Rubin
Observatory

Or

the good

the bad

and

the ugly.





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)



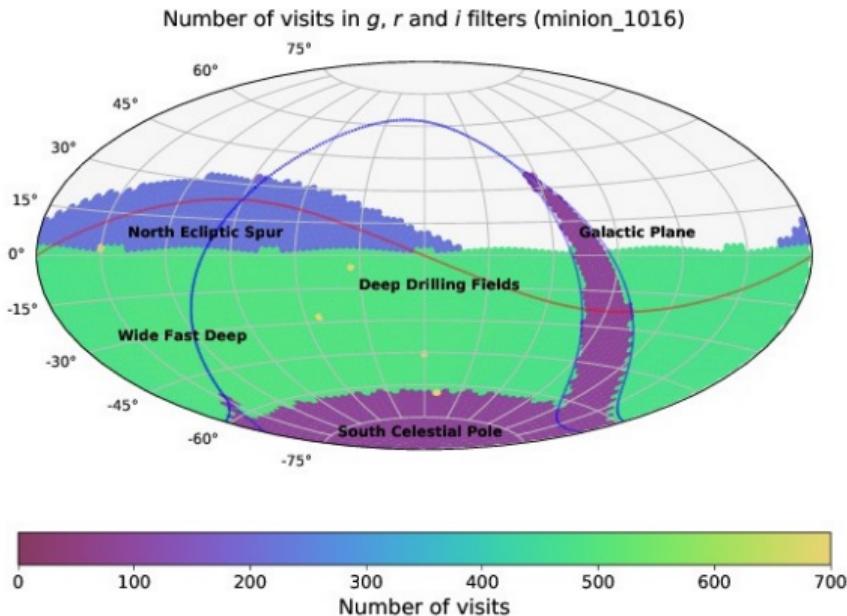
LSST:uniform sky survey

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 Ivezić et al. (2019)-arXiv:0805.2366



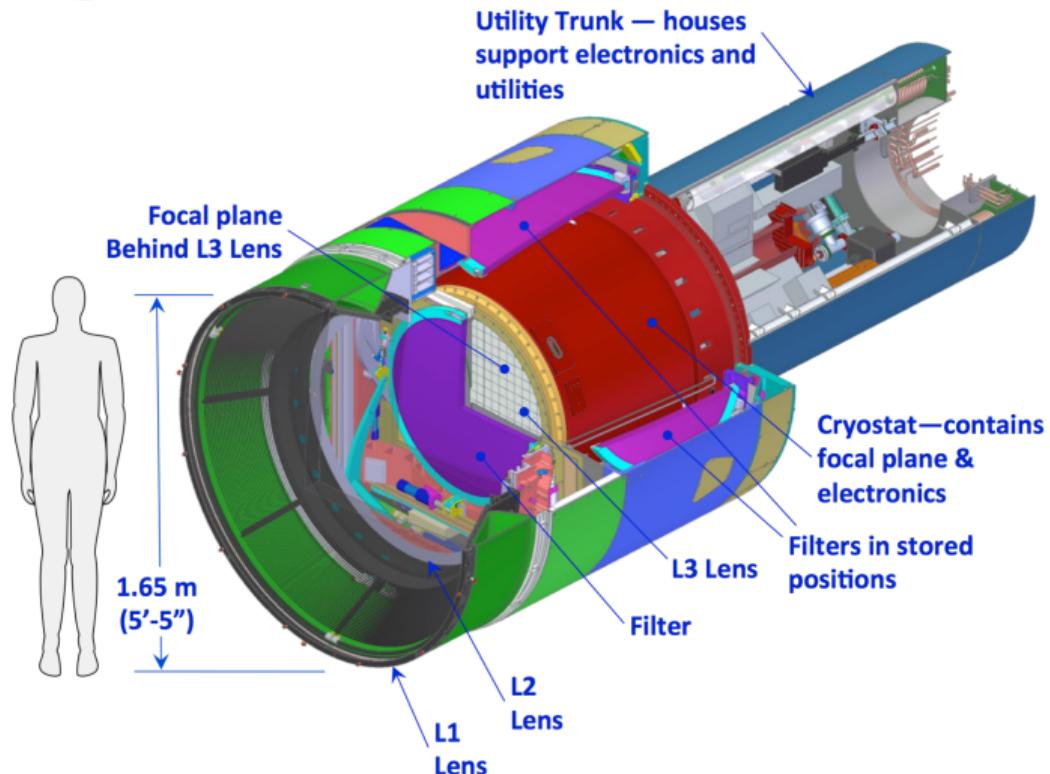
**10-year simulation of LSST survey:
number of visits in u,g,r band (Aitoff
projection of eq. coordinates)**





LSST Camera

Rubin
Observatory



The largest astronomical camera:

