

Tool for Navigating Provenance Information

Provenance Challenge Workshop

13 - 14 September 2006

Vikas Deora, Arnaud Contes and Omer Rana

Cardiff University

Contents

- Grid Provenance Project
- Tool Overview
- Tool for Provenance Visualization
- Screen shots
- Demo
- Further information

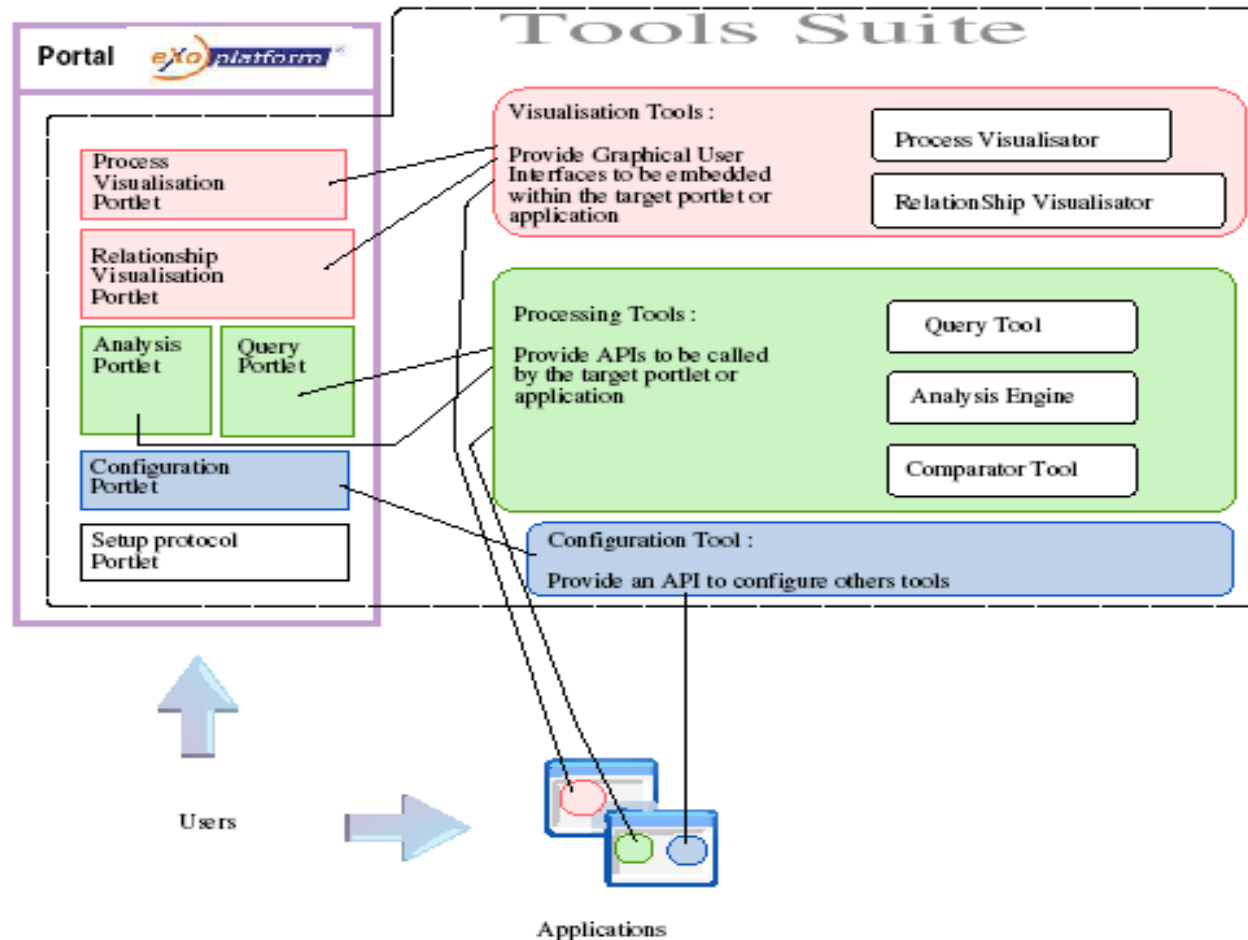
Grid Provenance Project

- **IBM United Kingdom Limited**
 - Alexis Biller
 - John Ibbotson
 - Neil Hardman
- **University of Southampton**
 - Luc Moreau
 - Paul Groth
 - Simon Miles
 - Victor Tan
 - Sheng
- **University of Wales, Cardiff**
 - Arnaud Contes
 - Omer Rana
 - Vikas Deora
- **German Aerospace Center (DLR)**
 - Andreas Schreiber
 - Guy K. Kloss
 - Tahjis
- **Universitat Politecnica de Catalunya**
 - Javier Vazquez
 - Sergio Alvarez
 - Steven Willmott
- **MTA SZTAKI Computer and Automation Research Institute, Hungarian Academy of Sciences**
 - Laszlo Varga
 - Tamás Kifor

Tool Overview

- The tool provide support to navigate through and analyze provenance information based on the use of a portal framework (eXo in our case).

Tools Overview



Concepts

- A set of provenance-aware actors involved in a process generate data about the execution. The data produced is composed of a set of *p-assertions*.
- Such set of p-assertions provide the description of the physical process.
- A p-assertion can be used to record one of the following events: an *interaction* between two actors, a *relation* between two events, or the state of an actor at a particular moment. In our system, interaction and relationship p-assertion are presently used.

Tools for Provenance Visualization

- The portal and portlets provide users with a set of tools to navigate through and analyze a set of p-assertions that represent an executed process.
- Interaction with a portal is made available using a Web client (browser).
- On receiving a user request to re-construct a process, the portlet interacts with local and external Provenance Store to retrieve all the p-assertions related to a particular process execution.

