

PROVENANCE

Scope

The concept of Provenance is already well understood in the study of fine arts where it refers to the trusted, documented history of some work of art. Objects that do not have a trusted, proven history may be treated with some scepticism by those that study and view them. This same concept of Provenance may also be applied to information generated within a computer system; particularly when the information is subject to regulatory compliance and verification. The scope of the Provenance project is to provide a framework that allows information created within a computing environment to be proven and trusted. By this we mean that the information's history is documented in a way that can be inspected, validated and reasoned about by authorised users that need to ensure that information controls have not been altered, abused or tampered with.

Advances

The Provenance project has delivered an architecture, methodology and set of open specifications to support the use of Provenance in a distributed computing environment. In addition, the architecture is supported by a set of tools to navigate, query and reason over the Provenance documentation captured and stored in a database. The architecture is tested by developing a reference implementation and deploying two test applications for aerospace component simulation and organ transplant management.

Positioning in global context

Provenance as a research activity has so far been confined to academic projects; particularly in e-science. The aim of this project is to move the work from its research base into industrial and business sectors where benefits can be achieved in terms of ROI, reduced cost and increased agility for IT applications developed to operate in an environment that is increasingly regulated by governmental, industrial and other external organisations.

Contribution to standardization and interoperability issues

The Provenance project has developed a set of open specifications that can be used as input to an open standards body. The project has defined a preliminary set of nine specifications including a methodology, glossary and architectural description. These specifications detail the Provenance data model, recording and querying interfaces, linking between Provenance stores, securing Provenance documentation and a documentation style for transforming information within a Provenance store.

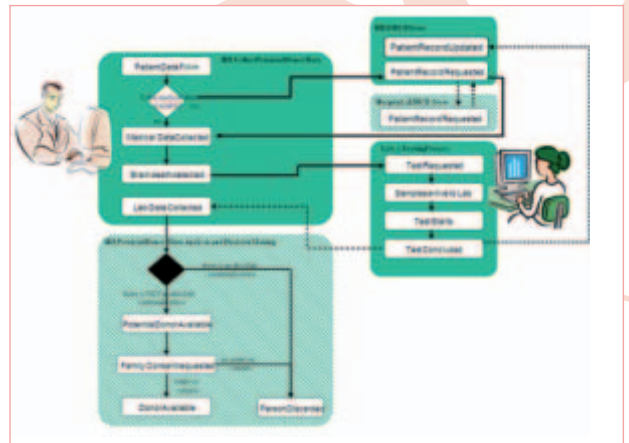
By proposing a set of open specifications for Provenance, the results of the project are made available to a wide

community of users within all sectors of industry, business and academia. A major objective of the project is to lower the cost of adopting and implementing Provenance-aware applications in all of the identified sectors. Openness and standardisation allow for Provenance to become a reusable, interoperable component in distributed IT systems, thereby lowering the cost of its adoption.

Target users / sectors in business and society

Users benefiting from the results of the Provenance project are:

- **Individuals and user and consumer associations** will benefit from increased trust in electronic information generated on their behalf.
- **Regulatory authorities** will benefit from an open architecture that allows regulatory compliance to be verified.
- **Research communities** will benefit from a common understanding of Provenance that will allow results shared between research groups to be verifiably reproduced based on a common architecture.
- **System developers, designers and suppliers** will be motivated to see Provenance as a business enabler and to integrate its concepts and solutions into their systems and products.



Overall benefits for business and society

Businesses operate in a regulated environment that is intended to prevent malpractice and protect the consumer. Regimes of corporate governance define the control processes that businesses must follow in order to meet their statutory obligations. Provenance provides a framework to let business show that these processes have been complied with.

Current mechanisms for regulatory compliance adopted by businesses involve closed, proprietary solutions that are not transparent and interoperable between different companies, business sectors or technologies. By

developing an open, standardised architecture, the Provenance project provides businesses with the flexibility to develop open and interoperable solutions to meet their regulatory obligations.

