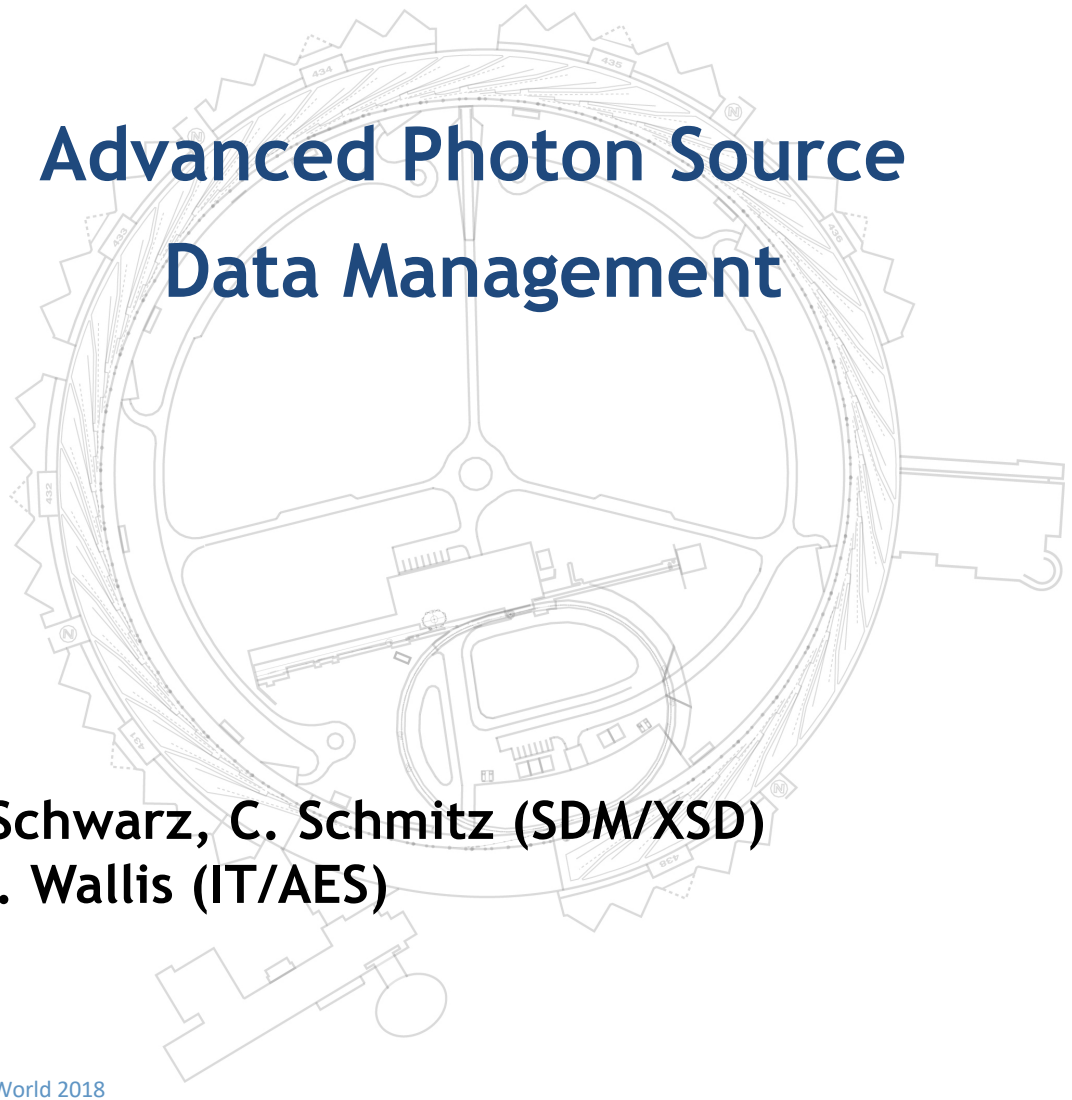
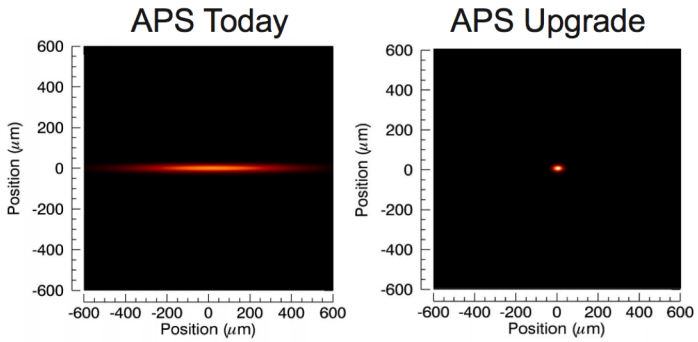
Advanced Photon Source Data Management



