DM Leadership Team Virtual Face-to-Face Meeting, 2019-02-26 to -28

Logistics

Date

26 Feb 2019 – 28 Feb 2019

Location

- This meeting will be conducted on BlueJeans: nobody is expected to travel to participate in it.

Browser	Room System	Phone Dial-in
https://bluejeans.com/293724745/	1. Dial: 199.48.152.152 or bjn.vc 2. Enter Meeting ID: 293724745 -or- use the pairing code	Dial-in numbers: - +1 408 740 7256 - +1 888 240 2560 (US Toll Free) - +1 408 317 9253 (Alternate Number) Meeting ID: 293724745

Slides

- Please upload slides to Confluence in advance of your presentation (to avoid issues with screen-sharing, etc).

Attendees

- Wil O'Mullane
- Tim Jenness
- Frossie Economou
- Unknown User (gcomoretto)
- Simon Krughoff
- Jim Bosch
- Yusra AlSayyad
- Margaret Gelman
- Fritz Mueller
- Kian-Tat Lim
- Colin Slater
- Jeff Kantor
- Joel Plutchak
- Unknown User (mbutler)
- Eric Bellm
- Gregory Dubois-Felsmann
- Robert Gruendl
- Zeljko Ivezic

Apologies

- Leanne Guy (DESC meeting - unless the dates for this meeting are changed)

- [Robert Lupton](#) (DESC and an integration meeting)

Agenda

Day 1: 2019-02-26			
Time (Project)	**Topic**	**Chair**	**Notes**
09:00	[Welcome](#)	[Wil O'Mulla ne](#)	• Review meeting agenda and logistics. • There is ongoing discussion at the senior level on requirements on minimum values for L1PublicT. Agreed that there's no need to push this at the DM level until those discussions have converged. • Next JSR is tentatively scheduled for the last week of August (2 weeks post-JSR).

09:15	[Many-user LSP testing](#)	[Gregory Dubois-Felsmann](#)	• Discuss milestones and procedures for arranging testing of LSP deployments by substantial numbers of users (DM staff? Sci Collab members?)
			• To what extent is the existing (or some future) LSP a "production" service for the use of DM staff / other project members / outsiders, vs. a construction project?
			
			• Other datasets, beyond the list in [Gregory Dubois-Felsmann](#) slides, which might be considered for testing:
			   ○ Kepler (& K2?)
			   ○ Something for alert production?
			      ■ This could mean diffims / DIASources / DIAObjects, and/or it could mean realistic tools for working with the alert stream / alert DB.
			      ■ No consensus that the former needs anything beyond the diffim that will be carried out in DRP.
			      ■ We could explore things like UI to the alert filtering service using the ZTF alert stream.
			   ○ Solar system objects?
			      ■ Most LSP requirements on SSOs were deleted in the replan.
			   ○ Agreed:
			      ■ We will plan on using LSST-processed HSC data, DR1 for now and DR2 when available
			      ■ For externally-sourced processed data we will continue to use WISE/NEOWISE, and...
			      ■ We will add Gaia DR2, as this dataset is now essential both for internal use and as context for anyone trying to do scientific work with survey data.
			      ■ There is no requirement to continue to support the 2013 SDSS processing, but we'll keep it around until it becomes impractical. (This was not clearly brought out in the wrap-up, though.)
			      ■ Requesting any further datasets will be handled by RFC.
			• Note that ingesting these datasets still requires data model work; see discussion tomorrow.
			• Can we expose some of this data by TAP queries to other archives, rather than copying the data to our own systems?
			• Discussion around general "providing a sandbox" testing vs. focused tests.
			• "Victims of our own success" — everybody wants to use notebooks. How are we supporting that? What are the resource implications?
			   ○ Also impacts the priority of technical work.
			   ○ Current tutorials are not providing useful technical input; they are just outreach to the community. Agreement from [Wil O'Mullane](#) that not every science collaboration should have effort allocated to support tutorial sessions (but e.g. the PCW will get supported).
			   ○ Discussed support for "internal users" — Stack Club and/or DM developers. What are the expectations here? Suggestion that T/CAMs can intercede on the part of their developers, and hence there's less pressure (and less reputational risk) on LSP support. (Do we agree with this?)
			
