

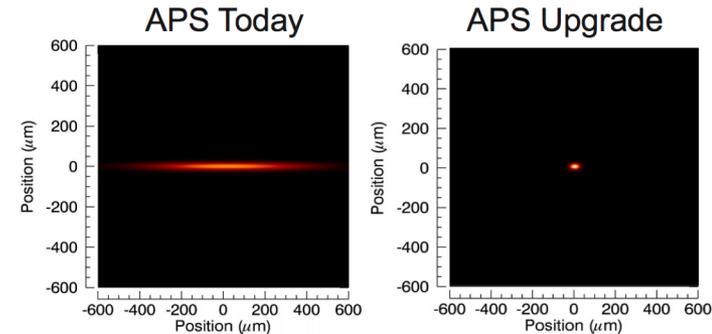
# Advanced Photon Source Data Management

**S. Veseli, N. Schwarz, C. Schmitz (SDM/XSD)  
R. Sersted, D. Wallis (IT/AES)**



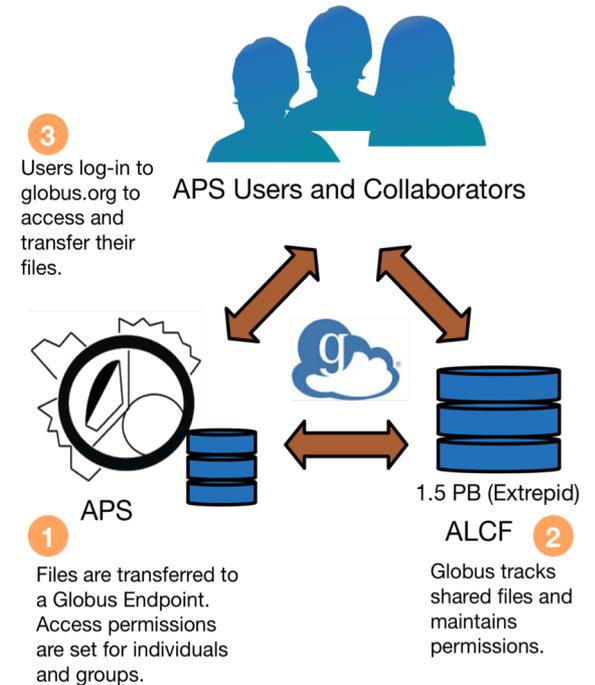
# Growing Beamline Data Needs

- X-ray detector capabilities are constantly improving: bigger frames, higher frame rates => more raw data
- APS Upgrade: Higher brightness => more x-rays can be focused onto a smaller area => more raw data in greater detail and less time
- APS 1ID:
  - Today: 4x Hydra GE detector, 8MB frame, 8 fps => 256MB/s data rate ( 1TB/hr)
  - Production Data Rates (including overhead, on a good week): 10TB/day, 60TB/week
  - Near future (1-3 years): 2x 2923 Dexela (or similar), 23MB frame, 26 fps => 1.2GB/s data rate
  - Near future: Pilatus 2M (or similar/larger), 9.5MB frame, 250 fps => 2.3GB/s data rate
- APS 8IDI:
  - 2010-2016: ANL/LBL FCCD, 2MB frame, 100fps, compressed in real-time with 10% non-zero pixels on the average => about 200MB/s data rate
  - 2016-Today: Lambda 750K, 1.5MB frame, about 10% non-zero pixels, 2000 fps => about 300 MB/s data rate
  - Production Data Rates: 8-10 TB/week (max), 1-2TB/week (average)
  - Future: research on VIPIC (>1MHz frame rate) and UFXC (50 kHz frame rate)

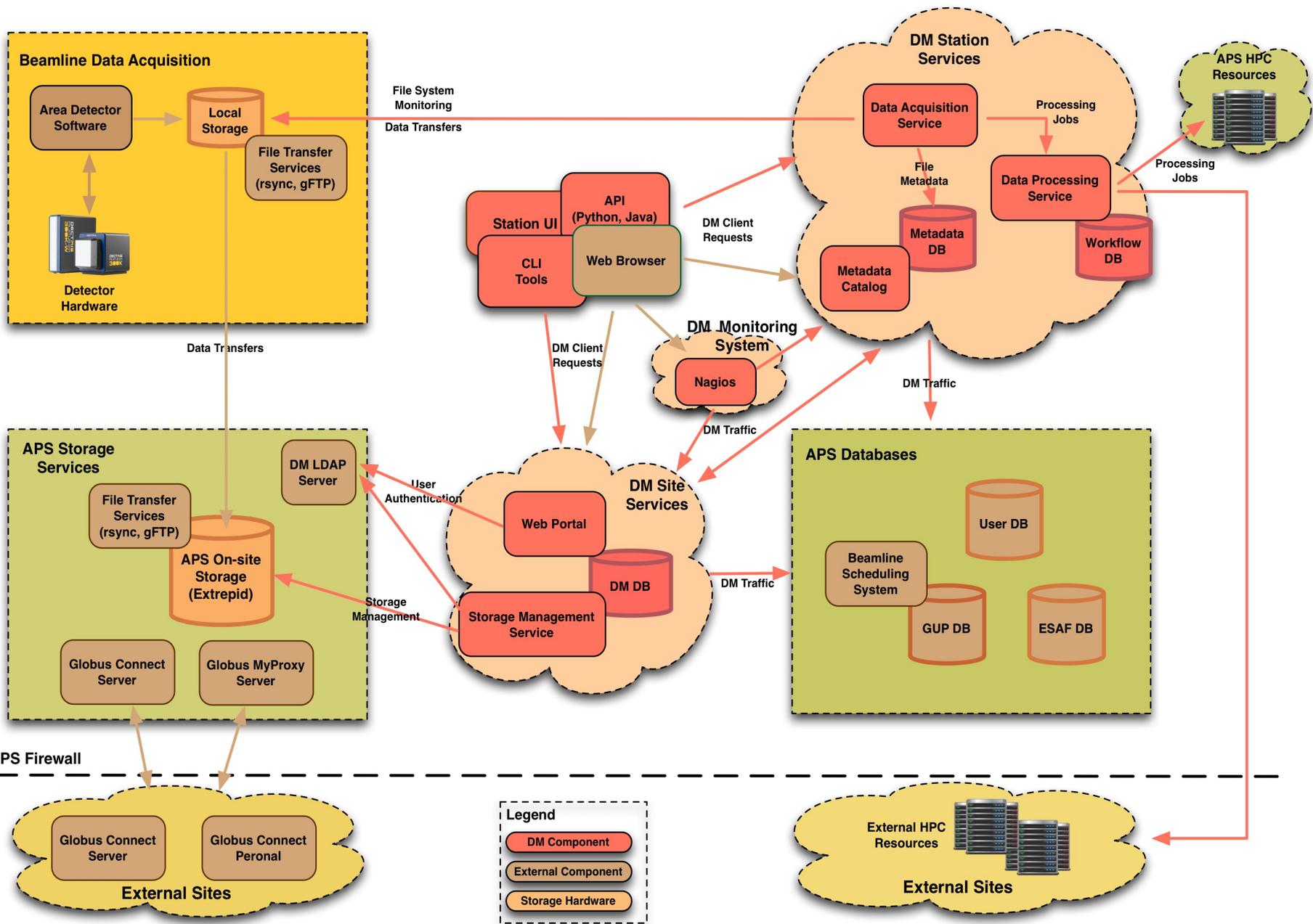


# APS Data Management

- Specific data management needs typically vary from beamline-to-beamline, mostly depending upon the types of detectors, X-ray techniques, and data processing tools in use
- However, most of the data management requirements are related to a set of tasks common to all beamlines:
  - Storage area management (e.g. movement of acquired data from local storage to a more permanent location, data archival, etc.)
  - Enabling users and applications to easily find and access data (metadata and replica catalogs, remote data access tools)
  - Facilitating data processing and analysis with automated or user-initiated processing workflows
- APS Data Management (DM) project strives to help with these tasks

