What does Open Science Require of Standards and Infrastructure?

Rebecca Ringuette - Heliophysics Digital Resource Library

Oral presentation given at the 2023 IHDEA meeting, JHU/APL, USA

What is Open Science?

Word cloud created using multiple definitions of Open Science:

- UNESCO
- NASA
- NSF
- US White House
- Europe
- Japan
- Africa
- Vietnam

"Making scientific knowledge and collaborations open."

Opening Scientific Knowledge

- Scientific knowledge is embedded in our
  - publications, data, software, metadata,
  - meetings, discussion boards,
  - technologies, web apps, user interfaces,
  - textbooks, courses, tutorials,
  - and many other types of resources.
- **Climbing the ladder** towards opening our scientific knowledge requires:
  - An increasing interweaving of technology into our methods of human interaction,
  - balanced with dynamic needs for privacy and security.

Image Credit: https://science.nasa.gov/science-red/s3fs-public/atoms/files/SMD%20Open-Source%20Science%20Guidance%20v2%2020230407.pdf

What capabilities does open science demand of our scientific knowledge?

- Searchable through a reduced set of highly capable, fully connected, and intuitive search interfaces.
- Understandable at multiple levels according to the capability of the person.
- Validatable with the items provided by the authors.
- Reusable by those outside of the original group.
- Begin with FAIR, but don't stop there.
  - Findable
  - Accessible
  - Interoperable
  - Reusable

What do these capabilities demand of Heliophysics infrastructure?

- FAIR metadata and data (including licensing and PIDs).
- FAIR software, publications, and other resources.
- Fully connected resources and infrastructure elements.
- Streamlined and semi-automated curation processes.
- Preparation for **increased demand for curation**.
- Preparation for curation of new types of objects.
- Development of new types of resources to increase understanding.
- Partnerships between infrastructure entities to increase curation efficiency.
- Orchestrated metadata sharing to increase discoverability of our resources.

Where are the current standards missing or too rigid?

- Missing: Standard definitions for levels of validation in peer-review.
- Missing: Standards for software curation and archival.
- Missing: Collaboration agreements between infrastructure elements (e.g. publishers and repositories).
- Too rigid: Multiple curation levels for each item type (e.g. from CoreTrustSeal.org: https://zenodo.org/record/8083359).
- Too rigid: More flexible minimum curation requirements based on curation level chosen or required.
- Missing: Guidance on what items contributors should include.
- Missing: Motivation for contributors to improve metadata or include more items.

Ways Forward

- Validation studies for publications, data, and software.
- Develop methods and standards for software curation and archival.
- Account for a spectrum of compliance, ability, and funding.
- Develop these standards WITH the community.
- Include the ‘why’ in the standards and curation development processes.
- Build upon others’ work to streamline metadata curation interfaces.
- Create user working groups and fund their participation.
- Research the barriers the community encounters in the scientific process to understand how to improve (e.g. the broken peer-reviewing process, proposal reviews, metadata curation).

Let’s get started

Announcing the **Defining the Spectrum of Open Science in Software** workshop!

- May 29-31, 2024 - hybrid
- In collaboration with pyOpenSci, PyHC, and other community leaders in software development.
- Goal is to work with the community to define the spectrum of possible tasks relevant to softwares of different categories.
  - Software intended for limited-use (e.g. scripts associated with publications)
  - Software intended for wide use (e.g. mission software, PyHC packages)
  - Modeling and simulation software (both limited-use and wide-use)
- SPD-41a as the minimum requirement, pyOpenSci+PyHC approaching ideal.
- Produce a draft rubric for possible use in ranking NASA proposals’ OSDMPs.