Project no: 511085

STREP

Thematic Priority

Provenance: Final Periodic Report for months 19 - 27

Deliverable D1.1.4

"Enabling and Supporting Provenance in Grids for Complex Problems"

Period covered: 01-03-2006 to 30-11-2006

Date of preparation: 30-11-2006

Start date of project: 01-09-2004

Duration: 27 Months

Project coordinator name: John Ibbotson

Project coordinator organisation name: IBM United Kingdom Limited

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Members of the PROVENANCE consortium:

- IBM United Kingdom Limited United Kingdom (IBM)
- University of Southampton United Kingdom (UoS)
- University of Wales, Cardiff United Kingdom (UWC)
- Deutsches Zentrum fur Luft- und Raumfahrt e.V. Germany (DLR)
- Universitat Politecnica de Catalunya Spain (UPC)
- Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézet Hungary (STA)

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Foreword

This document has been edited by John Ibbotson (IBM) based on input from project partners.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Table of Contents

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

1 Executive Overview

This Project Report provides details of the Provenance project for the final period (March 2006 to November 2006). It describes the workpackages that were active during that period, which milestones, deliverables or other tangible achievements were completed and relates these to the overall project objectives. Since this is the last reporting period of the project, it also details the work done to promote the project and produce the final set of deliverables under the project contract.

1.1 Management Report Overview

The Project Report is structured with the following sections:

General Management: Describes the overall project objectives and which objectives were achieved during this reporting period. It also itemises the major achievements of the project so far.

Workpackage Progress: Provides a detailed breakdown of all workpackages active during this reporting period together with milestones achieved and deliverables produced. It also details any deviations from the workpackage plan and where appropriate, details of any knowledge dissemination activities associated with the workpackage.

Consortium Management: Provides a description of the project management for the reporting period. It includes details of any plan deviations together with reasons and corrective actions. It also highlights any issues identified by the project coordinator.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

2 General Management

2.1 Project objectives

The overarching aim of the Provenance project is:

> *To design, conceive and implement an industrial-strength open provenance architecture for Grid computing, and to deploy and evaluate it in complex grid applications (aerospace engineering and organ transplant management).*

Specifically, the objectives of the project are:

1. To specify the contents of provenance in relation to workflow enactment.
2. To design and implement a scalable and secure distributed co-operation protocol to generate provenance data in workflow enactment.
3. To conceive and implement tools to navigate, harvest and reason over provenance data, also in a scalable and secure manner.
4. To design and engineer a scalable and secure software architecture to support provenance generation and reasoning.
5. To deploy and evaluate the provenance system in two different grid applications, namely aerospace engineering and organ transplant management.
6. To propose a draft provenance specification for input to an open standardisation process thereby contributing to the standardisation efforts in this area within the Grid and Web Services architecture domains.

2.2 Objectives for the Period

According to the project plan provided in the Provenance Technical Annex, objectives 4, 5 and 6 would be achieved at the end of the project (month 27).

To meet objective 4, the Provenance project has provided:

1. A technology independent provenance architecture that supports both security and scalability
2. An open specification that addresses security and scalability considerations and builds on existing open standards
3. A reference implementation that builds on the Globus GT4 toolkit and makes use of the Grid security specifications such as WS-SecureConversation
4. A Client Side Library that builds on the GT4 toolkit to provide secure communications with a remote provenance store

To meet objective 5, the Provenance project has provided:

1. A reference implementation of the provenance architecture that has been deployed and integrated with the Aerospace and OTM applications
2. Evaluation of the applications showing how the provenance architecture offered capabilities that did not previously exist.
3. In addition to the two applications, the architecture was successfully deployed on other workflow enactment engines including Functional Magnetic resonance Imaging (FMRI).