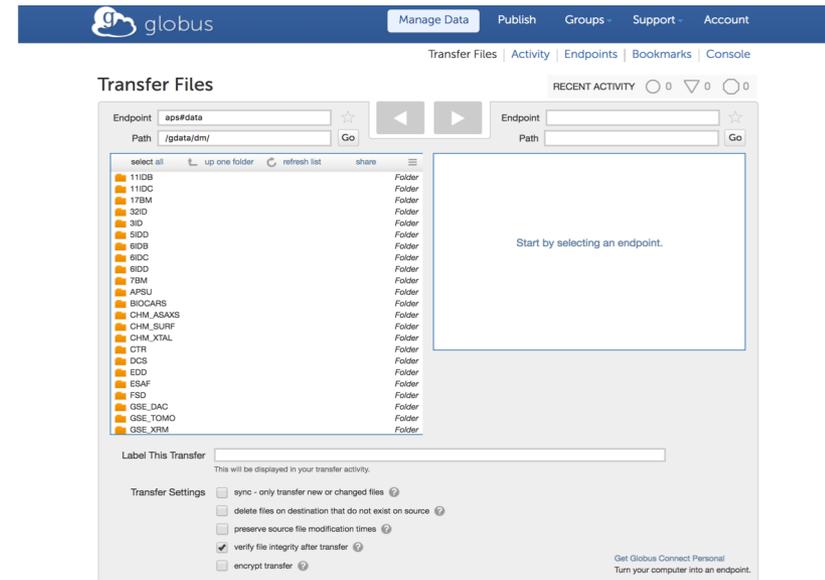


# APS Data Management System



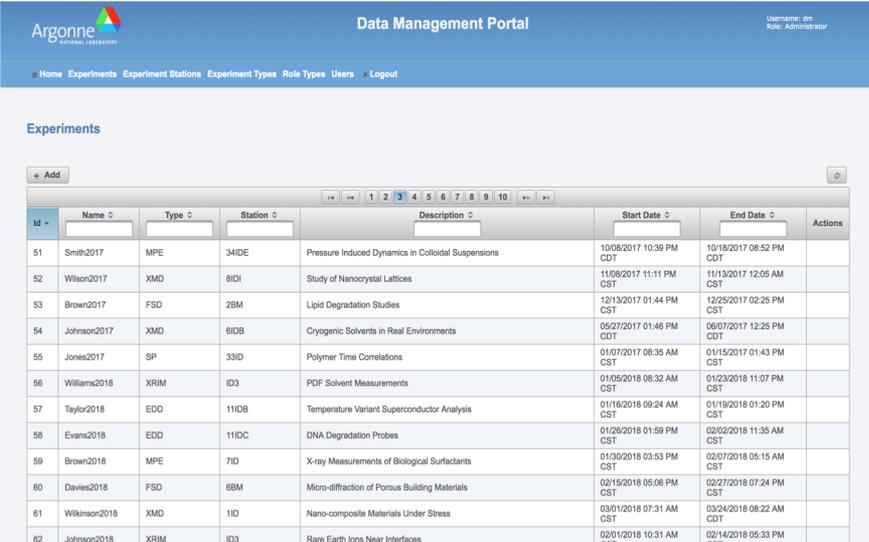
# How Does Globus Fit In?

- Globus Connect Server: provides remote data access to APS on-site storage (via the aps#data endpoint)
- Globus MyProxy OAuth Server: handles aps#data endpoint authentication
- GridFTP servers and clients: used by DM software for internal data transfers between beamline (local) storage and APS on-site storage, for transfers between local storage and HPC resources, etc.
  - Issue: Support for Open Source Globus Toolkit ended in early 2018



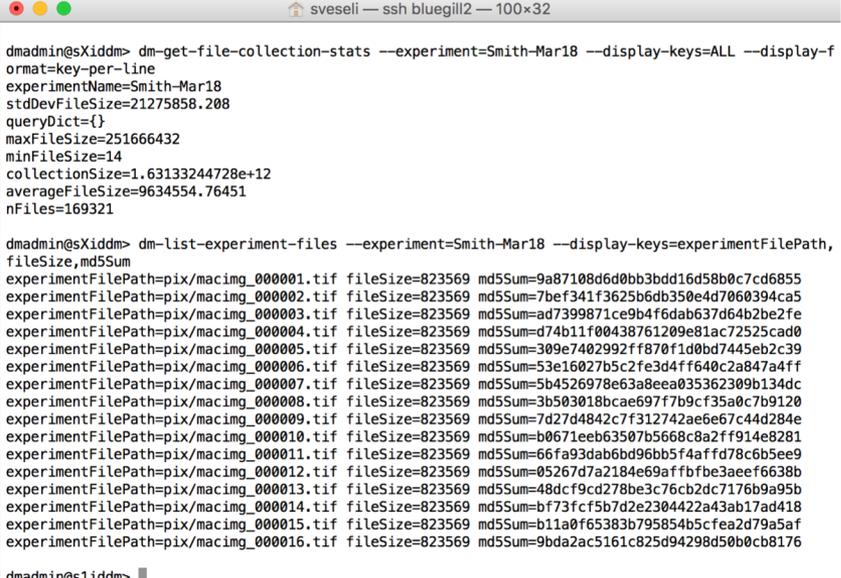
# Services

- Site Services:
  - DM Database (PostgreSQL)
  - Storage Management Service
  - Web Portal
  - Automated user account synchronization utilities
- Station (Beamline Deployment) Services:
  - Data Acquisition Service
  - Metadata Catalog
  - Processing Service
- All services are available via REST interfaces



The screenshot shows the Argonne Data Management Portal interface. At the top, there is a navigation bar with the Argonne logo and the text "Data Management Portal". Below the navigation bar, there is a breadcrumb trail: "Home Experiments Experiment Stations Experiment Types Role Types Users Logout". The main content area is titled "Experiments" and contains a table with the following columns: "Id", "Name", "Type", "Station", "Description", "Start Date", "End Date", and "Actions". The table lists 12 experiments, each with a unique ID, name, type, station, description, start date, and end date.

