Knowledge Annotations in Scientific Workflows:An Implementation in Kepler

Aída Gándara1, George Chin2, Paulo Pinheiro da Silva1, Terence Critchlow2, Chandrika Sivaramakrishnan2, Signe White2

1The University of Texas at El Paso
2Pacific Northwest National Laboratory

PNNL-UTEP Research

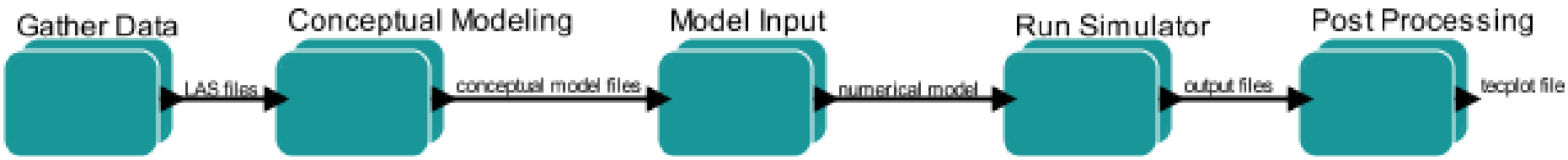
- Collaborative Team:
  - SciDAC Scientific Data Management Center at Pacific Northwest National Laboratory
  - Cyber-ShARE Research Center at UTEP
  - August – October 2010
- Collaboration Purpose
  - To help groundwater scientists at PNNL manage collaborative data that is traditionally generated during a research effort but not preserved as part of the research effort

Research Goals

- Generic Goals
  - Understand collaborative research processes before developing a workflow for it
  - Understand needs for documenting research collaboration
- Specific Goal
  - Use the Kepler Scientific Workflow System as a way of understanding a research process at PNNL

Case Study

- Subsurface Flow and Transport Analysis
  - Typically members include: project manager and several team members.
  - Each step requires expertise, e.g., groundwater scientists use STOMP and other software
  - Collaboration between steps

