Data Management and Software Centre

Mark Hagen
Head of DMSC

Mark.Hagen@esss.se
www.europeanspallationsource.se
Lithuanian Academy of Sciences, March 4th 2014
What is DMSC?

- Data Management and Software Centre (DMSC)

- A Division of ESS Science Directorate...
  - Just like Neutron Technologies, Neutron Instruments etc.
  - Two campuses: ESS Lund & ESS Copenhagen
    (Universitetsparken, Københavns Universitet)
  - DMSC building to be constructed in Copenhagen

- Responsibility: design, develop & implement for the ESS instruments:
  - Software (user control interfaces, data acquisition, reduction & analysis)
  - Hardware (servers, networks, workstations, clusters, disks, pfs etc.)
What is the ESS?

Proton Accelerator
- Energy: 2.5 GeV
- Frequency: 14 Hz
- Current: 50 mA

Target Station
- Solid Rotating W
- He or Water Cooled
- 5MW average power
- >22 beam ports

Instruments
- 22 Instruments in construction budget

5 times brighter than SNS
30 times brighter than ILL
What will DMSC do for ESS?

- DMSC is involved.... *From idea to publication*
- Software and servers for the user proposals & the review
- Coordination (scheduling) of experiments and equipment
- The operation of the instruments/experiments...
  - Controlling the components of the instrument
  - Acquiring the data from detectors & meta-data from instrument/samp. env. etc.
  - Streaming the data (publish/subscribe)
  - Recording/archiving/cataloguing the data
  - Carrying out the data reduction
- Software for Data Analysis – Modeling & Simulation
- Finally a Publications Database – track the results from ESS
DMSC Organization

DMSC

Data Systems & Technologies
- Copenhagen Data Centre
- DMSC servers in Lund
- Clusters, Workstations
- Disks, Parallel File System
- Networks (inc. Lund – CPH)
- Data transfer & Back-Up
- External Servers

Inst. Control, Data Acq. & Reduction
- Instrument Control User Interfaces
- EPICS read/write
- Streaming data (ADARA)
- Data reduction (MANTID)

Data Management
- File writers (ADARA)
- Data Catalogues
- Workflow Management
- Post-Processing
- ---- Reduction
- ---- Analysis
- Messaging Services
- Web Interfaces

Data Analysis & Modeling
- MCSTAS support + dev.
- Instrument Integrators
- Analysis codes (e.g. SANSview, Rietveld,…)
- MD + DFT Framework

User Office Software
- User Database
- Proposal System
- Training Database
- Publications Database

Initially one work package (2 work units)
General Framework + Customization

**Generic Data Framework**
- Event mode data for later reprocessing/filtering
- Stream the data ➔ on the fly data processing (live view)
  ➔ on the fly file creation
  ➔ to a location where appropriate post-acq. resources are available
- Create standard HDF5 data files (NeXus or other)
- (Where possible) Automate data reduction & analysis
- Catalogue the data & meta-data for fast processing

**Don’t re-invent the wheel**
➔ Existing projects: ICAT, MANTID, ADARA – Join these

**Customize for ESS Instruments**
- User Control Interface, detector type, IOC’s for sample environment etc.
Data Acquisition, Reduction & Control

**Generic Framework**
- Instrument Control
- Data Acquisition
- Data Reduction & Visualization

**INTEGRATED CONTROL SYSTEMS**
- EPICS
- Timing
- Control Box

**NEUTRON TECHNOLOGIES**
- Choppers
- Motion Control
- Sample Environment
- Fast Sample Environment
- Detectors & Monitors
- Bulk Data Readout (Detector Group)

**DATA MANAGEMENT & SOFTWARE CENTRE**
- Data Aggregator & Streamer
  - (ADARA)
  - Automated Data Reduction

**Instrument Control Room**
- User Control Interface
- Data Analysis Interfaces

**Lund Server Room**
- Automated Data Reduction

**Copenhagen Server Room**
- Automated Data Reduction
Data Acquisition, Streaming & Reduction

**EPICS**

- Used by ESS accelerator/target, SLS, Diamond, US light sources, to be used by ISIS & SNS

**ADARA**

- Publish/subscribe software & protocol for streaming data (neutron + meta)

**Mantid**

- Data reduction framework in Python & C++ developed by ISIS & SNS

Data Management

**PanData**

- ICAT data cataloguing software developed under NMI3 by PanData collaboration of 19 European facilities (+ SNS in US)
Data Analysis & Modeling

- Data on disk is useless!
  - It is published results from the data that makes progress
- Need to ensure that ESS users have access to
  - appropriate software packages for data analysis
  - the necessary computational resources to exploit the software to obtain those results
  - analysis software during experiment to influence the data taking strategies
- Roll out in-sync with instruments

Structure of Nanomaterials


Polarized SANS demonstrated that these nanoparticles have uniform nuclear structure but core-shell magnetic structure.

Required development of both data reduction and data analysis methods and tools.
Integration of Analysis with Advanced Modeling Techniques

Molecular Dynamics & Density Functional Theory (DFT) Techniques

Jose Borreguero (SNS/NDAV) Mike Crawford (Dupont) & Niina Jalarvo (Julich): BASIS experiment / MD simulation studies of Methyl rotations in methyl-Polyhedral oligomeric silsesquioxanes (POSS)

Optimization of dihedral potential governing the rotational jumps in methyl hydrogens. Preliminary simulations indicate that the dihedral potential barrier must be decreased ~37% in order to match experimental data.
DMSC will not be a “supercomputer” centre

Data (disk) storage:
- Back of the envelope ➔ 2 – 3 PBytes/yr
- Spectrum of file sizes: ~100MByte - ~10’s GByte – ~1TByte
- Fast disk (200MByte/s) & Parallel File System (10GByte/s)

Cluster(s) for data reduction & (modest) data analysis - ~2048 cores
Architecture – CPU, GPU... visualization cluster/server

Data download servers – sftp & gridftp

Remote login capability for ESS users:
- Re-reduce data using cluster
- Data analysis software available for users

Software development servers – repositories, bug trackers, build servers
During the Construction Years

- Primary goal: To be ready for hot commissioning of first ESS instruments in 2020.
- Includes significant amount of NRE (Non-Recurrent Engineering) for subsequent ESS instruments.
- Establish the basic facilities at both campuses.
- Front loaded with instrument-centric work + scope to grow analysis in an integrated way.
Conclusion

- DMSC’s mission is the computational (software/hardware) infrastructure for the ESS *data chain* - instrument control, data acquisition, reduction & analysis

- Generic framework customized for each of the instruments – utilizes data streaming to account for large data files/live processing

- Don’t re-invent the wheel – work with collaborators ISIS, SNS, PanData (+ others?) to develop/customize existing software – EPICS, ADARA, MANTID, ICAT...

- Customize for each of the ESS instruments in-sync as they roll out

- In-sync with instrument roll out work with instrument teams to ensure analysis software is available (doesn’t necessarily mean develop)

- Longer term goal to develop new data analysis methods for ESS instruments
QUESTIONS