Id	Name	Type	Station	Description	Start Date	End Date	Actions
51	Smith2017	MPE	34IDE	Pressure Induced Dynamics in Colloidal Suspensions	10/08/2017 10:38 PM CDT	10/18/2017 08:52 PM CDT	
52	Wilson2017	XMD	8IDI	Study of Nanocrystal Lattices	11/08/2017 11:11 PM CST	11/13/2017 12:05 AM CST	
53	Brown2017	FSD	28M	Lipid Degradation Studies	12/13/2017 01:44 PM CST	12/25/2017 02:25 PM CST	
54	Johnson2017	XMD	6I0B	Cryogenic Solvents in Real Environments	06/27/2017 01:46 PM CDT	06/07/2017 12:25 PM CDT	
55	Jones2017	SP	33ID	Polymer Time Correlations	01/07/2017 08:35 AM CST	01/15/2017 01:43 PM CST	
56	Williams2018	XRIM	ID3	PDF Solvent Measurements	01/05/2018 08:32 AM CST	01/23/2018 11:07 PM CST	
57	Taylor2018	EDD	11IDB	Temperature Variant Superconductor Analysis	01/16/2018 09:24 AM CST	01/19/2018 01:20 PM CST	
58	Evans2018	EDD	11IDC	DNA Degradation Probes	01/26/2018 01:59 PM CST	02/02/2018 11:35 AM CST	
59	Brown2018	MPE	7ID	X-ray Measurements of Biological Surfactants	01/30/2018 03:53 PM CST	02/07/2018 05:15 AM CST	
60	Davies2018	FSD	68M	Micro-diffraction of Porous Building Materials	02/15/2018 05:06 PM CST	02/27/2018 07:24 PM CST	
61	Wilkinson2018	XMD	1ID	Nano-composite Materials Under Stress	03/01/2018 07:31 AM CST	03/24/2018 08:22 AM CST	
62	Johnson2018	XRIM	ID3	Rare Earth Ions Near Interfaces	02/01/2018 10:31 AM CST	02/14/2018 05:33 PM CST	



The screenshot shows a terminal window with the following content:

```
sveseli — ssh bluegill2 — 100x32

dmadmin@sXiddm> dm-get-file-collection-stats --experiment=Smith-Mar18 --display-keys=ALL --display-format=key-per-line
experimentName=Smith-Mar18
stdDevFileSize=21275858.208
queryDict={}
maxFileSize=251666432
minFileSize=14
collectionSize=1.63133244720e+12
averageFileSize=9634554.76451
nFiles=169321

dmadmin@sXiddm> dm-list-experiment-files --experiment=Smith-Mar18 --display-keys=experimentFilePath,
fileSize,md5Sum
experimentFilePath=pix/macimg_000001.tif fileSize=823569 md5Sum=9a87108d6d0bb3bdd16d58b8c7cd6855
experimentFilePath=pix/macimg_000002.tif fileSize=823569 md5Sum=7bef341f3625b6db350e4d7060394ca5
experimentFilePath=pix/macimg_000003.tif fileSize=823569 md5Sum=ad7399871ce9b4f6da6b37d64b2be2fe
experimentFilePath=pix/macimg_000004.tif fileSize=823569 md5Sum=d74b11f00438761209e81ac72525cad0
experimentFilePath=pix/macimg_000005.tif fileSize=823569 md5Sum=309e7402992ff870f1d0bd7445eb2c39
experimentFilePath=pix/macimg_000006.tif fileSize=823569 md5Sum=53e16027b5c2fe3d4ff640c2a847a4ff
experimentFilePath=pix/macimg_000007.tif fileSize=823569 md5Sum=5b4526978e63a8eea035362309b134dc
experimentFilePath=pix/macimg_000008.tif fileSize=823569 md5Sum=3b503018bcae697f7b9cf35a0c7b9120
experimentFilePath=pix/macimg_000009.tif fileSize=823569 md5Sum=7d27d4842c7f312742ae6e67c44d284e
experimentFilePath=pix/macimg_000010.tif fileSize=823569 md5Sum=b0671eeb63507b5668c8a2ff914e8281
experimentFilePath=pix/macimg_000011.tif fileSize=823569 md5Sum=66fa93dab6bd96bb5f4affd78c6b5ee9
experimentFilePath=pix/macimg_000012.tif fileSize=823569 md5Sum=05267d7a2184e69aff7bfe3aee6f6638b
experimentFilePath=pix/macimg_000013.tif fileSize=823569 md5Sum=48dcf9cd278be3c76cb2dc7176b9a95b
experimentFilePath=pix/macimg_000014.tif fileSize=823569 md5Sum=bf73fcf5b7d2e2304422a43ab17ad418
experimentFilePath=pix/macimg_000015.tif fileSize=823569 md5Sum=b11a0fe5383b795854b5cfea2d79a5af
experimentFilePath=pix/macimg_000016.tif fileSize=823569 md5Sum=9bda2ac5161c825d94298d50b0cb8176

dmadmin@s1iddm> |
```



# User Interfaces

- Web browser access to DM Web Portal, Nagios web pages, beamline Metadata Catalog, and Globus Online (for remote data access)
- Python REST services are accessible via DM Python and Java APIs
- Processing Service provides OMQ interfaces
- Extensive set of command line tools
  - Built on top of Python APIs
  - Session based
  - Fully scriptable
  - Online usage documentation (accessible via the `-h|--help` options)
- DM Station GUI (C. Schmitz)
  - Implemented in PyQt
  - Uses Python REST APIs
  - Easiest way to start using the system

**Wilkinson2018 Settings**

Name: Wilkinson2018  
 Start Date: 3/1/18 End Date: 3/24/18  
 Type: XMD  
 Description: Nano-composite Materials Under Stress

Badge	First	Last	Email
1 678912345	Laura	Wilkinson	wilkinson@u...
2 891234567	John	Gray	jgray@anl.gov
3 789123456	Amanda	Greene	agreene@anl...
4 912345678	Kathy	Parkins	kparkins@un...
5 123456789	Patrick	Elmer	pelmer@iit.edu

```

Usage:
  dm-6bm-daq --experiment=EXPERIMENTNAME --data-directory=DATADIRECTORY
  [--duration=DURATION]
  [--dest-directory=DESTDIRECTORY]
  [--upload-data-directory-on-exit=UPLoADDATADIRECTORYONEXIT]
  [--upload-dest-directory-on-exit=UPLoADDESTDIRECTORYONEXIT]
  [--process-hidden]
  [--process-existing]
  [--workflow-name=WORKFLOWNAME --workflow-owner=WORKFLOWOWNER
  [--workflow-job-owner=WORKFLOWJOBOWNER]
  [--workflow-args="key1:value1 key2:value2 ..."]
  ]
  [--skip-plugins=SKIPPLUGINS]
  [--type=TYPENAME]
  [--description=DESCRIPTION]
  [--start-date=STARTDATE]
  [--end-date=ENDDATE]
  [--users=USERS]
  [--esaf-id=ESAFID]
  [--proposal-id=PROPOSALID]
  [--run=RUNNAME]
  [key1:value1, key2:value2, ...]
    
```

Description:  
 Run DAQ for experiment on station 6BM. If experiment does not exist, it will be added to the DM database. If list of users or proposal id is specified, this command will also add roles for all users listed on the proposal.

