

Science Platform and PDAC development and deployment in 2017

DRAFT

Reference material

- Science Platform presentation from DM Science session at March 2017 JTM
- SUIT (but also Science Platform) Vision document draft

Planning kickoff

At the March 2017 JTM we had two sessions with substantial Science Platform content:

- The DM Science session on the morning of Tuesday 07 Mar 2017
- A "decision-making" session on Tuesday afternoon, primarily focused on medium-term planning

Based on the discussion at the decision-making session, [Gregory Dubois-Felsmann](#) circulated a note to the DMLT on 20 Mar 2017 with a list of 2017-era development actions related to the Prototype Data Access Center (PDAC) and the Science Platform. Wil has now encouraged us to proceed with planning based on that list, with some prioritizations and amendments.

We need to complete this planning with some urgency for two closely-related reasons: to prepare for the [DM review in July](#), and to actually guide F17 development priorities. In some cases the JTM-derived plan skeleton does not exactly match previously established elements of the DM replan, so some fine-tuning will be needed, as well as sign-off from T/CAMs and their groups that the scope of work is realistic for their 2017 plans.

The list from Gregory's email, as revised based on Wil's feedback, is below, in a tabular form which we should fill in with some additional detail: rough schedules, coarse-grained assignments to teams, and JIRA epics for the work, to the extent that they have already been created, and a coarse prioritization code (1 - Critical, 2 - Normal, 3 - Stretch) where Wil's feedback or other considerations have suggested one.

Some of the elements in the original list break down into several units of work, in some cases by different teams. These will at least be shown as different epics, and we may divide the table rows as well to clarify the work packages.



(As we reach completion of the mapping of work to JIRA epics, we should flip this whole process around and build a table directly out of the epics by performing a JIRA query based on a distinctive component or label.)



Science Verification and Validation (a/k/a "QA") issues



We still also have to resolve what will be done for what we used to call "QA" this year. It is not entirely clear who has the ball at the moment to define this. Whatever is decided will need to be integrated with the Science Platform development plans for 2017.



Major work areas

General Area	Item	Comments	Priority	Team(s)	Estimated/desired initial delivery	Epic(s)
Portal	Deploy a light-curve viewing and analysis application in the Portal aspect of PDAC	An initial version of this was done for the JTM, but updates will be released in April/May based on user testing at IPAC and in PDAC.	NORMAL	SUIT/IPAC	2017-03, major update by 2017-05	



Portal	Begin to deploy LSST-specific online documentation through the Portal aspect	Framework for contextual documentation in place, awaits writing of LSST-focused content. IRSA-derived content available for many screens.	NORMAL	SUIT/IPAC		<div> DM-9368 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div> <div>(in broader epic</div> <div> DM-8727 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div> <div>)</div>
--------	--	---	--------	-----------	--	--

PDAC Data, Portal	Deploy the WISE primary mission Object-like and ForcedSource-like catalogs in PDAC, along with the associated image metadata tables, integrated with the existing Stripe 82 data to permit spatial joins in Qserv.	WISE primary mission (and AllWISE) catalog data have been loaded into Qserv, but are not yet supported in the Portal. Work remains to be done on loading the image metadata, supporting spatial joins between WISE and SDSS/S82, and a number of other details. The necessary storage space is available already in the PDAC ("integration cluster") hardware at NCSA.	CRITICAL	DB-DAX /SLAC, SUIT /IPAC		<div>DAX:<div><div></div><div>DM-8473</div><div>- Jira project does not exist or you don't have permission to view it.</div></div></div> <div>SUIT:<div><div></div><div>DM-8493</div><div>- Jira project does not exist or you don't have permission to view it.</div></div></div>
-------------------	--	--	----------	--------------------------	--	--

Portal	Improve Portal workflows	Covers improvements to both the functional workflows (e.g., allow direct query of ForcedSource data from a selected Object) and the general UX.	NORMAL			<div> DM- 10463 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div> <div>General epic:</div> <div> DM- 10432 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div>
PDAC Data	Attempt to maintain backward compatibility with the 2013 Stripe 82 processed image data for as long as this does not produce significant costs for Science Pipelines.	At the moment there are no specific functional issues in PDAC arising from backward-compatibility breakage, but not all Science Pipelines functionality is available now when reading in calibrated images from the 2013-era stack.	STRETCH	Science Pipelines (if needed), DB-DAX/SLAC (as it affects components like the cutout service)		

