# 2018-10-23 to 25 Science Platform Workshop


- Location
- Attendees
- Remote access
- Block Agenda
    - Tuesday 23 Oct 2018
    - Wednesday 24 Oct 2018
    - Thursday 25 Oct 2018
    - Action Items
    - Pre-Meeting Planning
- Documents
- Logistics
    - Hotel
    - DINNER / NIGHTLIFE / SHOPPING
    - TRANSPORTATION


## Location

The NCSA building is located at 1205 W. Clark St., Urbana, IL.  We will be meeting in Room 1040.

## Attendees

1. Gregory Dubois-Felsmann
2. Unknown User (xiuqin)
3. Trey Roby
4. Tatiana Goldina
5. Fritz Mueller
6. Frossie Economou
7. Simon Krughoff
8. Brian Van Klaveren
9. Kian-Tat Lim
10. Kenny Lo
11. Leanne Guy (one day only - probably Tuesday - TBD))
12. Unknown User (cbanek)
13. Colin Slater
14. Adam Thornton (remote)
15. Angelo Fausti (remote)
16. Loi Ly  (remote)
17. Unknown User (cwang) (remote)
18. David Shupe (remote, Weds and Thurs only)
19. James Basney (local)
20. Unknown User (awithers) (local)
21. Unknown User (mcarras2) (local)
22. Hsin-Fang Chiang
23. Steve Pietrowicz (local)

Remote attendance will be possible.

## Remote access

Remote access will be by Webex: https://ncsaevents.webex.com/ncsaevents/j.php?MTID=mae5edd6bd6eb3d96f1bcfca39c1a1c07

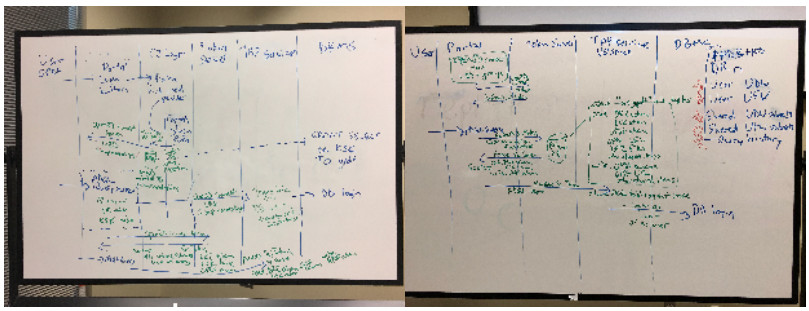
# Block Agenda

### Tuesday  23 Oct 2018

| Time | Location | Topic |
|---|---|---|
| 10am-12pm | NCSA 1040 | User workspaces: use cases and requirements Leanne Guy Gregory Dubois-Felsmann |
| 12pm-1:30pm | | Lunch on your own |
| 1:30pm-3pm | NCSA 1040 | • architecture discussion to make sure we can address all the listed use cases<br>• Deployment of LSP<br><br>Gregory Dubois-Felsmann |

| | | |
|---|---|---|
| 3pm-3:30pm | | coffee break |
| 3:30pm-5pm | NCSA 1040 | LSP review charge and prep-work tasks discussion Leanne Guy Gregory Dubois-Felsmann |

## Wednesday 24 Oct 2018

| Time | Location | Topic |
|---|---|---|
| 9am-10:30am | NCSA 1040 | • Demo of Firefly JupyterLab extensions (work in progress) Trey Roby<br>• Demo of object footprints overlay using Firefly Gregory Dubois-Felsmann Unknown User (xiuqin)<br>• Demo of TAP service based on CADC app Unknown User (cbanek) Frossie Economou<br>• Demo of visualization using DASK in LSP Simon Krughoff |
| 10:30am-11am | | coffee break |
| 11am-12:30pm | NCSA 1040 | • Next to DB processing workflow discussion Fritz Mueller Colin Slater<br>• Next-2-DB Slides |
| 12:30pm-1:30pm | | Lunch on your own |
| 1:30pm-3pm | NCSA 1040 | talk on SciToken and Oauth2 James Basney Unknown User (mbutler)<br> |
| 3pm-3:30pm | | coffee break |
| 3:30pm-5pm | NCSA 1040 | (This session was in practice used for an extended discussion of SciTokens and the possible role of capability-based authorization. The original agenda items below were moved to Thursday:)<br><br>VO standards LSST DM will implement Gregory Dubois-Felsmann<br><br>Development status report from DAX team Brian Van Klaveren<br><br>list of tasks to do Fritz Mueller |

