

Infrastructure - Story #1378

Adding new object formats shouldn't require buildout

2011-02-21 16:48 - Chris Jones

Status:	Closed	Start date:	2011-03-29
Priority:	Normal	Due date:	
Assignee:	Robert Waltz	% Done:	100%
Category:	d1_schemas	Estimated time:	0.00 hour
Target version:	Sprint-2011.35-Block.5		
Story Points:			
Description			
Subtasks:			
Task # 1451: Add an object format service to Metacat			Closed
Task # 1452: Replace ObjectFormat calls with ObjectFormatService calls in Metacat			Closed
Task # 1453: Proxy the Metacat ObjectFormatService as a D1 CN service			Closed
Task # 1454: Replicate state changes in the object format list across CNs			Closed
Task # 1455: Update D1 libraries to use CN object format service			Closed
Task # 1464: Write unit tests for listFormats() and getFormat()			Closed
Task # 1593: Add an ObjectFormatDisk Cache to d1_common_java			Closed
Task # 1594: Add an ObjectFormatCache to d1_libclient_java			Closed
Task # 1638: Migrate ObjectFormatService in Metacat to CNCOREImpl			Closed
Related issues:			
Related to Infrastructure - Story #582: Implement an object format registry		Closed	2010-10-06
Related to Infrastructure - Story #1038: need a stub ObjectFormat service at ...		Closed	

History

#1 - 2011-02-21 16:58 - Chris Jones

- Assignee set to Chris Jones
- Category set to d1_schemas

In order to add a new object format (mime type), the dataoneTypes.xsd schema needs to be modified and new classes must be generated from it, which requires a full buildout of the CNs.

Either an easier mechanism needs to be put in place to add in new types, or the object format list needs to be exhaustive. All mimetypes registered at:

<http://www.iana.org/assignments/media-types/index.html>

could be listed in the XSD, which would reduce the frequency of updates to the XSD dramatically. Otherwise, an object format lookup service could also be used to match file types.

#2 - 2011-02-21 17:04 - Chris Jones

- Target version changed from Sprint-2011.07 to Sprint-2011.10-Block.2

#3 - 2011-03-07 18:01 - Dave Vieglais

- Position deleted (34)
- Position set to 1
- Position changed from 1 to 170

#4 - 2011-03-11 16:28 - Chris Jones

- Status changed from New to In Progress

After looking at how to create a dynamic enum in Java to reduce how much code would need alteration, although it's possible, it is contrary to the spirit of enums (immutable) and may only work on the Sun JVM. Instead, all classes using object format will need to change to using getter methods that access a backing properties file, and client code will need to query a REST service for object format information.

Proposal:

The following REST interface would provide object type information for named object types in the DataONE system:

GET: RETURNS
/types/object/ : xml or json representation of all d1 object formats
/types/object/{fmtid} : xml or json representation of the object format

A full response would include a hierarchical structure including the D1 id of the object format, along with the important identifiers that are used for this object format in other systems (doctype, namespace, mime, puid, gfid, udfr).

Example 1 (Metadata from schema):

```
/types/object/EML_2_0_0/                :  
  
eml://ecoinformatics.org/eml-2.0.0  
application/xml  
x-fmt/522  
fmt/eml-2.0.0  
fmt/eml-2.0.0  
  
/types/object/EML_2_0_0/doctype        : ""  
/types/object/EML_2_0_0/namespace      : "eml://ecoinformatics.org/eml-2.0.0"  
/types/object/EML_2_0_0/mime           : "application/xml"  
/types/object/EML_2_0_0/puid           : "x-fmt/522"  
/types/object/EML_2_0_0/gfid           : "fmt/eml-2.0.0"  
/types/object/EML_2_0_0/udfr           : "fmt/eml-2.0.0"
```

Example 2 (Metadata from DTD):

```
/types/object/FGDC_STD_001_1_1999/     :  
FGDC-STD-001.1-1999  
  
application/xml  
x-fmt/319  
fmt/fgdc-std-001.1-1999  
fmt/fgdc-std-001.1-1999  
  
/types/object/FGDC_STD_001_1_1999/doctype : "FGDC-STD-001.1-1999"  
/types/object/FGDC_STD_001_1_1999/namespace : ""  
/types/object/FGDC_STD_001_1_1999/mime     : "application/xml"  
/types/object/FGDC_STD_001_1_1999/puid     : "x-fmt/319"  
/types/object/FGDC_STD_001_1_1999/gfid     : "fmt/fgdc-std-001.1-1999"  
/types/object/FGDC_STD_001_1_1999/udfr     : "fmt/fgdc-std-001.1-1999"
```