			☑ [Wil O'Mullane to discuss plans for future LSP demos and Tutorials (which take more effort) with Leanne Guy 25 Mar 2019](#)

10:00	LSP review preparations	[Gregory Dubois-Felsmann](#)	• Gather whatever cross-DMLT input is needed for the upcoming LSP review. • Prepare for review prep "focus session". • [List of topics](#) written by [Leanne Guy](#) & [Gregory Dubois-Felsmann](#) ; [Frossie Economou](#) & [Gregory Dubois-Felsmann](#) now converting that into a full agenda and preparing document pack. • Focused document writing days and dry-runs coming up.

10:15	Renaming	Frossie Economou	- Frossie wants a single name that she can use in code for "this thing" (being the Notebook Aspect of the Science Platform). - Could be "Jellybean". - Note that Leanne Guy is already in the process of a grand rebranding. - Gregory Dubois-Felsmann Also note that we are not planning to expose either "SUIT" or "DAX" - the equivalents for the other two Aspects - in the UI or in the user manual. The names, or their software-prefix equivalents, e.g., `dax_`, will appear in implementation documentation and in code. - The immediate aim here is just to have a name we can use in code and reporting; ultimate branding of the product release to the public will follow from Leanne Guy 's process (or some equivalent). ☑ All DMLT: Send suggestions for renaming "Jellybean" to Simon Krughoff 28 Feb 2019 ☑ Simon Krughoff — create a poll for the new name for Jellybean on on #dm-camelot, with a closing date of Monday 04 Mar 2019 . 28 Feb 2019
10:30	Break		
11:00	QAWG report	John Swinbank	- Provide a brief overview of the QAWG report, DMTN-085. - Please review DMTN-085 in advance of the meeting.
11:30	Review of QAWG recommendations	John Swinbank	- There are some 40 recommendations from the QAWG. - It seems likely that some of those recommendations correspond to work which is already planned; some should be added to the plan; and some should be ignored. - We will review each QAWG recommendation in turn and decide: - Whether this is something that the DMLT wishes to schedule work to address as a matter of urgency; - If so, which T/CAM is responsible for managing the work; - The delivery timescale which is necessary to make this exercise useful. - Assuming 3 minutes per recommendation means we'll need 2 hours for this exercise. <hr> - See DMLT Discussion of QAWG recommendations. ☑ Gregory Dubois-Felsmann and John Swinbank 25 Mar 2019 Re: QAWG-REC-20 - we should revisit RFC-243 - Getting issue details... STATUS and capture what is now actually being done in that respect (e.g., via Hsin-Fang Chiang's regular runs), what parts can be achieved by improving the documentation of and access to the output data from those runs, and what parts might still require substantive work (e.g., regularly running the SDM-Standardization afterburner on those outputs). ☑ Jim Bosch and Simon Krughoff should talk about the development path for metrics re: QAWG-REC-34. I.e. usually a plot motivates a metric not the other way around. How are metrics captured in this workflow?
12:30	Break		
13:00	Review of QAWG recommendations (continued)	John Swinbank	Continuation of the above session.
14:00	EFD transport and transformation decision	Kian-Tat Lim	- Given SQR-029, can we commit to technologies for transport, transformation, long-term storage, and user query of EFD data? - Please review SQR-029 in advance of the meeting. <hr> - A bunch of technical/implementation questions and discussion, which are not captured here. - We note that changes to the technical baseline must ultimately be accepted by the DM-CCB. - The DMLT suggest that this work should continue. ☑ Frossie Economou — cost implementation of DM-EFD based on SQR-029. 21 May 2019 ☑ Wil O'Mullane , Kian-Tat Lim , Unknown User (mbutler), and Gregory Dubois-Felsmann — revisit design of EFD-ETL on the basis of the DM-EFD design. 12 Jul 2019 ☑ Frossie Economou Document results of DM-EFD discussion with Unknown User (mbutler) 03 Jul 2019
14:30	Close		