## Thursday 25 Oct 2018

| Time | Location | Topic | Possible breakout |
|---|---|---|---|
| 9am-10:30am | NCSA 1040 | Identity & Access Management and Security (James Basney, Unknown User (awithers) )<br><br>1. Reminder of the role of the new federated-ID software/service<br>2. User management (password change, etc.)<br>3. A+A token passing diagrams for<br>  a. Notebook login and Python access to TAP via pyro<br>  b. TOPCAT access to LSST data<br><br>(The morning agenda items below were more or less covered in the extended Wednesday afternoon session on SciTokens, etc.: )<br><br>1. login page supplied by NCSA for portal access<br>2. login and get the token for DAX | |

| | | | |
|---|---|---|---|
| 10:30am-11am | | Coffee break | |
| 11am-12:30pm | NCSA 1040 | 1. VO standards LSST DM will implement Gregory Dubois-Felsmann (30 min)<br>a. (See also Unknown User (cbanek)'s DMTN-090)<br>2. Development status report from DAX team Fritz Mueller (15 min)<br>3. DAX task list in light of this week's discussions Fritz Mueller (15 min)<br>4. LSP instance / deployment / staging discussion (30 min)<br>a. Operations concepts for the various instances<br>b. What deployment platforms are **required** to be supported? Only in-house K8s? Cloud?<br>c. Ref.: Science Platform Instances - a re-evaluation | |
| 12:30pm-2pm | | Lunch and tour NCSA facility (we can have lunch delivered for those who want to visit the facility) | |
| 2pm-3pm | | 1. DM-10 (Portal) descope discussion (30 min)<br>2. wrap up, and establish timeline for agreed upon tasks (30 min) | |
| 3pm-3:30pm | | coffee break | |
| 3:30pm-5pm | NCSA 1040 | hack session: DAX set up WebDAV and SUIT access it from portal | |

## Action Items

| Session | Action Items |
|---|---|
| User workspace requirements | • Need a new term to refer to user-created tables that are do not bring the full spectrum of "Level 3" requirements; i.e., user uploads a random catalog, what do we call that?<br>◦ Context: Gregory suggested that a user could make a "level 3" data product on their home computer, but Frossie suggests this shouldn't be conflated with products stored on the DAC<br>• Need to clearly specify that we support a MyDB-type storage of arbitrary data, does not have to be related-in-scientific-meaning to the LSST data products<br>• Determine how tables will be uploaded to MyDB, either by Portal or Notebook or external clients.<br>• There was a discussion of "direct access to consolidated database", beyond capabilities of IVOA protocols. Need to follow up on whether this is mediated by DAX services or if some form of more direct connection to consolidated DB is acceptable.<br>• Science platform requirements say MyDB reqs are "still in development"<br>• Scope effort for providing user-facing crossmatches/comparisons with other surveys, beyond those that are already in-use for Pipelines validation.<br>• What are the sizes of user-generated data in existing archives? Our 10% for users is probably too small.<br><br>☑ Leanne Guy will ask Gaia archive for user data volumes.<br><br>• What is the story that we will tell reviewers for how external groups will leverage more compute/storage, of their own acquisition, beyond the 10%? What development requirements does such a system impose on us, and what decisions do we need to make to support this "beyond 10%" capability? Do we have appropriate requirements in the baseline, and to support evolving this N-years into operations?<br><br>☑ Leanne Guy will write a tech note listing some "Level 3"/"MyDB" science use cases and present first version at SST/DMLT F2F Nov 2018 05 Nov 2018 |
| Arch and Blockers | • Raise at DMLT the question of whether Google is a required deployment target.<br>• Raise at DMLT whether the parquet support should be added to the portal.<br>• Gregory Dubois-FelsmannColin Slater and Michelle Gower will discuss what type of story we should write up w.r.t. submitting jobs to the batch system from the notebook.<br><br>☑ Kian-Tat Lim will write some initial options for resource measurement and metering.<br><br>• Unknown User (malone12) to write up options for LSP application logging infrastructure 30 Nov 2018 |
| Next-to-DB | • Having parquet replication instead of replication inside of qserv has an impact on operational recovery; Are we willing to accept this?<br>• Implicit in the parquet+dask/spark scenario is that the data would be stored separately from the compute resources. Are the network resources sufficient to support this? |