Visualization Portlet

- The visualization portlet displays two graphs:
 - process graph and
 - relationship graph
- The above are based on interaction and relationship p-assertions respectively.

Process Graph

- By capturing all the interactions that take place between actors involved in the computation of some data, one can replay an execution, analyze it, verify its validity or compare it with another execution.
- A crucial element of an interaction p-assertion is information to identify a message uniquely. Such information allows us to establish a flow of data between actors.

Relationship Graph

- While matching interaction p-assertions denote a flow of data between actors, relationships explain how data flows inside actors. Relationship p-assertions are directional since they explain how some data was computed from other data.

Screen shot – Process Visualization

The screenshot displays a web browser window with the URL `http://protactinium.cs.cf.ac.uk:8888/portal/faces/private/javiervazquez/Visualization?portal:componentId=c2212784&portal:action=changeTab&objectId=ProcessPorlet`. The page title is "Home | Visualization | Search Site | Site Map". Below the navigation bar, there are several tabs: "GetProvenanceTracePorlet", "LoadCurrentTraceFromFilePorlet", "TextVisualizationPorlet", "ProcessPorlet" (which is circled in blue), and "RelationshipPorlet".

The main content area shows a network diagram with six nodes: EHCRS1, EHCRS2, OTM:CollectPatientData, ehcrauth, EHCRS3, and OTMA3. The nodes are interconnected with numerous directed edges, representing a complex process flow. A blue arrow labeled "Change visualization" points to a "Change View" panel on the right side of the interface.

The "Change View" panel contains the following options:

- View 1: Hierarchical
- View 2: Short Distributed
- View 3: Extended Distributed
- View 4: Cluster
- View 5: Compact Tree
- View 6: Radial tree
- View 7: Tree
- Reset

Below the "Change View" panel, there are three other panels:

- Display/Hide:** Multiple Interactions, Interaction Label, Timeline
- Zoom:** Zoom - Selected Interactions, Zoom In - All Interactions, Zoom Out - All Interactions, Reset
- Archive:** Save Graph, Retrieve Graph
- Handle Iteration:** Group, UnGroup, Collapse, Expand, Expand All

At the bottom of the browser window, the footer text reads: "Copyright © 2000-2005 eXo Platform SARL".

Screen shot – Relationship Visualization

The screenshot displays a web application interface for relationship visualization. At the top, there is a navigation bar with links for Home, My Portal, Search Site, and Site Map. Below this is a toolbar containing several portlets: Get Provenance Trace Portlet, Load Current Trace From File Portlet, Text Visualization Portlet, Process Portlet, and Relationship Portlet (which is circled in blue). The main visualization area shows a grid of small thumbnails representing different relationship P-assertions. A zoomed-in view of a specific relationship is shown below, featuring three nodes: 8347, 8349, and 8345. Node 8347 is labeled as the 'Local P-assertion Id'. Node 8345 is labeled as the 'Relationship'. The relationship between 8347 and 8345 is labeled 'inResponseTo', and the relationship between 8349 and 8345 is labeled 'causedBy'. The label 'Interactions' is also present near node 8345. On the right side of the interface, there is a 'Change View' panel with a list of view options: View 1: Hierarchical, View 2: Short Distributed, View 3: Extended Distributed, View 4: Cluster, View 5: Compact Tree, View 6: Radial tree, View 7: Tree, View 8: Circular, and Reset. Below this is a 'Display/Hide' panel with options for Multiple Relationships and Relationship Label. At the bottom of the Change View panel is a 'Zoom' panel with a checked box for 'Zoom - Selected', and options for Zoom In - All Interactions, Zoom Out - All Interactions, and Reset. A 'Handle Iteration' panel is also visible at the bottom right. Annotations with arrows point to various elements: 'Thumbnail view of relationship P-assertions' points to the grid of thumbnails; 'Zoom performed' points to the zoomed-in view; 'Selective zoom for identified critical relationships' points to the 'Zoom - Selected' checkbox; 'Local P-assertion Id' points to node 8347; 'Relationship' points to node 8345; 'inResponseTo' and 'causedBy' point to the arrows between nodes; and 'Interactions' points to the label near node 8345.

Screen shot – Provenance Challenge

exo platform® Try it, Adopt it...

Home | Search Site | Site Map Welcome: provenance

Home

GetProvenanceTracePortlet LoadCurrentTraceFromFilePortlet Provenance Stores List TextVisualizationPortlet ProcessPortlet RelationshipPortlet

RelationshipPortlet

```

    graph TD
      Root["376:PARAM_alignwarp_result:object"]
      AnatomyImage["374:anatomyImage:"]
      AnatomyHeader["374:anatomyHeader:"]
      ReferenceImage["374:referenceImage:"]
      ReferenceHeader["374:referenceHeader:"]
      Root -- alignwarp --> AnatomyImage
      Root -- alignwarp --> AnatomyHeader
      Root -- alignwarp --> ReferenceImage
      Root -- alignwarp --> ReferenceHeader
  
```

PAssertions

PAssertions View

```

<void property="view">
  <string>sender</string>
</void>
</object>
</void>
<void property="result">
  <object class="java.net.URI">
    <string>http://www.ipaw.info/challenge/warp1.warp</string>
  </object>
</void>
</object>
</java>
          
```

OK

Change View

Display/Hide

Zoom

Zoom - Selected

Zoom In - All Interacti...

Zoom Out - All Interac...

Reset

Handle Iteration

Java Applet Window

Copyright © 2000-2005 eXo Platform SARL

Provenance Challenge Demo

- Demo

Further information

- <http://sprocket-comsc.grid.cf.ac.uk:8080/portal>
- Publication in Web Intelligence 2006 Conference