PDAC Data, Portal	Deploy the WISE and NEOWISE Source-like catalogs in PDAC, when they become available for download from IPAC-IRSA.	<p>As of 06 Jun 2017 all the primary-mission single-epoch Source-like tables were available for bulk download from IRSA.</p> <p>Bulk download of the NEOWISE (reactivation mission) Source-like tables from IRSA is not yet available, but work on this is in progress (outside the IPAC LSST group). We show this as a stretch goal only because of this dependency - once the data become available this is actually an important thing to add to PDAC because of the ability to exercise Source-like semantics in the Portal and the database.</p> <p>The necessary storage space is available already in the PDAC ("integration cluster") hardware at NCSA.</p>	STRETCH	DB-DAX /SLAC (load data), SUIT /IPAC (support in Portal)	<p>As of 05 May 2017 - IRSA has now done a significant fraction of the work required. Still working on proper packaging (e.g., including column metadata in the export files, computing checksums). No firm schedule, but completion during Summer 2017 seems likely.</p> <p>As of 10 May 2017 the 4-band cryo Source table (one of the six pieces of the Source table for the full mission) has been released. By 06 Jun 2017 the 3-band and 2-band post-cryo data were also released for bulk download.</p>	<div>DAX work:</div> <div> DM- 10740 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div> <div>SUIT work:</div> <div> DM- 8762 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div>
-------------------	---	---	---------	--	---	---


PDAC Data, Infrastructure	Support access to the HSC public release input data as well as outputs of 2017 Science Pipelines processing of that data.	<p>This may well include an additional "release" step to make outputs available in PDAC after they have been processed. The details of this remain to be worked out.</p> <p>For SUIIT, the main goal is to access a data set as close as possible to the final LSST data set: "We want to learn all the possible connections among all the tables and images so we can design and implement the system to provide a UI that aids users more effectively to explore and analyze the data."</p>	NORMAL	DB-DAX /SLAC and Science Pipelines for creating /ingesting data; SUIIT /IPAC for developing a Portal more specific to LSST's expected data		<div>DAX work:</div> <div> DM-10917 - Jira project does n't exist or you don't have permission to view it.</div> <div>SUIT work:</div> <div> DM-8721 - Jira project does n't exist or you don't have permission to view it.</div>
---------------------------	---	---	--------	--	--	---

Portal	Integrate the SUIT Portal application(s) with the NCSA Authentication & Authorization system prototype. Support user logins and begin to provide persistent Portal state across logins.		CRITICAL	SUIT/IPAC, AAIM/NCSA		<div> DM-10434 - Jira project does n't exist or you don't have permission to view it.</div> <div>(first step; in broader epic</div>
APIs (DAX)	Deploy DAX-metaserv for all loaded datasets in PDAC.	Column metadata for the Summer 2013 SDSS Stripe 82 catalog outputs appears to have been lost sometime in the intervening years. It is in the process of being manually regenerated by the DB-DAX team.	NORMAL	DB-DAX /SLAC		<div> DM-8727 - Jira project does n't exist or you don't have permission to view it.</div> <div>)</div>


WISE catalog metadata is available from IRSA (not yet including UCDS) and will be ingested by the DB-DAX group.

Priority "critical" because a lot of the Portal development depends on the availability of a metadata service.

SDSS:





DM-10374 - Jira project does not exist or you don't have permission to view it.



DM-10435 - Jira project does not exist or you don't have permission to view it.

WISE:

						<div> DM- 10918 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div>
APIs (DAX), Portal	Add asynchronous-query support to DAX and integrate it into the way that SUIIT interacts with DAX.	Essential development to support both: <ul style="list-style-type: none">• Long-running queries• Portal-Notebook integration	CRITICAL	DB-DAX /SLAC, SUIIT /IPAC	DAX work:	<div> DM- 4451 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div>



DM-

10772

- Jira

proje

ct

does

n't

exist

or

you

don't

have

permi

ssion

to

view

it.

SUIT work:



DM-

10500

- Jira

proje

ct

does

n't

exist

or

you

don't

have

permi

ssion

to

[view](#)

it.

(in broader epic



DM-

8727

- Jira

proje

ct

does

n't

exist

or

you

don't

have



permi



ssion



to


[view](#)

it.