- 2800 kg
- 3.2 Gpix



Site as imagined and in March 2019

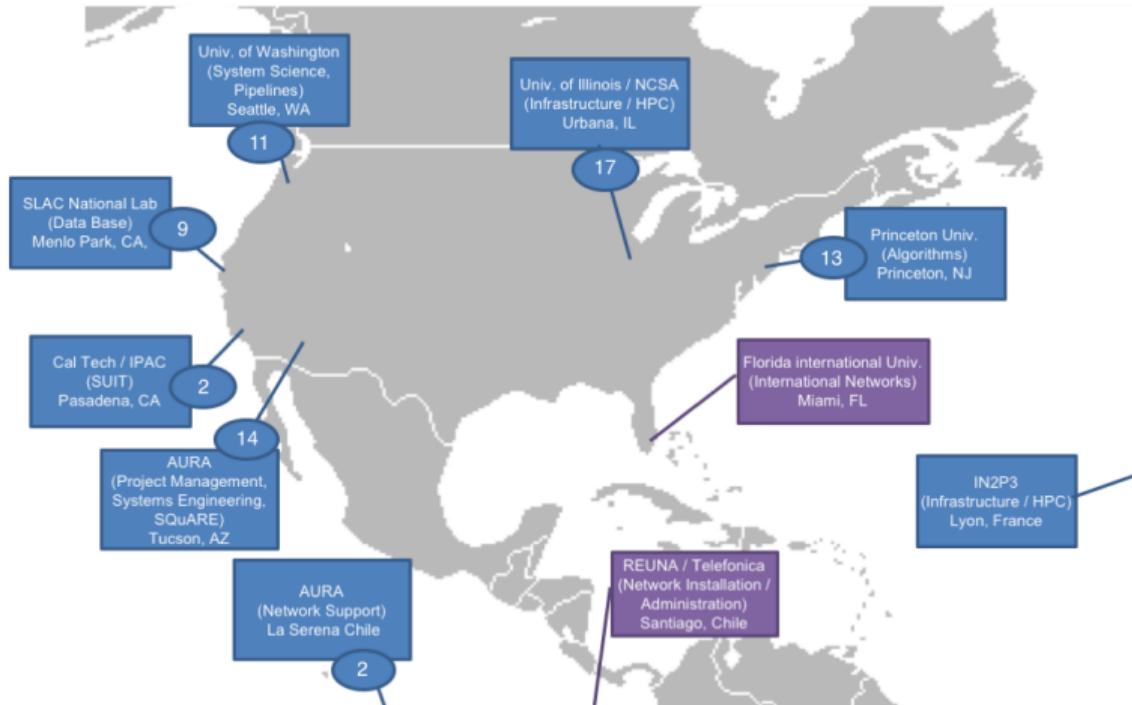
Rubin
Observatory





Data Management Mission Statement

Rubin
Observatory



DM's mission:
Stand up operable, maintainable, quality services to deliver high-quality LSST data products for science, all on time and within reasonable cost.

Development is distributed across the Americas.
Plus we have partners like IN2P3 (France).

About 100 individuals \approx 80 FTE



Communication is difficult and important I

Have a code of conduct.

- Internal

- Data Management Leadership meeting every Monday (30 minutes) - 4 times a year longer 2.5 days (2 now virtual with moderators and 5 hour days to match time zones).
- Gaia DPAC executive similar (2 times physical meeting per year) - technical focus.
- Both Rubin and DPAC have **newsletters** excellent and (DPAC) well contributed to
- (Few) Focused working groups and working meetings
- DPAC First consortium meeting Nov 2015; **10 years perhaps late** .
- Rubin have one every year **it seems too much**
- Though stressful personally **Rubin putting all software under one leader is good** - **Done rather late** hence stress .
- As for any project cost of entry for new people is very high — **no obvious solution**

- ESA policy initially to reduce contact between DPAC and Astrium (now AirbusDS who constructed Gaia) **not good**



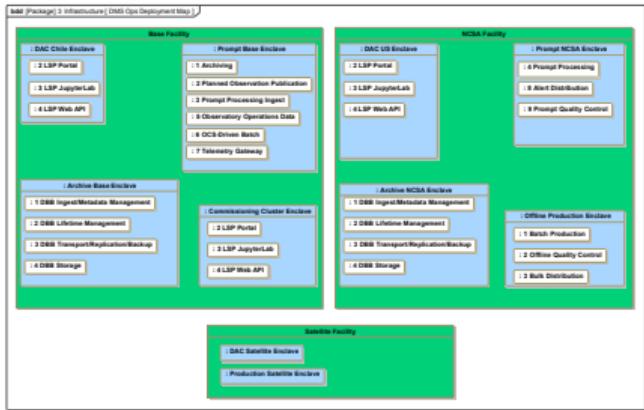
Communication is difficult and important II

– External

- Perhaps could have had a better DPAC website - Rubin is somewhat ok .. **would really like a git backed Pull Request driven site .. but we have Drupal..**
- ESA PR also not great (ok as they point out they have a tiny fraction of NASA budget)
- A bit better on Rubin (also fraction of NASA budget).
- **DPAC Publication policy was dealt with very late**
- **Publication policy clearly in place on Rubin**



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)

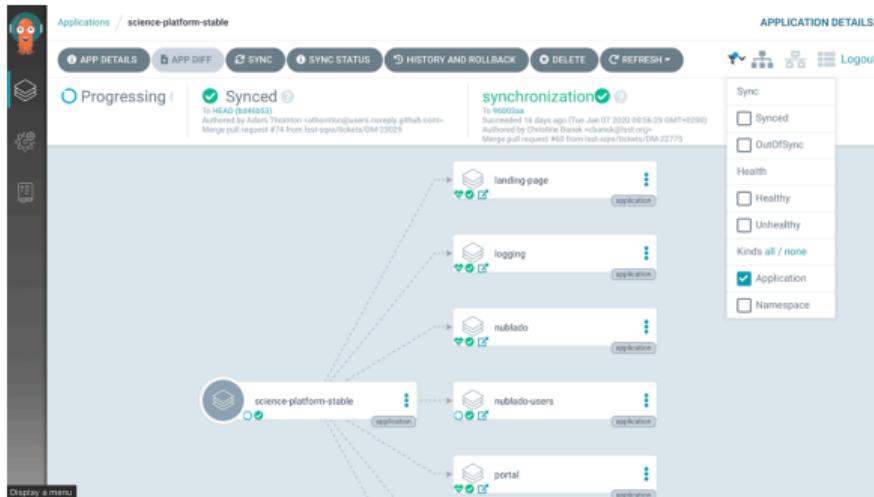
About $\frac{1}{2}$ million lines of code (C++/python) all open source on github

diagram K.T. Lim



Kubernetes-based services with ArgoCD

- Kubernetes provides powerful container orchestration and management of resources available to services
- ArgoCD provides a framework for deploying and monitoring Kubernetes-based services with configuration management via GitOps
- Vault provides secrets management
- Combined, these three allow us to have automated, reproducible deployments



