

Provenance

Enabling and Supporting
Provenance in Grids
for Complex Problems



IBM Hursley, University of Southampton, German Aerospace Centre,
University of Wales, Cardiff, Universitat Politecnica de Catalunya,
MTA SZTAKI

www.gridprovenance.org

Provenance

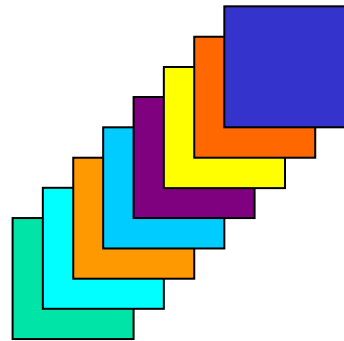


The Provenance Problem

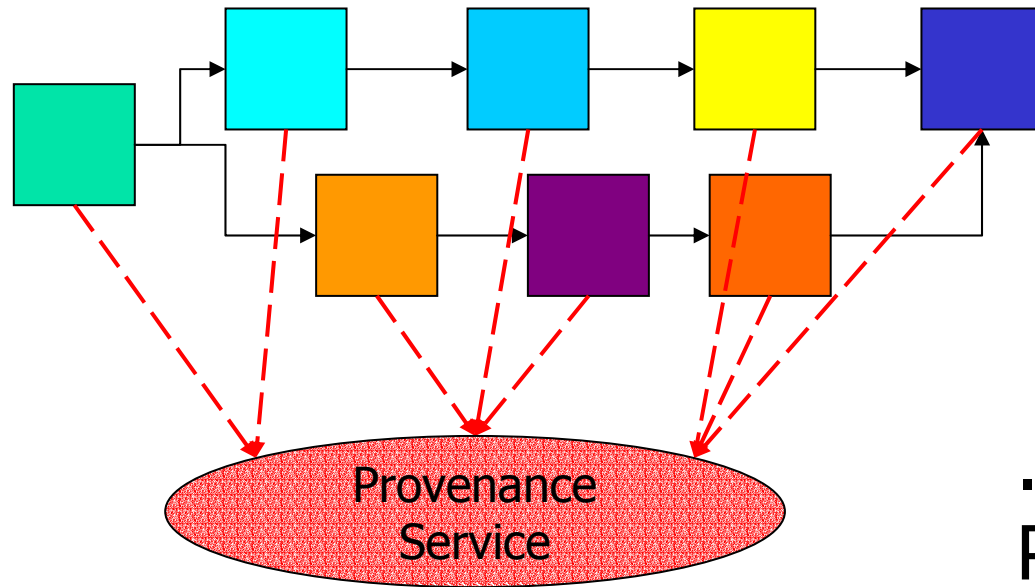
- Grids allow Virtual Organisations (VOs) to be created and destroyed
- VOs are intended to achieve a particular task
 - Run a simulation
 - Analyse some results
 - Perform a complex computation
 - Implement a business process
 - Allow researchers to collaborate
- VOs could be in existence for
 - A few seconds, A few hours, Indefinitely
- Workflow is used in the Grid community to choreograph the services that make up a task
 - For example
 - Computation and data management in experimental results analysis
 - Sequence of processing for Financial Transactions
 - Managing patient healthcare
 - Re-ordering process for supermarket supply chain
 - Customer call handling in a call centre
 - Information collection for pharmaceutical regulatory approval
- Data and Processes within a task are archived and audited on an ad-hoc basis
 - No standards and generic IT architectures available to support Provenance
- *How can we determine the origin and authenticity of results produced by a particular task as represented by its workflow ?*



Organising Services with Workflow



Services available on a Grid



..... organised into a Workflow

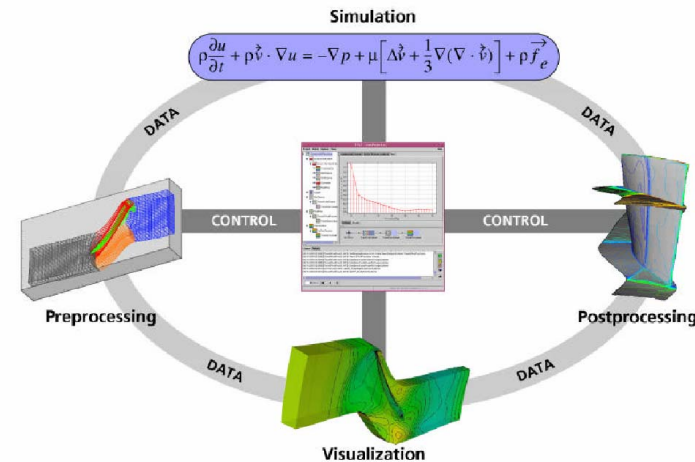
..... contributing Provenance information

Provenance in Aerospace Engineering

- Requirement:
 - Maintain a historical record of sub-system simulation results
 - Maintain these records for up to 99 years to meet regulatory requirements in some countries
- No support currently available but demand for this is increasing



TENT - Software Integration and Workflow Management





Provenance in Organ Transplant Management

- Decision support systems for organ and tissue transplant
 - Rely on a wide range of data sources:
 - Patient data
 - Doctors' and surgeons' knowledge
 - Regulatory frameworks
- A heavily regulated domain
 - European, national, regional and site specific rules govern how decisions are made.
 - Application of these rules must be ensured
 - Be auditable
 - May change over time
 - May be different between geographical areas
- Provenance will allow previous decisions to be tracked and analysed
 - Aim is to maximise the efficiency in matching organs and therefore the recovery rate of patients
 - A means of identifying and promoting best practice





Provenance Exploitation

- **Target Users:**
 - System Developers and Integrators
 - Requiring Provenance data management for regulatory compliance
 - Requiring standards-based system architecture and components
 - End Users
 - Requiring transparent management of Provenance data
 - Regulators
 - Requiring information management systems that meet compliance regulations

- **Project Deliverables:**
 - A scalable, secure Provenance architecture based on Grid standards
 - Draft proposals for submission to standards body
 - Reference implementation available through OMII in UK
 - Software tools to navigate, harvest and reason over Provenance data and processes
 - Deployment of the architecture in two application domains
 - Aerospace engineering
 - Organ transplant management

- **Project Exploitation:**
 - Lower the cost of Provenance adoption by standardisation
 - Ensure compliance with legislation and regulation
 - Retrospective deployment within Grid and Web Services
 - Transition from research project to market offerings
 - The enterprise world is increasingly regulated (Sarbanes-Oxley, Basel 2, FDA)
 - Enterprise computing increasingly needs Provenance capabilities
 - Increasing need to link academic research and industrial exploitation of results



Partners

- IBM United Kingdom Ltd
 - Coordination, Security, Scalability, Integration
- University of Southampton
 - Architecture, Security, Scalability
- German Aerospace Centre
 - Application: aerospace engineering
- University of Wales, Cardiff
 - Tools and setup
- Universitat Politecnica de Catalunya, Barcelona
 - Application: organ transplant management
- MTA SZTAKI, Budapest
 - Application: organ transplant management

- Contact:
 - john_ibbotson@uk.ibm.com