**globus** Manage Data Publish Groups Support Account

Transfer Files | Activity | Endpoints | Bookmarks | Console

Endpoint: appdata Path: /gdata/ Go

select all up one folder refresh list share

- 110B Folder
- 110C Folder
- 178M Folder
- 303 Folder
- 3D Folder
- 50D Folder
- 50S Folder
- 50C Folder
- 50D Folder
- 50E Folder
- 50F Folder
- 50G Folder
- 50H Folder
- 50I Folder
- 50J Folder
- 50K Folder
- 50L Folder
- 50M Folder
- 50N Folder
- 50O Folder
- 50P Folder
- 50Q Folder
- 50R Folder
- 50S Folder
- 50T Folder
- 50U Folder
- 50V Folder
- 50W Folder
- 50X Folder
- 50Y Folder
- 50Z Folder
- 50AA Folder
- 50AB Folder
- 50AC Folder
- 50AD Folder
- 50AE Folder
- 50AF Folder
- 50AG Folder
- 50AH Folder
- 50AI Folder
- 50AJ Folder
- 50AK Folder
- 50AL Folder
- 50AM Folder
- 50AN Folder
- 50AO Folder
- 50AP Folder
- 50AQ Folder
- 50AR Folder
- 50AS Folder
- 50AT Folder
- 50AU Folder
- 50AV Folder
- 50AW Folder
- 50AX Folder
- 50AY Folder
- 50AZ Folder
- 50BA Folder
- 50BB Folder
- 50BC Folder
- 50BD Folder
- 50BE Folder
- 50BF Folder
- 50BG Folder
- 50BH Folder
- 50BI Folder
- 50BJ Folder
- 50BK Folder
- 50BL Folder
- 50BM Folder
- 50BN Folder
- 50BO Folder
- 50BP Folder
- 50BQ Folder
- 50BR Folder
- 50BS Folder
- 50BT Folder
- 50BU Folder
- 50BV Folder
- 50BW Folder
- 50BX Folder
- 50BY Folder
- 50BZ Folder
- 50CA Folder
- 50CB Folder
- 50CC Folder
- 50CD Folder
- 50CE Folder
- 50CF Folder
- 50CG Folder
- 50CH Folder
- 50CI Folder
- 50CJ Folder
- 50CK Folder
- 50CL Folder
- 50CM Folder
- 50CN Folder
- 50CO Folder
- 50CP Folder
- 50CQ Folder
- 50CR Folder
- 50CS Folder
- 50CT Folder
- 50CU Folder
- 50CV Folder
- 50CW Folder
- 50CX Folder
- 50CY Folder
- 50CZ Folder
- 50DA Folder
- 50DB Folder
- 50DC Folder
- 50DD Folder
- 50DE Folder
- 50DF Folder
- 50DG Folder
- 50DH Folder
- 50DI Folder
- 50DJ Folder
- 50DK Folder
- 50DL Folder
- 50DM Folder
- 50DN Folder
- 50DO Folder
- 50DP Folder
- 50DQ Folder
- 50DR Folder
- 50DS Folder
- 50DT Folder
- 50DU Folder
- 50DV Folder
- 50DW Folder
- 50DX Folder
- 50DY Folder
- 50DZ Folder
- 50EA Folder
- 50EB Folder
- 50EC Folder
- 50ED Folder
- 50EE Folder
- 50EF Folder
- 50EG Folder
- 50EH Folder
- 50EI Folder
- 50EJ Folder
- 50EK Folder
- 50EL Folder
- 50EM Folder
- 50EN Folder
- 50EO Folder
- 50EP Folder
- 50EQ Folder
- 50ER Folder
- 50ES Folder
- 50ET Folder
- 50EU Folder
- 50EV Folder
- 50EW Folder
- 50EX Folder
- 50EY Folder
- 50EZ Folder
- 50FA Folder
- 50FB Folder
- 50FC Folder
- 50FD Folder
- 50FE Folder
- 50FF Folder
- 50FG Folder
- 50FH Folder
- 50FI Folder
- 50FJ Folder
- 50FK Folder
- 50FL Folder
- 50FM Folder
- 50FN Folder
- 50FO Folder
- 50FP Folder
- 50FQ Folder
- 50FR Folder
- 50FS Folder
- 50FT Folder
- 50FU Folder
- 50FV Folder
- 50FW Folder
- 50FX Folder
- 50FY Folder
- 50FZ Folder
- 50GA Folder
- 50GB Folder
- 50GC Folder
- 50GD Folder
- 50GE Folder
- 50GF Folder
- 50GG Folder
- 50GH Folder
- 50GI Folder
- 50GJ Folder
- 50GK Folder
- 50GL Folder
- 50GM Folder
- 50GN Folder
- 50GO Folder
- 50GP Folder
- 50GQ Folder
- 50GR Folder
- 50GS Folder
- 50GT Folder
- 50GU Folder
- 50GV Folder
- 50GW Folder
- 50GX Folder
- 50GY Folder
- 50GZ Folder
- 50HA Folder
- 50HB Folder
- 50HC Folder
- 50HD Folder
- 50HE Folder
- 50HF Folder
- 50HG Folder
- 50HH Folder
- 50HI Folder
- 50HJ Folder
- 50HK Folder
- 50HL Folder
- 50HM Folder
- 50HN Folder
- 50HO Folder
- 50HP Folder
- 50HQ Folder
- 50HR Folder
- 50HS Folder
- 50HT Folder
- 50HU Folder
- 50HV Folder
- 50HW Folder
- 50HX Folder
- 50HY Folder
- 50HZ Folder
- 50IA Folder
- 50IB Folder
- 50IC Folder
- 50ID Folder
- 50IE Folder
- 50IF Folder
- 50IG Folder
- 50IH Folder
- 50II Folder
- 50IJ Folder
- 50IK Folder
- 50IL Folder
- 50IM Folder
- 50IN Folder
- 50IO Folder
- 50IP Folder
- 50IQ Folder
- 50IR Folder
- 50IS Folder
- 50IT Folder
- 50IU Folder
- 50IV Folder
- 50IW Folder
- 50IX Folder
- 50IY Folder
- 50IZ Folder
- 50JA Folder
- 50JB Folder
- 50JC Folder
- 50JD Folder
- 50JE Folder
- 50JF Folder
- 50JG Folder
- 50JH Folder
- 50JI Folder
- 50JJ Folder
- 50JK Folder
- 50JL Folder
- 50JM Folder
- 50JN Folder
- 50JO Folder
- 50JP Folder
- 50JQ Folder
- 50JR Folder
- 50JS Folder
- 50JT Folder
- 50JU Folder
- 50JV Folder
- 50JW Folder
- 50JX Folder
- 50JY Folder
- 50JZ Folder
- 50KA Folder
- 50KB Folder
- 50KC Folder
- 50KD Folder
- 50KE Folder
- 50KF Folder
- 50KG Folder
- 50KH Folder
- 50KI Folder
- 50KJ Folder
- 50KK Folder
- 50KL Folder
- 50KM Folder
- 50KN Folder
- 50KO Folder
- 50KP Folder
- 50KQ Folder
- 50KR Folder
- 50KS Folder
- 50KT Folder
- 50KU Folder
- 50KV Folder
- 50KW Folder
- 50KX Folder
- 50KY Folder
- 50KZ Folder
- 50LA Folder
- 50LB Folder
- 50LC Folder
- 50LD Folder
- 50LE Folder
- 50LF Folder
- 50LG Folder
- 50LH Folder
- 50LI Folder
- 50LJ Folder
- 50LK Folder
- 50LL Folder
- 50LM Folder
- 50LN Folder
- 50LO Folder
- 50LP Folder
- 50LQ Folder
- 50LR Folder
- 50LS Folder
- 50LT Folder
- 50LU Folder
- 50LV Folder
- 50LW Folder
- 50LX Folder
- 50LY Folder
- 50LZ Folder
- 50MA Folder
- 50MB Folder
- 50MC Folder
- 50MD Folder