<div align="center">

Day 2: 2019-02-27

</div>

09:00	The future of middleware	Fritz Mueller	- Following an impressive demo in early February, what are the next steps for middleware development? - A rough timeline was presented for BG2 deprecation, which relies on effort being allocated by various T/CAMs. Have those resources been secured? Is this timeline now solid? - What longer-term development effort is needed? Where will long term maintenance responsibility lie? - And how does that represent to available developer effort / funding? <hr> - Are the S19 budgets that Fritz mentions still valid? - Yes, but note that this leaves Simon Krughoff fully budgeted for the next 3 months. - There will also be a request for resources from AP. - Commissioning team involvement? - Depends mostly on CPP and obs_lsst. - Need to capture the need to get Butler access to prompt data on distributors for AP. - This does not block conversion of AP pipeline tasks to PipelineTasks. - Aiming to have an early milestone for PipelineTask design review; after that, interfaces should be stable and porting can start in earnest. - General agreement that there should be a "portathon" at the PCW; likely policy is that beyond that there will be no *new* code developed in the Gen2 middleware (although old cold will likely be supported until ~November). - Futures: - General agreement that the tight loop between DRP and LDF is essential; not a strong appetite to have Architecture in the loop. - The DRP team suggests that the weight of development should move to NCSA; that could be supported by a new hire at NCSA, who wouldn't necessarily need to jump directly into the middleware "czar" role. ☑ Fritz Mueller — organize a meeting to discuss integration of G3 middleware with data distributors for AP. (Doodle poll sent to stakeholders: https://doodle.com/poll/75mzug8vd23be8bw3kymsinw/private)15 Apr 2019 ☑ Fritz Mueller , Unknown User (gcomoretto) — note in v17 release notes that the Gen2 Butler release will be removed following a Pipelines release in ~November. 28 Feb 2019 (https://pipelines.lsst.io/releases/notes.html#release-v17-0-deprecate-gen2) ☑ Wil O'Mullane , John Swinbank , Fritz Mueller , Unknown User (mbutler) — continue to iterate on hiring and staffing support for middleware development by convening for a chat at the JDR. 09 Aug 2019
10:00	Standardization of obs_ package design	Simon Krughoff	- We will use this session as a very focused discussion around timelines for how versioned calib repositories, yaml camera, and general Gen3 improvements will be integrated. - The result of this session will be a set of broad guidelines for how to order effort in this area. It's expected that Simon Krughoff and Tim Jenness will do the work. - In addition, we would like to identify a specific obs package to act as the exemplar. Notes taken by Simon Krughoff are here.<hr> - Identify an obs package to use as "the exemplar" to convert to new standards. - Note this is using the **current** implementation as the reference, but using this package as the basis of work to develop a future reference implementation. "Patient 0". - Simon & Tim will then handle that, then we can figure out how to convert other packages. - obs_subaru? obs_lsst? obs_decam? - Discussion of the "special" aspects of all the various packages — test stands, version cameras, etc; all of them have different concerns. - Also calibrations from external pipelines on DECam and CFHT. - Consensus seems to be that obs_decam is the right choice. - Some concern that it might be more efficient to simply fork and drop Gen2 support; Tim & Simon to play this by ear. - YAMLCamera is a fact of life, but it needs more testing before we rely on it, and there is ongoing development and its schema needs to be formalized. - Four tasks to convert and update obs_decam: - Conversion to YAMLcamera - Versioned cameras - Integration of user generated calibs - Deal with config overrides in obs_ packages. - The useful end product is a document describing how an obs_ package works. - This is not blocked on other work; Tim, Simon and their T/CAMs can schedule this based on their availability. ☑ Leanne Guy — work with Robert Lupton to extract a todo-list for YAMLCamera/obs_lsst as well as an ideas to whether he is planning to act on it. 25 Mar 2019
10:30			Break