APIs (DAX)	Advance the migration of the DAX services toward more VO-flavored (and VO-compliant) forms.	This is, essentially, the "DAX v1" work.	NORMAL	DB-DAX /SLAC	Most of the work to take advantage of this in the Portal will be in S18.	<div> DM- 2107 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div> <div> DM- 4863 - Jira proje ct does n't exist or you don't have permi ssion to view it.</div>
------------	---	--	--------	--------------	--	---

Notebook, Infrastructure	Deploy on the integration cluster (i.e., PDAC) hardware a JupyterHub service, integrated with the NCSA A&A system, supporting the launch of JupyterLab-beta (and possibly traditional Jupyter) notebooks with their IPython kernel server processes running with the user's NCSA identity and with read/write access to the NCSA LSST GPFS service, and with access to the PDAC DAX services.		NORMAL	SQuaRE /Tucson and MW&Inf /NCSA, <i>boundaries to be determined</i>		<div>SQuaRE work:<div><div></div><div>DM-10744</div><div>- Jira project does not exist or you don't have permission to view it.</div></div></div>
Notebook	Deploy the ability for the IPython kernel process to come up in an environment with a pre-installed LSST stack.		NORMAL	SQuaRE /Tucson		<div>NCSA work:<div><div></div><div>DM-9984</div><div>- Jira project does not exist or you don't have permission to view it.</div></div></div>

Notebook, Infrastructure	Ensure that DAX query services are available and readily usable in the PDAC Jupyter service.	<p>This is primarily a matter for the deployment architecture, to ensure that the Jupyter backend IPython kernel processes have network access to the DAX services.</p> <p>It may also include work on supporting users in exploiting those services from Python. What this means is TBD.</p>	NORMAL	Infrastructure /NCSA, DB-DAX/SLAC		
APIs (SUIT)	Develop extensions to the afw.display interfaces to expose existing table-image overlay and related SUIT functionality through an LSST-aware interface.		NORMAL	SUIT/IPAC		<div><p>DM-8770 - Jira project does not exist or you don't have permission to view it.</p></div>
Notebook, APIs (SUIT)	Develop usable early versions of JupyterLab-flavored table and image visualization widgets.	<p>The intent here is primarily to ensure that we understand and can work with the JupyterLab widget interface, which is different from the original Jupyter 3.0 widget model, and to do this in a way that delivers some basic functionality useful for developers.</p>	NORMAL	SUIT/IPAC		<div><p>DM-8770 - Jira project does not exist or you don't have permission to view it.</p></div>

APIs (DAX)	Integrate the DAX services and the underlying databases with the authentication and authorization system.	This is actually of critical importance, but is not expected to be possible in the F17 cycle.	<div>STRETCH</div> <div>for F17,</div> <div>CRITICAL</div> <div>for S18</div>	DB-DAX /SLAC, AAIM /NCSA		 <p>DM-2512 - Jira project does not exist or you don't have permission to view it.</p>
APIs (DAX)	Add the ability to create user tables through the DAX services.	If this goal is not realistic in 2017, a fallback would be to allow the creation of user databases in a more bare-bones, less integrated way. (E.g., as previously supported on the lsst-dev cluster.)	<div>STRETCH</div> <div>for doing this the DAX way,</div> <div>NORMAL</div> <div>for doing it the bare-bones way</div>	DB-DAX /SLAC		
Portal (first)	Begin cyber-security vetting of the PDAC / Science Platform components.	Note that the existing PDAC components have not yet been subjected to security review, and are therefore being kept behind an NCSA VPN. The goal is to bring several of these components out onto the public Internet. This requires both security review and the incorporation of authentication and authorization into the components that are exposed to the Internet.	NORMAL	Infrastructure /NCSA for strategy /screening, development teams for responding to issues.		
DM Documentation	Prepare a documentation package defining the Science Platform's expected capabilities, requirements, functional design, and deployment architecture.		CRITICAL	Everyone; core documents currently being developed by Unknown User (ciardi) and Gregory Dubois-Felsmann	Required for DM review, document deadline is 30 Jun 2017	
DM Project Science / Project Management	Perform verification tests of the F17 work, tied to requirements.		NORMAL	DM Subsystem Scientist, DM-SST to define tests		