| | |
|---|---|
| Authentication and Authorization | • Brian Van Klaveren will update DMTN-094 to describe a proposed design.<br>• Brian Van Klaveren aim to bring this up at the November IVOA, if possible.<br>• Kian-Tat Lim to raise at DMLT that the DMTN-094 proposal would require a token server to be built and is not yet assigned to a WBS.<br>• Colin Slater to raise the A&A/security usage patterns at SST, develop a document describing the user stories.<br>• In the Thursday A&A session, Kian-Tat Lim proposed to write up the identity-based authentication system as an initial sufficient system to build. The more sophisticated capabilities-based system could be developed as an evolution of that, closer to or during operations.<br>• DMTN-094 update should include a comparison to what other archives are implementing.<br>• Kian-Tat Lim Gregory Dubois-Felsmann to write up options for standardizing the names and meanings of capabilities. Proposed methods are: 1) change-controlled document 2) shared library 3) shared service.<br>• After we codify our DMTN-094-level design, we need to collaborate (via IVOA) on improving the authentication capabilities of non-LSST clients (topcat, pyvo). Arch team responsibility? |

## Pre-Meeting Planning

Suggested topics for discussion

| Topic | Requested by | Time required (estimate) | Notes |
|---|---|---|---|
| Deployments of all three aspects within one instance | Frossie Economou<br><br>Fritz Mueller | | • Identify the issues before the F2F if any<br>• Resolve them and have a documented deployment procedure<br>• PDAC cluster move into the k8s commons in Oct.? More details needed |
| Connect all three aspects | Unknown User (xiuqin) Frossie Economou | | • Search result user gets at portal is accessible through NB, and vice versa.<br>• Search submitted by DAX API is accessible through portal and NB. |
| A+A | Unknown User (xiuqin) | | • Understand and clarify the different types of tokens before F2F<br>• Portal and DAX resolve the token passing issues<br>• login page supplied by NCSA for portal access<br>• login and get the token for DAX<br>• user management (password change, etc.) |
| WebDAV API for workspace | Unknown User (xiuqin) | | • relationship with A+A needs to be clarified |
| Python access to LSST data (next to database process) | Gregory Dubois-Felsmann | | • afw.table data access through butler, enable panda compatible format (but it may be not realistic to provide it in time)<br>• access to Parquet formatted data?<br>• 2 use cases: after burner data process with LSST pipeline; just get the data as close as to community standard as possible and so the community tools can be used<br>• Fritz Mueller Colin Slater will give a status on this |
| Tour of NPCF facility | Joel Plutchak or Unknown User (mbutler) can coordinate if this is desired. | 1.5 hours | If desired, a tour of the National Petascale Computing Facility can be arranged. ~45 minute tour with 10-15 minutes travel each way. |
| Perspectives from elsewhere in NCSA:<br>• Lessons learned from DES data access infrastructure (tools, usage, etc)<br>• National Data Service (NDS) as a scientific gatewa | | | |

# Documents

- LSP Requirements: LDM-554
- LSP Design: LDM-542
- Data Access White Paper: Document-5373
- For reference: original text file with agenda contents draft for this workshop

# Logistics

## Hotel

Home2 Suites by Hilton Champaign/Urbana
2013 S. Neil St.
Champaign, IL 61820

Erin Carlson will be doing a rooming list. The hotel is self-pay upon check-out. As long as a TR is placed for this meeting, she will add you to the list for the dates of travel.