To meet objective 6, the Provenance project has provided:

1. An open specification philosophy described through a white paper
2. An extensive open provenance specification consisting of twelve documents circulated in the public domain
3. Contributions of provenance scenarios to the OGSA Data Scenario activities

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

4. A reference implementation which is a realisation of the open specification and released to the community under an open source license

2.3 Major Achievements during the Period

The following are considered the major achievements and activities during this six month period:

1. Following the approval of a requested project extension of 3 months at the annual review in November 2005, the project coordinator obtained formal approval from the project officer for the extension until the end of November 2006.
2. A project face to face meeting was held at DLR Cologne during May 31st to June 2nd 2006
3. A management face to face meeting was help at UPC in Barcelona during November 6-7 2006 to finalise the project deliverables and discuss arrangements for the final project review to be held in January 2007.
4. Regular developer telephone conferences were initiated to connect the reference implementation and application developers to ensure an efficient integration process. This led to a series of 3 face to face developer meetings held at the University of Southampton (July 19-21 and August 21-24 2006) and the University of Wales Cardiff (October 9-11 2006)
5. Following a first external release of the reference implementation software at the end of February 2006, further releases were made at the end of August and November 2006
6. The implementation of the security and scalability workpackage specifications have been merged into the reference implementation deliverable
7. A Client Side Library (CSL) has been developed by the University of Southampton allowing application developers to communicate securely with a remote provenance service. This CSL will support communication with provenance stores that support the standard interfaces defined by the open provenance specification.
8. The Architecture workpackage WP3 produced an open provenance specification which was made public on the project website at the beginning of September 2006 for community consultation. This consists of a set of 12 documents specifying different aspects of the developed Provenance Architecture.
9. The project provided leadership in the organisation of the IPAW06 conference held in Chicago in May 2006. Several project partners contributed papers to the workshop.
10. Arising from the IPAW06 workshop, a Provenance Challenge was launched to bring together the provenance research community. This challenge was hosted under the auspices of GGF18 in Washington and brought together contributions from 17 teams from across the world. A second challenge has been agreed which will be co-located with HPDC in June 2007.
11. The Aerospace Application workpackage WP7 partner DLR has completed the deployment of the reference implementation to its demonstration application. This has been tested and evaluated resulting in the evaluation report deliverable.
12. The OTM Application workpackage WP8 partners UPC and SZTAKI have completed the deployment of the reference implementation to its demonstration application which also integrates an Electronic Healthcare Record (EHCR) system based on open standards. This has been tested and evaluated resulting in the evaluation report deliverable.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

3 Workpackage Progress

The following sections describe in more detail the activities in each of the non-management workpackages WP2 to WP10. The project management activities are described in section 3 – Consortium Management.

3.1 WP2: Requirements

This workpackage finished before the reporting period.

3.2 WP3: Architecture

3.2.1 Objectives

The architecture will be described in several ways in order to address the concerns of its multiple stakeholders, end-users, developers and system managers. Different views of the architecture will be adopted to this end: a logical view addressing functional requirements, process and physical architectures taking into account non-functional requirements and physical deployment, and development architecture identifying modules and libraries at the level of software development. We will adopt an iterative design process, deriving a first architecture definition (strawman) from the technical requirements, iterating it into a final architecture definition, using feedback from the different application specific studies, tool design, and security and scalability analyses. The architecture will ultimately be defined into a standardisation proposal.

3.2.2 Progress

We structure our report on the progress of WP3 according to the deliverables that were produced: architecture specification (D3.2.1), Methodology (D3.2.2) and Open Provenance Specification (D3.3.1).

Architecture (D3.2.1). Building upon our pre-prototype and our logical architecture, this reporting period has seen the finalisation of the architecture. The architecture is a technology-independent description of the key components expected in provenance systems, their relationships and their behaviour. The project decided that it was better for the architecture to be specified in a technology-independent manner because Provenance systems may be embedded in different execution environments; hence, it is important to avoid cluttering the essence of such architecture, with technology-specific terms. Despite this technology-independent view, where appropriate, we highlight how the architecture can be considered within Service Oriented Architecture and workflow enactment engine scenarios to address the emphasis on these areas expressed in the Technical Annex of the original project proposal. The specific instantiation of the architecture for Grid and Web Services environments is the object of the standardisation proposal.

Architecture Views and Stakeholders. In deliverable D3.2.1, we cover the logical and process architectures of provenance systems. Specifically, the logical architecture identifies key roles and their interactions, whereas the process architecture discusses distribution, scalability and security. The other views of architecture, namely development and physical architectures, are addressed by the implementation of its different components, and in particular the Provenance Store (WP9, D9.3.3) and the Client-Side Library (WP9, D9.3.3).