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

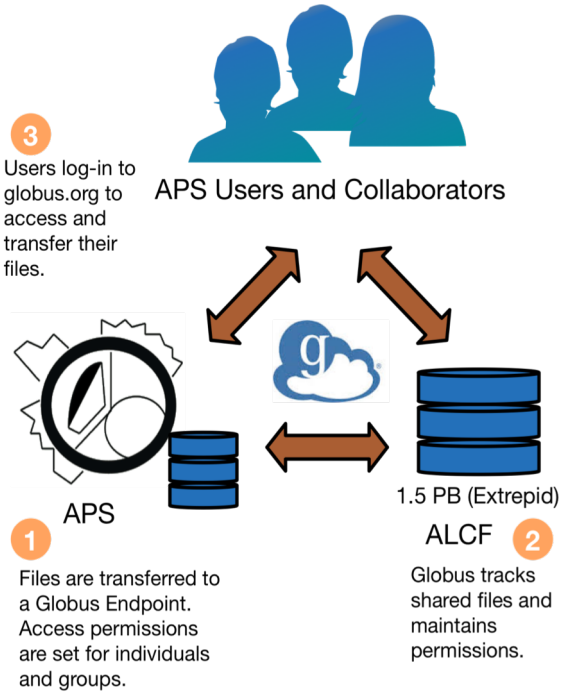


- APS 1ID:
  - Today: 4x Hydra GE detector, 8MB frame, 8 fps => 256MB/s data rate (1TB/hr)
  - Production Data Rates (including overhead, on a good week): 10TB/day, 60TB/week
  - Near future (1-3 years): 2x 2923 Dexela (or similar), 23MB frame, 26 fps => 1.2GB/s data rate
  - Near future: Pilatus 2M (or similar/larger), 9.5MB frame, 250 fps => 2.3GB/s data rate

