



Report to ESIP Products & Services Cluster

Re: Data Stewardship &
Preservation Cluster Metadata
Testbed Activities

Recent activities: 2009-2010

- Persistent Identifiers
 - the problem
 - the use cases
 - the recommendations, in theory
 - but, really??

Recent activities: 2009-2010

- Persistent Identifiers
 - the problem
- Uniquely and unambiguously identify the data of interest
- Find and access the data regardless of where they currently reside or how the responsibility for managing the data has changed over time
- facilitate management of data over time
- facilitate data citation in publications

Recent activities: 2009-2010

- Persistent Identifiers
 - the use cases:
suitability of each to
provide:
- Unique *Identifiers* for Earth science data objects
- Unique *Locators* for the objects
- Service as *Citable Locators*
- Unique identification of the *scientific contents* of data objects *if the data were reformatted*

Recent activities: 2009-2010

- Persistent Identifiers
 - the preliminary recommendations
in theory
- Unique ID = UUID
- Unique Locator = URI family
- Citable Locator = DOI most common
- Scientific ID = nada

Persistent Identifiers considered:

- DOI
- ARK
- Handle
- PURL
- LSID
- OID
- URI/URL/URN
- XRI
- UUID
- In terms of:
- Naming syntax
- Technical value of the scheme
- User value
- Value to archives or data centers
- Use case support.

Recent activities: 2009-2010

- Article soon to be published:

“On the utility of identification schemes for digital earth 5 science data: an assessment and recommendations”, Duerr, Ruth E., et al, Earth Science Informatics, DOI 10.1007/s12145-011-0083-6

Planned activities: 2011

Persistent identifiers:

- Add to evaluation of 9 Identifier schemes by assessment based on implementation
- Determine if ID schemes and/or best practices can be recommended

Planned activities: 2011

- With at least 2 datasets of different levels of complexity, address implementation issues
- Assign IDs to each component in each of 2 datasets from each of 9 ID schemes

Some potential implementation issues:

- Capability to uniquely data objects at needed levels of granularity
- Applicability for mechanisms to locate, i.e., resolve to a location as part of a citation
- Support services from external organizations, if any, such as

Support services provided?

- Maintenance of persistence and uniqueness of _identification_ of data object?
- Reliance upon descriptive MD to aid unique identification of data object?
- Incorporation & ongoing maintenance of descriptive MD used within various citation formats?
- Mechanisms and efficacy of same for:
 - Establishing and maintaining relationships among related objects (subsets, supersets / collections, hierarchies and webbed (non-hierarchical) links
 - Noting, managing changes in data object (versioning, obsolescence)

Stay tuned...

- Data Stewardship & Preservation cluster wiki, ID pages: http://wiki.esipfed.org/index.php/Interagency_Data_Stewardship/Identifiers