

Connecting the literature with on-line data

**Guenther Eichhorn, Alberto Accomazzi, Carolyn S. Grant,
Edwin A. Henneken, Michael J. Kurtz, Donna M. Thompson
and Stephen S. Murray**

Harvard Smithsonian Center for Astrophysics,
60 Garden Street, MS-67, Cambridge, MA 02138, USA
email: gei@cfa.harvard.edu

Over the last few years there has been considerable progress in linking the published scholarly literature with on-line data. This will greatly help with data discovery and aid the efforts to the VO.

In an initial effort to provide the means for such linking, the Astrophysics Data Center Executive Committee (ADEC), a collaboration of NASA data centers in the USA, has worked with the American Astronomical Society (AAS) and the University of Chicago Press, the publisher of the journals of the AAS, to establish a system that allows authors to specify data sets that they used in an article. This information is then used by the journal publisher to link to the on-line data. It is also forwarded to the Astrophysics Data System (ADS) and to the data centers to provide similar links between the literature and the data in these other systems.

We have developed the software infrastructure to handle the all aspects of this system from registering data center's data holdings, through automatic verification of the data set identifiers, to persistent linking from the journal articles to the on-line data.

The ADEC agreed on a format for the data set identifiers that is compatible with current VO identifier structures. Data set identifiers have the form:
ADS/FacilityId#PrivateId .

The AuthorityId string 'ADS' has been specified. This simply recognizes the current role of ADS in managing the namespace used for these identifiers, in the absence of a community-wide namespace granting authority. It does not suggest nor imply that ADS controls or manages the dataset itself.

The ResourceId token will be interpreted as a Facility. An ever-growing list of facilities is maintained by ADS. Data centers should contact ADS should they need to register new entries.

The PrivateId string can be anything that the data center desires, with the provision that the identifiers string as a whole should abide by the general syntax of a URI, as required by the IVOA identifiers specification.

Data centers who wish to participate in this effort, should register with the ADS. While it is expected that the appropriate metadata will one day be made available by a public VO registry, its format and access methods are at this time not available. As an intermediate solution to the problem, we require that the data centers maintain a simple profile which will provide ADS with the necessary metadata.

The data center profile is simple XML document that lists the data center name and description, the name and email address of the person responsible for the maintenance of the profile, the URL of the web service to be used for dataset verification, and the list of facilities that the datacenter has data for.