



# Introduction to Kepler

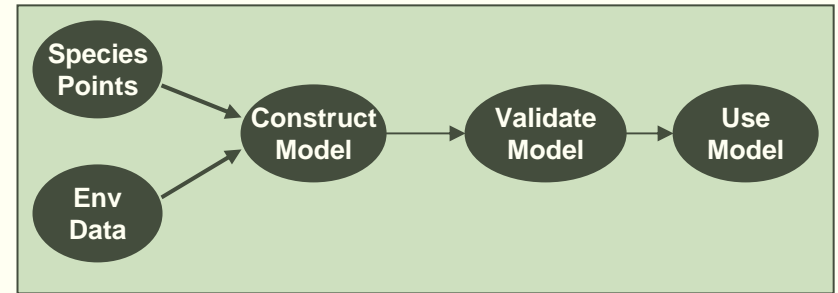
Deana Pennington  
University of New Mexico  
January 3, 2005



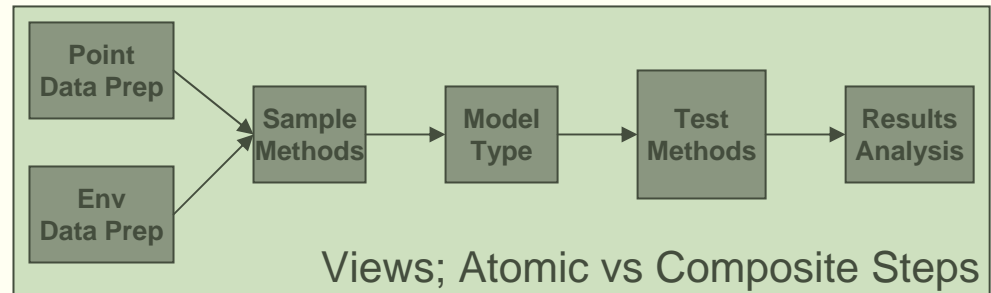


# Hierarchical Workflows

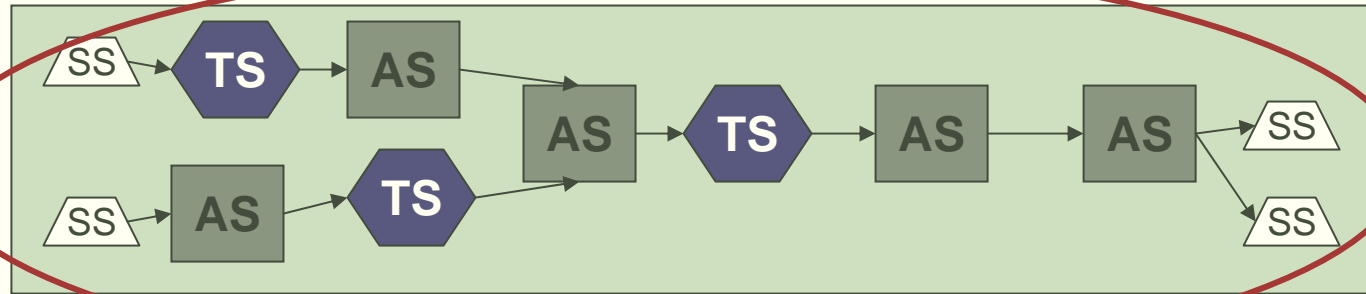
Conceptual Workflow



Abstract Workflow



Executable Workflow



C

Concept

AS

Analysis Step

TS

Transformation Step

SS

System Step





# Technology-Enabled Workflows

- ❑ Automated analysis environment (today)
- ❑ Sharing technologies (grid technologies, metadata, data models) (Tues/Wed)
- ❑ Concept formalization (ontologies) (Thurs)





# Automated Workflows

- Scripts
  - Visual modeling
- } Single platform  
Single environment
- Workflows:
    - Cross-platform
    - Cross-environment
    - Distributed data & analyses