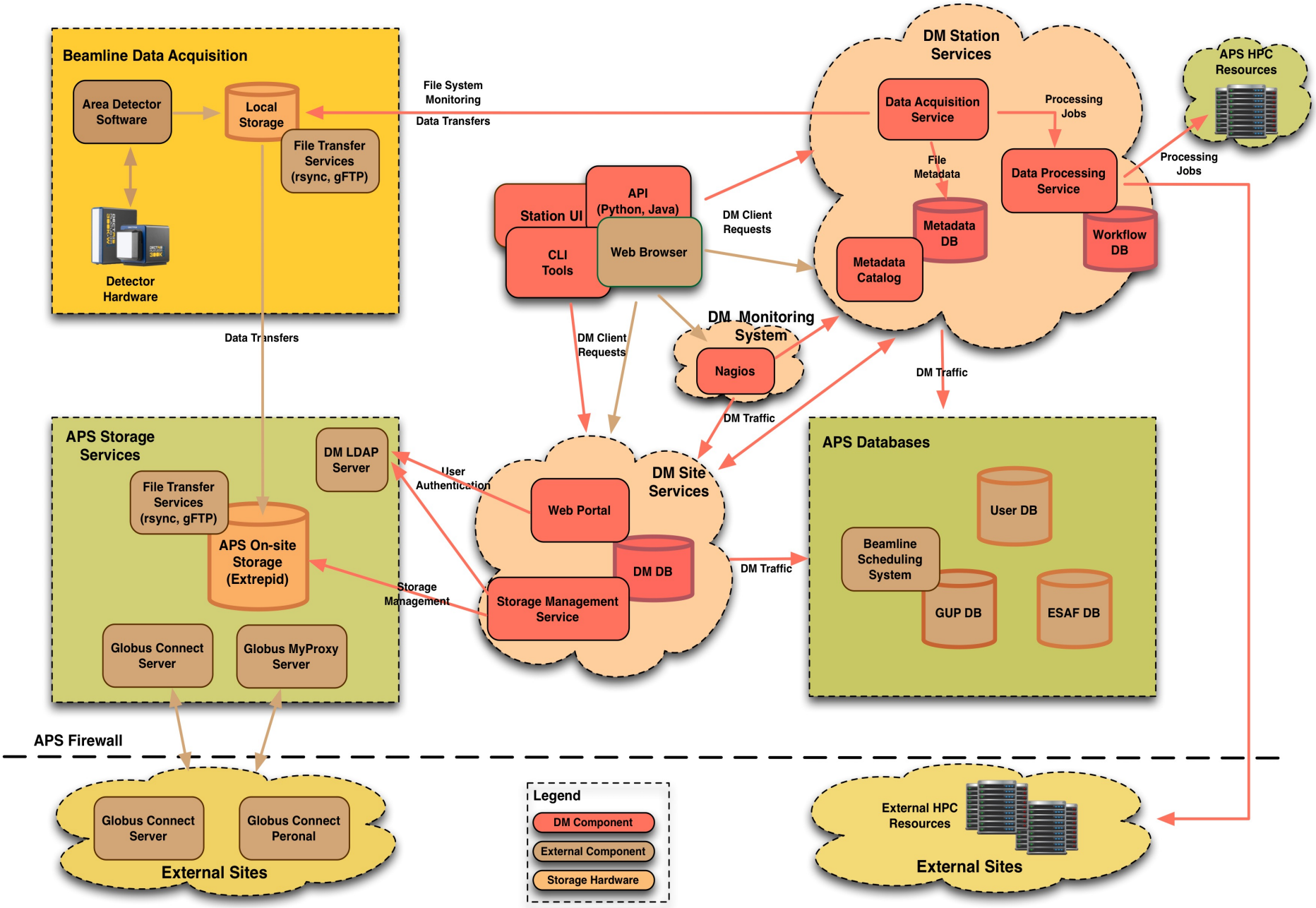
- APS 8IDI:
  - 2010-2016: ANL/LBL FCCD, 2MB frame, 100fps, compressed in real-time with 10% non-zero pixels on the average => about 200MB/s data rate
  - 2016-Today: Lambda 750K, 1.5MB frame, about 10% non-zero pixels, 2000 fps => about 300 MB/s data rate
  - Production Data Rates: 8-10 TB/week (max), 1-2TB/week (average)
  - Future: research on VIPIC (>1MHz frame rate) and UFXC (50 kHz frame rate)

