DataONE R Client - Feature #3639

Create DataPackage from ID of metadata or data objects

2013-03-05 20:09 - Matthew Jones

Status:	Closed		Start date:	2013-03-05
Priority:	Normal		Due date:	2014-09-10
Assignee:			% Done:	100%
Category:			Estimated time:	0.00 hour
Target version:	DataONE R Client 1.1.0			
Milestone:	None		Story Points:	
Description				
Related issues:				
Related to Infrastructure - Story #3448: add (client) methods to traverse obs...			**Closed**	

History

#1 - 2013-03-05 20:09 - Matthew Jones

- Due date set to 2013-03-05

- Start date set to 2013-03-05

- Target version set to DataONE R Client 1.1.0

#2 - 2013-03-05 20:13 - Matthew Jones

Currently DataPackage instances can be created with the ID of the package of interest. However, users often have the ID of a metadata object or data object -- provide a methods in D1Client to get DataPackage given the ID of one of its components. The difficult part of this is to determine which package to instantiate when an object is a member of multiple packages. One common case will be when there are multiple versions of a package due to edits in other files in the package; in this case, the best algorithm is probably to return the most recent revision in the package obsolescence chain that contains the id of interest. Other cases where an object is contained in multiple packages that are not in the same versioning chain are harder to decide on what to return.

#3 - 2013-03-06 19:58 - Rob Nahf

- Due date changed from 2013-03-05 to 2013-03-06

Yes, unfortunately the number of packages that a data / metadata object can be part of only increases over time, and membership is out of the control of the object owner, which further complicates what is meant by the "latest package" for a given object. So, we'd have to apply some limits so that the package that gets returned is easily understood and expected.

I think that if we can rely on the obsoletes chain of resource maps to meaningfully define what replaces what, we can move forward. Users could decide whether they want the latest package containing that id, or the latest package containing that id or it's replacement.   In the latter case, I'm not sure what to do if the object is archived instead of replaced.

I think for the first, we follow the obsoletes chain of the object to the latest, then find it's package with a solr query, and for the second, we do the solr query, then follow the resourceMap's obsoletes chain to the latest.

#4 - 2013-03-06 20:06 - Rob Nahf

fyi, in anticipation of this need, I put in place a client-side method in d1_libclient_java to build an ObsoletesChain for a given object. D1Client.listUpdateHistory(Identifier pid).

Ben relates that he has tried a client-side implementation for Morpho and it is slow, so a server-side impl is probably needed. But in the meantime, what's already in libclient v1.2.0 (and included in dataonelibs project) should get the job done.

#5 - 2014-09-10 20:53 - Matthew Jones

- Status changed from New to Closed

- Due date changed from 2013-03-06 to 2014-09-10

Moved issue to github: https://github.com/DataONEorg/rdataone/issues/38