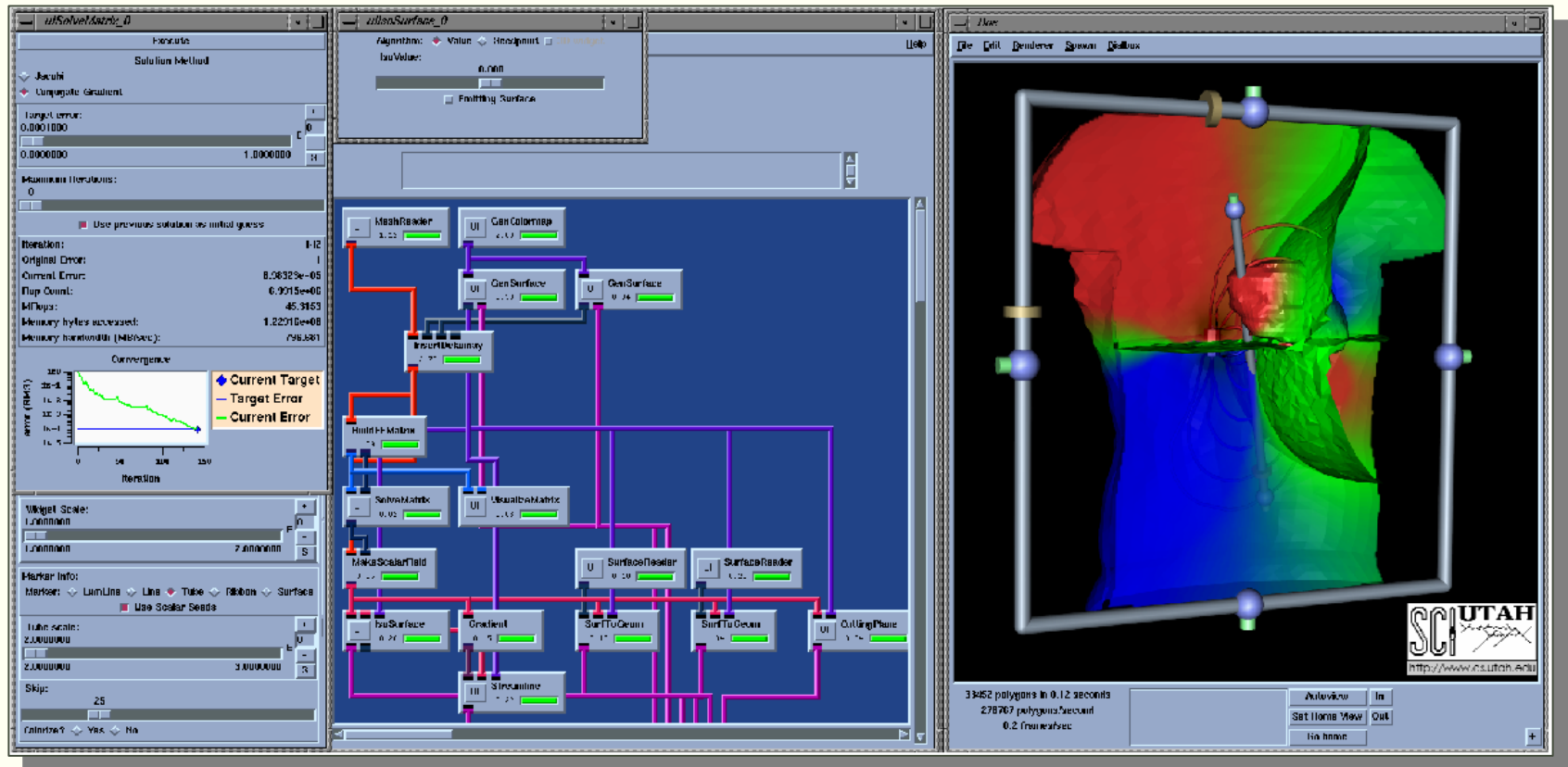
# Visual Modeling

- Benefits
  - Built on scripting approaches
    - Documentation
    - Reproducibility
    - Reusability
  - Visually intuitive
    - Closer match to abstract workflows
- Shortcomings
  - Learning curves
    - Programming logic/syntax
    - System
    - Terminology





