

DM Leadership Team Virtual Face-to-Face Meeting, 2020-05-12/14

- [Logistics](#)
 - [Date](#)
 - [Location](#)
 - [Attendees](#)
- [Agenda](#)
 - [Day 1: 2020-05-12](#)
 - [Day 2: 2020-05-13](#)
 - [Day 3: 2020-05-14](#)
 - [Attached Documents](#)
- [Action Item Summary](#)
- [Pre-Meeting Planning](#)

Logistics



COVID-19

Due to the ongoing COVID-19 situation, this meeting will be **virtual**. Please do not attempt to travel to Seattle — or to anywhere else — to participate.

Date

12 May 2020 – 14 May 2020

Location

Browser	Phone Dial-in
https://washington.zoom.us/j/95408203481	Meeting ID: 954 0820 3481
Please see #dm-camelot and/or e-mail for the password.	Dial by your location: <ul style="list-style-type: none">• +1 253 215 8782 US (Tacoma)• +1 669 900 6833 US (San Jose)• +1 720 928 9299 US (Denver)• +1 971 247 1195 US (Portland)• +1 213 338 8477 US (Los Angeles)• +1 346 248 7799 US (Houston)• +1 602 753 0140 US (Phoenix)• +1 669 219 2599 US (San Jose)• +1 301 715 8592 US (Germantown)• +1 312 626 6799 US (Chicago)• +1 470 250 9358 US (Atlanta)• +1 470 381 2552 US (Atlanta)• +1 646 518 9805 US (New York)• +1 646 876 9923 US (New York)• +1 651 372 8299 US (St. Paul)• +1 786 635 1003 US (Miami)• +1 267 831 0333 US (Philadelphia)

Attendees

- [Wil O'Mullane](#)
- [Leanne Guy](#)
- [Unknown User \(mbutler\)](#)
- [Colin Slater](#)
- [Eric Bellm](#)
- [Fritz Mueller](#)
- [Frossie Economou](#)
- [Unknown User \(gcomoretto\)](#)
- [Gregory Dubois-Felsmann](#)

- Jeff Kantor
- Jim Bosch
- Kian-Tat Lim
- Unknown User (npease)
- Robert Gruendl
- Robert Lupton
- Simon Krughoff
- Tim Jenness
- Yusra AlSayyad
- Zeljko Ivezić
- John Swinbank

Agenda

Day 1: 2020-05-12				
Time (Project)	Topic	Coordinator	Pre-meeting notes	Running notes
Moderator: Leanne Guy				
09:00	Welcome	Wil O'Mullane	<ul style="list-style-type: none">• Introductory remarks• Review agenda and code of conduct	
09:10	Project news	Wil O'Mullane		<ul style="list-style-type: none">• All ME20-03 variance narratives are now complete.
09:30	Middleware status update	Tim Jenness / Robert Gruendl	<ul style="list-style-type: none">• Timeline for deprecating Gen 2.• Current development activities.• Staffing plans moving forward.	<ul style="list-style-type: none">• Command line utility:<ul style="list-style-type: none">◦ In the short term, will provide high-level repository management.◦ Ultimately, will provide limited query capability on the repository.◦ There is a document describing this.• July date for registry stability is of general interest.• Expect things to get busier in ~September, as Gen3 moves into general use.• Key transition point for developers is reprocessed HSC RC2 data being available in Gen3.• Quantum graph generation is slow, but reprocessing one tract or patch is fast enough.• Jim believes that scaling quantum graph generation to PDR2 in ~6 months is plausible.• What is the next tall pole after quantum graph generation?<ul style="list-style-type: none">◦ There's nothing obvious.◦ Ingest may be a bit slower than Gen2.• Shared users are no longer an unsolved problem, but will be handled by naming conventions rather than technical mechanisms.• There was much discussion of the definition of the visit; it's not clear that this is really conclusive.• Not clear that adding new resources will really help with deadlines over the next couple of months, because the ramp up time is too long.• However, they would be useful for later in the year.
10:30	Break			
Moderator: Wil O'Mullane				