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

In developing the ideas contained within the architecture, the requirements of the following stakeholders are considered, End Users, Developers and System Managers, by adopting several views of the architecture as follows:

- The logical architecture addresses the needs of users by defining the services and interfaces that users can interact with the architecture (cf. Chapters 6 and 7).
- The process architecture address non-functional aspects such as security and distribution allowing system managers to deploy the architecture (cf. Chapters 4 and 5).
- The development architecture provides developers with the means to build upon and adapt the architecture to their needs (cf. Chapter 8).

Furthermore, other documents produced by the project have contributed to fulfilling the projects obligations for the three sets of stakeholders identified in the technical annex. For example, the methodology addresses the needs of users and developers. The open specifications address the needs of developers and the design implementation addresses the needs of system managers who must deploy a realized vision of the architecture in their domain.

Software engineering and Iterative Process. The architecture is the outcome of a rigorous software engineering process undertaken by the project. We gathered requirements from multiple projects in WP2, drafted the architecture, got feedback from partners, refined architecture on basis of experience, repeated last two steps multiple times as project progressed. In order to clearly identify design decisions, and their relationship with captured requirements, design decisions are marked by a specific symbol and a cross-reference to original requirement (produced by WP2, D2.2.1) and analysis appears in the margin. In the online version of this document, the link can simply be followed by clicking on the requirement reference; for the paper version, a page number is also provided for convenience.

While designing the architecture, we have adopted an iterative process, as illustrated by the following table summarizing the dates for internal drafts, reviews, and final versions. We relied on TWiki to capture discussions on the architecture: over 70 pages of discussion have been produced over the architecture specification. We note, however, that Deliverable D3.2.1 is a public deliverable that reflects the final view of this iterative process.

Chapters	Milestone Draft by	Review by	Final by
Logical architecture frozen: 2, 3	24/6/05	8/7/05	15/7/05
Functional architecture frozen: 4, 5	7/10/05	21/10/05	28/10/05
Requirement analysis: 9	30/11/05	10/12/05	20/12/05
Final architecture frozen: 6, 7, 8, 10	5/2/06	18/2/06	21/2/06
Final Deliverable: 2 to 10	15/09/06	15/10/06	31/10/06

While the project work plan identified February as the date to produce the final architecture specification, it kept on involving in minor ways. Indeed, after February, we produced the standardisation proposal, partners (IBM and UoS) implemented elements of the architecture (Provenance Store and CSL) and partners (UoS, Cardiff, UPC, SZTAKI, DLR) made use of the architecture in their applications and tools design. In order to manage the queries and feedback from partners about the architecture, we produce a Provenance FAQ, answering key questions about the architecture. Doing so, two specific refinements had to be brought to the architecture: mechanisms to pass provenance contexts in application headers and p-assertion definition. The former led to the introduction of `ExposedInteractionMetadata` (now explicit in the p-structure, see View in Fig 6.10), while the latter p-assertion definitions are now used in all specifications. A number of questions raised in the FAQ did not mandate changes to the architecture, but instead led us to specify a

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

number of schemas that could usefully be shared between partners (e.g. name of actor, type of actor, time at which an interaction occurs). Such schemas were produced and made available through the FAQ.

Methodology and Best Practice (D3.2.2). Producing a complete architectural specification for provenance systems is clearly useful, but by itself, may be inadequate for application designers to make their applications provenance-aware. Discussion with the application owners (DLR, UPC, STA) and other UK and EU projects, such as myGrid and SIMDAT, clearly indicated the needs for a document that explains how the architectural concepts can be put in practice in a concrete application. Hence, we decided to formulate an explicit provenance methodology PRIME (as approved by the mid-term review). We see such a methodology as a key instrument for the dissemination of provenance ideas. A first draft of the methodology document was produced by the end of July 2005, and the final version was produced in September 06.
Through this methodology document, we showed best practice of provenance, illustrating how to make the two applications (OTM and Aerospace Engineering) provenance-aware. We again adopted the Twiki as the means to collect comments, suggestions and feedback for the document.