# SCIRun: Problem Solving Environment for Large-Scale Scientific Computing



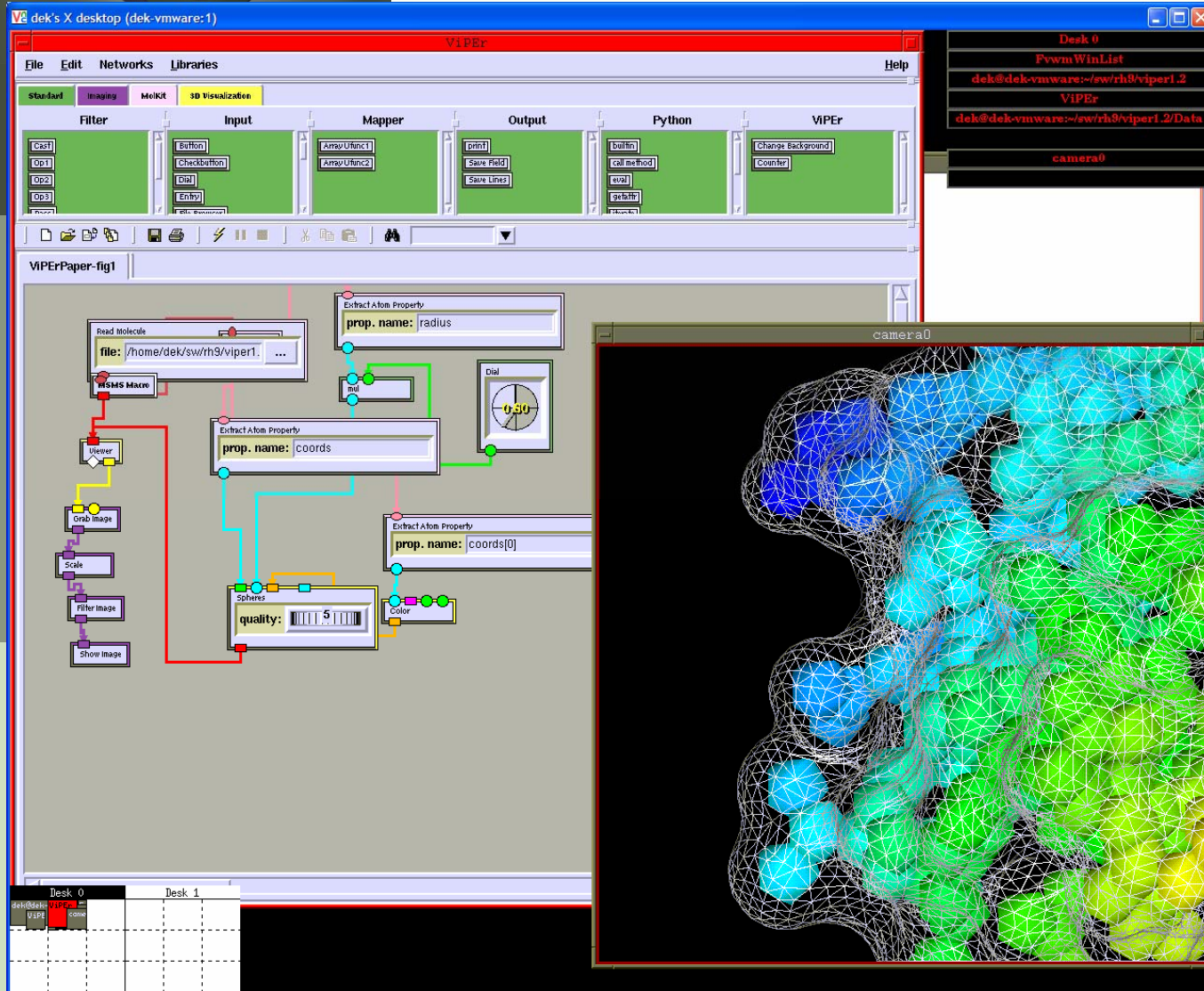
- ❑ SCIRun: PSE for interactive construction, debugging, and steering of large-scale scientific computations
- ❑ Component model, based on generalized dataflow programming

Steve Parker (cs.utah.edu)