Frossie Economou frossie@lsst.org

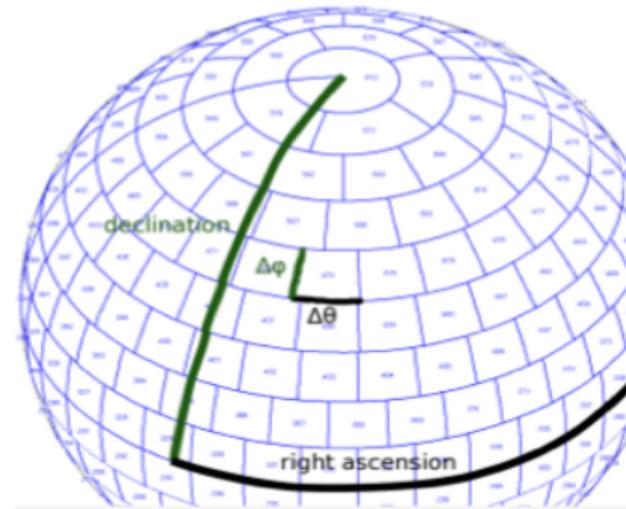


Astronomy catalogs tend to be highly structured, tabular, somewhat predictable access.

- 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**

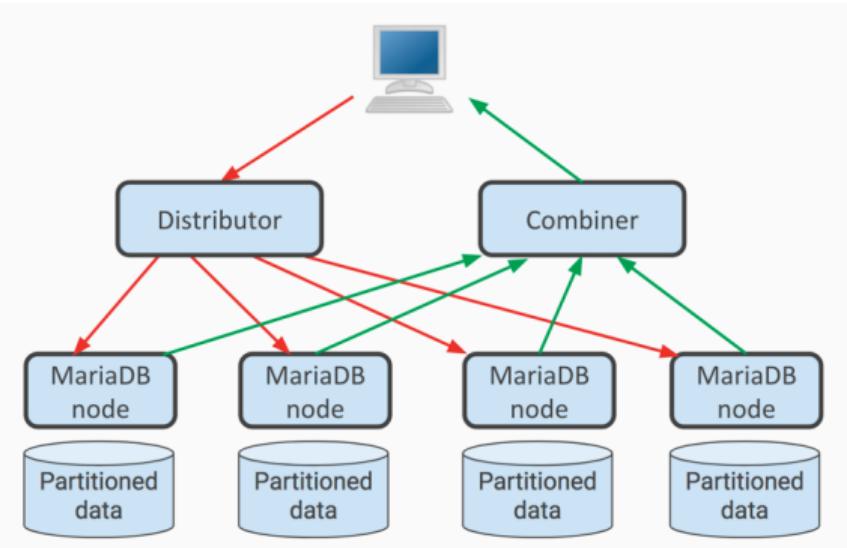
Built on project at SLAC, leverage existing tech within Stanford (MariaDB, MySQL Proxy, XRootD, Google protobuf, Flask)

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





Guidelines, tools, conventions

Rubin
Observatory

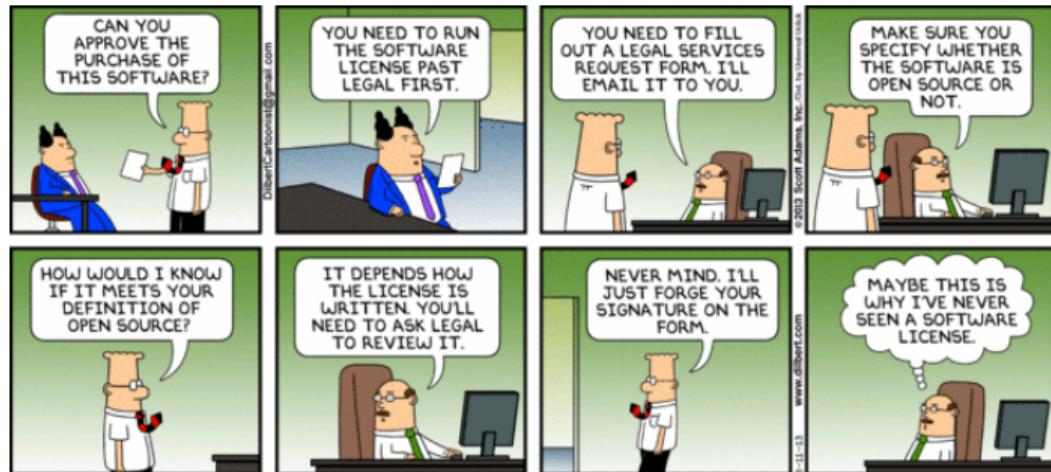
- Its great having extensive guidelines - it was also something super on Gaia
- DPAC had full engineering Guide from early on (WOM-011) - how to Mantis, SVN, Java guide ...
- <http://developer.lsst.io> is a full developer guide - everything from git commit messages to style guides for Python and C+.
- <http://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.
 - RubinObs lacks something like BAS-003





Software (and data) licensing

- Protect intellectual property — grant use to the consortium
- Often forgotten or not well dealt with - or worse ignored!
- DPAC agreed to LGPL (WOM-019) - some institutes e.g. ESA, do not allow GPL code
- LGPL in ESA involved lawyers and directors and time **Now ESA have own open license.**
- Rubin GPL .. would prefer APL.
- **Gaia Data license only after DR1** (open with attribution).



You may use up to seven (7) cartoons per year at no costs as part of our fair use policy.

Have only mentioned Data license on Rubin



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 ?
- Photometry challenging - will Gaia help ..
- Everything is blended!!

