

Discovery and Provenance Metadata for Persistent Data Objects

Arnold Rots
SAO

Individual Data Products

- High-level data products
 - Associated with publications
 - Complementary to publications
 - Unrelated to publications
- Valuable lower-level datasets
- Catalogs, tables, numbers
- Press releases
- Pretty pictures
- ...

Three Crucial Issues

- Persistence and preservation
 - We are well aware of this aspect
 - But we have paid scant attention to:
- **Discovery**
- **Provenance**

Wanted: Metadata (a new standard)

- To facilitate data discovery
- To record the provenance trail

- Also needed:
 - Data products registry
 - Existing registry is a registry for collections/repositories
 - That is an appropriate entity
 - But for isolated data products we need a special dedicated registry of metadata and identifiers

Data Discovery Metadata (req'd)

- Can be derived from existing resource metadata:
 - Identity:
 - Title, ShortName, Identifier
 - General Content Metadata:
 - Subject, Description, Reference, ReferenceURL, Type, ContentLevel
 - Specific Content Metadata:
 - Facility, Instrument, Coverage, Resolution, Format, Rights
 - Data and Metadata Quality Assessment:
 - DataQuality, ResourceValidation, Uncertainty

Type values – existing and new

- Observation
 - Image
 - Cube
 - Light curve
 - Catalog
 - Value
 - Library
 - Survey
 - Artwork
 - Historical
- Object
 - Mosaic
 - Spectrum
 - Event list
 - Table
 - Value pair
 - Simulation
 - Animation
 - Facsimile
 - Other

Provenance (Curation) Metadata

- Required when applicable
 - **Publisher** PublisherID
 - **Creator** Creator.Logo
 - **Contributor**
 - **Date**
 - **Relationship** RelationshipID
- Highly desirable
 - **Version**
 - **Contact** (name, address, email, phone)

Relationship values

- Proper provenance metadata:
 - **Primary** (new)
 - **Derived-from**
 - **Copy-of** (new)
- Other existing values
 - Mirror-of
 - Service-for
 - Served-by