- 50ME Folder
- 50MF Folder
- 50MG Folder
- 50MH Folder
- 50MI Folder
- 50MJ Folder
- 50MK Folder
- 50ML Folder
- 50MN Folder
- 50MO Folder
- 50MP Folder
- 50MQ Folder
- 50MR Folder
- 50MS Folder
- 50MT Folder
- 50MU Folder
- 50MV Folder
- 50MW Folder
- 50MX Folder
- 50MY Folder
- 50MZ Folder
- 50NA Folder
- 50NB Folder
- 50NC Folder
- 50ND Folder
- 50NE Folder
- 50NF Folder
- 50NG Folder
- 50NH Folder
- 50NI Folder
- 50NJ Folder
- 50NK Folder
- 50NL Folder
- 50NM Folder
- 50NO Folder
- 50NP Folder
- 50NQ Folder
- 50NR Folder
- 50NS Folder
- 50NT Folder
- 50NU Folder
- 50NV Folder
- 50NW Folder
- 50NX Folder
- 50NY Folder
- 50NZ Folder
- 50OA Folder
- 50OB Folder
- 50OC Folder
- 50OD Folder
- 50OE Folder
- 50OF Folder
- 50OG Folder
- 50OH Folder
- 50OI Folder
- 50OJ Folder
- 50OK Folder
- 50OL Folder
- 50OM Folder
- 50ON Folder
- 50OO Folder
- 50OP Folder
- 50OQ Folder
- 50OR Folder
- 50OS Folder
- 50OT Folder
- 50OU Folder
- 50OV Folder
- 50OW Folder
- 50OX Folder
- 50OY Folder
- 50OZ Folder
- 50PA Folder
- 50PB Folder
- 50PC Folder
- 50PD Folder
- 50PE Folder
- 50PF Folder
- 50PG Folder
- 50PH Folder
- 50PI Folder
- 50PJ Folder
- 50PK Folder
- 50PL Folder
- 50PM Folder
- 50PN Folder
- 50PO Folder
- 50PP Folder
- 50PQ Folder
- 50PR Folder
- 50PS Folder
- 50PT Folder
- 50PU Folder
- 50PV Folder
- 50PW Folder
- 50PX Folder
- 50PY Folder
- 50PZ Folder
- 50QA Folder
- 50QB Folder
- 50QC Folder
- 50QD Folder
- 50QE Folder
- 50QF Folder
- 50QG Folder
- 50QH Folder
- 50QI Folder
- 50QJ Folder
- 50QK Folder
- 50QL Folder
- 50QM Folder
- 50QN Folder
- 50QO Folder
- 50QP Folder
- 50QQ Folder
- 50QR Folder
- 50QS Folder
- 50QT Folder
- 50QU Folder
- 50QV Folder
- 50QW Folder
- 50QX Folder
- 50QY Folder
- 50QZ Folder
- 50RA Folder
- 50RB Folder
- 50RC Folder
- 50RD Folder
- 50RE Folder
- 50RF Folder
- 50RG Folder
- 50RH Folder
- 50RI Folder
- 50RJ Folder
- 50RK Folder
- 50RL Folder
- 50RM Folder
- 50RN Folder
- 50RO Folder
- 50RP Folder
- 50RQ Folder
- 50RR Folder
- 50RS Folder
- 50RT Folder
- 50RU Folder
- 50RV Folder
- 50RW Folder
- 50RX Folder
- 50RY Folder
- 50RZ Folder
- 50SA Folder
- 50SB Folder
- 50SC Folder
- 50SD Folder
- 50SE Folder
- 50SF Folder
- 50SG Folder
- 50SH Folder
- 50SI Folder
- 50SJ Folder
- 50SK Folder
- 50SL Folder
- 50SM Folder
- 50SN Folder
- 50SO Folder
- 50SP Folder
- 50SQ Folder
- 50SR Folder
- 50SS Folder
- 50ST Folder
- 50SU Folder
- 50SV Folder
- 50SW Folder
- 50SX Folder
- 50SY Folder
- 50SZ Folder
- 50TA Folder
- 50TB Folder
- 50TC Folder
- 50TD Folder
- 50TE Folder
- 50TF Folder
- 50TG Folder
- 50TH Folder
- 50TI Folder
- 50TJ Folder
- 50TK Folder
- 50TL Folder
- 50TM Folder
- 50TN Folder
- 50TO Folder
- 50TP Folder
- 50TQ Folder
- 50TR Folder
- 50TS Folder
- 50TT Folder
- 50TU Folder
- 50TV Folder
- 50TW Folder
- 50TX Folder
- 50TY Folder
- 50TZ Folder
- 50UA Folder
- 50UB Folder
- 50UC Folder
- 50UD Folder
- 50UE Folder
- 50UF Folder
- 50UG Folder
- 50UH Folder
- 50UI Folder
- 50UJ Folder
- 50UK Folder
- 50UL Folder
- 50UM Folder
- 50UN Folder
- 50UO Folder
- 50UP Folder
- 50UQ Folder
- 50UR Folder
- 50US Folder
- 50UT Folder
- 50UU Folder
- 50UV Folder
- 50UW Folder
- 50UX Folder
- 50UY Folder
- 50UZ Folder
- 50VA Folder
- 50VB Folder
- 50VC Folder
- 50VD Folder
- 50VE Folder
- 50VF Folder
- 50VG Folder
- 50VH Folder
- 50VI Folder
- 50VJ Folder
- 50VK Folder
- 50VL Folder
- 50VM Folder
- 50VN Folder
- 50VO Folder
- 50VP Folder
- 50VQ Folder
- 50VR Folder
- 50VS Folder
- 50VT Folder
- 50VU Folder
- 50VV Folder
- 50VW Folder
- 50VX Folder
- 50VY Folder
- 50VZ Folder
- 50WA Folder
- 50WB Folder
- 50WC Folder
- 50WD Folder
- 50WE Folder
- 50WF Folder
- 50WG Folder
- 50WH Folder
- 50WI Folder
- 50WJ Folder
- 50WK Folder
- 50WL Folder
- 50WM Folder
- 50WN Folder
- 50WO Folder
- 50WP Folder
- 50WQ Folder
- 50WR Folder
- 50WS Folder
- 50WT Folder
- 50WU Folder
- 50WV Folder
- 50WW Folder
- 50WX Folder
- 50WY Folder
- 50WZ Folder
- 50XA Folder
- 50XB Folder
- 50XC Folder
- 50XD Folder
- 50XE Folder
- 50XF Folder
- 50XG Folder
- 50XH Folder
- 50XI Folder
- 50XJ Folder
- 50XK Folder
- 50XL Folder
- 50XM Folder
- 50XN Folder
- 50XO Folder
- 50XP Folder
- 50XQ Folder
- 50XR Folder
- 50XS Folder
- 50XT Folder
- 50XU Folder
- 50XV Folder
- 50XW Folder
- 50XX Folder
- 50XY Folder
- 50XZ Folder
- 50YA Folder
- 50YB Folder
- 50YC Folder
- 50YD Folder
- 50YE Folder
- 50YF Folder
- 50YG Folder
- 50YH Folder
- 50YI Folder
- 50YJ Folder
- 50YK Folder
- 50YL Folder
- 50YM Folder
- 50YN Folder
- 50YO Folder
- 50YP Folder
- 50YQ Folder
- 50YR Folder
- 50YS Folder
- 50YT Folder
- 50YU Folder
- 50YV Folder
- 50YW Folder
- 50YX Folder
- 50YY Folder
- 50YZ Folder
- 50ZA Folder
- 50ZB Folder
- 50ZC Folder
- 50ZD Folder
- 50ZE Folder
- 50ZF Folder
- 50ZG Folder
- 50ZH Folder
- 50ZI Folder
- 50ZJ Folder
- 50ZK Folder
- 50ZL Folder
- 50ZM Folder
- 50ZN Folder
- 50ZO Folder
- 50ZP Folder
- 50ZQ Folder
- 50ZR Folder
- 50ZS Folder
- 50ZT Folder
- 50ZU Folder
- 50ZV Folder
- 50ZW Folder
- 50ZX Folder
- 50ZY Folder
- 50ZZ Folder