Examples of use

The Provenance project has tested the architecture in two application areas:

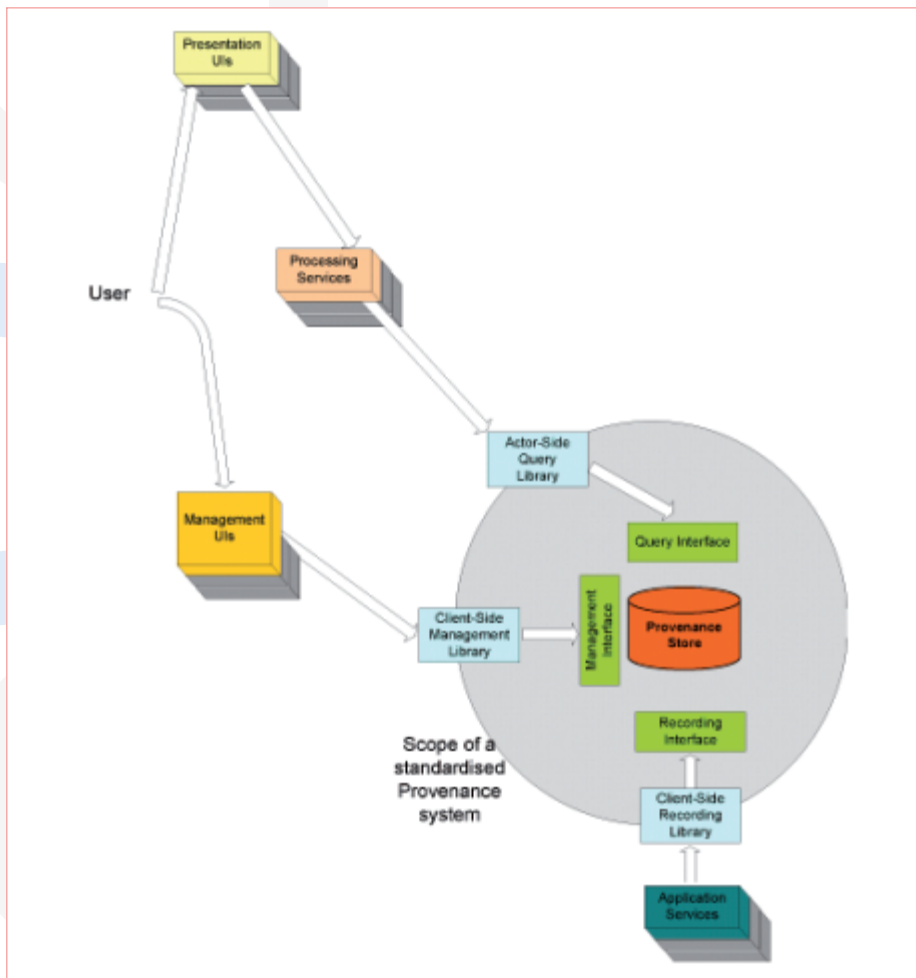
- The management of aerospace component simulation data.
- A proof of concept application for managing organ transplants and compliance with national healthcare regulations.

In addition to these applications, the project anticipates the use of the architectural framework in other business sectors not being investigated as part of the project such as:

- Regulation of financial markets and services.
- Submission of pharmaceutical information for drug approval.
- Compliance with import/export regulations in distributed supply chains.
- Compliance with rules of evidence in legal IT systems.
- Traceability of ingredients in food chains.

Achievements

All public results including published papers and architectural descriptions are available from the project website www.gridprovenance.org including the reference implementation and descriptions of the two application scenarios. The open specifications are available together with all other public deliverables.



title
Enabling and supporting provenance in grids for complex problems

contract number
511085

type of project
Specific Targeted Research Project

contact point
John Ibbotson
IBM UNITED KINGDOM LIMITED, UK
E-mail: john_ibbotson@uk.ibm.com

project website and partner list
<http://www.gridprovenance.org/>

EC contribution
1 981 996 €

start date
01/09/2004

duration
27