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

Image HSC collaboration, Robert Lupton



Catalog extraction

```
File Notebook Editor Terminal Console Help Hub Save/Exit
TERMINAL SESSIONS
  terminal/2 SHUTDOWN
  terminal/1 SHUTDOWN
KERNEL SESSIONS
  simon.ipnb SHUTDOWN
TERMINAL SESSIONS
  terminal/2 SHUTDOWN
  terminal/1 SHUTDOWN
  simon.ipnb SHUTDOWN
In [10]: butler = daf_persistence.Butler('singlechip_sample')
exp = butler.get('calexp', visit=410877, ccd=28, filter='r')

In [11]: import lsst.afw.geom as afw_geom
bbox = afw_geom.Box2I(afw_geom.Point2D(1024, 1024), afw_geom.Extent2D(512,512))
sources = butler.get('src', visit=410877, ccd=28, filter='r')
overlay_masks(exp, bbox=bbox, sources=sources)

In [12]: x = sources.getFxFlux()
y = sources.getNdxFlux() / sources.getFxFlux()

frossie@jld-lab: ~
```

```
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 LSST environment in a terminal.

If you want to add your own configuration to LSST stack startup in the
LSST ipython kernel, create a sourceable shell fragment in
$HOME/.notebooks/_user_setups and it will be sourced during kernel
startup.

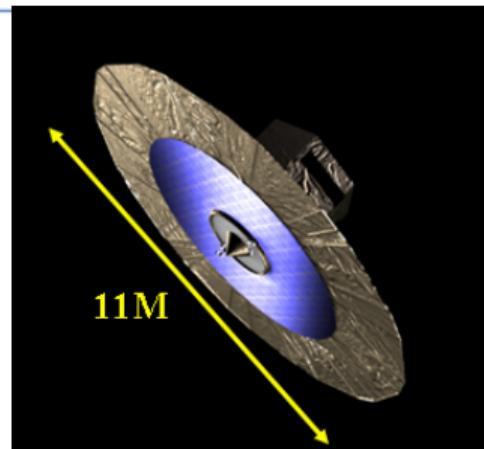
[frossie@jld-lab: ~]$
```

- 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
- This helps users understand how catalogs are produced.



