Data Curation and Preservation Interest Group

Raffaele D'Abrusco[1], **Anne Raugh**[2]



[1] Center for Astrophysics | Harvard & Smithsonian
[2] University of Maryland

Data Curation and Preservation IG

The main goals of the DCP IG are to **collect, formulate and share best practices** and engage IVOA member projects **in the long-term curation and preservation of astronomical data**.

Charter priorities

- Identification of at-risk data and data collections.
- Engaging with the community in curation and preservation activities.
- Coordination with strategic partners (astronomical data centers, research libraries and publishers).
- Review of technology relevant to all aspects of curation and preservation.
- Support for the implementation of TRUST principles and archive certification.

Recent documents

- Endorsed note on Data Origin:
  https://www.ivoa.net/documents/DataOrigin/

- Best practices on the creation of Digital Object Identifiers (DOIs) in astronomical archives:
  https://www.ivoa.net/documents/Notes/DOI4Archives/

E-mail: datacp@ivoa.net Webpage: https://wiki.ivoa.net/twiki/bin/view/IVOA/IvoaDCP

New Questions

- Doubling down on Data Provenance
  - Within the IVOA
    - → Do we need to document provenance aspects that are most useful/valuable for IVOA users?
  - Elsewhere
    - → Can we find a common inter-disciplinary/inter-repository approach to provenance?
    - → Can we learn something interesting from provenance for data from other disciplines?

New Questions

- Doubling down on Data Provenance
  - Within the IVOA
    - → Do we need to document provenance aspects that are most useful/valuable for IVOA users?
  - Elsewhere
    - → Can we find a common inter-disciplinary/inter-repository approach to provenance?
    - → Can we learn something interesting from provenance for data from other disciplines?

- DOIs
  - Furthering the investigation of some advanced DOI topics
    - → Can we establish guidelines for the creation of DOIs associated with catalog-based queries?
    - → Do we need to keep monitoring emerging DOI community-driven usage practices?
    - → How can we help improve the coordination with publishers and digital libraries (SciX) for the adoption of data citation best practices?

New Questions

- Doubling down on Data Provenance
  - Within the IVOA
    - → Do we need to document provenance aspects that are most useful/valuable for IVOA users?
  - Elsewhere
    - → Can we find a common inter-disciplinary/inter-repository approach to provenance?
    - → Can we learn something interesting from provenance for data from other disciplines?

- DOIs
  - Furthering the investigation of some advanced DOI topics
    - → Can we establish guidelines for the creation of DOIs associated with catalog-based queries?
    - → Do we need to keep monitoring emerging DOI community-driven usage practices?
    - → How can we help improve the coordination with publishers and digital libraries (SciX) for the adoption of data citation best practices?

- The rise of new technologies
  - Methods for the exploration of multi-modal datasets
    - → Can we expand data curation to include tracked associated/linked literature?
  - Large Language Models (LLMs)
    - → Can we take advantage of LLMs to enable preservation and speed up the curation of (old, unstructured, messy) astronomical data?
    - → Do we need to adopt specific practices to curate data **for** LLM-powered algorithms?

CDP @ 2026 Fall InterOperability Meeting

No CDP session!

Relevant sessions

- **Knowledge Discovery** (Friday Nov 6 @16:00 AWST)
- **Distributed Services and Protocols** (Saturday Nov 7 @09:00 AWST)
- **AI & Applications** (Sunday Nov 9 @09:00 AWST)

We will hold a meeting open to all IVOA members soon after the end of the InterOp to discuss in details the DCP priorities, their actionable aspects and receive input from the community.

If you are interested in attending, subscribe to the DCP mailing list, or just get in touch.