DataONE R Client - Task #853

Story # 927 (Closed): R Client needs DataPackage model and implementation

add R client DataPackage constructors from data or science metadata

2010-09-14 19:17 - Matthew Jones

Status:	Closed		**Start date:**	2010-10-07
Priority:	High		**Due date:**	
Assignee:	Rob Nahf		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	DataONE R Client 1.0.0			
Milestone:	None		**Story Points:**	

Description

Currently the R client assumes it is getting a data object, and so the constructor for [[DataPackage]] adds the retrieved object to the [[DataObject]] data list. Instead, need to inspect the [[ObjectFormat]] of the incoming object form system metadata, and use this to determine how to construct the [DataPackage]. Also, in some cases we may be able to use lazy loading to grab the new parts of the data package as needed -- ie, only download science metadata or data objects as needed by calls to [[DataPackage]] accessor methods. This might be based on the size of the object (fine to cache small objects).

Under the new design (ORE packaging), the behavior probably changes somewhat - given a pid, it needs to make sure the format of the pid is that of a ResourceMap (RDF/XML), and throw an exception if not so.

Lazy loading leads back to expanding the underlying Java libclient D1Object class to be able to represent the object bytes in a flexible manner - download should hold a "D1RemoteDataSource", for example.

History

#1 - 2010-10-06 20:04 - Matthew Jones

- Target version deleted (CCI-0.6)

- Milestone set to CCI-0.6

- Parent task set to #927

#2 - 2010-10-07 12:25 - Dave Vieglais

- Start date set to 2010-10-07

- Tracker changed from Bug to Task

#3 - 2012-10-11 16:37 - Dave Vieglais

- Target version changed from Sprint-2012.39-Block.5.4 to Sprint-2012.41-Block.6.1

#4 - 2012-10-19 22:48 - Chris Jones

- Milestone changed from CCI-0.6 to None

- Assignee changed from Matthew Jones to Rob Nahf

Assigning to Rob.

#5 - 2012-10-24 18:23 - Matthew Jones

- Target version changed from Sprint-2012.41-Block.6.1 to Sprint-2012.44-Block.6.2

#6 - 2012-11-14 18:30 - Rob Nahf

- % Done changed from 0 to 40

- Description updated

#7 - 2012-12-12 18:58 - Rob Nahf

- Target version changed from Sprint-2012.44-Block.6.2 to Sprint-2012.50-Block.6.4

#8 - 2012-12-12 19:16 - Rob Nahf

For a given pid, getD1Object can return either a data object, metadata object, or resourceMap object (the packaging mechanism)
The desired behavior is probably to return a DataPackage, in order to get the appropriate functionality, but that requires downloading all of the member objects, which could be unexpectedly memory and time intensive.

Implementing this should probably wait for a refactor in d1_libclient_java to lazy-load (proxy) the content of the member objects, which mitigates this issue. Alternatively, an option could be added to "download all", or provide a list of identifiers of which to download, as well as create a downloadMembers method to allow a more interactive approach.

#9 - 2013-02-22 18:58 - Matthew Jones

- Status changed from New to Closed

- % Done changed from 40 to 100

- translation missing: en.field_remaining_hours set to 0.0

Agreed -- this feature should probably wait until the underlying libclient supports lazy loading. So, the support for DataPackages can be considered complete for the purposes of release 1.0.0. Closing.

#10 - 2013-02-22 19:02 - Matthew Jones

- Target version deleted (DataONE R Client 1.0.0)

#11 - 2013-02-22 19:09 - Matthew Jones

- Target version set to DataONE R Client 1.0.0