Kepler - Bug #3573

Support for importing file contents automatically using CollectionSource

10/27/2008 11:32 AM - Timothy McPhillips

Status:	New		**Start date:**	10/27/2008
Priority:	Normal		**Due date:**	
Assignee:	Timothy McPhillips		**% Done:**	0%
Category:	general		**Estimated time:**	0.00 hour
Target version:	3.X.Y		**Spent time:**	0.00 hour
Bugzilla-Id:	3573			

Description

The CollectionComposer and CollectionReader actors extend CollectionSource to read XML representations of the input to a COMAD workflow and translate them into data tokens, metadata tokens, collection delimiters, etc. Presently all data read in by CollectionComposer must be contained in the XML that is provided either as a parameter value to CollectionComposer or as an external file to CollectionReader. However, many workflows use data from other files and this data currently must be read and parsed by explicit actors elsewhere in the workflow. The input to a workflow would be clearer, and workflows simpler and more transparent, if files could be referred to in the XML processed by CollectionSource, and if CollectionSource were to automatically include the contents of these files in the workflow input.

A simple first step would be to enable CollectionComposer to read in text files either as a TextFile collection containing a single StringToken holding the contents of the file, or a TextFile collection containing one StringToken for each line of the text file. (Existing COMAD workflows demonstrate the usefulness of both approaches).

A second step would be to allow one to register format-specific parsers for CollectionSource to use when reading particular types of files. For example, a FASTA file parser could be plugged in that would create a FASTA collection filled with (e.g., DNA) Sequence tokens, and a Nexus file parser could create a Nexus collection containing CharacterMatrix, WeightVector, and phylogenetic Tree tokens.

Related issues:

Blocked by Kepler - Bug #3574: Support for importing directory contents using...	**New**	**10/27/2008**

History

#1 - 03/27/2013 02:23 PM - Redmine Admin

Original Bugzilla ID was 3573