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 $G < 20$

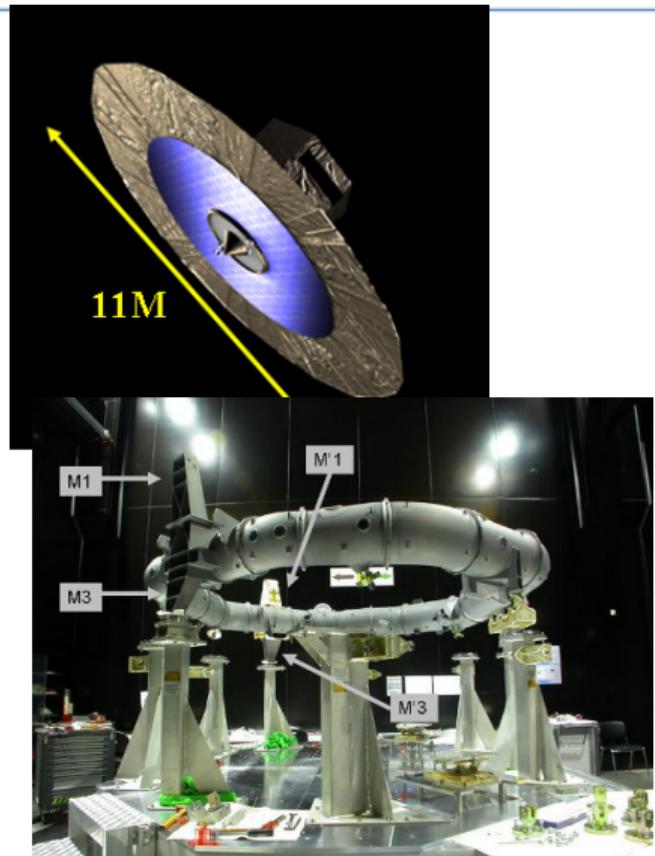


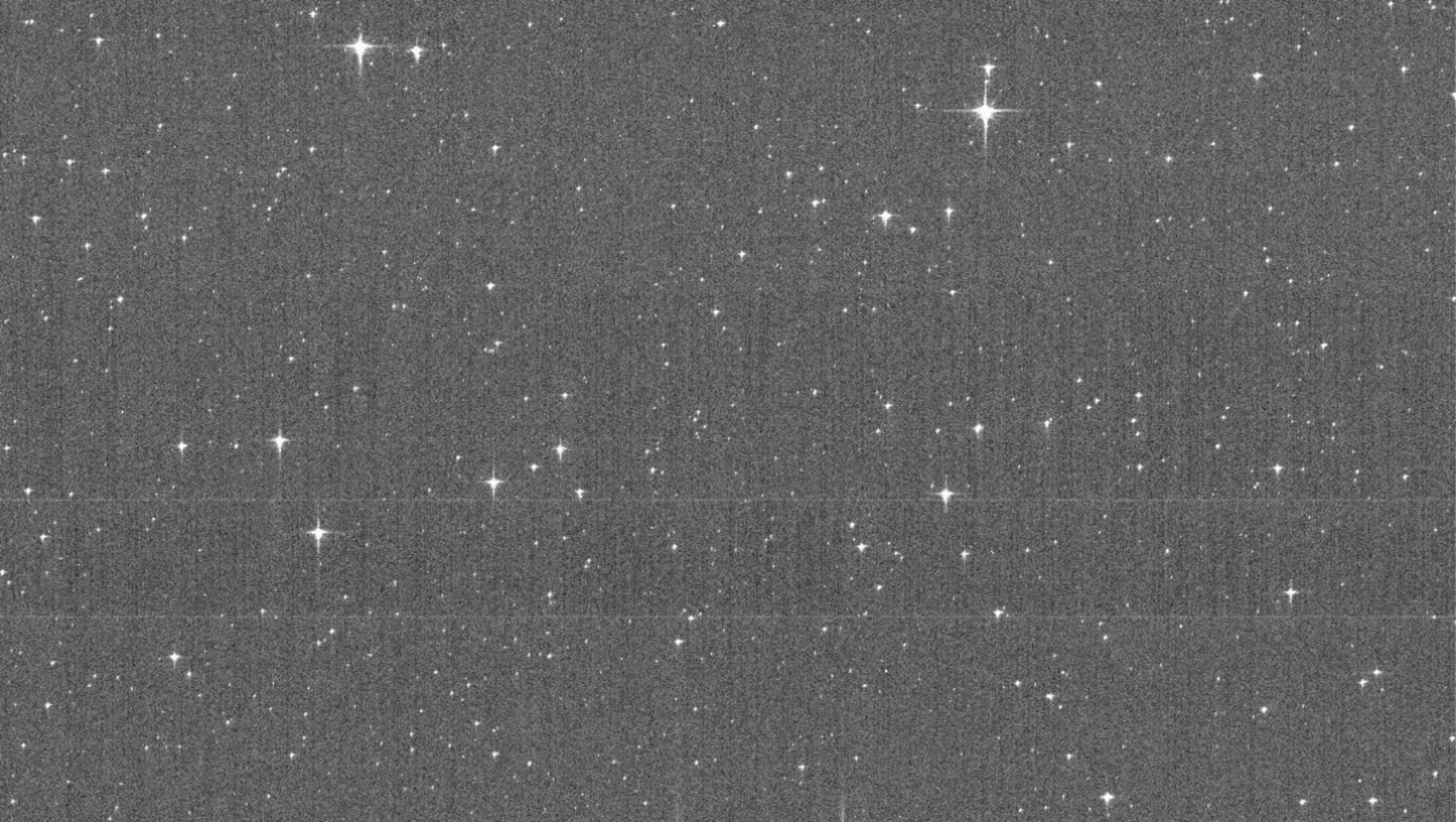


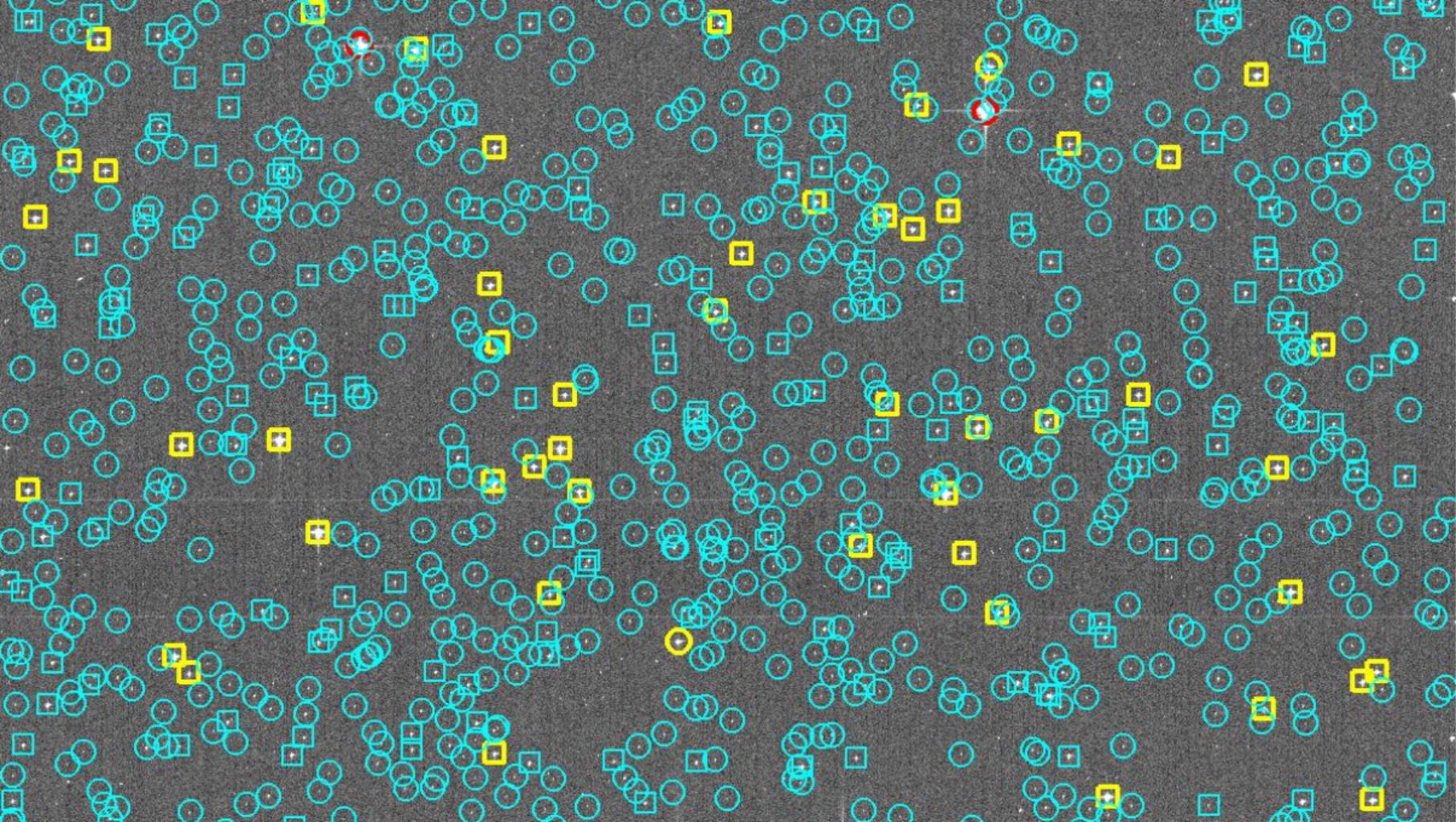
Satellite overview

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.