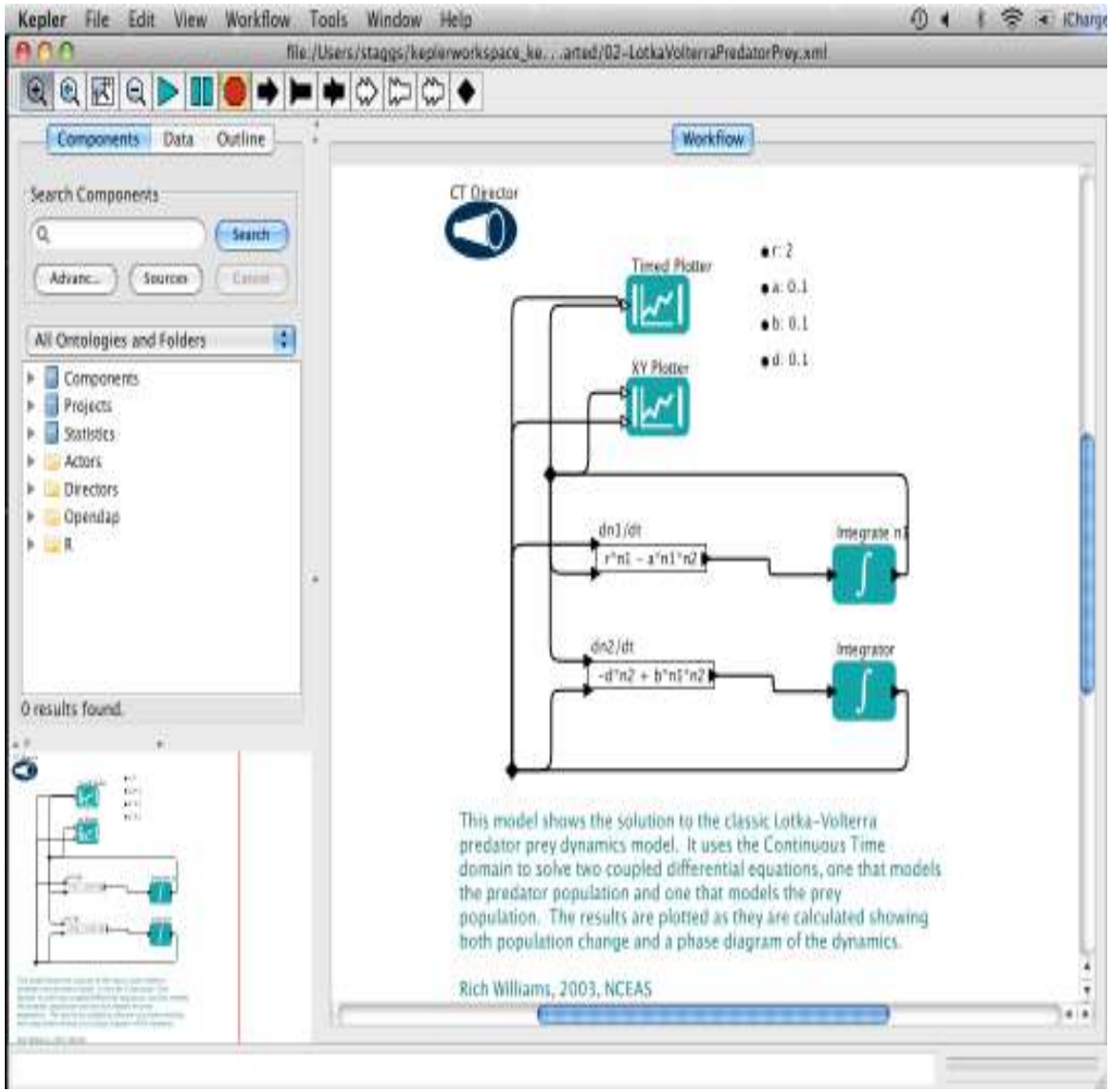


Some Observations

- At some point, scientists seek to understand the “hows” and “whys” of scientific results
- Scientists keep journals and notes of what worked and what did not, e.g., decisions, assumptions and constraints
- Much of this information is needed for final reports and publications

Scientists often need to capture their notes about ad hoc processes, not processes predefined in a workflow

Kepler Scientific Workflow System



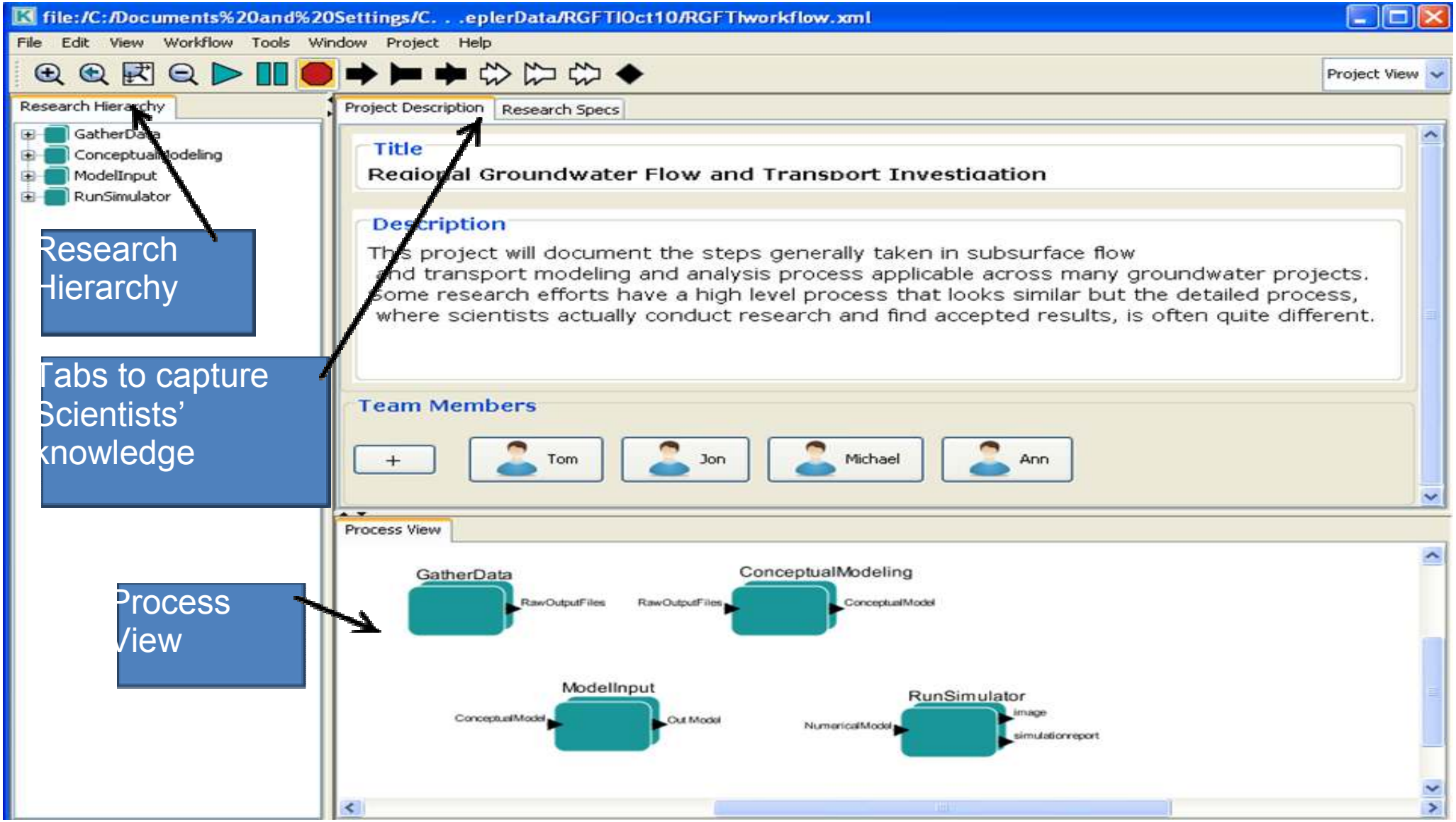
- Collect sufficient information to document a scientific process
- Support reproducing results
- Help collect provenance