Open Provenance Specification (D3.3.1). The Open Provenance Specification is a major effort to define an open specification for provenance systems in the context of Web Services. Two fundamental considerations influenced our design.

1. While the architecture was specified as a monolithic document, it was decided that a standardisation proposal should be composed of a collection of individual documents, each putting forward a standalone specification for a specific aspect of the architecture. Such a decision was based on trends in Web services specifications (e.g., SOAP specified in several documents, or OGSI was revisited in a family of WSRF specifications).
2. Second, there exists a number of options for which standardisation would bring benefit: standardisation at the data level, standardisation at the interface level, and standardisation at the API level, respectively offering data inter-operability, service inter-operability, and programming inter-operability. In our Standardisation white-paper [Moreau-Ibbotson 06], we opted for the first two as primary focus for an open provenance platform.

The documents of the open provenance specifications are split into four groupings: 1) supporting documents, 2) core specifications, 3) generic profiles, 4) specific technology profiles. Supporting documents introduce the specification philosophy and provide a glossary of terms. Core specifications define key aspects of the provenance model and recording and querying interfaces. Generic profiles describe non-core aspects of the model. Finally, we provide an example of one specific technology binding – a tying in of an aspect of the model to an implementation technology.

The production of the different documents was staged across a period of 6 months, with internal reviews, using the Twiki.

Document	Sent for Review	End of Review	Final
Data Model Linking Glossary Process Doc Query (XQuery)	30/6	16/7	31/7

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Documentation style Recording Provenance Query	17/7	31/7	15/8
Secure P-Structure Overview	3/7	15/8	31/8

Specification Style. When we specified the scope of our specification effort, we also adopted a style for these documents. First, even though we consider the specifications as a major project deliverable, we did not use or "corporate" logo on these documents, to enforce the message that this is an open specification. Second, the documents do not refer to EU Provenance in any form, and adopt the terminology "specification" as opposed to "standardisation proposal". Third, we adopted a "GGF-like" front page, and a document structure reminiscent of some OASIS proposals; in particular, the introduction identifies goals, requirements and non-requirements. Fourth, each document is given the following status statement, reflecting its open nature "*This document provides information to the community regarding the specification of XXX and has the status of a working draft. The document does not define any standards or technical recommendations. Distribution is unlimited.*"

Towards Standardisation. Finally, while the aim of the project was not to start a standardisation process per se, we set the ground for such a standardisation activity in different ways.

1. We organised a discussion on standardisation at the IPAW (International Provenance and Annotation workshop, also organised by the Southampton). To raise interest with the GGF community, we made sure that the IPAW workshop was a GGF-sponsored event.
2. During this discussion, it became clear that the different parties were not ready to initiate a standardisation process, but they were willing to cooperate in order to understand provenance systems better.
3. We made several presentations on the benefit of an open approach to provenance, including at Harvard, Chicago, ISI, Rennes (CoreGrid), Kent, Cetraro (HPC'06).
4. We contributed to a GGF document, the OGSA Data scenarios v0.13, which now contains two provenance scenarios.
5. In response to the community interests on provenance, we launched the first Provenance Challenge, which we had hosted under the auspice of GGF18 in Washington. The community participation was outstanding, resulting in contributions from 17 teams across the world. A special issue of the Journal Computation and Concurrency: Practice and Experience is in preparation.
6. The community has agreed to participate in a second provenance challenge, likely to be co-located with HPDC in June'07, where we aim at integration of systems.
7. In these events, we participated and submitted papers by which we have been advertising the benefits of the Open Provenance Specification.
8. While most participants at IPAW and the challenge were from academia, we note Microsoft strong interest in provenance and their involvement in these two events. We also attended a Microsoft sponsored workshop on the data lifecycle, contributed our vision on open provenance architecture, and had a "call to arms" for agreed definitions for Provenance systems (cf. our paper "A tower of Babel. Towards a concept map").
9. IBM followed a more industrial approach to support standardisation effort, which is reported in their exploitation documentation.
10. Our recently accepted Communications of the ACM paper is a highly visible publication, which also advocates provenance systems and the benefits of an open provenance approach.