11:00	Calibration products	John Swinbank	<ul style="list-style-type: none"> Brief review of the DMTN-148 proposals. Is this document acceptable to the DMLT? What are the remaining open questions? How will we resolve them? 	<ul style="list-style-type: none"> ✓ DMTN-148 is almost there suggest 2 weeks review by DMLT. 15 Jun 2020 ✓ John Swinbank to setup feedback system with Chris Waters on calibration note (DMTN-148).21 May 2020 <p>This should be baselined (change controlled)</p> <p>Robert asks when we will start "acting on this" - e.g when could it be used for LATISS on the mountain. On going work from Andres and Merlin - where is the ingest and validate..</p> <p>KT last stage getting from production system via OODS to summit to be used for ISR on summit. Certified and transferred to where its needed.</p> <p>Jim - Good to separate operations concerns (how its used on the mountain) from about the code and how we implement. DMTN-111 could have the summit details. Tim - no agreement on every curated calibration had class somewhere, one end - other is the certification</p>
11:15	Plans for IVOA and Python interfaces to time series data. (archived slides)	Gregory Dubois-Felsmann / Eric Bellm	<ul style="list-style-type: none"> Follow-up on discussion within the DM-SST. 	<ul style="list-style-type: none"> Headline: The DMLT agrees that the story we tell the community is that our data model is effectively two tables, and users will need to join them themselves. General agreement about using PyVO and Pandas. Are DIAForcedSources included? <ul style="list-style-type: none"> The same considerations apply mapping DIAObject to DIAForcedSource. Our feature computation may be based on DIASources or DIAForcedSources; a recommendation from Eric will be forthcoming. Adding support for e.g. non-detection upper limits in feature computation is a possible, and may make the inputs to feature computation more complex. However, this should not be unmanageable. How tightly coupled is the AP pipeline with the database? Is this a technical risk? <ul style="list-style-type: none"> Reconstructing data structures from the AP pipelines based on VO interfaces would be challenging. The details of feature computation are well abstracted and testable; they are not tightly coupled. Plugins are implemented for feature computation below the task level; the master task takes a Pandas data frame as input. None of these proposals are changes to previous promises made to the community. In terms of announcements to the community, we suggest that this should be rolled into discussion of capabilities available for DP0. <ul style="list-style-type: none"> Some discussion of a PST-SciCollab talk if necessary. <ul style="list-style-type: none"> ✓ Eric Bellm — update time-series technote to contain a discussion of the way in which data will be presented to users. 01 Jul 2020 <p>On ticket branch at https://dmtn-118.lsst.io/v/DM-19593/index.html</p> <ul style="list-style-type: none"> ✓ Gregory Dubois-Felsmann — update the Science Platform design documentation to reflect that data access services should be tested with PyVO. 01 Jul 2020
12:30	Break			
Moderator: Wil O'Mullane				

13:00	Rebaselining & project schedule	Wil O'Mullane	<ul style="list-style-type: none">What's our current understanding of rebaselining?How will DM respond to slips in the overall project schedule?	<ul style="list-style-type: none">Calabrese coordinating mail pickup in Tucson.Services which were used in commissioning /integration are easy to define as "done".Would be good to get a statement of thanks from project leadership to DM staff.Aim to make "blurring" between construction and commissioning a positive opportunity.Also look for opportunities in the deliveries to ops (but be careful that this is not blurring).The details of financing, ramps, etc through FY22 & FY23 will have to be addressed on a case-by-case basis, depending on guidance from construction project management and the agencies.Comments from Victor:<ul style="list-style-type: none">Covid 19 costs are not an appropriate use of current funding (baseline or contingency).We will therefore adopt a new baseline; the so-called "over target baseline".The earlier we do this, the riskier it will be and the less accuracy it will have.Currently seems like NSF will accept late replanning (October/November).This is not an opportunity for us to reinstate previously-accepted descopes.This information can be shared with the rest of the project.Some concern expressed that operational priorities are different from construction priorities; we should be clear that staff transitioning do so in the project's interest, rather than just because it is financially expedient.Victor is petitioning the agencies for a minimal status review this year.The drawback of waiting longer for a rebaselining is that we have to live with bad metrics until it kicks in; might be an issue for reviews.Not clear what the rebaselining process will be: could imagine an FDR-like process, but it's not clear that will be practical.Concern raised that DM may be able to reach completion on close to original timescale.Expectation is a 12 month delay with a cost of \$3.5M per month. Do not believe there is a serious risk of this not being approved at the moment. Also do not believe there is a serious risk of being forced to accept technical compromises.
14:30 (at latest)	Close			
Day 2: 2020-05-13				
Moderator: Gregory Dubois-Felsmann				