# Viper/Vision/VIPUS



Source: Keith Jackson,  
David Konerding,  
Michel Sanner



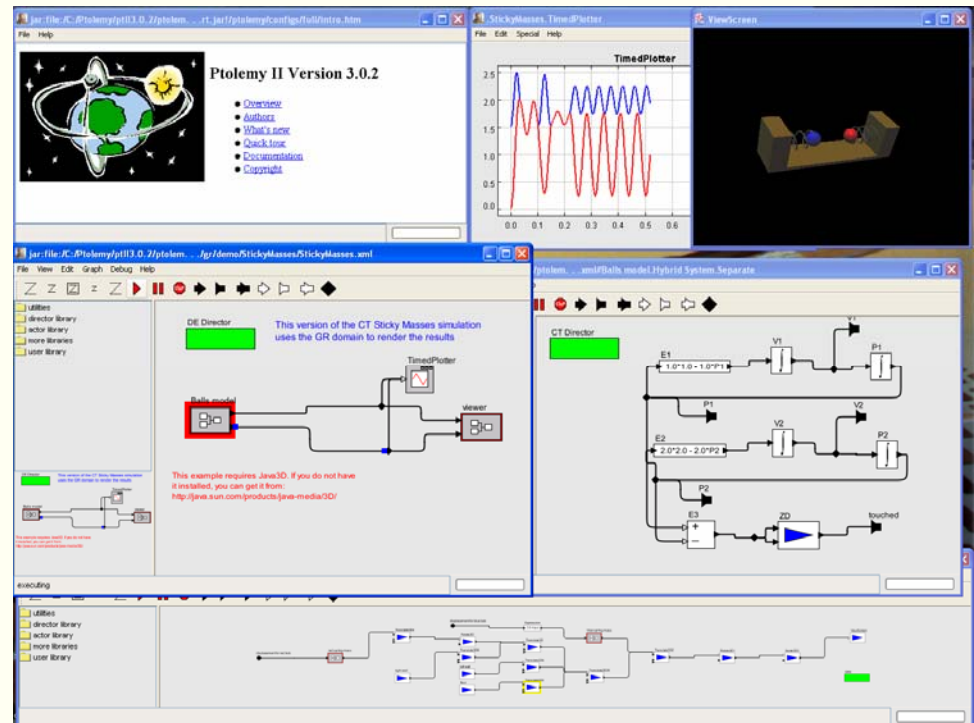




# Starting point: Ptolemy II



- Electrical engineering community
- Large mathematical library



## Ptolemy II - Heterogeneous Modeling and Design in Java

The Ptolemy project studies modeling, simulation, and design of concurrent, real-time, embedded systems. The focus is on assembly of concurrent components. The key underlying principle in the project is the use of well-defined models of computation that govern the interaction between components.

