

Butler Discussion 2015-12-02

Date and Location

02 Dec 2015, 11am-12pm pacific.

ls.st/sup

Attendees






- UW: [Russell Owen](#), [Simon Krughoff](#)
- Princeton/DRP: [Jim Bosch](#), [John Swinbank](#), [Paul Price](#)
- IPAC: [Unknown User \(jar\)](#), [Gregory Dubois-Felsmann](#), [Unknown User \(xiuqin\)](#), [Unknown User \(ciardi\)](#)
- SQUARE: [Frossie Economou](#)
- NCSA: [Greg Daues](#), [Steve Pietrowicz](#), [Margaret Gelman](#), [Matias Carrasco Kind](#), [Hsin-Fang Chiang](#)
- Architecture: [Kian-Tat Lim](#), [Tim Jenness](#)
- SLAC: [Nate Pease](#), [Brian Van Klaveren](#), [Fritz Mueller](#), [Andy Salnikov](#), [Vaikunth Thukral](#), [Unknown User \(kelsey\)](#), [Jacek Becla](#)

Goals

- To start getting on the same page with regards to butler, with focus on what butler really is, what it does, what its roles are, what APIs it has, and most importantly what it will and will not do in the future.

Discussion items

It'd be great if each team could cover features you think you "must have" in butler that are not available today, and features that would be nice-to-have but are not critical.

Time	Item	Who	Notes
5min	Intro	K-T or NateP or Jacek	
10min	UW		<ul style="list-style-type: none">• Easily specify new outputs from tasks (no need to edit obs_ packages)<ul style="list-style-type: none">• Nate Pease:  DM-4189 - Butler: provide API so that a task can define the output dataset type WON'T FIX• butler.get should fail immediately if the data is not available• Easily iterate over partial IDs for all dataset types<ul style="list-style-type: none">• Nate Pease:  DM-4532 - define and implement mechanism to iterate over partial datalds TO DO• Simplify the way IDs are normalized (?), and document the system better; also preferably make it easier to know when you have an ID that has been normalized or not.<ul style="list-style-type: none">• Nate Pease:  DM-4539 - Improve datald transformation TO DO• Handling of bitemporal calibration products including camera descriptions. N.b. calibration products can be lots of different things: objects, images, telemetry data, sky model, etc.<ul style="list-style-type: none">• Nate Pease:  DM-4168 - productize "Data repository selection based on version" TO DO• Butlerized access to logs, configs, stack version, etc.<ul style="list-style-type: none">• Nate Pease: Need more info. I created a topic thread on CLO. Won't do this. It's going to be done another way, see the topic thread.
10min	Princeton		DRP Butler Requirements / Feature Requests
10min	IPAC		SUI / Level 3 and SuperTask feedback on the Butler
10min	SQUARE		<ul style="list-style-type: none">• isolate in-memory from IO, architecture for plugging other protocols (like S3)<ul style="list-style-type: none">• Nate Pease:  DM-4555 - Add initial butler support for remote GET WON'T FIX
10min	NCSA		

Architecture	<p>Tim Jenness : I feel strongly that we should have two persistable file format options designed in to the butler (FITS and HDF5 say) so that we can be sure we have a flexible design that is not wedded to a single file format.</p> <p>Nate Pease:</p> <ul style="list-style-type: none"> DM-4543 - Add HDF5 support to butler dispatch mechanism TO DO DM-4542 - Refactor butler dispatch (object serialization) to be pluggable. DONE DM-4541 - Implement HDF5 serialization for AFW objects TO DO (not butler work, has an AFW component tag)
Discussion	<ul style="list-style-type: none"> Carry technical discussions via community Next meeting: Tuesday Dec 8, 12:30pm, via bluejeans. Focus: (a) check if created stories adequately cover all requests, (b) finalize prioritization

DRP Butler Requirements / Feature Requests

SUI / Level 3 and SuperTask feedback on the Butler

Random notes

Butler (not-done) stories:

