FAIR data at the NIST Center for Neutron Research

Brian Maranville (NIST)
https://orcid.org/0000-0002-6105-8789

2021 FAIR Data User Workshop @ MagLab
2021-10-12
NCNR: a neutron-beam user facility

"a national resource for industry, universities, and government agencies."

22 Scattering Instruments
(8 additional Physics/Analytical Chemistry)

- Crystallography
- Reflectometry
- Small-Angle Neutron Scattering (SANS)
- Spectroscopy

https://www.nist.gov/ncnr
NCNR user program (per year*)

- 774 beamtime proposals
- 2769 users

Institutions:
- 17 NIST Divisions and Offices
- 38 U.S. Government Laboratories
- 42 U.S. States + DC + Puerto Rico
- 50 U.S. Corporations
- 181 U.S. Universities

*figures from 2019 NCNR annual report
CHRNS: Center for High Resolution Neutron Scattering (NSF)

• Instruments (266 proposals in 2019):
  1. CHRNS Very Small Angle Neutron Scattering (vSANS)
  2. CHRNS Multi-Axis Crystal Spectrometer (MACS)
  3. CHRNS High-Flux Backscattering Spectrometer (HFBS)
  4. CHRNS Neutron Spin Echo Spectrometer (NSE)
  5. CHRNS White Beam Reflectometer (CANDOR)

• Support for NCNR proposal system
• Non-Equilibrium Science initiative
• FAIR data initiative (beginning FY2021)
Raw data: open by default

• Oldest: NG3SANS (1989-02-08)
• Total number: 4,425,906 files
• Total size: 1.2 TB
• DOI: https://dx.doi.org/10.18434/T4201B
  • Points to: https://ncnr.nist.gov/pub/ncnrdatas

NCNR Data Management Plan:

All NCNR raw data within the scope stated above are made available to the public following an experiment unless the researcher ... opts out ...
NCNR and FAIR data: up to 2020

- Open raw data
- Limited search by metadata (file date, name, description)
- Live instrument data display since 2008
- Online data reduction: 2017 (reflectometry)
- Custom web-based data viewers
  - Column-format (text)
  - NeXus (HDF5)
Data and Metadata Lifecycle (before)
CHRNS (NSF) FAIR data sprint

During unplanned reactor outage of 2021

Accelerate FAIR initiatives

<table>
<thead>
<tr>
<th>ORCiD integration</th>
<th>DOI per experiment</th>
<th>search data by metadata</th>
<th>capture of process metadata</th>
<th>traceable data reduction workflows</th>
</tr>
</thead>
</table>
ORCiD integration

Proposal system:
- Federated login
- User disambiguation

Instrument Control:
- ORCiDs pulled from proposal
- Re-transmitted to data.nist.gov

Experiment DOI record in data.nist.gov

Future:
- User identity management for remote computing
FAIR sprint contributors

**Instrument control software team**
- **Steve Pheiffer**, Nick Siebenlist, Natasha Shmunis, Chirag Parikh, Fawaz Joseph
- Sample metadata taxonomy and management
- Initiating DOI creation on new experiments
- Capturing process metadata

**NCNR center office**
- Siddharth Khosla, Alan Munter
- Integration of ORCiD identity management across facility

**Scientific staff**
- Tyler Martin, Katie Weigandt, William Ratcliff...
- Development of taxonomies
- Identifying incentives
- Curating process metadata
Citeable data: DOI per experiment


Automated pre-publish through API on experiment start

Metadata populated from proposal:
- Participants, including ORCiD
- Title, Description
- Activity dates/times
- Links to raw data DOI, search functions, etc.
Define metadata for sample

Flexible JSON record with unique ID

Sample associated with measurement

Associations published with data

Document database from records

Public search API

Human-friendly search tools

Searchable metadata
Capture process metadata

Ensure that workflows in instrument control software include

<table>
<thead>
<tr>
<th>Measurement intent</th>
<th>Sample environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Background</td>
<td>• Temperature</td>
</tr>
<tr>
<td>• Transmission</td>
<td>• Magnetic Field</td>
</tr>
<tr>
<td>• Scattering ...</td>
<td>• Frequency ...</td>
</tr>
</tbody>
</table>

Add process variables to searchable metadata
Traceable data workflows

- Track history of processing steps
- Embed or export processing recipes
- Support replay
Data and Metadata Lifecycle (after)
Public data accessibility

- data.nist.gov (expt. DOI)
- Data Repo
- Metadata DB

API

Web search
Nice things you can do with FAIR data...
Online data reduction

- [https://reductus.nist.gov](https://reductus.nist.gov)
- Data accessed directly using data repo and API
- Anonymous, stateless
- Python backend
- Feed-in from other tools e.g. experiment browser
Upgrades to raw data repository web view

- DOI direct link
- SHA-256 integrity hash
- "Download All" button
- Future: File type (mime)
- Future: Viewer links
Public Data viewers

- **Live data** (all instruments)
- CHRNS VSANS viewer
- CHRNS CANDOR viewer
- Column-data viewer
- NeXus inspector