

# Data Curation & Preservation

Thursday 2011-05-19T07:30:00 (UTC)  
Arnold Rots

# Program

- Norman Gray:  
Long-term URIs in astronomy
- Sébastien Derriere:  
SIMBAD URIs
- Arnold Rots & Alberto Accomazzi:  
Persistent Data Identifiers in Astronomy
- Pierre Le Sidaner:  
OAIS and IVOA
- José Ruiz:  
Update on Wf4Ever

# Persistent Identifiers (main topic)

- Norman and Arnold emphasized the need for persistent URIs/identifiers for datasets
- The IVOA identifiers issued under ADS naming authority work and provide the right set of characteristics: unique, persistent, verifiable, resolvable –  
<ivo://ADS/Sa.CXO#obs/123>
- Need to establish formal registry to replace current more informal ADS operation
- Alternative is to use DOIs through DataCite; this would provide registry functions
- Both approaches offer same functionality; this will be worked out and we expect no problems either way

# SIMBAD URIs

- Sébastien presented a review of the issues involved in assigning URI-type identifiers to objects in SMBAD
- Objects have multiple identities
- The dynamic nature of the SIMBAD records presents a particular problem:
  - **Objects merge**
  - **Objects split**
- Note that this is closely related to the problems associated with crossmatch algorithms: it is not absolute, but represents a probability – which may change at any time as new information becomes available

# OAIS and Similar Standards

- Pierre urged archives participating in the IVOA to comply with OAIS archiving models
- Noted importance of provenance (including context) information and authentication
- Noted that IVOA protocols contain information on exchange format, but not the native format of the data
- And urged to move toward certification as trusted repositories, following the auditing standards currently being developed; this has been mentioned before and my guess is that our funding agencies will start requiring it at some point – may as well be proactive

# Wf4Ever

- José presented an update on Workflows Forever
- As digital collaborations are proliferating we need to get serious about curating and preserving collaborative digital research objects
- Preservation of workflows is a complex matter
- Publication of research objects is desirable, but how?