From Kepler getting started guide, the Lotka-Volterra Workflow

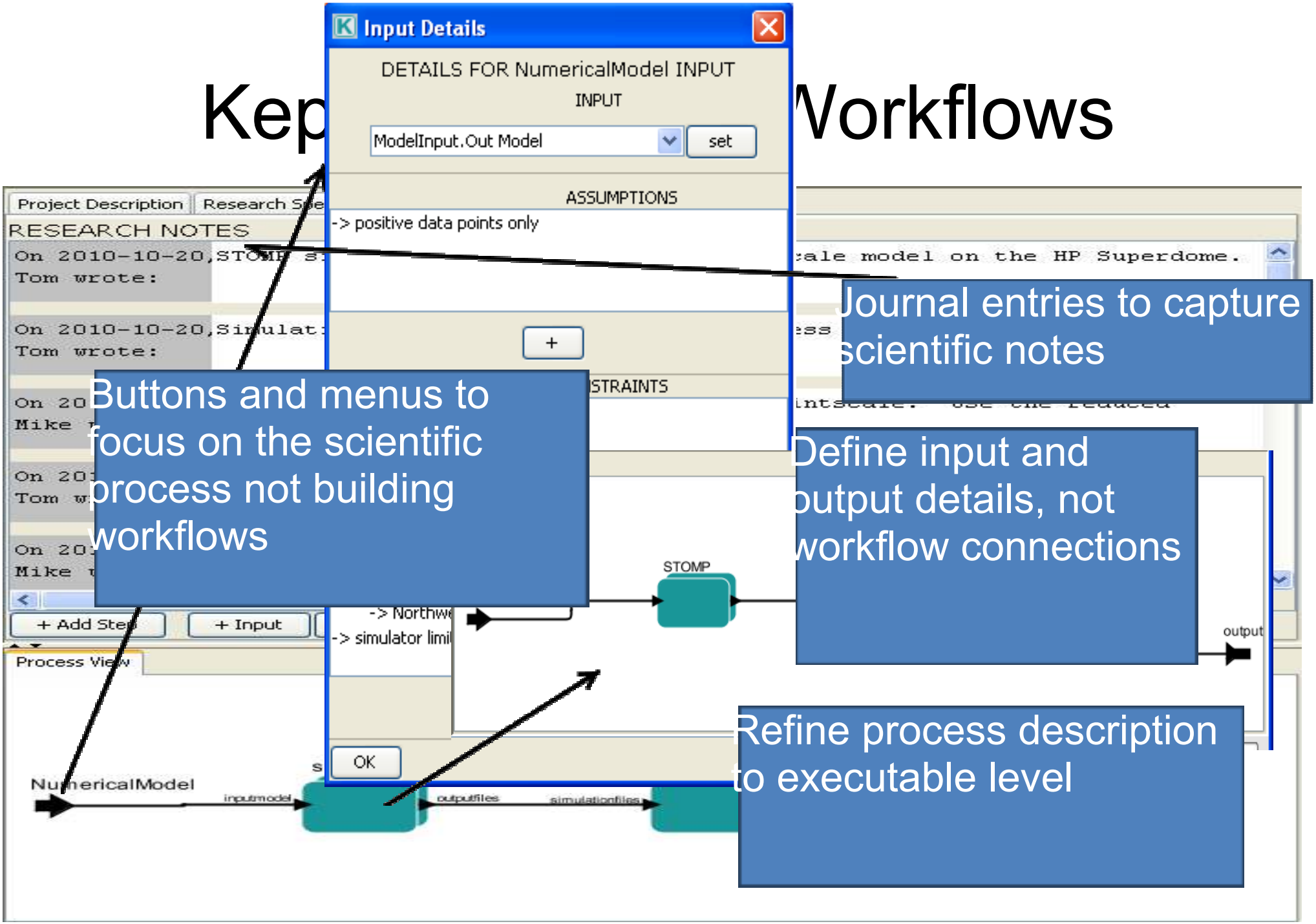
Knowledge-Annotated Scientific Workflows :design principles

1. Scientists describe their research: build workflow from information
2. Align with scientific research process: reduce duplication and alteration of process
3. Leverage workflow to manage annotations: annotations relate to actors and connections in workflow

Knowledge-Annotated Kepler Workflow System



Kep Workflows



Results

- Scientists do not add workflow components
  - steps, journal entries, inputs/outputs, assumptions, constraints, comments …
- Various views of the data:
  - Research summary report
  - Process traversal (forward and back through inputs/outputs)
  - Status of a step
  - RDF output that links to workflow information in SIOC

Current Status

- kadm.jar with features identified in research
- Embedded in UTEP CyberShARE tools for use by environmental scientists and geoscientists
- Building RDF specific to research teams with annotations, workflows and data
- Evaluation of process and data

Conclusions

- Workflows not always intuitive
- Some scientists feel workflows are too rigid
- This research has presented an alternative method for scientists to create and annotate an ad hoc scientific workflow

Contact

Aída Gándara
The University of Texas at El Paso
agandara1@miners.utep.edu

George Chin
Pacific Northwest National Laboratory
George.Chin@pnnl.gov