



dataset acquisition
accessibility & annotation
e-research technologies

DMQ5 – Kepler

Building an Intelligent Storage
Framework using Kepler.

Tristan King



Purpose of DMQ5

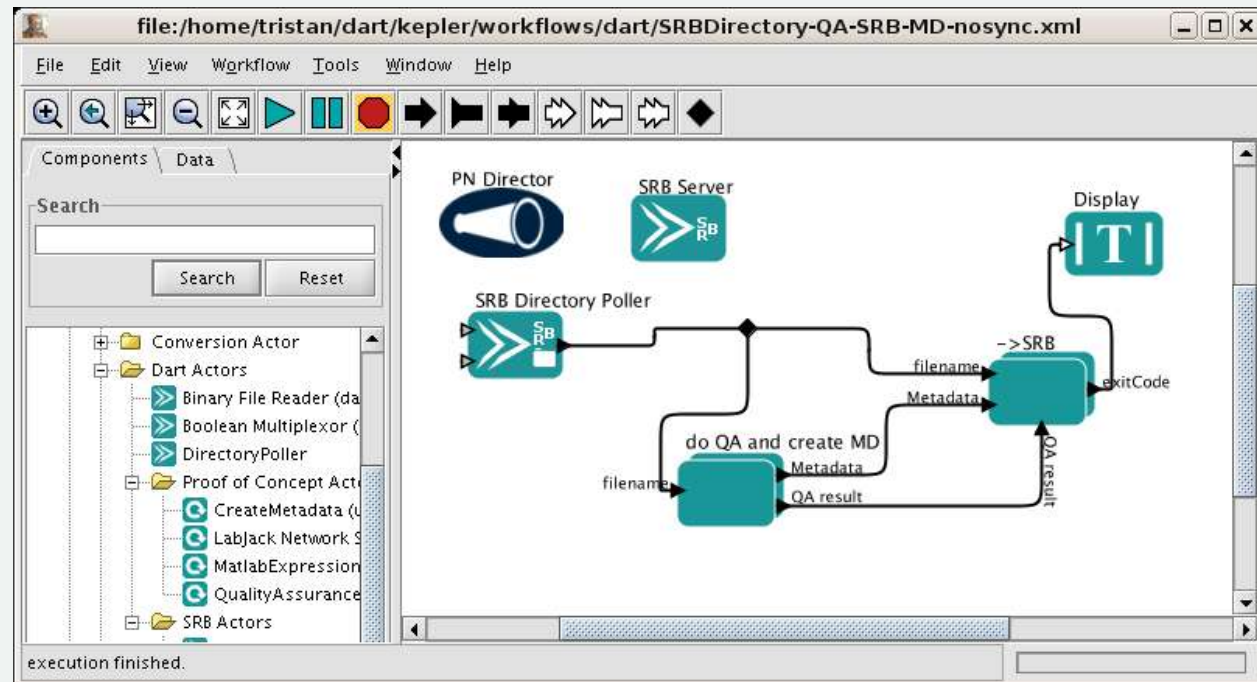
- Improve the intelligence of the storage framework.
- Implement triggers in SRB.

Current Direction

- Original Plan was to modify the SRB source
- Decided to use Kepler to preform this task.

Kepler

- Scientific Workflow system.
- Visual Programming
- Actor Orientated.
- Grid friendly



DMQ5 - Kepler

3



What we've done with Kepler

- Built proof of concept actors and workflows.
 - Designed Directory Polling system
 - Redesigned Kepler SRB actors
 - Designed Proof of concept Actors
 - Quality Assurance
 - Metadata creation
 - Built workflow using these components.



Directory Polling

- Watch a Directory for new or modified files.
- Works on both local filesystems and SRB.
- Problems
 - How can we tell when a file has finished copying to the filesystem?
 - Temp solution: make sure files haven't been modified for a specific amount of time.
 - Is this good enough?



SRB Actors

- Preform tasks involving access to SRB.
- Why re-write them?
 - Current Kepler Actors didn't work in required environment.
- Current actors:
 - SRB server
 - SRB Directory Poller
 - Smv, Sput
 - SRB File Reader
 - SRB Add Metadata

DMQ5 - Kepler

6



Proof of concept Actors

- Built to Prove Quality Assurance and Metadata Creation could be preformed using kepler.
- Concept scenario:
 - data files containing a random number of alphabet characters are created.
 - Quality Assurance
 - make sure an instance of every letter of the alphabet appears in the file.
 - Metadata Creation
 - count the number of times each letter appears.



Proof of concept Workflow

- Directory Poller watches a raw data directory on SRB.
- When new files are found:
 - Perform Quality Assurance and Metadata Creation
 - If QA test passes:
 - move data file to storage directory
 - write metadata
 - If QA test fails:
 - move data file to trash directory



What's Next?

- Explore Grid Functionality in Kepler.
- Work with CIMA to develop a replacement for their data manager.