T	Key	Summary	Assignee	Reporter	P	Status	Resolution	Created	Updated	Due
	DM-13903	CFITSIO is unable to open fits file when there is a blank character in the path	Fritz Mueller	Dominique Boutigny		TO DO	Unresolved	Mar 27, 2018	Feb 18, 2020	
	DM-11288	butler._addParents(self, repoDataList) ignores mapperArgs	Unassigned	Robert Lupton		INVALID	Done	Jul 16, 2017	Jul 26, 2017	
	DM-10556	The butler incorrectly caches dataTypes such as calexp_sub	Unassigned	Robert Lupton		INVALID	Done	May 17, 2017	May 22, 2017	
	DM-9036	butler PickleStorage is not py2 /py3 cross-compatible	Kian-Tat Lim	John Parejko		WON'T FIX	Done	Jan 18, 2017	Jan 06, 2019	
	DM-7046	Prototype python cache for weak_ptr and weakref objects.	Nate Pease	Nate Pease		WON'T FIX	Done	Jul 27, 2016	Oct 21, 2016	
	DM-6896	remove ButlerFactory and change all the packages that use it to use Butler() directly	Unassigned	Nate Pease		WON'T FIX	Done	Jul 05, 2016	Jan 27, 2020	
	DM-5608	plan and RFC for "data repository based on version"	Unassigned	Nate Pease		WON'T FIX	Done	Mar 24, 2016	Sep 26, 2019	
	DM-5518	Create proposal & RFC for Butler API to define output dataset type	Unassigned	Fritz Mueller		WON'T FIX	Done	Mar 17, 2016	Feb 18, 2020	
	DM-5460	Allow butler to use python classes without readFits	Unassigned	Nate Lust		WON'T FIX	Done	Mar 15, 2016	Oct 02, 2019	
	DM-4944	butler should transparently allow files to be compressed or not	Unassigned	Nate Pease		INVALID	Done	Jan 28, 2016	Nov 30, 2016	
	DM-4698	Add initial butler support for remote PUT	Brian Van Klaveren	Nate Pease		WON'T FIX	Done	Dec 17, 2015	Feb 21, 2018	
	DM-4555	Add initial butler support for remote GET	Brian Van Klaveren	Nate Pease		WON'T FIX	Done	Dec 04, 2015	Feb 21, 2018	
	DM-4554	Basic Remote Butler functionality	Unassigned	Nate Pease		WON'T FIX	Done	Dec 04, 2015	Jun 02, 2016	

	DM-4553	Move butler from daf_persistence to daf_butler	Unassigned	Nate Pease	WON'T FIX	Done	Dec 04, 2015	Feb 18, 2020
	DM-4551	Butler policy settings via pipe tasks argument parser	Unassigned	Nate Pease	TO DO	Unresolved	Dec 04, 2015	Nov 20, 2017
	DM-4549	Investigate creating a VO data link URL from butler datald	Unassigned	Nate Pease	TO DO	Unresolved	Dec 04, 2015	Aug 30, 2019
	DM-4548	Create DAX URL from butler datald	Unassigned	Nate Pease	WON'T FIX	Done	Dec 04, 2015	Feb 21, 2018
	DM-4543	Add HDF5 support to butler dispatch mechanism	Unassigned	Nate Pease	TO DO	Unresolved	Dec 04, 2015	Dec 07, 2018
	DM-4539	Improve datald transformation	Unassigned	Nate Pease	TO DO	Unresolved	Dec 03, 2015	Nov 20, 2017
	DM-4532	define and implement mechanism to iterate over partial datalds	Unassigned	Nate Pease	TO DO	Unresolved	Dec 03, 2015	Nov 20, 2017

Showing 20 out of 30 issues

Action items

- Nate Pease, turn the requirements discussed / input received into JIRA stories
- Identify core architectural issues
- Design core architectural issues
- Write a comprehensible high level design document
- Create interface definition (*we will defer remote butler & URL interface definition*)
 - Application Interface
 - Configuration Interface
 - Back-end interface. Include what can be plugged in, and where.
 - transport
 - serialization
 - storage
- Nate Pease and Kian-Tat Lim - propose prioritization of all butler stories
- Work on top-priority item for each group:
 - Princeton: **DM-4181** - Butler: add support for skymap based datalds **TO DO** (epic S16) ([Serge Monkewitz](#) is working on kind-of-related **DM-3472** - Implement Basic spatial lookups for the Butler **DONE**)
 - SUI: Multiple repositories: finish defining, create story, work on partial implementation TBD.
 - DM-4625** - Design for butler support of multiple repositories **DONE**
 - Implement **DM-4682** - Implementation of multiple repositories v1 **DONE** (epic W16) (might get broken into 2 stories and delivered over 2 sprints)
 - UW: **DM-4168** - productize "Data repository selection based on version" **TO DO** (epic Butler W16), also "if time": **DM-4180** - Butler: provide API so that a task can define the output dataset type **WON'T FIX** (epic S16)
 - Tuscon: define interface for **DM-4555** - Add initial butler support for remote GET **WON'T FIX** (epic S16, assigned to BVK)
 - NCSA: (nothing at this time)