Label This Transfer

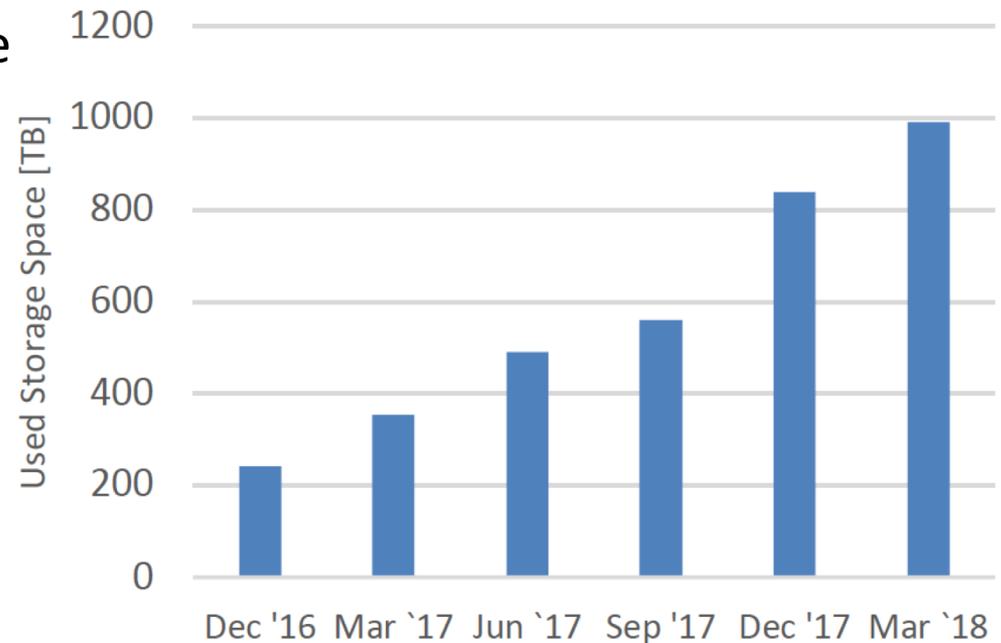
Transfer Settings

- sync - only transfer new or changed files
- delete files on destination that do not exist on source
- preserve source file modification times
- verify file integrity after transfer
- encrypt transfer

# System Usage

- December 2016: 5 beamline deployments, 250TB of storage space used
- March 2018: 21 beamline deployments, close to 1PB of storage space used
- Three month average growth of storage space used: 75TB/month

## APS Storage Usage



- We are currently using more than 75% of available storage space
- Assuming 75TB/month increase, we have 3-4 months before we run out of space



# System Usage: Processing @ 8-ID-I (S. Narayanan)

- 8IDI uses X-ray Photon Correlation Spectroscopy technique (XPCS) for the studies of equilibrium fluctuations and fluctuations about the evolution to equilibrium in condensed matter in the Small-Angle X-ray Scattering (SAXS) geometry
- SPEC software is used for instrument control and data acquisition
- For every raw data file SPEC scripts can start DM processing job based on one of the implemented workflows
- Batch processing workflow:
  - 1) Run a custom shell script to prepare processing environment.
  - 2) Copy raw data file to APS HPC cluster using GridFTP (the *globus-url-copy* command).
  - 3) Append XPCS metadata to the data file by running a custom 8-ID-I utility.
  - 4) Submit a processing job to the SGE batch scheduler via the *qsub* command. This job runs a custom 8-ID-I processing executable.
  - 5) Monitor batch job by running a shell script that interacts with SGE via the *qacct* command.
  - 6) Copy resulting output file into designated beamline storage area using GridFTP (the *globus-url-copy* command).
- Jobs are monitored via static web pages generated by a cron job running DM utilities



# Future Plans

- Develop Data Acquisition Service plugins that handle integration with external cataloging and data publishing systems (DOE Data Explorer, Materials Data Facility)
- Storage hardware replacement (purchase approved recently)
- Enable beamline managers to organize their experiment data in storage in a manner that best fits their beamline
- Further develop functionality offered in the DM Station GUI:
  - Improve file metadata and collection views
  - Add workflow and processing job management capabilities
- Enhance DM system monitoring infrastructure:
  - Develop service capabilities for self-diagnosing error or warning conditions and issuing alarms.
  - Improve support for measuring performance (e.g., data transfer rates, file processing rates, etc.)
- Further develop beamline management functionality available in the DM Web Portal
- Develop standard set of workflow definitions that can be reused on different beamlines for automating processing pipelines (need use cases)
- Develop policy engine for automated management of experiment data in storage, archiving of old data, etc.
- Improve documentation, write user guide
- New beamline deployments?
- Archival system?



# Conclusions

- The DM system has grown significantly over the last couple of years, in terms of both its usage and capabilities
- The software can be customized and extended to serve individual beamline needs related to data management
- In particular, it can be used to fully automate data acquisition and processing pipelines

# Acknowledgements

- APS IT group (R. Sersted and D. Wallis) for their work on building and maintaining the APS On-site Storage and virtual machines used to host beamline DM services
- APS IS group (F. Lacap and Y. Huang) for their help and cooperation with accessing APS databases
- APS SDM group (B. Frosik and A. Glowacki) for many useful discussions
- APS 1-ID (J.-S. Park and P. Kenesei) and 8-ID-I (S. Narayanan) beamlines for their patience, support and help with system testing and troubleshooting since the early prototype versions.
- All DM users



# Additional Slides



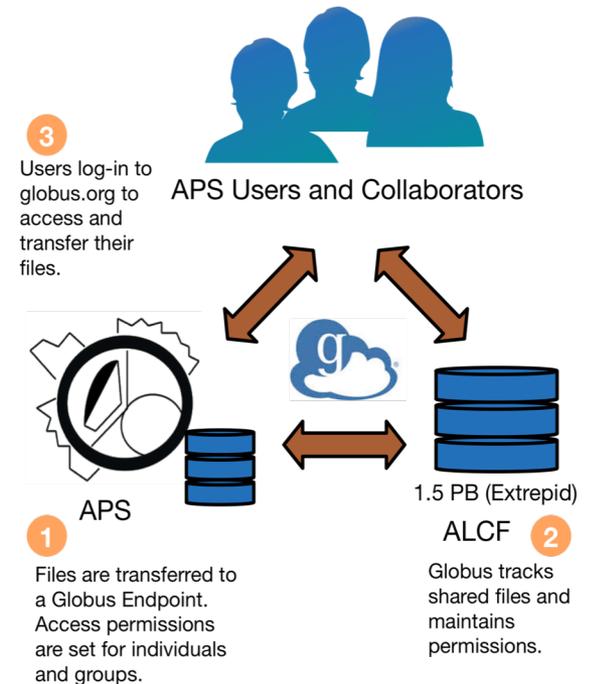
# Data Management Challenges?