11:00	OPS Rehearsal #1	Robert Gruendl	- Status for plans/pre-work for first OPS rehearsal (LDM-503-09). - LDM-643 (now has outline of OPS rehearsal #1) - Test Plan should also be ready before F2F meeting. - Please review LDM-643 in advance of the meeting. <hr> - Expectation that QA will be manual, although it may be possible to automate some of it. - No consensus that we need a new Jira project to support this activity. - Software releases (e.g. addressing problems discovered) will be part of a subsequent rehearsal; they are not covered here. - Agreed to run processCcd (including ISR, etc) *as well* as Calibration Products Production. - Because that enables us to generate KPMs. - We expect that this is broadly equivalent to what Hsin-Fang is already doing. - Running CPP is probably not interesting on the simulated data, but is likely to generate problem reports so it's agreed we should do it anyway. - Note that master calibs and a reference catalog are already available on lsst-dev. - Can we agree on date? - End-of-April is the preliminary target. - Depending on staff availability: identify key players, then choose date. - Avoid LSP review. - Need "sniff testing" of "precursors" prior to the rehearsal; that also drives dates. ☑ Wil O'Mullane , Robert Gruendl — set names to roles in ops rehearsal #1; converge on date. Report to DMLT telecon. 11 Mar 2019 ☑ Robert Gruendl , Simon Krughoff — agree on dataset for ops rehearsal #1; report to DMLT telecon. 11 Mar 2019 ☑ Wil O'Mullane, Robert Gruendl — review all dates for ops rehearsal #1 at (or in follow-on to) DMLT telecon; this will cover everything except pipelines, which will follow. 11 Mar 2019
11:30	Release maintenance	Wil O'Mullane on behalf of Leanne Guy	- Including back-porting of bug fixes to stable releases for science users. <hr> - Deprecation blocking on adding the appropriate package to the Conda environment; will not be in v17. - LDM-672 (in prep) is policy (see also LDM-294); DMTN-106 is process. - There will be an RFC on these documents following further discussion. - There's a request for requirements on the process, and then to enable each product provider to show how their product meets those requirements. - Concern that we should still aim for common tooling for manageability. - Concern over the cost of Conda support. - Discussion of the semantics of versions; request to use semantic versioning. ☑ Frossie Economou , Wil O'Mullane , Leanne Guy — ensure that LDM-672 captures a list of requirements on the release process. 25 Mar 2019
12:00	SUIT summary	Unknown User (xiuqin)	- Progress report on SUIT development activities - Development plan for next 2 months, dependencies, blockers (if any) - Possible plan for team ramping down <hr> - From next month, Unknown User (xiuqin) plans to attend LSP meetings and T/CAM meetings, but not DMLT. ☑ Frossie Economou , Gregory Dubois-Felsmann — discuss development of Portal help system and the larger full-LSP document delivery context 11 Mar 2019
12:30	Break		