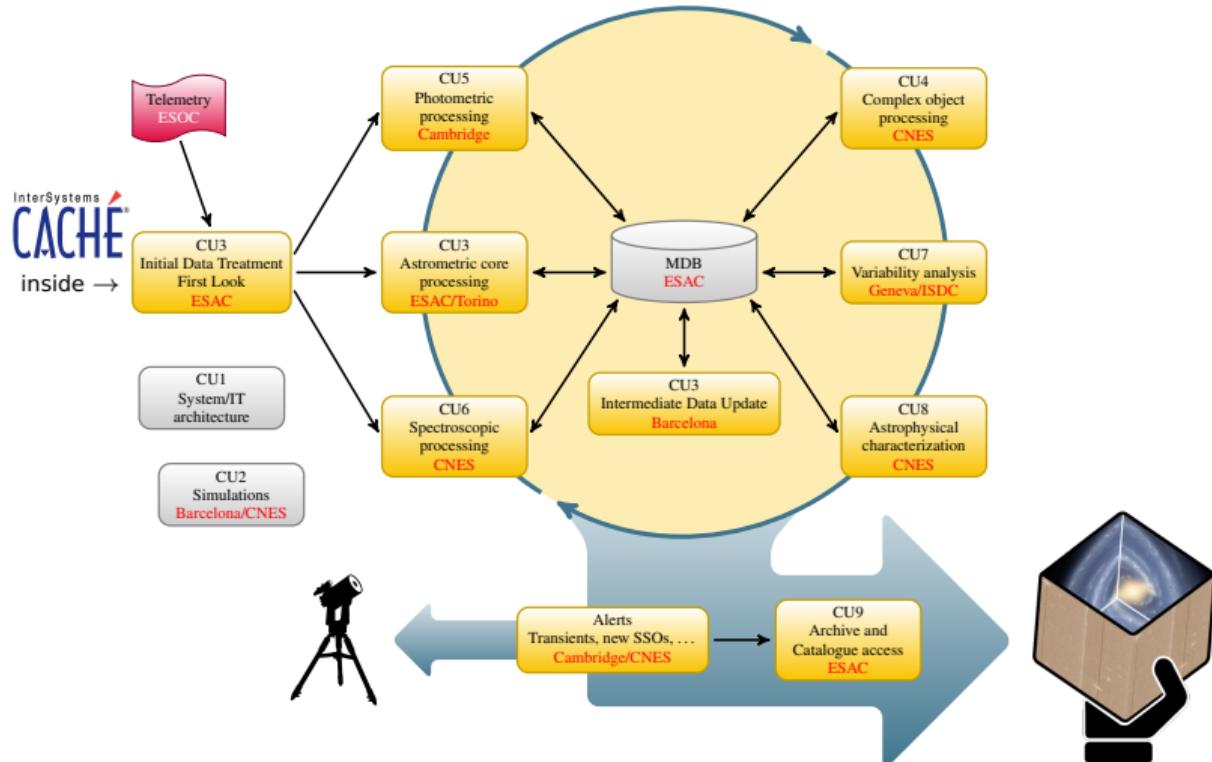
- Astrometric centroid of the CCD image to be determined to an accuracy of 1% of the pixel size!
 - There will be 10^{12} images $\approx 100\text{TB}$ downlink need to handle $\approx 1\text{PB}$
 - At 1 millisecond each that is ≈ 30 years
- Reconstructed attitude is required to order $10 \mu\text{arcsec}$
 - 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



CU=Coordination Unit

Daily 50 – 100GB
compressed

5yrs 1PB

(see also O'Mullane et al. (2006),
O'Mullane et al. (2009))



Parameters and data models

Rubin
Observatory

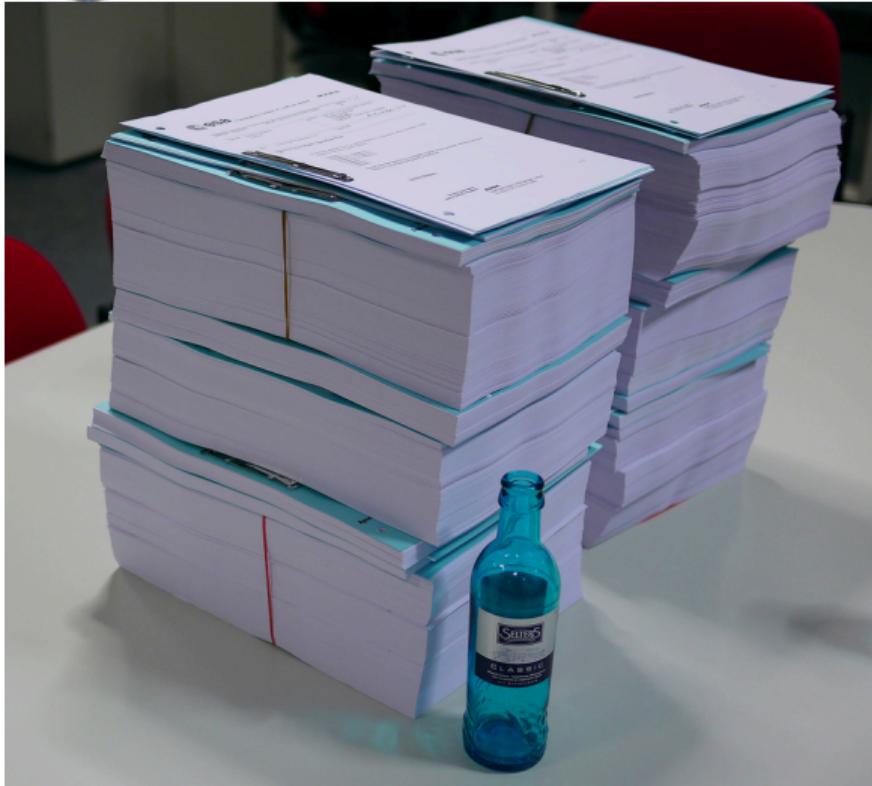
- Avoid different values of constants in peoples code ...
- The Gaia Parameter Database was set up early on for this (de Bruijne et al., 2005)
 - all constants in one place; web searchable configuration controlled (Only updated by Jos De Bruijne)
 - published as constants for Java (can also do C, Fortran...) 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

