Kepler - Bug #4160

EML - Dynamic data retrieval

06/15/2009 02:48 PM - ben leinfelder

Status:	Resolved		**Start date:**	06/15/2009
Priority:	Normal		**Due date:**	
Assignee:	ben leinfelder		**% Done:**	0%
Category:	data access		**Estimated time:**	0.00 hour
Target version:	2.0.0-alpha1		**Spent time:**	0.00 hour
Bugzilla-Id:	4160			

Description

The existing EML actor is able to get the latest revision of a datapackage.
We need to augment EML data access and allow the actor to emit a *set* of datatables rather than just one.
A downstream actor can stack them as needed - hopefully that's manageable for the 2a1 release.
I've discussed with Jim, and I believe we have a decent approach.

History

#1 - 06/17/2009 12:37 PM - ben leinfelder

new data output format emits information about each datatable (including the pointer to the cache file).
This allows R actor downstream to read and combine similar datasets into a complete set (annual tables, for example).
Combined with the existing "check for latest revision" this should get us close to the "dynamic" data needs for TPCs

#2 - 06/18/2009 08:38 AM - ben leinfelder

testing this new output format with R actor downstream - looks promising, but deferring to Jim's R expertise to get the read.table() call corrected.
Running into "more columns than column names" type of errors when constructing the dataframe for each table.

#3 - 06/18/2009 05:15 PM - ben leinfelder

looks like we got this going in terms of getting the existing tables.
next up: check for latest revision!

#4 - 06/19/2009 12:42 PM - ben leinfelder

Jim has this working in the riverflow tpc.
And I've recently made the "check for latest" feature of EML work as it needs to.
Looking pretty good for dynamic data.

#5 - 03/27/2013 02:25 PM - Redmine Admin

Original Bugzilla ID was 4160