APS Data Management

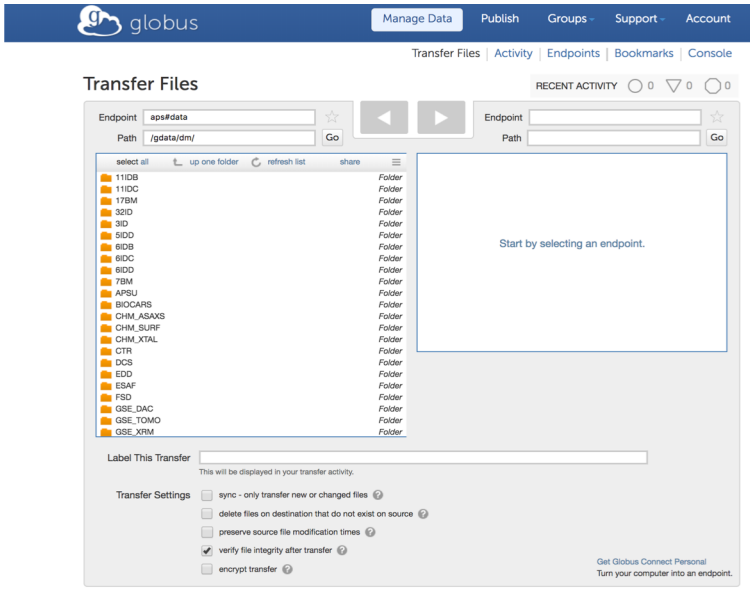
- Specific data management needs typically vary from beamline-to-beamline, mostly depending upon the types of detectors, X-ray techniques, and data processing tools in use
- However, most of the data management requirements are related to a set of tasks common to all beamlines:
  - Storage area management (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



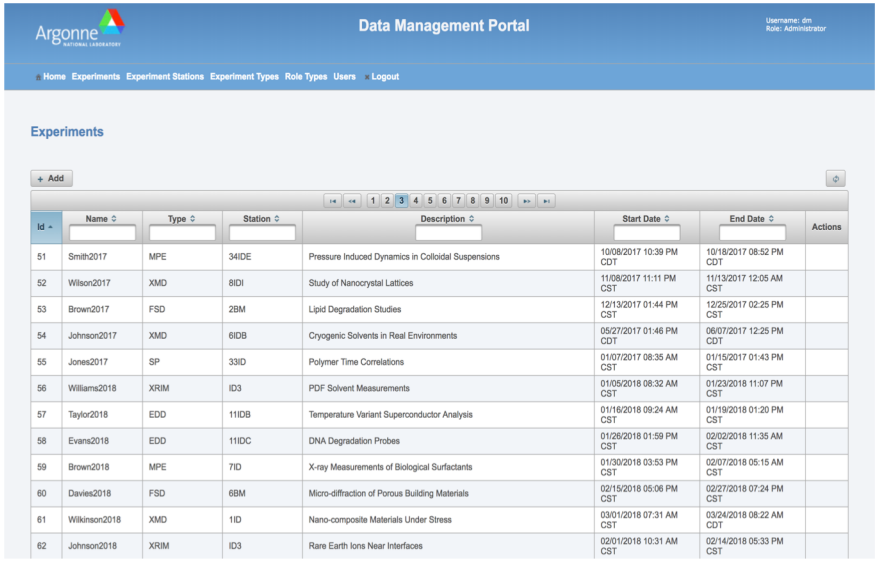
How Does Globus Fit In?