13:00	Next steps in data model toolchain	Yusra AlSayyad and Colin Slater	"SDM standardization" is part of a long chain of transfer of data organization / schema metadata from the point of generation in AP/DRP code (and image ingest) through databases to external service via TAP/SIAv2 and use in the Portal. - A complete architecture for this is not yet RFCd. <hr> - Concrete milestone: ingest HSC data. - Note that Object tables resulting from the current code are already ingestible (modulo concerns to come later). - Concerns are that some effort is needed to reconcile the DPDD and what the pipelines are actually producing. - The same basic model is meant to be used for image metadata (exposure, visit, coadd patch); this needs to be harmonized with Butler Gen3 development. The Gen3 database schema should be migrated to be derived from a Felis representation. - The DPDD does not currently specify the content of such tables (it just says they should exist). Something equivalent to such a specification is probably needed in order to define some effective requirements on this metadata (in part this is necessary in order to guarantee that our metadata can support the generation of ObsCore and CAOM2 representations). - The mid-2019 deadlines of the Gen3 project for convergence on its DB schema must be kept in mind. - Aiming to get an end-to-end processing through HSC data to TAP_SCHEMA into Portal/Topcat is a high priority; this should be associated with a level 3 milestone (or milestones; not necessarily a new one). - Qserv ingest improvements to support this are expected by the end of the cycle. ☑ Fritz Mueller with Gregory Dubois-Felsmann and Colin Slater 11 Mar 2019 — ensure that linkages between SDM support for image metadata tables and the Butler Gen3 development timeline are understood / ticketed. (See Image and Visit metadata - linking the Gen3 Registry and the Science Data Model for developing thoughts here. Meeting to be held with stakeholders post LSP-review.)
13:30	Butlers at the Summit and control+processing scripts	Kian-Tat Lim	CSCs at the Summit will be using DM code and are likely to expect to use Data Butler interfaces. Are we ready to support this? Scripts for the Script Queue are going to combine commands to CSCs with data processing. How and where should that processing occur? Where should the data being processed live?<hr> - Desirement has been articulated to execute DM code "directly in the script queue". - Assertion is that e.g. CBP scripts could be run in OCS Controlled Batch to meet these goals. - We "hope" that this is not much more overhead. - But this will be required in ~July this year, before the OCS Controlled Batch service exists. This is "a worry"; the script queue machine will need access to a Data Butler. - Note that no database services are expected on the summit; Butler G3 repositories will need SQLite (except the DBB, which will use the Consolidated DB). - How to handle script execution for AuxTel (before OCS Batch)? - Can the AuxTel have an OODS Butler? - If you're willing to go the Base to get the data. - Seems unlikely that networks are a limiting factor. - K-T is reluctant to allow direct access to the script queue. - Seems convenient to have the script queue access the Camera Diagnostic Cluster Butler. - Commissioning Cluster is not due until 2020 at earliest. ☑ Kian-Tat Lim — produce a DMTN describing DM summit services. 25 Mar 2019 ☑ Robert Lupton — confirm agreement with DMTN-111. 25 Mar 2019 ☑ Kian-Tat Lim with Frossie Economou Unknown User (mbutler) and Gregory Dubois-Felsmann 11 Mar 2019 — define LSP service level to be provided in the Commissioning Cluster circa 2020 — added some text to LDM-148; submitted as **RFC-581** - Getting issue details... STATUS
14:00	Review action items & plans for next meeting	Wil O'Mullane	- Confirm dates of next meetings. Provisionally: - May 21–23 at NCSA. - Conflicts with LSST@Asia. - Aiming for 4—6 June instead. - Will be in person. - November 5–7 at SLAC. - Tentatively agreed this is F2F; could still change and make it virtual. - Next DMLT telecon will be 11 Mar 2019 . <hr> - Agreed that vF2F worked reasonably well. - Some concern about the technology; could try Zoom or WebEx or something. - Consensus that we wouldn't want to do a virtual meeting every time, but once or twice a year seems appropriate. ☑ All DMLT — Confirm your availability to Wil O'Mullane for 4—6 June DMLT meeting at NCSA. 28 Feb 2019 ☑ Fritz Mueller — check availability of SLAC guest house for 2019-11-5/7 DMLT F2F.
14:30			Close

Day 3: 2019-02-28

09:00	Networks, Summit and Base Data Center Status and Planning	Jeff Kantor	• Brief overview of Summit, Summit - Base, and Base - LDF Network status and schedule
			• Brief overview of Summit and BDC construction status and move-in schedule
			• Update on planned deployments of DM equipment to BDC, visits, tests/rehearsals in FY19 (and IT support required)
			
			• There is already wifi at the base & summit for AURA credentials (an LSST sign-on, **not** and AURA ID); others have to sign up in advance.
			   ◦ "All of the connectivity you would need by June" (of this year).
			   ◦ None of this connects to the controlled networks carrying EFD traffic; can access them through a bastion.
			• There is a visitor network at La Serena; have to register a MAC address in advance; not planning to provide a visitor network at the summit.

10:00	Open-ended discussion on deliveries and visits to Chile	Jeff Kantor	Summit - Base IT-related Deliveries and Visits
11:00	Close		

Attached Documents

File	Modified
PDF File 2019-02-26 — DMLT — QAWG.pdf QAWG summary & recommendations	Feb 25, 2019 by John Swinbank
PDF File EFD Decision.pdf	Feb 26, 2019 by Kian-Tat Lim
PDF File DMLT-F2F-20190227-OPSRehearsals.pdf	Feb 27, 2019 by Robert Gruendl
PDF File DM and Summit.pdf Added explicit proposal	Feb 27, 2019 by Kian-Tat Lim
Microsoft Powerpoint Presentation DMLT-LSP-testing-20190226.pptx Presentation on LSP science-driven testing and demos	Feb 27, 2019 by Gregory Dubois-Felsmann
PDF File Gen3 Middleware Feb 2019.pdf	Feb 27, 2019 by Fritz Mueller
Microsoft Powerpoint Presentation SUIT_status20190227-2.pptx	Feb 27, 2019 by xiuqin
Microsoft Powerpoint Presentation Kantor Networks and Base.pptx	Mar 12, 2019 by Jeff Kantor

Download All

Action Item Summary

Task report

Looking good, no incomplete tasks.