09:00	<p>Plans for an interim data facility</p> <ul style="list-style-type: none"> 09:00-09:30: technical discussion: Google PoC slides 09:30-10:00: programmatic discussion (with Bob Blum) 	<p>Wil O'Mullane</p>	<ul style="list-style-type: none"> What's the current status of the Google POC? What's the timeline for an IDF decision (if one hasn't been made by the time of this meeting)? What do we need to do to prepare for the IDF? How does it impact the other tasks that we are working on? 	<ul style="list-style-type: none"> USDF FOA: <ul style="list-style-type: none"> Early FOA discussions seemed to discourage commercial entities from being involved; this wording has been softened/removed in later versions. Wil hopes to structure the FOA as infrastructure/middleware/execution; expect that commercial cloud vendors might be interested in the infrastructure part, but not the others. The decision making process is still TBD: expect that a DOE review committee will evaluate responses to the FOA and make a recommendation to the high echelons of the agency. We should not expect to have a resolution by the end of this calendar year. Relationship of the IDF to construction and commissioning: <ul style="list-style-type: none"> The IDF is intended <i>primarily</i> as a means for the pre-ops project to meet its milestones and prepare for operations, rather than as a service to the construction project. We expect that, at least until a new USDF site is chosen, construction and commissioning activities will continue at NCSA. Wil O'Mullane has budget which can be used to procure more resources in support of that if necessary. It may nevertheless be possible to use IDF resources for scale testing of DM services at a level beyond that which can be undertaken at NCSA. Concerns were expressed that the IDF as envisaged does not provide a clear transition route from the current NCSA infrastructure to future USDF infrastructure. While these are valid, we note that the transition to future USDF will include a transition from NCSA, not just from IDF. The IDF is sized to cover the activities described in DMTN-135. Proof-of-concept: <ul style="list-style-type: none"> POC activities go beyond those simply required to demonstrate the viability of the IDF (e.g. they include alert processing). However, these provide essential inputs for future rounds of decision making. However, the basic goal for the POC is to replicate regular processing which is currently being carried out at NCSA, but using Gen3 middleware. The POC is expected to produce meaningful results in early July; decision making on the IDF is expected by July. IDF & POC implementation: <ul style="list-style-type: none"> The platform for batch processing will be Condor on GCE. Alternative solutions (e.g. Airfoil) may be examined, but this is still at an early stage. Solutions which would irrevocably tie us to a particular cloud implementation are obviously unacceptable.
-------	--	--------------------------------------	--	--

10:00	APDB Update	Fritz Mueller	<ul style="list-style-type: none"> Action item from the previous DMLT vF2F to report on the status of the APDB. The DMLT acknowledges that this work may have been delayed by the focus on middleware development. 	<ul style="list-style-type: none"> Feedback from NCSA/Michelle is that a deep understanding of the structure of the data is essential, regardless of the database implementation chosen. We note that Andy Salnikov has already undertaken this analysis, but the DAX, AP and LDF teams are ready to collaborate further as needed. More test nodes are needed to fully understand scaling of the current system. As yet, there is no story about the (user-facing) PPDB. We encourage the DAX team to convert the information on DM-23881 to a technote at their convenience; the timing on this should be up to Fritz Mueller. <p>✓ Fritz Mueller — engage with the LDF and, as necessary, AP team to best understand the data structures required for the APDB. 06 Jul 2020</p>
10:30	Break			
Moderator: Simon Krughoff				

11:00	Status of Ingest	Fritz Mueller	<ul style="list-style-type: none"> • See Fritz Mueller's Slack message. • Should include: <ul style="list-style-type: none"> ◦ Status update on HSC RC2 ingest to Qserv. ◦ Discussion of longer term planning for ingest. What are the key questions? How do we get them answered? 	<ul style="list-style-type: none"> • Ingest: <ul style="list-style-type: none"> ◦ Expect "authorized ingesters" to be able to self-serve, but they will have to follow a (TBD) process. ◦ In principle, "bad" values in the data being provided to ingest should be fed back to the Pipelines developers as bugs. In practice, the processing systems are sufficiently in flux, and many of these are artefacts introduced by the ingest process, so this is not yet regularly happening. ◦ We do not have a written specification for the semantics of database contents (e.g. use of IEEE inf, NULL vs NaN, etc), despite some memories of previous (undocumented) agreements. • SDM: <ul style="list-style-type: none"> ◦ Extensive discussion, but much of it seemed to retreat ground that we have visited before. ◦ We discussed the right level of detail for the DPDD, and whether it needs to be radically (or slightly) redrafted. There was no conclusion to this. ◦ We agreed to prioritise the production of a DMTN describing the overall architecture being developed here. This is effectively refreshing the action item on Wil O'Mullane from our previous vF2F and now codified as DM-23658. We agreed that Wil and Colin Slater should be charged with making this happen. The aim here is to propose as comprehensive and concrete a system as possible for future DMLT discussion. ◦ We note that some relevant text exists at The Science Data Model and its Standardization(OBSOLETE). • Outstanding questions: <ul style="list-style-type: none"> ◦ We agreed that it is impossible for the DMLT to converge on answers to Fritz Mueller's questions as a group. ◦ Fritz Mueller should write up a proposed operational procedure for ingest to form the basis of future discussions. ◦ He should feel free to draw on expertise from across the project. <div> <input checked="" type="checkbox"/> Colin Slater — Augment LDM-153 to provide a description of the semantics of NULL, NaN, inf, and other database vocabulary DM-25926 - Getting issue details... <div>STATUS</div> </div> <p>06 Jul 2020</p> <div> <input checked="" type="checkbox"/> Fritz Mueller — Draft an technote describing operational procedures for database ingest. (Ticketed: DM-26341) 03 Aug 2020 </div> <div> <input checked="" type="checkbox"/> Wil O'Mullane & Colin Slater — Complete DM-23658. 06 Jul 2020 </div>
12:30	Break			
Moderator: Fro ssie Economou				