In their mid-term review, the reviewers recommended that "*the consortium follow two parallel tracks: GGF through the academic partners and Oasis through IBM. GGF is considered a low risk approach*

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

addressing mainly the research community. Trying with Oasis is high risk but could have very high industrial impact if successful." With Southampton and IBM efforts in this area, the project can declare success in involving the community towards a standardization process. The IPAW standardisation session and the provenance challenge resulted in an incredible momentum of 17 teams getting involved in such activities, and in angst to continue work towards system integration. This is a trans-national, community-driven, bottom-up approach, which also involves industrial partners such as Microsoft. The events we organised (IPAW and Provenance Challenge Workshop) were either GGF sponsored or hosted at GGF 18; and provenance scenarios are now explicitly in the OGSA Data scenario document.

FMRI Workflow Demonstrator. Finally, in order to prepare and participate to the provenance challenge, but also to test the architecture to its limits, UoS have implemented the challenge workflow, in the area of Functional Magnetic Resonance Imaging (fMRI). This has allowed us to test extensively project software, but also architectural principles, such as tracers and provenance queries. Our input is available on twiki.ipaw.info, and will appear in the Journal special issue that we are organising.

3.2.3 Deviations

Following the December Review, it was agreed that the methodology document would be made a formal project deliverable, and would supersede the internal milestone on Best Practice.

Section 5.1 "contribution to Standards" of the technical annex indicated that we were planning to submit a community best practice document on provenance to the Global Grid Forum. This submission was replaced by a contribution to the OGSA scenarios document, which was regarded as more efficient way of justifying the need for provenance in OGSA. We have also concentrated our efforts on the Provenance Challenge, at GGF in Washington, which was deemed a better opportunity to capitalise on the community interest in understanding provenance systems (and integrating them), and its relative reluctance to initiate a formal standardisation process. We also note that the reviewers invited us to engage with the GGF community, and we have done so systematically, for all the public events the project organised.

Section 5.1 "contribution to standards" also mentioned the possibility of integrating "service provenance" in the UDDI service directory framework. Our understanding of provenance of data has dramatically moved on since the conception of this proposal, and we no longer find that such a notion of "service provenance" is actually relevant to the topic of provenance. Hence, due to our better understanding of the context, we have not pursued this line of work because of its irrelevance.

Our extensive involvement in the Provenance Challenge was not planned, and in particular, our realisation of the Challenge FMRI workflow is a nice demonstrator of some of the more advanced architecture features.

3.2.4 Deliverables

D3.2.1 Architecture (Final): final architecture definition
D3.3.1 Standardisation Proposal
D3.2.2 Methodology

3.2.5 Milestones

No formal milestone in this period, but internal deadlines for draft documents, reviews and final documents as indicated above.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

3.2.6 Knowledge Dissemination

The following are key knowledge dissemination activities undertaken during this reporting period. The Final Project Documentation D1.2.1 and Exploitation Report D1.3.1 provide full details.

1. IPAW: Raising international profile of provenance activity
2. Provenance Challenge: Bottom-up, community-driven effort to standardisation
3. GGF-related events (IPAW, and Provenance Challenge), OGSA Data scenario document
4. Provenance tutorial to members of OntoGrid, NextGrid, myGrid, Harvard, DataMiningGrid.
5. Publications (IPAW06, SEM06, CCPE07, CACM07).
6. Demonstration at All Hands Meeting (AHM'06)

3.3 WP4: Security

3.3.1 Objectives

Deliverable D4.2.1 provides a specification for Provenance Services that includes actor authentication, actor authorisation, federated identity and process documentation non-repudiation and encryption. During this period of the project, the objectives are to implement this specification and deploy them as part of the reference implementation package developed in workpackage 9.

3.3.2 Progress

Security support in the reference implementation of the Provenance Service consists of four parts. These are:

1. Secure communications between the client and Provenance Service
    a. For a Provenance Service deployed in Tomcat
        i. HTTP Basic-Authentication
        ii. HTTPS
    b. For a Provenance Service deployed in a basic GT4 container
        i. WS Secure Conversation
2. Support for cross domain security by
    a. Root certificate installation
    b. Cross certification
3. Role Based Access Control (RBAC) to the Provenance Service using a customized Policy Decision Point
4. Provenance Architecture documentation style supporting p-assertion signing and encryption

The Provenance Service makes use of the Globus GT4 toolkit as Grid middleware. This means that the Security support is provided as a mixture of customized code, configuration scripts and instructions for systems administrators that deploy GT4 based infrastructures.

