(P)IDs in the Virtual Observatory



Markus Demleitner msdemlei@ari.uni-heidelberg.de



- What is identified?
- Identifying data collections
- Identifying data sets
- Permanent: ivoids? dois? handles?



Data Collections and Datasets

The VO has a two-layered architecture:

- 1. The first level of discovery are *Data Collections*, kept in a Registry
- 2. The second level are *Datasets* (e.g., spectra, images, perhaps database rows), discovered through services on top of the data collections.

Identifying Data Collections

Identifiers in the VO Registry are ivoids of the form: ivo://unique-authority/local/part

Authority uniqueness is ensured by registering Authorities (\sim data centres) in the Registry itself.

Resolution: Through the VO Registry (typically via relational tables).

For citeability, several providers also assign DOIs to data collections.

Identifying Datasets

Several VO protocols have the notion of publisher dataset identifiers (pubDIDs) and creator dataset identifiers (creatorDIDs). There are ivoids with query parts:

ivo://authority/svc/id?dataset-key

Resolution: Cut off query part to find the data collection, then use service type-specific ways to locate datasets within the data colelctions.

Some providers also have DOIs or handles for datasets; no overarching practice exists.



Ivoids contain the publisher's authority. When a data collection moves from one publisher to another, its ivoid with change, too.

The same reasoning applies to PubDIDs. CreatorDIDs would remain the same, but at least so far they are not used at all.

Actual PIDs: For data collections, DOIs certainly are the way to go. For datasets, everything is complicated, including the Why.