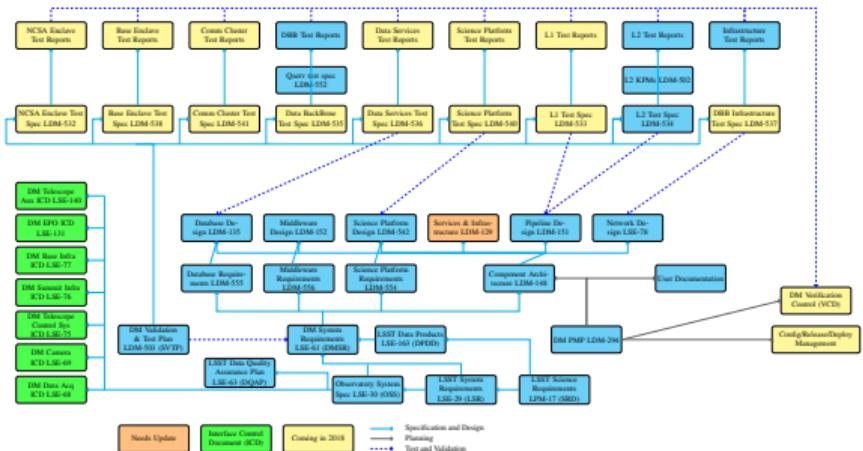


The FOP is followed by the spacecraft operators - the paper copy is just in case the computers fail - could be useful!

But we should avoid *write only* documents.



Have a standard: DPAC follows ECSS, Rubin MBSE



Rubin Obs DM Doctree (LDM-294)

Rubin less rigorous than
DPAC - but then its on the ground

European Cooperation for Space Standardization

- Standards need to be tailored
 - LaTeX Templates/examples provided
 - 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
 - Some still found it too heavy — other reports requested beyond the standard ones.
- DPAC had sufficient QA people (~ 1/CU) from the start (Rubin NONE)



Development tools

- All DPAC code and docs in Subversion, Rubin in GitHub.
 - Access control according to Group membership in the LDAP
- Centralized issue tracking (includes risks and actions)
 - DPAC - Mantis eventually Jira. Rubin Jira



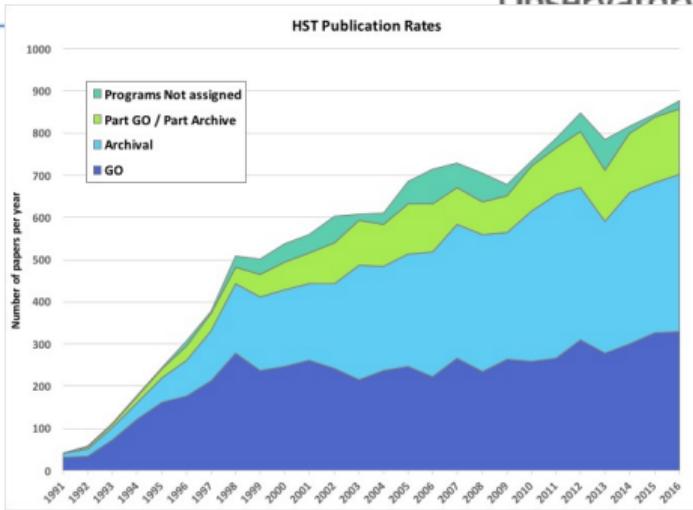
Development tools

- All DPAC code and docs in Subversion, Rubin in GitHub.
 - Access control according to Group membership in the LDAP
- Centralized issue tracking (includes risks and actions)
 - DPAC - Mantis eventually Jira. Rubin Jira
- Having one language is good (O'Mullane et al., 2011b) Gaia agreed on Java 2006, Rubin Python/C++ (2009 perhaps earlier)
 - 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... we have that on Rubin
 - virtual machines make some reasons for Java invalid
 - Rubin builds take a long time ..



The Era Of Surveys and Archives

Rubin
Observatory



<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)