13:00	Prompt processing	Kian-Tat Lim /Robert Gruendl	<ul style="list-style-type: none">What exactly are “prompt services” (in terms of the product tree, system architecture, etc)?What are the desires and use cases that have been advanced for an expanded scope “prompt processing” system?How practical is it for the DM construction team to meet those desires?If it is practical, what is the timeline and plan for doing so?	<ul style="list-style-type: none">This rather wide-ranging discussion provided more background material for further thinking than concrete decisions which can usefully be minuted.We discussed whether the “commissioning” use cases championed by Robert Lupton can be unified with the alert production use cases. There was no really concrete decision here.We note the requirement expressed by Robert Lupton for flexibility, and acknowledge that this is often more important than extremely high reliability in a commissioning situation.We further note the desire to provide uniform interfaces at all our various processing sites as far as is possible.We agreed that the best way to proceed is for Robert Gruendl to develop a prototype OCPS capable of executing pipelines based around the NCSA test stand. Unknown User (mbutler) agreed to provide staffing to make this possible. <div><input checked="" type="checkbox"/> Robert Gruendl — report on OCPS status to the DMLT. 06 Jul 2020</div>
Time permitting	OR2	Robert Gruendl	Operations Rehearsal #2 (preparations)	<ul style="list-style-type: none">Jeff Kantor is preparing for data transfers from ComCam on the basis of:<ul style="list-style-type: none">72 MB per image.30 images per minute for periods of up to ten minutes at peak rate.10 images per minute for periods of up to 2 hours on average.
14:30 (at latest)	Close			
Day 3: 2020-05-14				
Moderator Wil O'Mullane				
09:00	Team status	John Swinbank	<ul style="list-style-type: none">Each group please provide (~10 minutes total):<ul style="list-style-type: none">A brief retrospective on what's happened since our last meeting.Plans for the next few months.Let's go in reverse-WBS order for a change:<ul style="list-style-type: none">SQuaRE (Frossie Economou)Data Facility (Unknown User (mbutler))DAX (Fritz Mueller)Data Release Production (Yusra AlSayyad)Alert Production (John Swinbank)DM Science (Leanne Guy)Architecture (Kian-Tat Lim)	
10:30	Wrap-up	Wil O'Mullane	<ul style="list-style-type: none">Actions and next meetings.Virtual, 2020-11-16/19<ul style="list-style-type: none">This meeting will be virtual.Tucson, 2021-02-22/25.<ul style="list-style-type: none">MCR booked - does not seem to clash with anything	
11:00 (at latest)	Close			

Attached Documents

File	Modified
PDF File 2020-05-12 — DMLT — Calibration Products.pdf	May 09, 2020 by John Swinbank

PDF File DMLT Google PoC 2020 Status.pdf	May 13, 2020 by Kian-Tat Lim
PDF File DMLT-F2F-20200513_prompt.pdf	May 13, 2020 by Robert Gruendl
PDF File DMLT-vF2F-13052020-Ingest.pdf	May 13, 2020 by Leanne Guy
PDF File 200512_Timeseries_Interfaces.pdf	May 13, 2020 by Eric Bellm
PDF File 2020-05 APDB Update.pdf	May 14, 2020 by Fritz Mueller
PDF File Arch F20A Plans.pdf	May 14, 2020 by Kian-Tat Lim
PDF File DRPActivities202005.pdf	May 14, 2020 by Yusra AlSayyad
PDF File DM Science Plans F20A.pdf	May 14, 2020 by Leanne Guy
PDF File DAX End of S20 Status_Plans.pdf	May 14, 2020 by Fritz Mueller
PDF File dmlt_may_2020.pdf	May 14, 2020 by Frossie Economou
PDF File 2020-05-14 — AP F20A.pdf	May 15, 2020 by John Swinbank