- 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

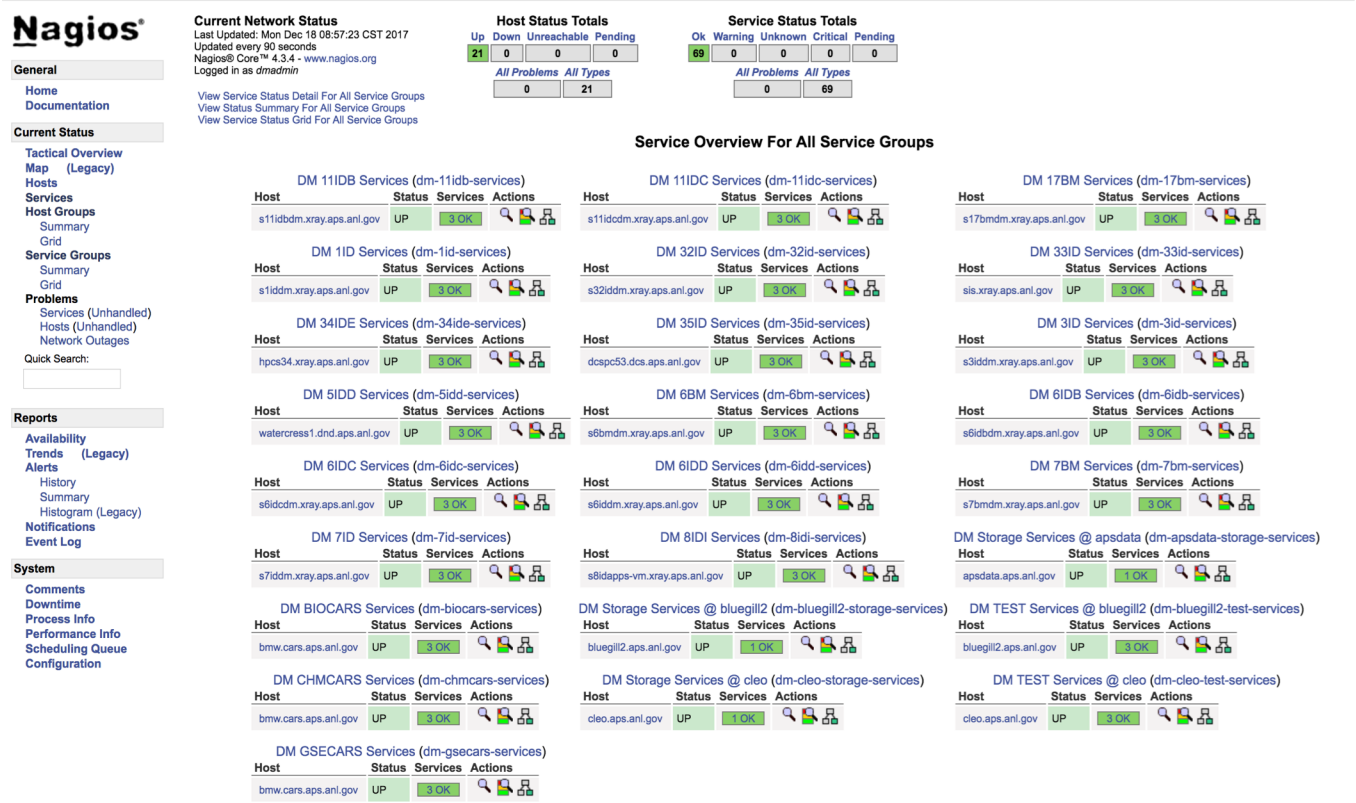
- Site Services:
  - DM Database (PostgreSQL)
  - Storage Management Service
  - Web Portal
  - Automated user account synchronization utilities



- Station (Beamline Deployment) Services:
  - Data Acquisition Service
  - Metadata Catalog
  - Processing Service
- All services are available via REST interfaces