Example 3 (data):

```
/types/object/TEXT_CSV/doctype         :  
  
text/csv  
x-fmt/18  
fmt/text-csv  
fmt/text-csv  
  
/types/object/TEXT_CSV/namespace        : ""  
/types/object/TEXT_CSV/mime             : "text/csv"  
/types/object/TEXT_CSV/puid             : "x-fmt/18"  
/types/object/TEXT_CSV/gfid             : "fmt/text-csv"  
/types/object/TEXT_CSV/udfr             : "fmt/text-csv"
```

#5 - 2011-03-14 18:57 - Chris Jones

After discussing the interface definition with Matt, we decided to:

- 1) Make the collection more flat to be consistent with other D1 REST URLs (we don't need the /types/object hierarchy)
- 2) Change the collection name to /formats to coincide with the D1 ObjectFormat class
- 3) Keep the collection name plural
- 4) Only return full object format records, as opposed to individual identifier leaf nodes to reduce calls to the CN service
- 5) Use the canonical identifier string as the D1 identifier string ()
- 6) Add a 'name' element with a human readable name for the object format
- 6) Set the content model of to ANY to allow for future hooks into other typing systems

So, the revised interface would be:

GET:	RETURNS:
/formats	: xml or json representation of all d1 object formats
/formats/{fmtid}	: xml or json representation of the d1 object formats

with the following examples:

GET:	RETURNS:
/formats/eml://ecoinformatics.org/eml-2.0.0 :	

Ecological Metadata Language,
version 2.0.0

/formats/FGDC-STD-001.1-1999	:
------------------------------	---

Federal Geographic Data Committee Content Standard,
version 001.1-1999

/formats/text/csv	:
-------------------	---

Comma Separated Values

The above will be added to the D1 architecture documents as the proposed interface.

#6 - 2011-03-15 16:24 - Matthew Jones

- Position set to 1
- Position deleted (171)
- Target version changed from Sprint-2011.10-Block.2 to Sprint-2011.11-Block.2

#7 - 2011-03-22 13:59 - Dave Vieglais

- Position deleted (11)
- Target version changed from Sprint-2011.11-Block.2 to Sprint-2011.12-Block.2
- Position set to 4

#8 - 2011-03-29 14:01 - Dave Vieglais

- Position deleted (13)
- Target version changed from Sprint-2011.12-Block.2 to Sprint-2011.13-Block.2
- Position set to 3

#9 - 2011-04-05 01:41 - Dave Vieglais

- Target version changed from Sprint-2011.13-Block.2 to Sprint-2011.14-Block.2
- Position set to 2
- Position deleted (6)

#10 - 2011-04-13 04:14 - Dave Vieglais

- Position deleted (24)
- Target version changed from Sprint-2011.14-Block.2 to Sprint-2011.15-Block.2
- Position set to 7

#11 - 2011-04-25 15:54 - Dave Vieglais

- Position set to 4
- Target version changed from Sprint-2011.15-Block.2 to Sprint-2011.17-Block.3
- Position deleted (25)

#12 - 2011-05-02 17:08 - Chris Jones

- Position set to 9
- Position deleted (10)
- Target version changed from Sprint-2011.17-Block.3 to Sprint-2011.18-Block.3

#13 - 2011-05-16 17:58 - Dave Vieglais

- Target version changed from Sprint-2011.18-Block.3 to Sprint-2011.20-Block.3

#14 - 2011-05-25 15:43 - Dave Vieglais

- Position set to 2
- Target version changed from Sprint-2011.20-Block.3 to Sprint-2011.21-Block.3
- Position deleted (26)

#15 - 2011-05-31 15:01 - Dave Vieglais

- Target version changed from Sprint-2011.21-Block.3 to Sprint-2011.22-Block.3
- Position deleted (12)
- Position set to 2

#16 - 2011-06-06 17:12 - Chris Jones

- Target version changed from *Sprint-2011.22-Block.3* to *Sprint-2011.23-Block.3*
- Position set to 21
- Position deleted (9)

#17 - 2011-06-27 01:17 - Dave Vieglais

- Target version changed from *Sprint-2011.23-Block.3* to *Sprint-2011.26-Block.4*
- Position set to 13
- Position deleted (23)

#18 - 2011-08-16 18:27 - Rob Nahf

- Target version changed from *Sprint-2011.26-Block.4* to *Sprint-2011.33-Block.4*
- Milestone set to *None*

#19 - 2011-08-24 15:43 - Chris Jones

- Assignee changed from *Chris Jones* to *Robert Waltz*

Assigning this to Robert since the one remaining task is assigned to him.

#20 - 2011-08-30 01:43 - Dave Vieglais

- Target version changed from *Sprint-2011.33-Block.4* to *Sprint-2011.35-Block.5*
- Position set to 3
- Position deleted (42)

#21 - 2011-08-30 17:31 - Robert Waltz

- Status changed from *In Progress* to *Closed*