Research data needs of the Photon and Neutron Science community IG

Session at the RDA 8th plenary

Container in Scientific Workflows & Data Preservation

research data sharing without barriers
rd-alliance.org

Agenda

Outline.

- Short introduction Photon & Neutron Science IG

- Jaroslaw Nabrzyski (Uni. Notre Dame): Container Strategies for Data & Software Preservation that Promote Open Science

- Raymond Osborn (ANL): cloud servers for neutron and x-ray data.

- Frank Schlünzen (DESY): Container at Photon and Neutron Facilities

- Discussions

Discussion topics.

- Sharing efforts to provide containerized scientific workflows

- What do we need and what could we do together?

- Future activities?

AOB.

research data sharing without barriers
rd-alliance.org

RDA
RESEARCH DATA ALLIANCE

Photon & Neutron Science IG

Chairs (any volunteers?).

- Brian Matthews STFC; Frank Schluenzen (DESY)
- Amber Boehnlein (JLAB); Thomas Proffen (ORNL)

Photon & Neutron Science IG

Chairs.

- Brian Matthews STFC; Frank Schluenzen (DESY)
- Amber Boehnlein (JLAB); Thomas Proffen (ORNL)

Representing.

- Light- & Neutron-Sources (exclusively in Europe & US)
- Investments of 10th of billions $
- Huge & increasing number of (large) instruments aka beamlines
- Well over 30.000 (known & active) users annually
- Creating billions of files and 10-100s of PB per yr
  - HDF5 is our prime choice
  - NeXus for well structured & standardized meta-data
  - ICAT as **the** standard data catalog

RDA
RESEARCH DATA ALLIANCE

Photon & Neutron Science IG

Specialties.

- Many scientific fields with volatile scientists (can't afford the data efforts)

- Data mining of the raw scientific data is an extremely rare use-case

- Re-use of raw scientific data almost equally rare

- Sharing of raw data relevant in some mostly collaborative applications

- … but quality control, validation & reproducibility is of course essential

RDA relations.

- Data mining, sharing, re-use not the main goal → don't fit in very well

- PaNSIG (rdnpnsc) has naturally difficulties adopting RDA outcomes

  - but some pieces are still highly valuable

  - believe we also do a good job advertising RDA

- Usually focus on specific topics very close to our research environment

- Scientific data formats, workflows, data management, policies etc

research data sharing without barriers
rd-alliance.org

Photon & Neutron Science IG

Policies.

- One of the P&N success stories, though it takes an extremely long time

- Increasingly (European) P&N facilities are adopting a common policy

  - Avoids policy as a "competitive advantage"

- So far adopted (or in the pipeline) by ILL, ISIS, HZB, ESRF, PSI, Eur.XFEL…

  - As well as communities; means we publish data & ensure persistence

Standardization.

- Major investment into standardization of formats (HDF5/NeXus)

- Adoption of standards by IUCr ($\rightarrow$ greatly facilitating inter-operability)

  - It's an extremely tedious process, sometime lacking behind HDF evolution

  - Diversity of scientific field doesn't help

  - Slowly getting there and there is no alternative

  - Takashi Onishi: prime importance

RDA
RESEARCH DATA ALLIANCE

Photon & Neutron Science IG

Container.

- Preservation of scientific software (reproducibility) crucial

- Sustainability of core scientific software indispensible

  - Dead standards are none

- Maintenance of complex frameworks a nuisance

- Uniform data exchange essential

- Believe that we can greatly benefit from joint efforts

RDA
RESEARCH DATA ALLIANCE

Photon & Neutron Science IG

Ads.

- NOBUGS 2016: https://indico.esss.lu.se/event/357/
  - RDA, Container, DAAS, …

- "Granule metadata and data flows in data sharing and publications":
  - Joint Session of the RDA/Codata Materials IG and the Photon & Neutron IG
  - Breakout Session 7. Saturday 17th 09:00-10:30
  - Also: several sessions on reproducibility!

- Computing and Software for Big Science   Springer
  - New data centric journal to be launched spring 2017