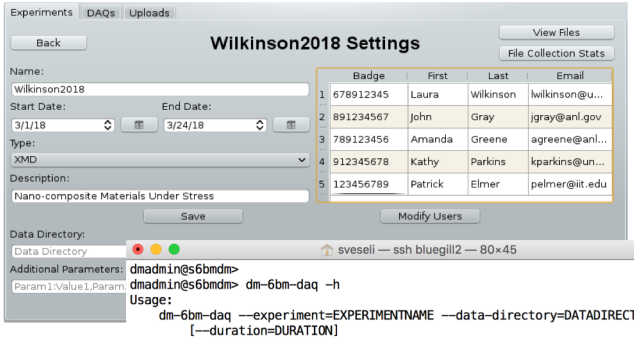
Monitoring

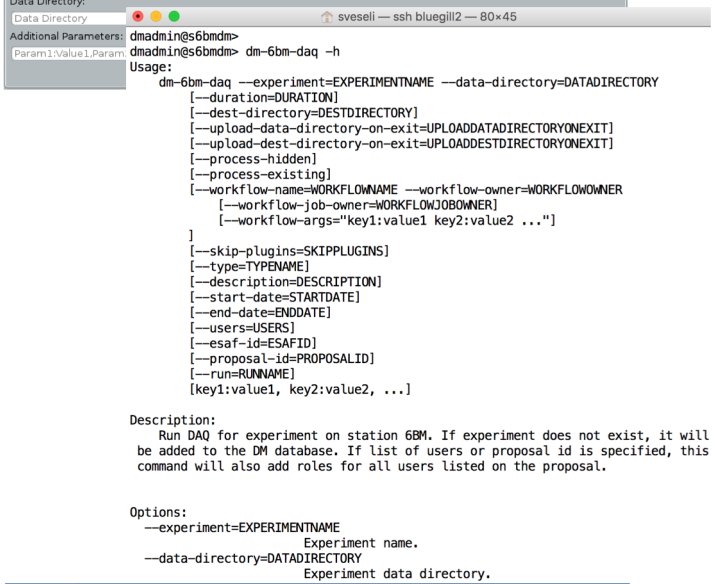
- Every DM service has a set of monitoring interfaces that enable external applications to find out about its state
- These are used by the custom Nagios plugins that provide up-to-date information about the health of the DM station deployments

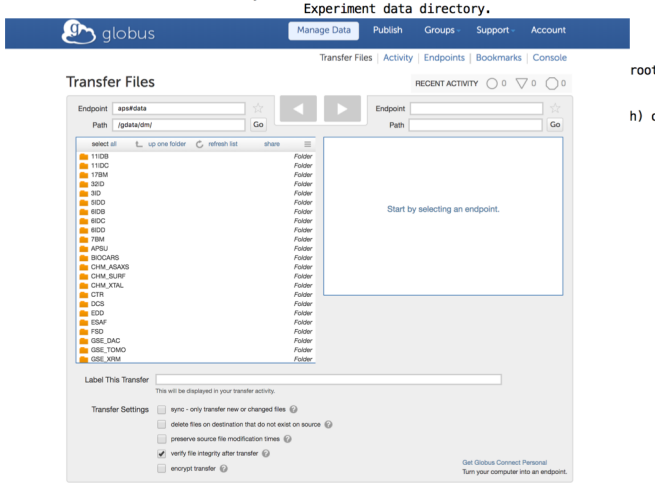


User Interfaces

- 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 0MQ 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