[Download All](#)

Action Item Summary

Description	Due date	Assignee	Task appears on
<input type="checkbox"/> Frossie Economou Will recommend additional Level 3 milestones for implementation beyond just the DAX-9 Butler provenance milestone. 15 Mar 2022	15 Mar 2022	Frossie Economou	DM Leadership Team Virtual Face-to-Face Meeting, 2022-02-15 to 17
<input type="checkbox"/> Kian-Tat Lim Convene a meeting with Colin, Tim, Robert, Yusra to resolve graph generation with per-dataset quantities (likely based on Consolidated DB work). 18 Mar 2022	18 Mar 2022	Kian-Tat Lim	DM Leadership Team Virtual Face-to-Face Meeting, 2022-02-15 to 17
<input type="checkbox"/> Frossie Economou Write an initial draft in the Dev Guide for what "best effort" support means 17 Nov 2023	17 Nov 2023	Frossie Economou	DM Leadership Team Virtual Face-to-Face Meeting - 2023-Oct-24
<input type="checkbox"/> Convene a group to redo the T-12 month DRP diagram and define scope expectations Yusra AlSayyad 30 Nov 2023	30 Nov 2023	Yusra AlSayyad	DM Leadership Team Virtual Face-to-Face Meeting - 2023-Oct-24
<input type="checkbox"/> Gregory Dubois-Felsmann Complete DMTN-105 defining the goal for "Prompt Products Release Ops" 11 Dec 2023	11 Dec 2023	Gregory Dubois-Felsmann	DM Leadership Team Virtual Face-to-Face Meeting - 2023-Oct-24
<input type="checkbox"/> Frossie Economou - you have some text for PSTN-017 02 May 2024	02 May 2024	Frossie Economou	DMLT Meeting - 2024-04-22
<input type="checkbox"/> DMLT Meeting - 2024-04-22 UKDF part in data facilities for PSTN-017 22 May 2024	22 May 2024		DMLT Meeting - 2024-04-22

<input type="checkbox"/> Richard Dubois USDF part in data facilities for PSTN-017 and distrib processing ? 22 May 2024	22 May 2024	Richard Dubois	DMLT Meeting - 2024-04-22
<input type="checkbox"/> Fabio Hernandez FrDF part in data facilities for PSTN-017 22 May 2024	22 May 2024	Fabio Hernandez	DMLT Meeting - 2024-04-22
<input type="checkbox"/> Tim Jenness - section on middleware for PSTN-017 22 May 2024	22 May 2024	Tim Jenness	DMLT Meeting - 2024-04-22
<input type="checkbox"/> Cristián Silva - section on summit/data acquisition for PSTN-017 22 May 2024	22 May 2024	Cristián Silva	DMLT Meeting - 2024-04-22

Pre-Meeting Planning

Topic	Requested by	Time required (estimate)	Notes
Status of the APDB	John Swinbank	30 mins	<ul style="list-style-type: none"> Action from previous DMLT F2F, delayed due to DAX focus on middleware.
Plans for IVOA and Python interfaces to query time series data in Prompt and DR data products.	Leanne Guy	30 mins (maybe 1hr, probably not)	<ul style="list-style-type: none"> Gregory Dubois-Felsmann and Eric Bellm .
Status on RC2 ingest to Qserv	Leanne Guy	30 mins	<ul style="list-style-type: none"> As requested at February DMLT
Management of calibration products	John Swinbank	30 mins	<ul style="list-style-type: none"> Following the February DMLT meeting, Christopher Waters has drafted DM-TN-148. Is the DMLT ready to sign off on that as our plan moving forwards? Leanne Guy requests that we also talk about product ownership.
Plans for an interim Data Facility	John Swinbank	1 hour	<ul style="list-style-type: none"> What's happening with the Google POC? What do we need to do to prepare for an iDF? <ul style="list-style-type: none"> When might it happen? Who needs to be involved?
Rebaselining	John Swinbank	1 hour	<ul style="list-style-type: none"> Everybody's talking about it, but what does it mean? Who will have to do what when? Can we use this opportunity to get ahead of whatever Victor/Kevin/etc will ask for, and make sure DM comes out of the rebaselining process in good shape?
Prompt processing	Unknown User (mbutler) , Robert Lupton , etc.	2 hours	<ul style="list-style-type: none"> What are the desires, use cases, requirements, plans, schedule, for an expanded scope "prompt processing" system, as requested by Robert Lupton? What are "prompt services", and what is the status of their product ownership?