- How do we organize data?
- Where do we store data?
- How do we store data?
- Do we store all data?
- How do we find data?
- How do we access data?
- How do we manage stored data?
- How can we ensure data integrity?
- Can we automate data acquisition and processing pipelines?
- What can we do about user data analysis and processing?
- ...



# A Bit Of History

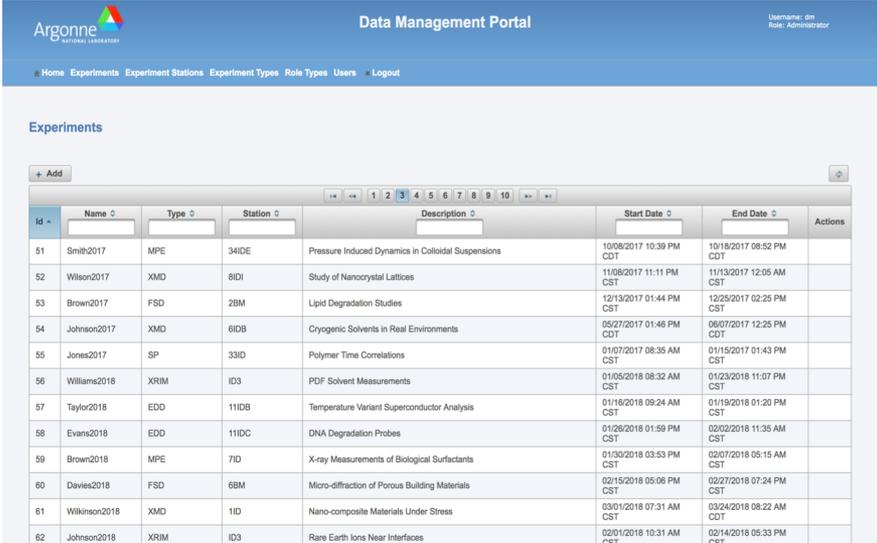
- 2013: Tao Fusion project (LDRD) acquired XSTOR storage (250 TB)
- September 2014: APS Data Management project started
  - Goal: provide APS users with means to easily access their data remotely using Globus Online
- October 2015: First successful software deployment at 6IDD (D. Robinson)



- January 2017: Transition to DDN storage (1.5PB) with high-performance GPFS file system, data redundancy, and 2x10Gbps network links
- March 2018: New VM cluster used exclusively for DM virtual machines
  - Total of 512GB RAM, 144 CPU cores (72x2 due to hyper-threading), 2x10Gbps network links

# Site Services

- DM Database (PostgreSQL)
  - Maintains information about users, experiments, and beamline deployments
- Storage Management Service (Python, CherryPy, SQLAlchemy)
  - Runs on the storage head node
  - Provides experiment management services (via REST interfaces)
  - Interacts with LDAP and APS Databases
  - Controls storage file system permissions, which enables data access for remote users
- Web Portal (Java EE Application, Glassfish, JPA, JSF, Primefaces)
  - Experiment management
  - Support for beamline deployments
- Automated utilities for synchronizing DM user information with APS User Database



The screenshot shows the Argonne National Laboratory Data Management Portal. The page title is "Data Management Portal" and the user is logged in as "dm Administrator". The main content area is titled "Experiments" and contains a table with the following data:

Id	Name	Type	Station	Description	Start Date	End Date	Actions
51	Smith2017	MPE	34IDE	Pressure Induced Dynamics in Colloidal Suspensions	10/08/2017 10:39 PM CDT	10/18/2017 08:52 PM CDT	
52	Wilson2017	XMD	8IDI	Study of Nanocrystal Lattices	11/06/2017 11:11 PM CST	11/13/2017 12:05 AM CST	
53	Brown2017	FSD	28M	Lipid Degradation Studies	12/13/2017 01:44 PM CST	12/25/2017 02:25 PM CST	
54	Johnson2017	XMD	6IDB	Cryogenic Solvents in Real Environments	05/27/2017 01:48 PM CDT	06/07/2017 12:25 PM CDT	
55	Jones2017	SP	33ID	Polymer Time Correlations	01/07/2017 08:35 AM CST	01/15/2017 01:43 PM CST	
56	Williams2018	XRIM	ID3	PDF Solvent Measurements	01/05/2018 08:32 AM CST	01/23/2018 11:07 PM CST	
57	Taylor2018	EDD	11IDB	Temperature Variant Superconductor Analysis	01/18/2018 09:24 AM CST	01/19/2018 01:20 PM CST	
58	Evans2018	EDD	11IDC	DNA Degradation Probes	01/26/2018 01:59 PM CST	02/02/2018 11:35 AM CST	
59	Brown2018	MPE	7ID	X-ray Measurements of Biological Surfactants	01/30/2018 03:53 PM CST	02/07/2018 05:15 AM CST	
60	Davies2018	FSD	68M	Micro-diffraction of Porous Building Materials	02/15/2018 05:06 PM CST	02/27/2018 07:24 PM CST	
61	Wilkinson2018	XMD	1ID	Nano-composite Materials Under Stress	03/01/2018 07:31 AM CST	03/24/2018 08:22 AM CDT	
62	Johnson2018	XRIM	ID3	Rare Earth Ions Near Interfaces	02/01/2018 10:31 AM CST	02/14/2018 05:33 PM CST	

# Station Services

- Each beamline deployment (“DM Station”) includes several Python services accessible via REST interfaces: DAQ Service, Metadata Catalog and Processing Service
- Data Acquisition Service
  - Responsible for data uploads and for monitoring local file storage
  - Customizable, plugin-based processing framework
  - Plugins handle file transfers, metadata cataloging, interaction with other services
- Metadata Catalog (MongoDB)
  - Metadata are arbitrary key/value pairs
  - Each experiment has its own file metadata collection
  - File metadata can be retrieved using command line or API tools, DM Station GUI, or via the Mongo Express application

```
sveseli — ssh bluegill2 — 100x32
dmadmin@sXiddm> dm-get-file-collection-stats --experiment=Smith-Mar18 --display-keys=ALL --display-format=key-per-line
experimentName=Smith-Mar18
stdDevFileSize=21275858.208
queryDict={}
maxFileSize=251666432
minFileSize=14
collectionSize=1.63133244728e+12
averageFileSize=9634554.76451
nFiles=169321

dmadmin@sXiddm> dm-list-experiment-files --experiment=Smith-Mar18 --display-keys=experimentFilePath,
experimentFileSum
experimentFilePath=pix/macing_000001.tif fileSize=823569 md5Sum=9a87108d6d0bb3bdd16d58b0c7cd6855
experimentFilePath=pix/macing_000002.tif fileSize=823569 md5Sum=7bef341f3625b6db350e4d7060394ca5
experimentFilePath=pix/macing_000003.tif fileSize=823569 md5Sum=ad7399871ce9b4f6dab637d64b2be2fe
experimentFilePath=pix/macing_000004.tif fileSize=823569 md5Sum=d74b11f0438761209e81ac72525cad0
experimentFilePath=pix/macing_000005.tif fileSize=823569 md5Sum=309e7402992ff870f1d0bd7445eb2c39
experimentFilePath=pix/macing_000006.tif fileSize=823569 md5Sum=53e16027b5c2fe3d4ff640c2a847a4ff
experimentFilePath=pix/macing_000007.tif fileSize=823569 md5Sum=5b4526970e63a8ea035362309b134dc
experimentFilePath=pix/macing_000008.tif fileSize=823569 md5Sum=3b503018bcae697f7b9cf35a0c7b9120
experimentFilePath=pix/macing_000009.tif fileSize=823569 md5Sum=7d27d4842c7f312742ae6e67c44d284e
experimentFilePath=pix/macing_000010.tif fileSize=823569 md5Sum=b0671eeb63507b5668c8a2ff914e8281
experimentFilePath=pix/macing_000011.tif fileSize=823569 md5Sum=66fa93dab6db96bb5f4afd78c6b5ee9
experimentFilePath=pix/macing_000012.tif fileSize=823569 md5Sum=05267d7a2184e69affbfb3aee6638b
experimentFilePath=pix/macing_000013.tif fileSize=823569 md5Sum=08cf9cd278be3c76cb2dc17176b9a95b
experimentFilePath=pix/macing_000014.tif fileSize=823569 md5Sum=bf73fcf5b7d2e2304422a43ab17ad418
experimentFilePath=pix/macing_000015.tif fileSize=823569 md5Sum=b11a0f65383b795854b5cfea2d79a5af
experimentFilePath=pix/macing_000016.tif fileSize=823569 md5Sum=9bda2ac5161c825d94298d50b0cb817e
dmadmin@s1iddm> |
```