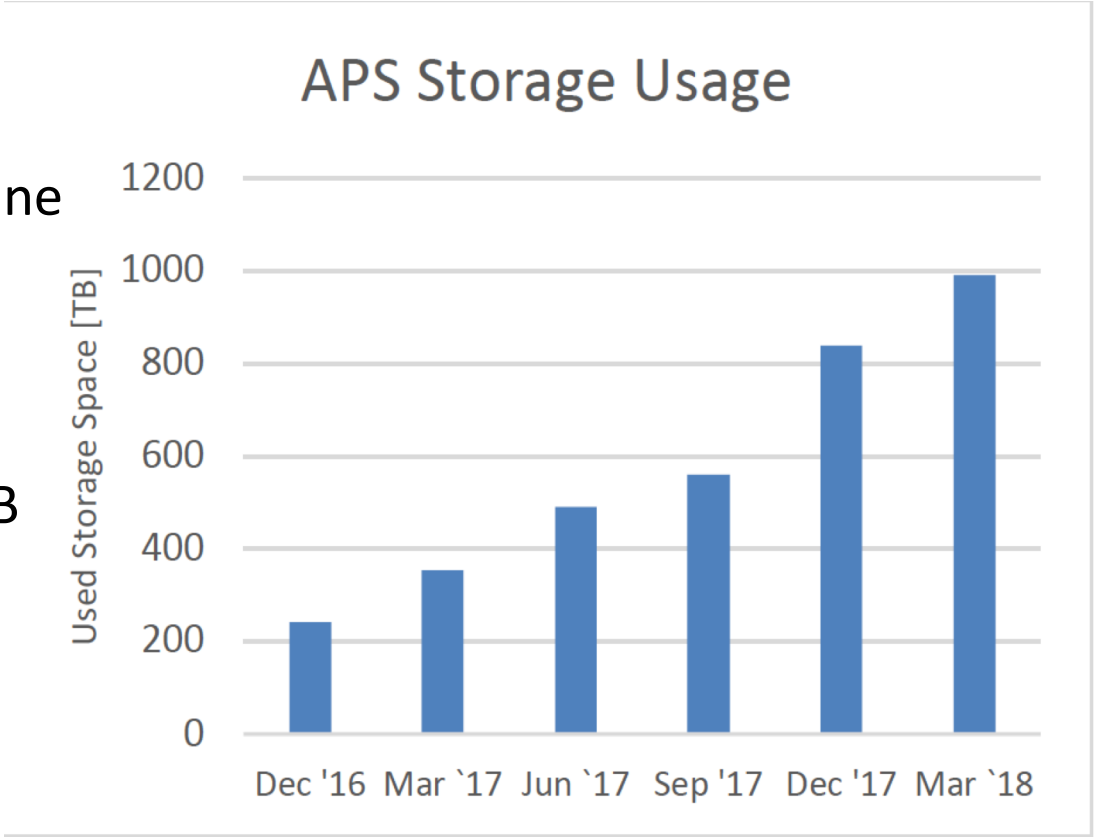






System Usage

- 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



- 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

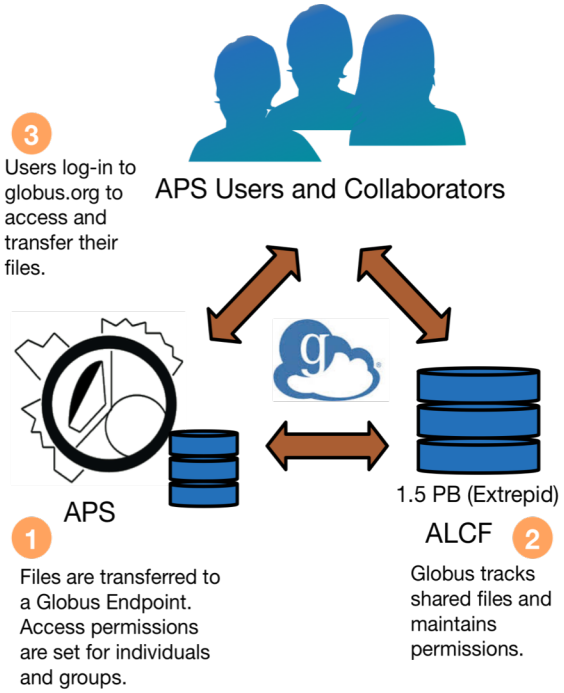
Additional Slides

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

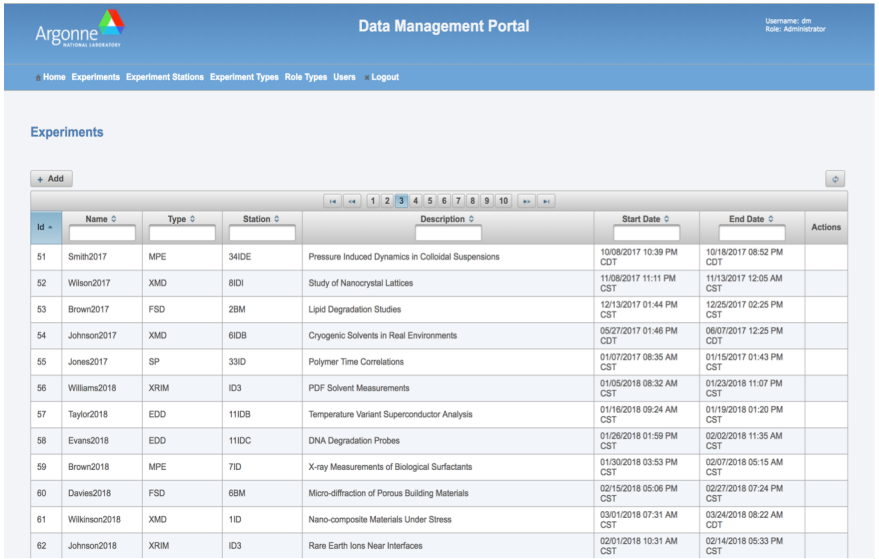
- 2013: Tao Fusion project (LDRD) acquired XSTOR storage (250 TB)
- September 2014: APS Data Management project started
  - Goal: provide APS users with means to easily access their data remotely using Globus Online
- October 2015: First successful software deployment at 6IDD (D. Robinson)



