Member Node Topics:
Identifiers

DataONE Member Node Forum
2016-09-15

DataONE is a collaboration among many partner organizations, and is funded by the US National Science Foundation (NSF) under a Cooperative Agreement. This material is based upon work supported by the National Science Foundation under Grant Numbers 0830944 and 1430508. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Identifiers

Overview

Identifiers are used to identify something within a collection of things.
Identifiers

Overview

Many identifier schemes in use:

- DOI  Digital Object Identifier
- ARK  Archival Resource Key
- LSID  Life Sciences Identifier
- UUID  Universally Unique Identifier
- LSRN  Life Science Resource Name

*DataONE:*

Character string with no whitespace, less than 800 characters, and unique in DataONE
Identifiers

Application Programming Interfaces

Many DataONE API methods require an identifier

resolve( ID ) -> Object Locations

getSystemMetadata( ID ) -> Sys Meta

get( ID ) -> object bytes
Identifiers

Escaping

Identifiers as parameters in URLs must comply with rules for URL construction ([RFC 3896](https://tools.ietf.org/html/rfc3896))

\[
PID = @some/reserved$characters! \\
\]

%40some%2Freserved%24characters%21
Identifiers

Escaping

Identifiers as parameters in URLs must comply with rules for URL construction (RFC 3896)

PID = @some/reserved$characters!

%40some%2Freserved%24characters%21

cn.dataone.org/cn/v1/resolve/%40some%2Freserved%24characters%21
Identifiers
Series Identifiers (SIDs)

An identifier for the latest version of a series
Identifiers

Series Identifiers (SIDs)

An identifier for the latest version of a series

Diagram:

- A (S1) → B (S1) obsoletes
- B (S1) → C (S1) obsoletes
- A (S1) obsoletedBy B (S1)
- B (S1) obsoletedBy C (S1)

Time
Identifiers

**PID s and SIDs**

Persistent Identifiers (PIDs):
Always resolve to exactly the same object

Series Identifiers (SIDs):
Always resolve to the latest version

- Share the same domain -> MUST be unique
- API methods are PID / SID agnostic
- Examine System Metadata to distinguish
Identifiers

Objects and Packaging

CiTO
The Citation Typing Ontology

System Metadata
Identifier
Checksum
Object Type
Size
Access Control
Replication Policy

Science Metadata

documents
isDocumentedBy

aggregatedBy

Resource Map

Data

OAI-ORE
Open Archives Initiative Object Reuse and Exchange
Identifiers

Search User Interface

2016. Physical trajectory profile data from glider sp011 deployed by University of California - San Diego; Scripps Institution of Oceanography in the Coastal Waters of California from 2016-06-02 to 2016-09-06 (NCEI Accession 0156569). NOAA NCEI Oceanographic Data Archive. (3E5AA5C1-1570-4DDD-82F1-02325B11B1BC).

2016. Physical trajectory profile data from glider sp039 deployed by University of California - San Diego; Scripps Institution of Oceanography in the Coastal Waters of California from 2016-02-18 to 2016-09-06 (NCEI Accession 0156570). NOAA NCEI Oceanographic Data Archive. (47B3BF61-847C-458E-AD78-A8810433DC05).


Daniel J. Metcalfe and Eric Vanderduys. 2015. Desert Uplands Plot Network: Vegetation Survey (Presence Data, Species Richness and Stem and Hollow Counts) for Plot WOUR15, Northern Queensland, Australia, 2013. TERN Australia. [link]
Identifiers
Generating and Reserving

Coordinating Node methods:

- **reserveIdentifier**
  
  POST /v2/reserve
  
  Reserves an identifier for future use

- **generateIdentifier**
  
  POST /v2/generate
  
  Creates a new identifier UUID scheme
  
  DOI generation coming
Identifiers

e.g. Generate Identifier

**CN `generateIdentifier()` method**

```python
import requests
URL = "https://cn-dev.test.dataone.org/cn/v2/generate"
p = requests.post(URL,
                  cert="/tmp/x509up_u501",
                  data={'scheme':'UUID'})
print( p.text )

<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<identifier
xmlns="http://ns.dataone.org/service/types/v1">
   urn:uuid:b8e0f0be-fd30-481c-8cbb-c5bddd1d4a4ca
</identifier>
```

[https://purl.dataone.org/architecture/apis/CN_APIs.html#CNCore.generateIdentifier](https://purl.dataone.org/architecture/apis/CN_APIs.html#CNCore.generateIdentifier)
Technical Resources

Education

- purl.dataone.org/architecture
- ...architecture/design/UseCases/06_uc.html
- dataone.org/previous-webinars

Communications

- mnforum@dataone.org
- developers@dataone.org
- irc.ecoinformatics.org #dataone

Sources

- repository.dataone.org/software/cicore
- github.com/DataONEorg

Tools

- search.dataone.org... search-dev... search-sandbox... search-stage...
- cn.dataone.org/cn/v2/diag/subject/sysmeta/object
- Browser, curl, Python, Java libraries
- examples.dataone.org