The elements of the Provenance Security support shown above are implemented as follows (more details are provided in the deliverable D9.3.3 which accompanies the Final Prototype deliverable from workpackage 9:

1. Secure communications are configured as part of the GT4 client and server setup
2. Cross domain security is configured by systems administrators. Information is provided as part of the reference implementation installation package

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

3. RBAC support is provided by developed Java code activated by configuration of the GT4 container
4. Signing and encryption of process documentation is implemented using the Provenance Client Side Library (CSL). This makes use of the documentation style features of the Provenance Architecture described in more detail in the workpackage 3 deliverables.

The features identified above have been implemented as part of the Final Prototype package in workpackage 9 and released internally to partners and externally via the project website.

3.3.3 Deviations

In deliverable D4.2.1, the GT4 Community Authorisation Service (CAS) was identified as the correct way to implement cross domain security for the project architecture. Subsequent investigation led us to the conclusion that this was not possible for a concrete implementation. This was confirmed by a note in the GT4 CAS Release Notes at http://www.globus.org/toolkit/docs/4.0/security/cas/WS_AA_CAS_Release_Notes.html which stated that "There currently is no support for CAS-based authorization for web services". We had to identify an alternative way to provide the same function. This led us to the decision to support certificates that are cross-certified amongst the collaborating security domains.

3.3.4 Deliverables

D4.3.1 Implementation of Security Specification is packaged as part of deliverable D9.3.3 Final Prototype.

3.3.5 Milestones

No formal milestone in this period, but internal deadlines for draft documents, reviews and final documents as indicated above.

3.3.6 Knowledge Dissemination

No specific knowledge dissemination activities were undertaken for this workpackage.

3.4 WP5: Scalability

3.4.1 Objectives

Deliverable D5.2.1 provides a specification for Provenance Services that includes a number of Architectural and Implementation scalability recommendations. During this period of the project, the objectives are to implement parts of this specification and deploy them as part of the reference implementation package developed in workpackage 9. The priority of implementation being based on the requirements of the application workpackages 7 and 8.

3.4.2 Progress

The only part of the Scalability specification implemented during this period was the support for large query result sets identified in section 5.1.1 of deliverable D5.2.1. This was implemented as the XPathFactoryPort in the Provenance Service interface. It allows the results of a query to be cached within the Provenance Service and retrieved using an iterator mechanism. Details of this implementation are provided in deliverable D9.3.3 Final Prototype.

Deliverable D5.2.1 made a number of recommendations about how architectural features could be implemented. These recommendations were analysed against the application requirements. We

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

concluded that some aspects although desirable, did not meet the needs of the demonstration applications from workpackages 7 and 8. Therefore the following specification recommendations were not implemented:

1. Section 5.1.3 – Repeated p-assertions. Not required.
2. Section 5.1.4 – Import and Export. Whilst not specifically implemented as ports, the function is available using the features of the eXist database used to implement the persistent store.
3. Section 5.1.5 – Non-XML Data. Not required.
4. Section 5.2.1 – Clustering and Non-Affinity. Supported by the GT4 toolkit which can be deployed in clustered configurations using Apache Tomcat. The Provenance Service is addressed by a single network endpoint which meets the non-affinity requirements.
5. Section 5.2.2 – Persistent Stores. The Provenance Service implementation uses OGSA-DAI data services which support persistent stores.

3.4.3 Deviations

As detailed in the previous section, five recommendations from the Scalability specifications were not implemented following an analysis of the specification recommendations against the demonstration application requirements.

3.4.4 Deliverables

D5.3.1 Implementation of Scalability Specification is packaged as part of deliverable D9.3.3 Final Prototype.

3.4.5 Milestones

No formal milestone in this period, but internal deadlines for draft documents, reviews and final documents as indicated above.

3.4.6 Knowledge Dissemination

No specific knowledge dissemination activities were undertaken for this workpackage.

