Capturing provenance information with Kieker.WorkflowMonitor

Peer C. Brauer
Wilhelm Hasselbring

Software Engineering Group
Christian Albrechts Universität
Agenda

- PubFlow
- An introduction to Kieker
- Collecting provenance data with Kieker
PubFlow
The project group

Software Engineering Group
Library of CAU
Computing Center of CAU

Kiel Datamanagement Team
Library of Geomar
Data and Computing Center Geomar
What is PubFlow about?

Publication Workflows
What is PubFlow about?

The goal of the project is it, to create a workflow environment for the work with scientific data based on established business workflow systems like Apache ODE, which increases the degree of automation in the publication process.
What is PubFlow about?

The goal of the project is it, to create a workflow environment for the work with scientific data based on established business workflow systems like Apache ODE, which increases the degree of automation in the publication process.
What is PubFlow about?

The goal of the project is it, to create a workflow environment for the work with scientific data based on established business workflow systems like Apache ODE, which increases the degree of automation in the publication process.
Observation
Observation

Pre-processing

Simulation & Analysis

Institutional Data Repository

> PubFlow
Observation ➔ Institutional Data Repository

Observation ➔ Pre-processing ➔ Simulation & Analysis ➔ Post-processing
Observation → Institutional Data Repository

Pre-processing → Visualization

Simulation & Analysis → Research Paper

Post-processing
Observation → Institutional Data Repository → Data Curation

Observation → Pre-processing

Pre-processing → Simulation & Analysis

Simulation & Analysis → Post-processing

Post-processing → Visualization

Visualization → Research Paper

Research Paper → Review

Review → Archival & Publication
Institutional Data Repository

Observation → Institutional Data Repository → Data Curation

Pre-processing → Institutional Data Repository → Visualization

Simulation & Analysis → Institutional Data Repository → Research Paper

Post-processing → Institutional Data Repository → Review

Archival & Publication

WDC
Crete, 28.05.2012

Observation → Institutional Data Repository → Data Curation

Observation → Pre-processing → Visualization → Archival & Publication

Observation → Simulation & Analysis → Research Paper → WDC

Observation → Post-processing → Review → Digital Library
the case study
Kieker

a short introduction
What is Kieker?

- Kieker is a modular monitoring framework
- developed by the Software Engineering Group
  CAU Kiel
- Kieker is open source (APL 2)
What is Kieker?

Kieker is distributed as part of SPEC® RG's repository of peer-reviewed tools for quantitative system evaluation and analysis

http://research.spec.org/projects/tools.html
What is the scope of Kieker?

- Kieker is for application monitoring - not for profiling
What is the scope of Kieker?

- Kieker is for application monitoring - not for profiling

Monitoring of an application during runtime
What is the scope of Kieker?

- Kieker is for application monitoring - not for profiling

measure application performance during development stage
Collecting provenance data with Kieker
requirements

...
Provenance

Kieker.WorkflowMonitor > Process "Hello World"

Inspect Process Run #212

- Process started
  Time 11:59 a.m. 2012-02-26
  - Process started by QUARTZ-Framework
- Sequence
  - start
  - assign1
  - end
- Process completed
  Time 12:00 a.m. 2012-02-26
  - Calling external web service

Success - Process exited without warnings
Provenance
<monitor:profile name="provenanceLight">
  <monitor:targetelement type="activity">
    <monitor:eventtype>activityEnabledEvent</monitor:eventtype>
    <monitor:eventtype>activityDisabledEvent</monitor:eventtype>
  </monitor:targetelement>
  ...
  <monitor:targetelement type="variable">
    <monitor:eventtype>VariableModificationEvent</monitor:eventtype>
    <monitor:eventtype>VariableReadEvent</monitor:eventtype>
  </monitor:targetelement>
  ...
  <monitor:event type="VariableModificationEvent">
    <monitor:fokus>varName</monitor:fokus>
    <monitor:fokus>newValue</monitor:fokus>
  </monitor:event>
</monitor:profile>