# Station Services

- Processing Service provides support for managing user-defined workflows, as well as for submitting and monitoring processing jobs based on those workflows

- DM workflow is a collection of processing steps executed in order:

- Workflow definitions are described as Python dictionaries
- Each processing step must be associated with an (arbitrary) executable
- Support for input/output variables
- Processing steps are automatically parallelized if possible

```
{
  'name': 'tomo-demo-01',
  'description': 'Tomo Demo 01 Workflow',
  'owner': 'pelmer',
  'stages':
  {
    '1-Startup':
    {
      'command': '/home/dm/workflows/tomo-01/start-tomo-01.sh $filePath $outputRootDir',
      'outputVariableRegexList': [
        'Created processing output dir: (?P<outputDir>.*)',
        'File number: (?P<fileNumber>.*)'
      ]
    },
    '2-Process':
    {
      'command': 'ssh txmtwo "source /home/dm/etc/dm.setup.sh > /dev/null;
        /home/dm/workflows/tomo-01/tomo-01.py $filePath $outputDir"'
    },
    '3-Upload':
    {
      'command': 'source /home/dm/etc/dm.setup.sh > /dev/null;
        dm-32id-upload --experiment=$experimentName --data-directory=$outputDir
        --dest-directory=processed/$fileNumber parentFile:$experimentFilePath'
    },
  },
}
```

- Processing Service can be used either standalone, or together with other DM Station services in support of fully automated beamline data acquisition and processing pipelines



# Software Installation

- Each beamline at APS has its own Data Management software installation visible to all beamline machines
- Deployment area contains DM software, support software packages, beamline databases, configuration files, various runtime and log files
- All beamlines are fully independent of each other, which works well in the APS environment (beamline machines typically have different maintenance cycles)
- The DM software has fully scripted installation, upgrade, and deployment testing processes, which reduces to a minimum maintenance overhead due to independent beamline software installations
- The DM services typically run on a designated beamline server machine, which can be either virtual or physical (VMs are preferred)
- Services are controlled via a standard set of control scripts suitable for the RHEL operating system used at APS
- Typically, beamline staff has no involvement with DM software installation and maintenance



# Support

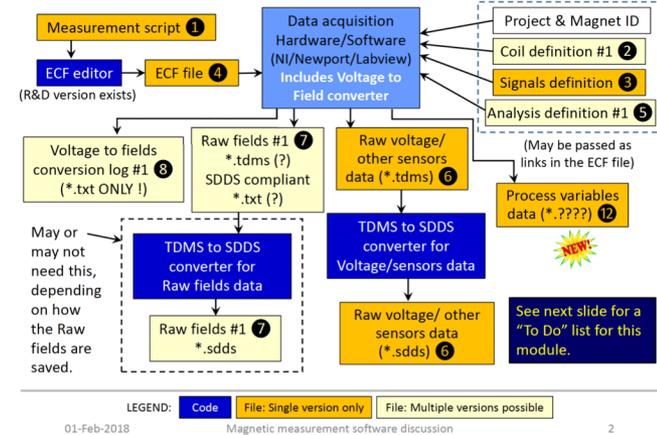
- Official “APS Data Distribution System” support policy:  
***“Although no guarantees are made as to the system’s availability, problems are addressed as soon as possible on a best-effort basis.”***
- Two email lists:
  - DM Users mailing list (dm-users@aps.anl.gov) is used for announcing new software releases, system outages, etc.
  - DM Admin mailing list (dm-admin@aps.anl.gov) can be used for system inquiries



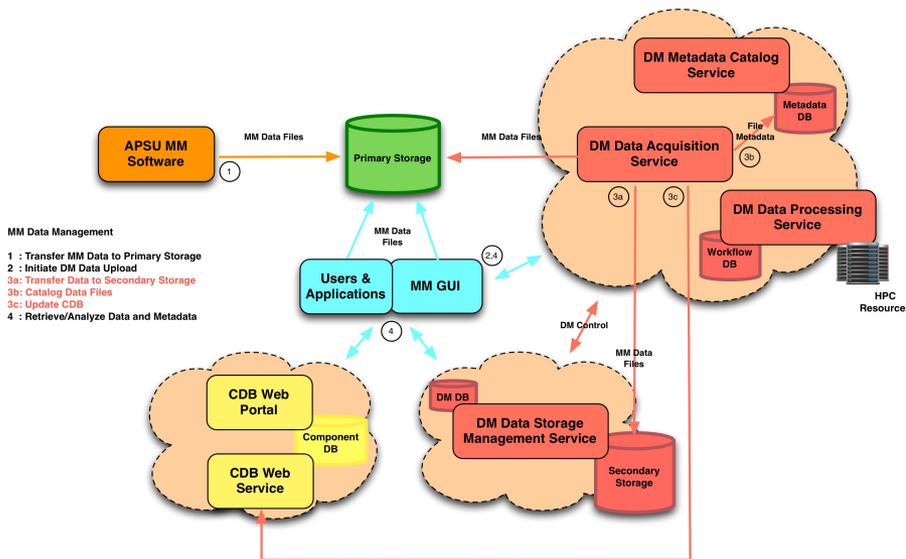
# System Usage: APSU MM Data (A. Jain)

- APSU uses Component Database (CDB) for identifying and tracking components needed to build the new facility
- APSU Magnetic Measurements (MM) software works with and creates many different types of files: various definition/configuration files, measurement and analysis files, scripts, raw measurement data, PV data, processed data, log files,...
- Over 1K magnets, each will require multiple measurements => MM software will generate large amounts of data and numerous data files

## Overview (Rev 1): Data Acquisition Module



## APSU MM Data Management



- DM data upload tools will associate MM experiment with corresponding magnet QR ID in CDB
- During the DM data upload, each MM data file will be processed as follows:
  - Metadata plugin stores file metadata in the APSU DM metadata database
  - CDB plugin adds file item to CDB, and also adds file metadata as item's property
- After the DM data upload, MM experiment item views on the CDB web portal contain links to corresponding magnet, allow downloading experiment files, etc. (CDB work: D. Jarosz)