Sumanariu Raluca

Group 258

Kepler – Workflow tool

1. What is Kepler?

Kepler is an application for analyzing and modelling scientific data. The Kepler Project's overall goal is to produce an open-source scientific workflow system that allows scientists to design workflows and execute them efficiently. Using Kepler's graphical interface and components, scientists with little background in computer science can: create executable scientific workflows, execute complex analysis on data, can exchange, archive and version the workflows.

2. Installation & usability

The system requirements for installing and using Kepler are the following: 300 MB of disk space, 512 MB of RAM minimum, 1 GB or more recommended, 2 GHz CPU minimum, Java 7 or greater.

The tool can be downloaded from: http://kepler-project.org

The application is cross-platform; it can run on either Windows, Mac or Linix.

3. History of the Kepler project

Kepler was founded in 2002 by researchers at the National Center for Ecological Analysis and Synthesis (NCEAS) at University of California Santa Barbara, the San Diego Supercomputer Center (SDSC) at University of California San Diego, and the University of California Davis as part of the Science Environment for Ecological Knowledge (SEEK) and Scientific Data Management (SDM) projects.

The Kepler software extends the Ptolemy II system developed by researchers at the University of California Berkeley. Although not originally intended for scientific workflows, Ptolemy II

provides a mature platform for building and executing workflows, and supports multiple models of computation.

Kepler is an open collaboration with many contributors from diverse domains of science and engineering, including ecology, evolutionary biology, molecular biology, geology, chemistry, computer science, electrical engineering, oceanography, and others.

Work was conducted with logistical support from the National Center for Ecological Analysis and Synthesis, the University of California Santa Barbara, and the State of California.

4. **Future Goals**

The Kepler project is an ongoing collaboration, and we will continue to refine, release, and support the Kepler software. The aim is to improve and enhance the Kepler scientific workflow system to yield a comprehensive, open, reliable, and extensible scientific workflow infrastructure suitable for serving a wide variety of scientific communities.

The goal of future Kepler development is to:

- enable multiple groups in a number of distinct disciplines to easily create, support, and make available domain-specific Kepler extensions;
- better support those crucial features that are needed by all disciplines;
- provide a wide range of deployment scenarios required by different disciplines and distinct research settings.

5. **Process Modeling – Components of a workflow**

- *Directors*

A director controls (or directs) the execution of a workflow, just as a film director oversees a cast and crew.

- *Actors*

Actors are the basic building blocks of workflows. Each actor is designed to perform a specific task. The actors take their execution instructions from the director. In other words, actors specify *what* processing occurs while the director specifies *when* it occurs.

- *Ports*

Each actor in a workflow can contain one or more ports used to consume or produce data and communicate with other actors in the workflow. Actors are connected in a workflow via their ports. Ports can be: input port (for data consumed by the actor), output port (for data produced by the actor) or input/output port (for data both consumed and produced by the actor).

- *Channels and Tokens*

Channels are used to pass data from one port to another. Each channel can transport a single stream of data.
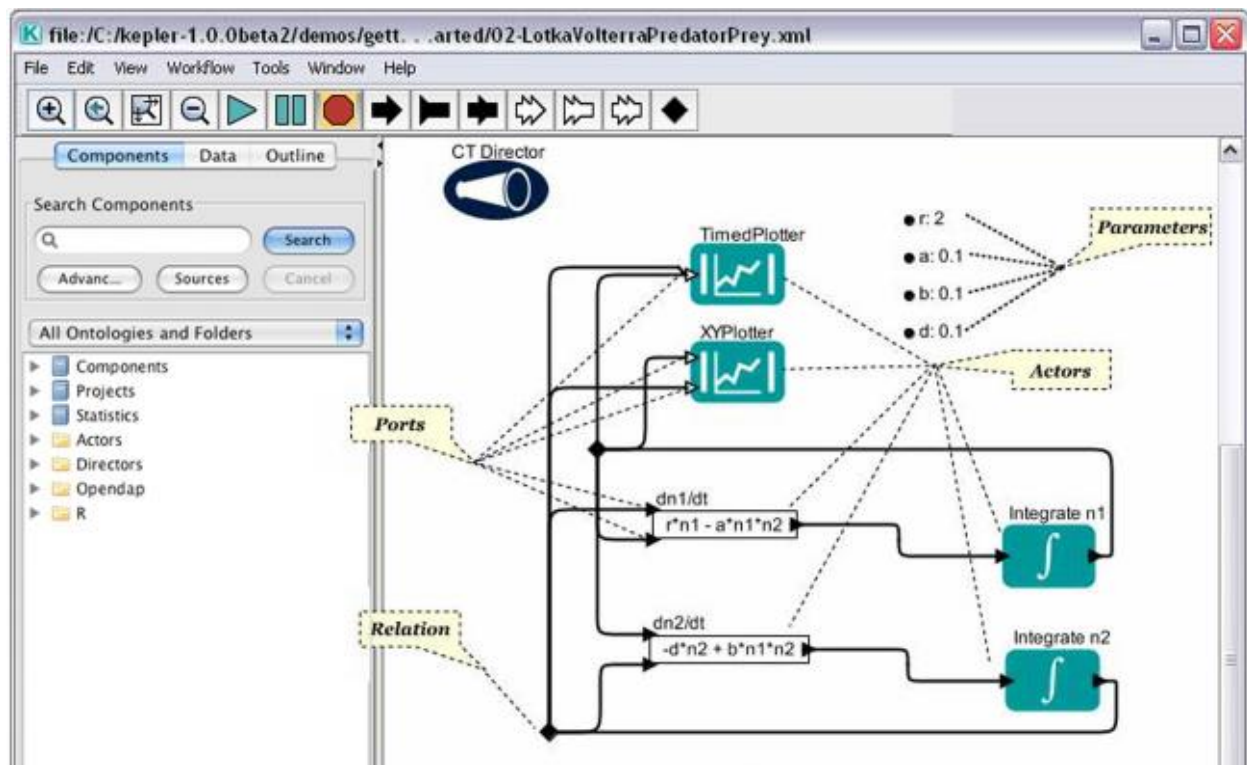
- *Data types*
- *Relations*

Relations allow workflows to "branch" a data flow. Branched data can be sent to multiple places in the workflow.

- *Parameters*

Parameters are configurable values that can be attached to a workflow (model parameters) or to individual directors or actors (coupled parameters).

Some of the components of a workflow are depicted below:

6. Advantages & Disadvantages

On the bright side we have the facts that: users with little background in computer science can create workflows with standard components, or modify existing workflows and there is no need of R programming knowledge, although it is recommended.

On the other hand, time investment is required in learning the Kepler environment and its approach to actor creation, because they bear little resemblance to the usual workflow components. Also, passing data as structures can lead to some execution time problems.

7. Bibliography

All information was taken from Kepler's Getting Started Guide (https://code.kepler-project.org/code/kepler-docs/trunk/outreach/documentation/shipping/2.5/getting-started-guide.pdf) and Kepler's User Manual (https://code.kepler-project.org/code/kepler-docs/trunk/outreach/documentation/shipping/2.5/UserManual.pdf).