3.5 WP6: Tools and Setup

3.5.1 Objectives

Workpackage 6 focused on the design and implementation of tools for navigating, accessing and reasoning over provenance data, for configuring the system. Over the period of 01/03/06 to 30/11/06, the focus in this workpackage has been on improving the implementation of our portal server and associated portlets, evaluating the scalability of the rule engine, evaluating access time and overall performance of the portal server, and developing specialist portlets in collaboration with the two application users. The main deliverable during this reporting period has been "Tools for Configuration" (D6.3.1).

3.5.2 Progress

Over the period of 01/03/06 to 30/11/06, the focus in this workpackage has been on:

1. Identifying and implementing mechanisms for enabling access to Provenance information from a range of different users – such as application end users, system administrators, and application administrators. The difference between these users was highlighted in deliverable D6.1.1. This has led to the development of specialist portlets in association with the two applications involved in the project – mainly the TENT application from DLR, and the

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

EHCR/OTM application for UPC and SZTAKI. This was achieved through a combination of face to face and weekly "skype" discussions between the developers involved in WP6, WP7 and WP8.

2. Better understanding how the visual output from the portal, and the use of the rule engine, could be used to answer particular provenance queries. Using p-assertions from WP7 and WP8, we were able to better understand the process involved in reaching particular decisions. Ultimately, the usefulness of the tools are restricted by the types of p-assertions that are available.

Deliverable (D6.3.1) was focused on the operations necessary to support configuration management, the types of configuration options that are made available within a particular provenance system, and mechanisms for the use of these configuration options within an application. A holistic view to configuration management was adopted, to highlight the types of configurations that would be necessary within a Provenance system. This was used as the starting point for highlighting the various aspects of a Provenance system that could be configured by different categories of users. Hence, it was recognized that configuration management could be supported for various components that make up a provenance system. These range from:

1. Configuring the overall architecture – i.e. identifying the components involved within a provenance system, such as the location and number of Provenance Stores, the recording actors, the location of the tool suite, etc.
2. Configuration of individual components that make up a provenance system, such as the Client Side Library, the tool suite, the Provenance Store, etc.
3. Configuration of the security mechanism being employed within the provenance system – such as role-based access control, use of digital certificates, username/password-based access management, etc.
4. Configuration of components within the tool suite – such as the Analysis and Navigation tools, etc.
5. Configuration of individual portlets within the Navigation tool – for instance, allowing users to modify the "view" on particular p-assertions that have been retrieved from one or more Provenance Stores. A user should also be provided with the ability to add new Provenance Stores and remove Stores that are no longer being used.
6. Configuration of the setup protocol – identifying the number and types of stages allowed for a particular provenance system (as described in Deliverable D6.2.1 from WP6).

A given type of configuration may only be undertaken by a particular category of user. For instance, application end users may not be allowed to modify the security mechanism being used – whereas system administrators may be allowed to modify any of the above configurations. The focus within D6.3.1 was primarily on configurations 4, 5, and 6 listed above.

The timeline for the production of the deliverable was as followings:

1. First version of Deliverable released to the project partners: August 18, 2006
2. Comments back from Project Partners: September 11, 2006
3. Second and final version of Deliverable released to the project partners: October 8, 2006 (with one week to add any additional comments to the second version)

Standardisation and Community Involvement: Members of workpackage 6 have been involved with the IPAW (International Provenance and Annotation workshop, organized by Southampton). We also participated in the first Provenance Challenge, organized at the Open Grid Forum in Washington DC. The participation generally involved close collaboration with the Southampton team, with a focus

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

on using Provenance tools to visualize p-assertions from the fMRI application that was part of the provenance challenge.

3.5.3 Deviations

Cactus is an open source problem solving environment designed for scientists and engineers. Its modular structure facilitates parallel computation across different architectures and collaborative code development between different groups. Cactus is used by numerous application communities internationally, including Numerical Relativity, Climate Modelling, Astrophysics, Biological Computing and Chemical Engineering. In order to demonstrate how the Provenance Tools could be used for a non-Provenance aware application (i.e. one that does not make use of a Client Side Library), a Cactus output from the AstroGrid-D simulation was used. This involved collaboration with Tom Goodale and members of the Cactus team at AEI/Max Planck institute in Germany. Tom Goodale was hired for 6 months on the project to demonstrate the use of Provenance Tools to astrophysicists and demonstrate how the tools could be used alongside Cactus.