**Principal Investigator**  
Edward A. Lee

**Technical Staff**  
Christopher Hylands  
Mary P. Stewart

**Postdocs and Researchers**  
Jörn Janneck  
Sonia Sachs

**Grad Students**  
Elaine Cheong  
Chamberlain Fong  
Jie Liu  
XiaoJun Liu  
Steve Neuendorffer

Brian Vogel  
Paul Whitaker  
Yuhong Xiong







# Kepler Contributors, Projects, Sponsors



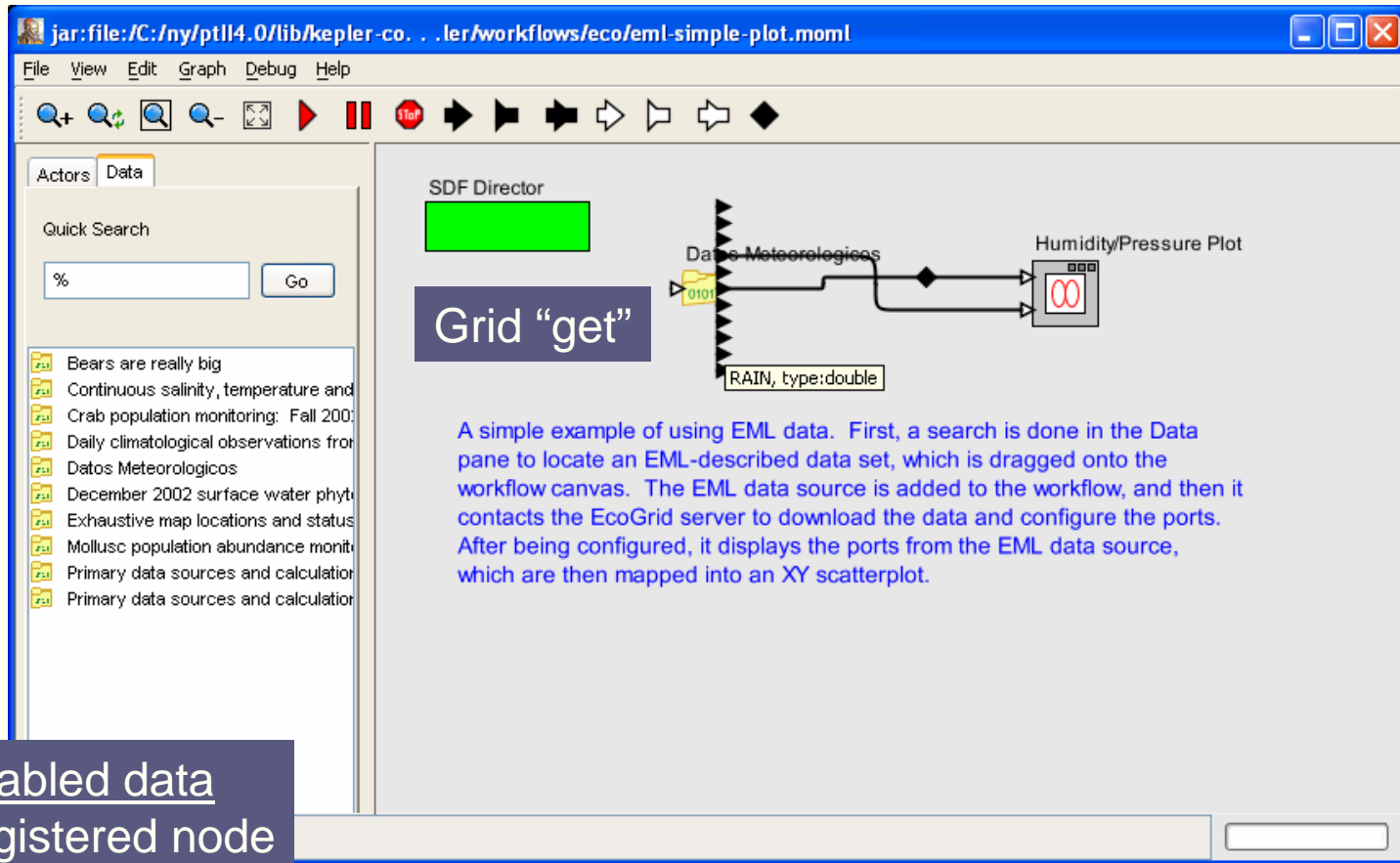
- Ilkay Altintas *SDM*
- Chad Berkley *SEEK*
- Shawn Bowers *SEEK*
- Tobin Fricke *ROADNet*
- Jeffrey Grethe *BIRN*
- Christopher H. Brooks *Ptolemy II*
- Zhengang Cheng *SDM*
- Dan Higgins *SEEK*
- Efrat Jaeger *GEON*
- Matt Jones *SEEK*
- Edward A. Lee *Ptolemy II*
- Kai Lin *GEON*
- Ashraf Memon *GEON*
- Bertram Ludaescher *BIRN, GEON, SDM, SEEK*
- Steve Mock *NMI*
- Steve Neuendorffer *Ptolemy II*
- Jing Tao *SEEK*
- Mladen Vouk *SDM*
- Xiaowen Xin *SDM*
- Yang Zhao *Ptolemy II*
- Bing Zhu *SEEK*
- ...



