ICALEPCS 2015
International Conference on Accelerator & Large Experimental Physics Control Systems

HDF5 and Data Formats Workshop

Nick Rees, Diamond Light Source
Elena Pourmal, The HDF Group
James Hester, ANSTO
Tobias Richter, ESS
Andy Götz, ESRF
David Schneider, LCLS
Daron Chabot, NSLS-2

Workshop overview

- Aim:
  - Discuss and inform about:
    - Storage of data and metadata, particularly with HDF5.
    - The merits of ontologies built on top of HDF5 (e.g. NeXus)
    - Use HDF5 in demanding applications (e.g. high speed detectors)
    - Developments for our community and how support them.
- Participants:
  - 51 registered
  - 16 Countries
  - 27 Institutes
  - 6 Continents

ICALEPCS 2015

International Conference on Accelerator & Large Experimental Physics Control Systems

Workshop program

- Structure:
  - Morning:
    - Tutorial by Elena Pourmal, Director of Technical Operations, The HDF Group.
  - Afternoon:
    - Presentations by lead participants.
    - Open discussion about way forward.
- Online resources:
  - Morning tutorial:
    - ftp://ftp.hdfgroup.uiuc.edu/pub/outgoing/epourmal/ICALEPCS2015/
  - Afternoon presentations:
    - http://controls.diamond.ac.uk/downloads/other/files/icalepcs_hdf5/
  - General information:
    - https://www.hdfgroup.org/

Take home messages

- HDF5 is the primary storage format for many people
  - All current and future NASA missions.
  - Much of the worlds meteorological data and models (netCDF4).
  - Much of the worlds HPC simulations.
  - Matlab (.mat) and other commercial software.
  - … as well as photon, neutron (NeXus, Data Exchange) and some astronomy data.
- HDF5 is more than just a file format
  - It has a data model that allows storage of a wide variety of data.
  - It provides API's & tools for very efficient data handling (but you can also hang yourself!).
- HDF5 is open source software largely supported by a private company
  - Open source software is not free.
  - The HDF Group are continually challenged by how to support their work.
  - Being used as a long term archival format adds sustainability complications.
- HDF5 has recently been enhanced for our community.
  - Better and more varied compression.
  - Single writer/multiple reader for data analysis while scanning
  - Virtual datasets for (for example) parallel compressed writing

Conclusion

- What was achieved?
  - Understanding about HDF5, and its new features.
  - Understanding of the relationship between HDF5 and ontologies
  - Understanding of the HDF5 support model.
  - Suggestions on sustainable support for an archival data format.
- What next?
  - A number of institutes volunteered to contribute to supporting HDF5 in the form of support agreements.
  - Agreed we need to foster the user community – starting with regular workshops at conferences like ICALEPCS or RDA.