- 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



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

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

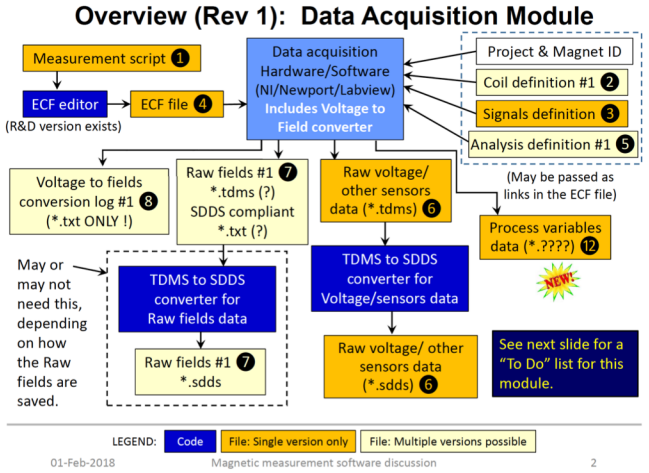
- Official “APS Data Distribution System” support policy:

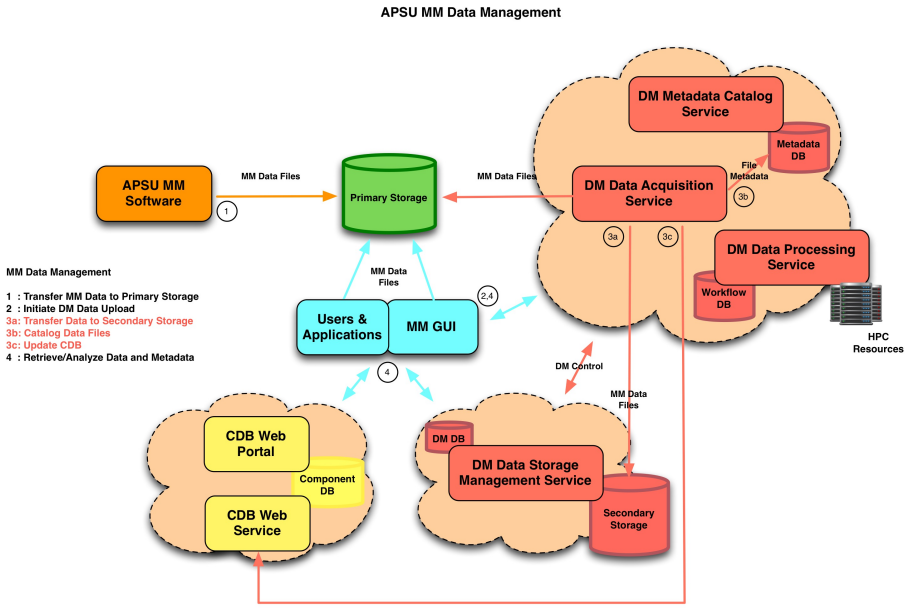
  “Although no guarantees are made as to the system’s availability, problems are addressed as soon as possible on a best-effort basis.”

- 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





- 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)