9.25.18- sent rooming list to hotel. Please let Erin know if there are changes to your TR and/or dates.

10.2.18- hotel rooming list with confirmation numbers

10.5.18- updated rooming list

# LUNCH

Food nearest NCSA:

- Daily Byte (sandwiches); NCSA Lobby
- Daily Byte II (sandwiches); Electrical Engineering Building lobby (1 block east)
- Einsteins (sandwiches); Siebel Center lobby (next building south)
- Beckman Café (cafeteria-style; sandwiches and hot entrees with vegetarian & vegan options); Beckman Institute (1 block east-northeast)
- Other (north and east on University Ave):
    - Burger King
    - Perkins
    - Dairy Queen
    - Taco Bell
    - Niro's Gyros

- Food Trucks (varies by day); Two blocks south near Mathews and Springfield
    - Cracked (egg sandwiches)
    - Jurassic Grill (burgers and such)
    - TK Cheesesteaks
    - Dragon Fire Pizza

Many more options slightly farther afield (3/4 to 1 mile):

- Campustown
- Southeast campus

## DINNER / NIGHTLIFE / SHOPPING

- Downtown Champaign has a variety of restaurants, bars, and clubs, as well as a movie theater or two, e.g.:
    - Black Dog Smoke and Ale House, Champaign
    - Radio Maria
    - Seven Saints
    - Big Grove Tavern
    - Maize at the Station
    - The Himalayan Chimney
- Downtown Urbana is smaller, but also has numerous restaurants and clubs, e.g.:
    - The Dancing Dog (100% vegan)
    - Silvercreek
    - Siam Terrace
    - Crane Alley
- South Neil near the hotel there is Harvest Market (grocery store with restaurant), Papa Del's (new location), Starbucks, Moe's Southwestern Grill, Scotty's Brewhouse, etc.
- North Prospect Avenue is the main twin cities shopping area, with chain restaurants, Portillio's, big box stores, the Market Place Mall, etc.
- There are a couple parks within walking distance of the hotel -- Mattis Park and the Boulware Trail to the west and Hessel Park to the north.

## TRANSPORTATION

### Airport

The University of Illinois Willard Airport is 5 miles south of the Home2suites in the village of Savoy. Taxi, Lyft, or Uber should be about $12 plus tip.

### Mass Transit

C-U buses generally cost $1 per one-way trip. iStops are campus stops where it is unnecessary to pay a fare to ride the Gold, Illini, Silver, Teal, and Yellow routes. iStop locations are indicated on individual route maps with an iStop symbol (an orange octagon with a white "i" inside) and are identified as such on bus stop signs at the stop location.

The Brown route is closest to the Home2 Suites with a stop at Fox Drive and State Street. Disembark at Springfield and Sixth; NCSA is ½ mile northeast. Route 9B is counter-clockwise (to NCSA), route 9A is clockwise (to the hotel).

A Gold route stop is a long block north of the hotel at State and Kirby. Disembark at the Ceramics building shelter on Goodwin Avenue; NCSA is 0.3 miles north. Take Gold East to NCSA, Gold West to the hotel.

Use the C-U MTD Trip Planner to investigate times and options, but beware of options having you board/disembark at Lot E-14 or St. Mary's and Oak. To get to/from the hotel you have to traverse the narrow St. Mary's Road under an overpass with no sidewalk or shoulder and plenty of morning/evening traffic.