

# Data Curation & Preservation Summary

IVOA Interop Meeting  
Pune, India

Monday 17 October 2011  
Bhaskara 2 | 6:00-17:30

# Schedule - Mon. 16:00-17:30

- Wf4Ever update - J.E. Ruiz
- World Data System - F. Genova
- Data Citation in Astronomy - A. Accomazzi
- Long-Term URIs - N. Gray
- Discussion

# Workflows

- Workflows capture the activity carried out on “Research Objects” (RO) and aims at creating a framework to support reusable, reproducible research
- As a preservation matter, workflows are a necessary component of the research lifecycle, so their preservation is essential in order to support reproducibility
- Wf4Ever goals: support archival, preservation and indexing of workflows, creation of scientific communities
- “Workflows and processes should benefit from the same benefits acquired by data”
- Now working on prototypes

# World Data System and Data Identifiers

- FG represents IAU at WDS Scientific council; other astro data centers should consider applying (IVOA can as well)
- WDS members are pushing for the adoption of DOIs as THE WDS-sanctioned persistent data identifier (PDID) system
- Recent Data Citation workshop:  
[http://sites.nationalacademies.org/PGA/brdi/PGA\\_064019](http://sites.nationalacademies.org/PGA/brdi/PGA_064019)
- CODATA WG on Data Citation Standards and Practices:  
<http://www.codata.org/taskgroups/TGdatacitation/>

# Data Citation in Astronomy

- Review of current practices, citations to “data” papers as well as catalogs (e.g. VizieR) and tables in journal articles
- Some material mentioned in papers is not “properly” archived and readily citeable in a well-established way, links are done via URLs
- For the sake of preservation, we need proper archiving policies and persistent identifiers for all citable nuggets: papers, journal tables, catalogs, datasets
- Dataset registration and minting of PDIDs (DOIs or similar)
- What gets a PDID? Who assigns it? Who maintains the registry? What long-term commitment can our community put forth?

# Long-Term URIs

- Focus is on URIs used to reference documents and resources NOT already covered in the literature or data archives
- Desire for persistence comes from the need to use URIs in SemWeb applications. Examples: FITS, IVOA SKOS vocabularies, SIMBAD objects, VOEvents
- Preservation goal ~ 100-1000 years
- Possible candidates: ARKs, EZIDs, PURLs, DOIs
- Why not get started with <http://purl.org/astronomy>

# Actions and Timeline

- Long-term URIs: NG will revise current draft and a note will be created in short order (1-2 months) reflecting our discussion, current assessment of available solutions, and possibly a recommendation for some URI sets
- Recommendations for PDIDs: FG will monitor CODATA WG developments, AA will engage UC3 on DataCite; NG to inquire/learn about use of PDIDs in HEP and DataCite in UK
- Discussed use of PDIDs in ObsCore tables in conjunction with the DataLink proposal
- ADASS BOF on Data Publication and Preservation (AA, SD, NG, CB) will assess requirements and impact of PID registration/use on astronomy archives and publishers
- An IVOA note developed between ADASS and next IVOA interop on PDIDs possible but not certain
- ADASS BOF on Scientific Workflows (AS, JR), IVOA note released