Pre-Meeting Planning

Suggested topics for discussion

Topic	Requested by	Time required (estimate)	Notes
Summit and Base Data Center Status and Planning	Jeff Kantor	1 hour	• Brief overview of Summit and BDC construction status and move-in schedule • Update on planned deployments of DM equipment to BDC, visits, tests/rehearsals in FY19 (and IT support required)

Networks Status and Planning	Jeff Kantor	1 hour	• Brief overview of Summit, Summit - Base, and Base - LDF Network status and schedule
Many-user testing of the LSP	Gregory Dubois-Felsmann	45 min	• Discuss milestones and procedures for arranging testing of LSP deployments by substantial numbers of users (DM staff? Sci Collab members?)
LSP review preparations	Gregory Dubois-Felsmann, Leanne Guy	2-8 hours	(NB I — John Swinbank — don't think it's useful for the whole DMLT to spend 8 hours on this! A smaller splinter session may be more effective.) Leanne Guy- I agree
QAWG report	John Swinbank	45 min	• DMTN-085 • NB Per DMLT discussion of 2019-01-28, we should expect this not just to be a 45 minute presentation of the report, but to grow into a larger & longer discussion of what we're going to do about it. Details TBD.
OPS Rehearsal #1	Robert Gruendl Wil O'Mullane	30 min	• Status for plans/pre-work for first OPS rehearsal (LDM-503-09). • LDM-643 (now has outline of OPS rehearsal #1) • Test Plan should also be ready before F2F meeting.
SUIT work summary	Unknown User (xiuqin)	15 min	• Progress report on SUIT development • Development plan for next 2 months, dependencies, blocker (if any) • Possible plan for team ramping down
Update on planning tooling to swap in interpolated pixel values for inspection.	Simon Krughoff	15 min	• What is the schedule for producing the said tooling.
The future of middleware	John Swinbank (but volunteering Fritz Mueller to chair... sorry Fritz!)	60 min (? Fritz to confirm!)	• Following an impressive demo in early February, what are the next steps for middleware development? • A rough timeline was presented for BG2 deprecation, which relies on effort being allocated by various T/CAMs. Have those resources been secured? Is this timeline now solid? • What longer-term development effort is needed? Where will long term maintenance responsibility lie?
Standardization of obs_ package design	Eric Bellm	20 minutes	• obs_lsst provides an updated vision of obs_ packages for the BG3 era. Can the relevant decisions be captured in a design document to guide potential updates to other obs_packages? • Gen3 middleware will also change how obs_ packages work; for that obs_subaru provides more of a prototype than obs_lsst. We need need to integrate these mostly-orthogonal changes.
Next steps in data model toolchain	Gregory Dubois-Felsmann and others (we could invite a non-DMLTer to report?)	30 min	• "SDM standardization" is part of a long chain of transfer of data organization / schema metadata from the point of generation in AP/DRP code (and image ingest) through databases to external service via TAP/SIAv2 and use in the Portal. • A complete architecture for this is not yet RFCd.
Release maintenance, including back-porting of bug fixes to stable releases for science users	Leanne Guy	20 mins	
EFD transport and transformation decision	Kian-Tat Lim	30 minutes?	Given SQR-029, can we commit to technologies for transport, transformation, long-term storage, and user query of EFD data?
Butlers at the Summit and control+processing scripts	Kian-Tat Lim	30 minutes?	• CSCs at the Summit will be using DM code and are likely to expect to use Data Butler interfaces. Are we ready to support this? • Scripts for the Script Queue are going to combine commands to CSCs with data processing. How and where should that processing occur? Where should the data being processed live?