E-Science Link-up Project



# Grid-enabled data



**Grid "get"**

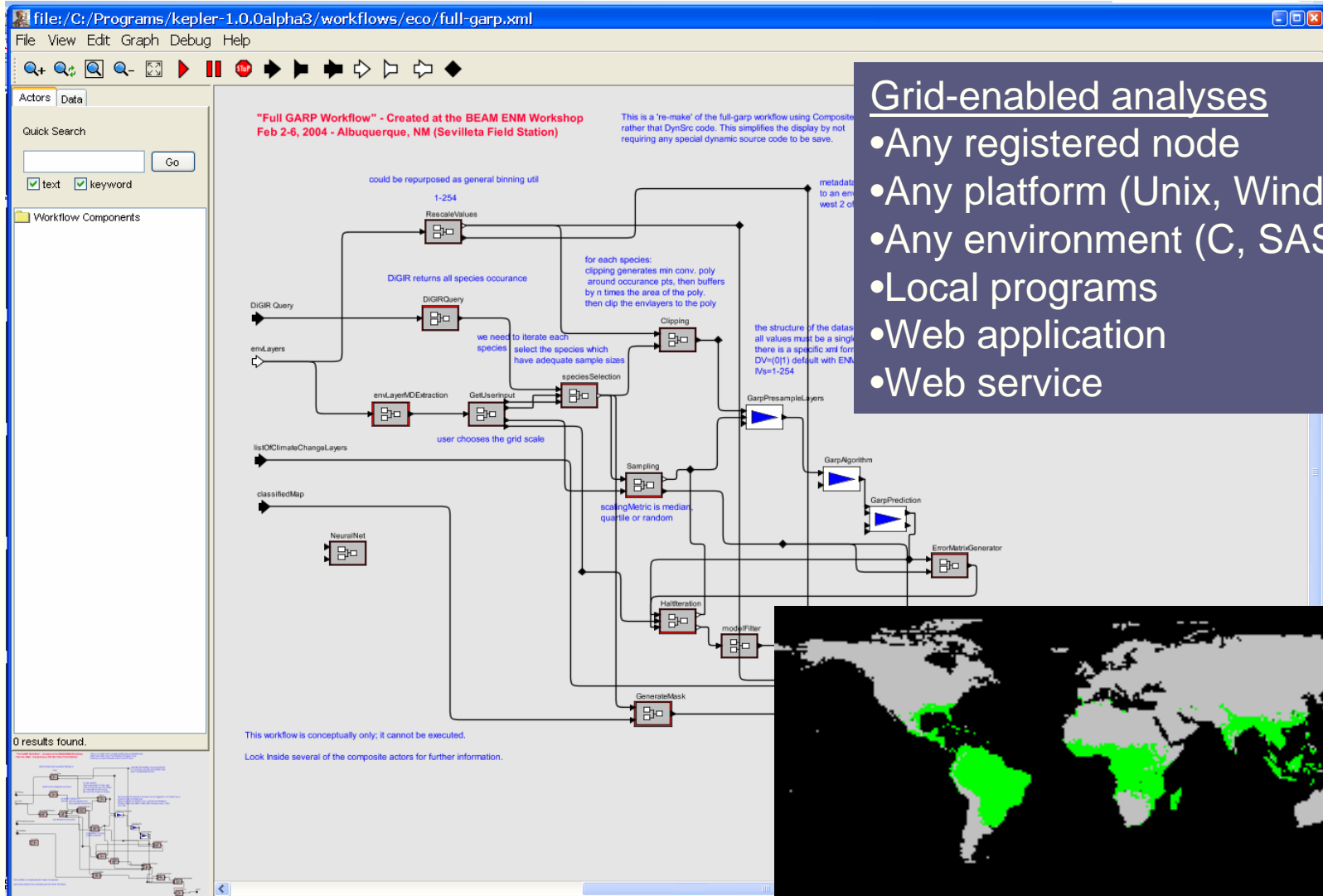
A simple example of using EML data. First, a search is done in the Data pane to locate an EML-described data set, which is dragged onto the workflow canvas. The EML data source is added to the workflow, and then it contacts the EcoGrid server to download the data and configure the ports. After being configured, it displays the ports from the EML data source, which are then mapped into an XY scatterplot.

## Grid-enabled data

- Any registered node
- Metadata driven
- Ontology-based



# Kepler Workflow System



## Grid-enabled analyses

- Any registered node
- Any platform (Unix, Windows, Mac)
- Any environment (C, SAS, GIS)
- Local programs
- Web application
- Web service