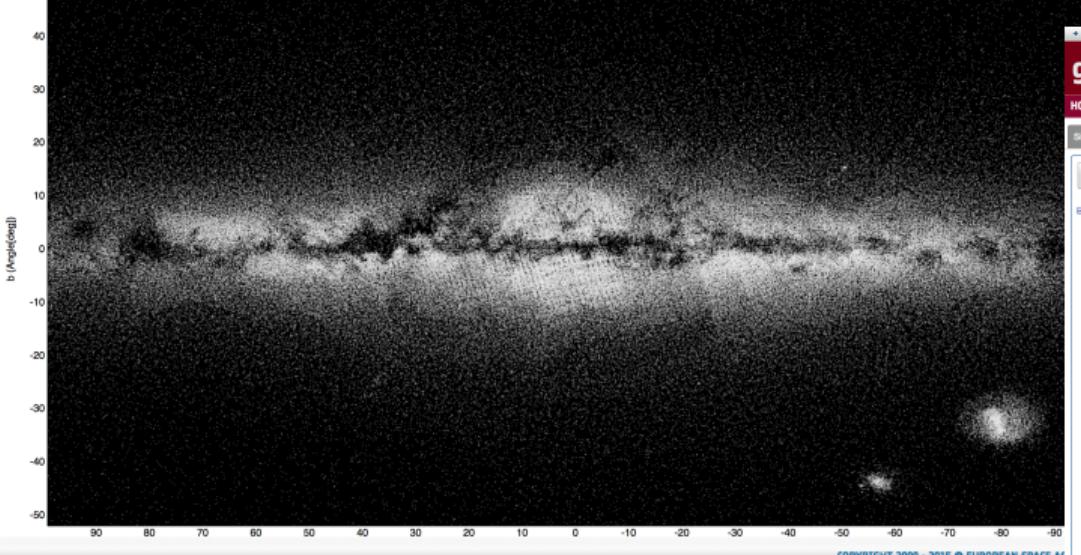


Gaia Archive

Rubin
Observatory

EUROPEAN SPACE AGENCY ABOUT ESAC

gaia archive visualization



COPYRIGHT 2000 - 2015 © EUROPEAN SPACE AGENCY



gaia archive

HOME SEARCH STATISTICS VISUALIZATION HELP DOCUMENTATION VOSPACE SHARE

Simple Form ADQL Form Query Results

Job name: Query examples

1 SELECT c."dist", a."id", a."hip", a."tyc1", a."tyc2", a."tyc3",
a."id_tyc1", a."id_tyc2", a."ra_mdeg", a."de_mdeg", Type a query to be executed as a job.
a."ep_ra_mdeg", a."ep_de_mdeg", a."ep_pm_mdeg", a."ep_ra_deg",
a."ep_de_deg", a."ep_pm_deg", a."ra", a."de", a."pm", a."ra_pm_mdeg",
a."de_pm_mdeg", a."pm_pm_mdeg", a."ra_deg", a."de_deg", a."pm_deg", a."ra_pm_deg", a."de_pm_deg", a."pm_pm_deg",
a."pflag", a."parallax", a."ecdm", a."prox", a."bt_mag", a."vt_mag",
a."ebt_mag", a."evt_mag", b."agn_id", b."de1950", b."de2000".
Ctrl+Space to query autocompletion

Reset Form Submit Query

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: VOTable Apply jobs filter Select all jobs Delete selected jobs

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

Thoughts on crafting big data systems

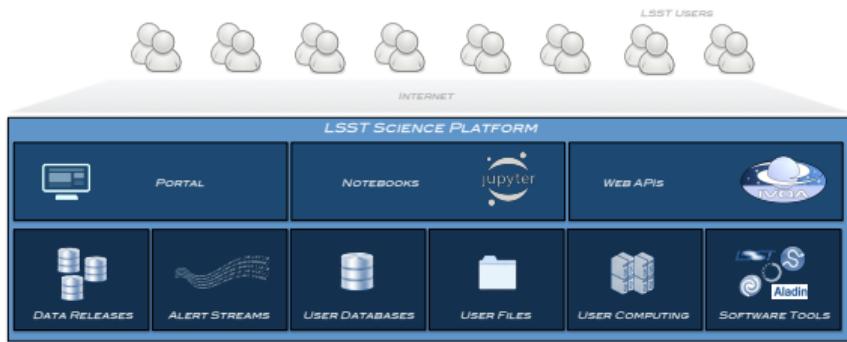
29

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)





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)



Its all team work on big projects

Rubin
Observatory



The End



Acronyms I

Acronym	Description
APL	Apache Public Licence
AURA	Association of Universities for Research in Astronomy
C	Specific programming language (also called ANSI-C)
CC	Change Control
CC-IN2P3	Centre de Calcul de l'IN2P3
CCD	Charge-Coupled Device
DM	Data Management
DMTR	DM Test Report
DOE	Department of Energy
DPAC	Data Processing and Analysis Consortium (Gaia)
ESA	European Space Agency
ESAC	European Space Astronomy Centre
EUPS	Extended Unix Product System
FTE	Full-Time Equivalent
GB	GigaByte
GPL	GNU Public License
HEALPix	Hierarchical Equal-Area iso-Latitude Pixelisation
HEASARC	NASA's Archive of Data on Energetic Phenomena
HSC	Hyper Suprime-Cam
HTM	Hierarchical Triangular Mesh
IN2P3	Institut National de Physique Nucléaire et de Physique des Particules



Acronyms II

IPAC	Infrared Processing and Analysis Center
IR	Infra Red
IT	Integration Test
LDM	LSST Data Management (Document Handle)
LSE	LSST Systems Engineering (Document Handle)
LSST	Legacy Survey of Space and Time (formerly Large Synoptic Survey Telescope)
LaTeX	(Leslie) Lamport TeX (document markup language and document preparation system)
MPP	Massively Parallel Process
MYDB	My Database
NASA	National Aeronautics and Space Administration
NCSA	National Center for Supercomputing Applications
NSF	National Science Foundation
PB	PetaByte
PR	Pull Request
PSF	Point Spread Function
QA	Quality Assurance
Qserv	Query Service, Proprietary LSST Database system
RAC	Resource Allocation Committee
RDBMS	Relational Database Management System
SDSS	Sloan Digital Sky Survey
SLAC	SLAC National Accelerator Laboratory (formerly Stanford Linear Accelerator Center; SLAC is now no longer an acronym)
SQL	Structured Query Language
TB	TeraByte
US	United States



Acronyms III

Rubin
Observatory

WISE	Wide-field Survey Explorer
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)
http	HyperText Transfer Protocol





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>
- 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>
- 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
- [WOM-019]**, O'Mullane, W., 2007, *CU1 Progress Report #1*,
GAIA-C1-PR-ESAC-WOM-019, URL <http://www.rssd.esa.int/cs/livelink/open/2803205>
- 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
- [WOM-011]**, O'Mullane, W., Hoar, J., Levoir, T., et al., 2011, *Software Engineering Guidelines for DPAC*,
GAIA-C1-UG-ESAC-WOM-011, URL <http://www.rssd.esa.int/cs/livelink/open/2760364>
- 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'Hágain, 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