Kepler - Bug #3574

Support for importing directory contents using CollectionSource

10/27/2008 11:49 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:	3574			

Description

A common workflow pattern is to take as input all of the files (or those of a particular type) in a directory on a researcher's computer system. For example, there are COMAD workflows that process all the FASTA files in a directory, creating a collection for each FASTA file and storing the contained DNA or protein sequences in the corresponding input collections.

Once the CollectionSource actor is able to automatically import the contents of files (see bug 3573), it will be extremely useful to refer to directories in the XML input to CollectionReader or CollectionComposer and have the actor import all of the files it finds there. Another useful feature would be the option of having CollectionSource descend into sub-directories, creating a nested collection for each and importing contained files into the corresponding subcollections. Whole directories of scientific data files could then easily serve as input to COMAD workflows.

These features eventually could make it much easier to stage data for input to a workflow run without requiring modification of the workflow specification itself.

Related issues:

Blocks Kepler - Bug #3573: Support for importing file contents automatically ...	**New**	**10/27/2008**

History

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

Original Bugzilla ID was 3574