# Director/Actor Metaphor

Director

Actor

Actor

Actor

Actors know HOW to act..know their part  
Directors know WHEN they should act

## Examples:

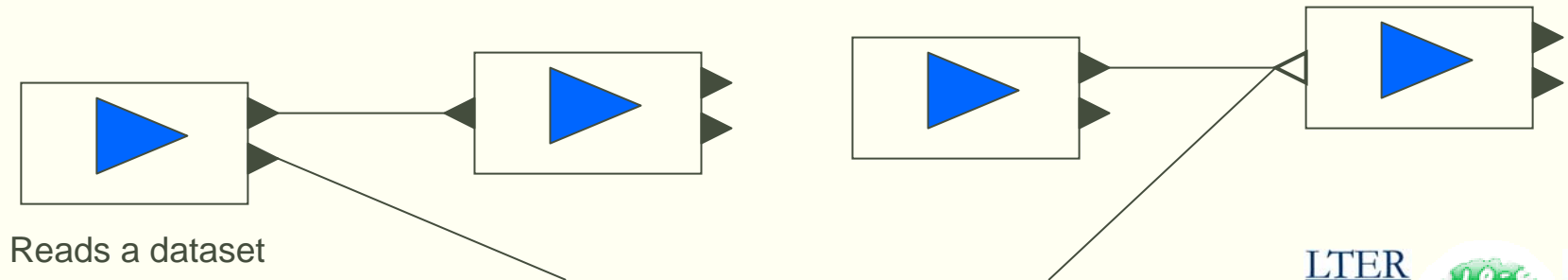
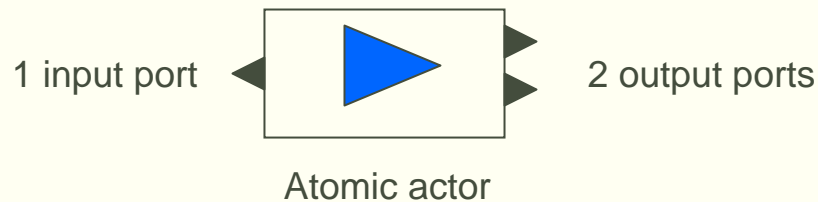
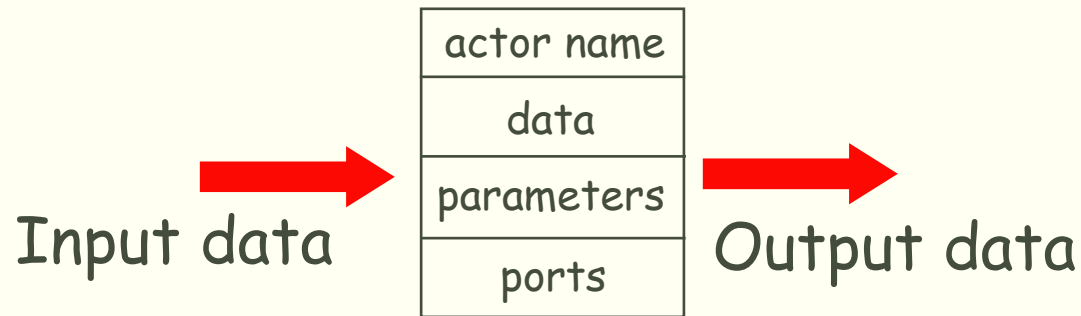
Process Network: procedural, single point in time

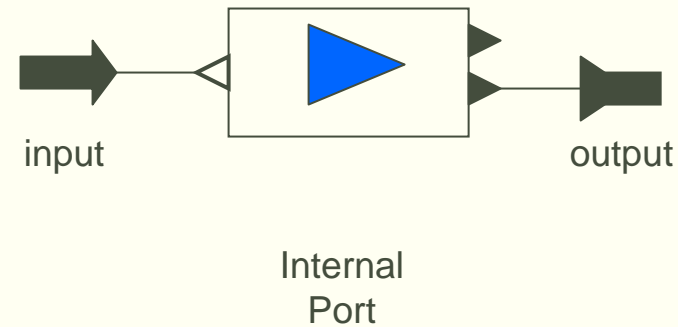
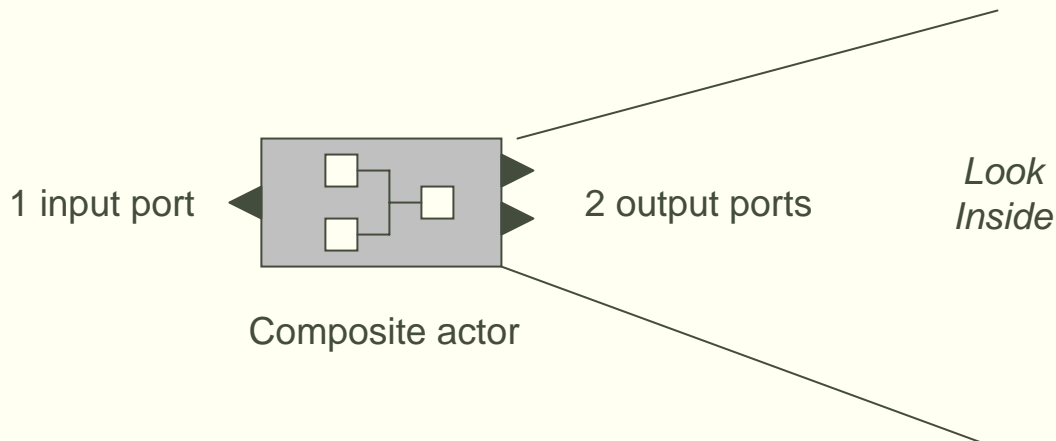
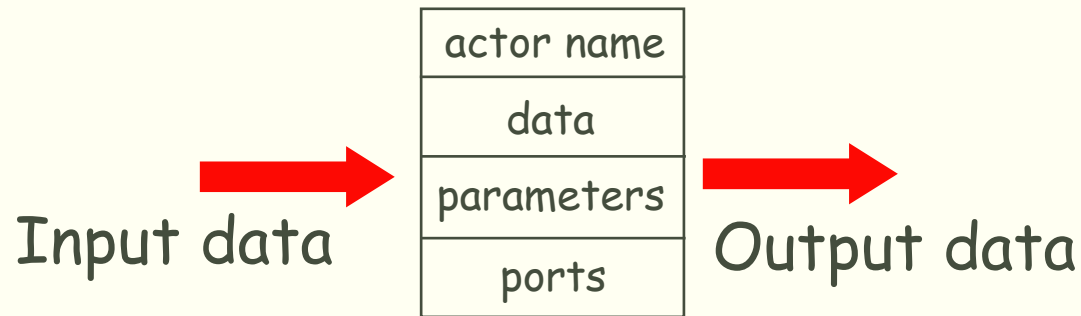
Synchronized Data Flow: subset of Process Net

Continuous Time: all points in time

- ❑ "Models of computation"
- ❑ "Behavioral polymorphism"



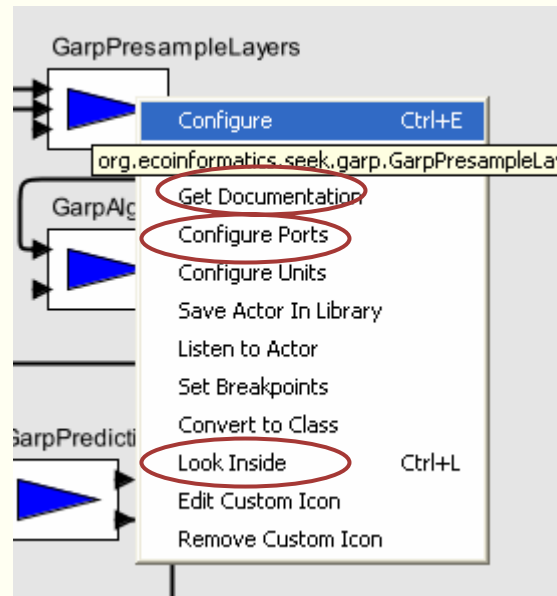








# Right-click menu





# Editing parameters

Double-click or right-click

**Edit parameters for Delimited Parser**

? outputFormat: column  
delimiter: """  
numberColumns: 12  
numberHeaderLines: 1  
firingsPerIteration: 1

Commit Add Remove Restore Defaults Preferences Help Cancel

0 to many



# Configuring Ports

Right-click

Name	Input	Output	Multiport	Type	Direction	Show Name	Hide	Units
input	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		DEFAULT	<input type="checkbox"/>	<input type="checkbox"/>	
output	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>		DEFAULT	<input type="checkbox"/>	<input type="checkbox"/>	
delimiter	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		DEFAULT	<input type="checkbox"/>	<input type="checkbox"/>	
numberColumns	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		DEFAULT	<input type="checkbox"/>	<input type="checkbox"/>	
numberHeaderLi...	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		DEFAULT	<input type="checkbox"/>	<input type="checkbox"/>	
stringOutput	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	{string}	DEFAULT	<input type="checkbox"/>	<input type="checkbox"/>	

Commit Apply Add Remove Help Cancel

User-defined

String  
Int  
Double  
{ } array

8, 4, 73, 12  
{string}  
1 element  
4 positions (0-3)  
{3}





# Procedure

---

- ❑ Open a new workflow
- ❑ Add a director
- ❑ Search for data (optional)
- ❑ Add data source (optional)
- ❑ Add an actor
- ❑ Edit parameters
- ❑ Add ports (if needed)
- ❑ Configure ports
- ❑ Add another actor
- ❑ Hook up input/output ports





## □ Kepler Exercises





# Acknowledgements

This material is based upon work supported by the National Science Foundation under awards 0225676 for SEEK and 0225673 (AWSFL008-DS3) for GEON and by the Department of Energy under Contract No. DE-FC02-01ER25486 for SciDAC/SDM and by DARPA under Contract No. F33615-00-C-1703 for Ptolemy. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

The National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant Number 0072909), the University of California, and the UC Santa Barbara campus.

The Andrew W. Mellon Foundation.

PBI Collaborators: NCEAS, University of New Mexico (Long Term Ecological Research Network Office), San Diego Supercomputer Center, University of Kansas (Center for Biodiversity Research)

Kepler contributors: SEEK, Ptolemy II, SDM/SciDAC, GEON