3.5.4 Deliverables

D6.3.1: Tools for Configuration.

3.5.5 Milestones

The milestone over this period were:
Design of Configuration Tool (end of month 20).

3.5.6 Knowledge Dissemination

The following activities were carried out during the reporting period:

Publications: The following publications have been accepted over this period (CCGRID06, IPAW06, SISS/EDOC06, MACE06, WI06, CACM07). Work is currently underway on publication within a special issue of "Multi-Agent Systems and Grid" journal.

Talks: A number of talks were given to members of the computer science, computational science and engineering and the arts and humanities community, to better illustrate how Provenance could add value to their existing work. More details can be found in the exploitation report. A collaborative workshop was also organised at the IST event in Helsinki. This was a joint event between the OntoGrid, IntelliGrid, DataMiningGrid and Provenance projects. 47 people commented and expressed interest in the workshop prior to the event via the IST Web site. The event also provided a good opportunity to demonstrate tools and identify how these tools could be used alongside other Grid projects.

Take up: The provenance tools have been demonstrated to members of the Triana group -- and a prototype that makes use of the Client Side Library within Triana, and subsequently analyses submitted p-assertions using the navigation tools from WP6 has been implemented. This forms the basis for collaboration with the astrophysics community in the future. Deliverable D6.3.1 also contains a description of how the Cactus toolkit can be used alongside Provenance tools. Work is currently underway to also better understand how the Provenance tools could be made available to existing Cactus users.

Tutorials: Provenance material was covered in a tutorial delivered at CCGRID06/GRIDASIA conference.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Demonstrations: These were given at the "Provenance Challenge" (organized alongside the Open Grid Forum, Washington DC, September 2006) and the "All Hands Meeting" (Nottingham, September 2006).

Teaching Material: Some of the contents of deliverables from WP3, WP6, WP7 and WP8 will be included in the CMP629 ("Distributed Multi-Agent Systems") course. This is a specialist MSc course within the Strategic Information Systems programme, and to be delivered over the period February to April 2007.

The Final Project Documentation D1.2.1 and Exploitation Report D1.3.1 provide full details.

3.6 WP7: Application 1 – Aerospace

3.6.1 Objectives

The purpose of this workpackage is to deploy the provenance architecture in an industrial grid-based application, so as to exhibit some of the key scenarios identified in WP2, and to evaluate the system in real life applications. This work package will focus on the aerospace engineering domain, and specifically the TENT distributed workflow management system designed by DLR.

Specific objectives for this reporting period:

- Deploying provenance architecture: This task will deploy provenance system in the TENT system and will implement application-specific support according to the specification D7.1.1.
- Evaluating provenance architecture: In this application context, the provenance architecture will be evaluated against the requirements identified by WP2.

3.6.2 Progress

1. The TENT application has been extended for Provenance recording by integrating the Provenance Client Side Library. This allows storing of some general workflow specific information for every workflow.
2. Based on a simple parameter variation workflow, a demonstrator has been build. This demonstrator allows testing the Provenance recording with a running TENT workflow.
3. Some sample queries have been defined to demonstrate Provenance queries on the demonstration workflow.
4. The demonstrator has been deployed for testing by creating a virtual machine.
5. Provenance recording has been added to the components of the high-fidelity simulation workflow from the SikMa project.
6. Provenance recording has been included in all deployments of TENT.

3.6.3 Deviations

There are no deviations from the plan accepted after the extension of the project.

3.6.4 Deliverables

D7.2.1 Evaluation report (end of month 23)
D7.3.1 Final Deployment (end of month 23)

3.6.5 Milestones

M7.1.1 Application-specific provenance generation implemented (end of month 17)

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

M7.2.X Frequent deadlines set by the management to monitor progress of Code and Unit Test (CUT) (months 12 –23)

3.6.6 Knowledge Dissemination

In the reporting period, a paper about the aerospace application has been published. For all other dissemination activities see the exploitation report.

3.7 WP8: Application 2 – Organ Transplant Management (OTM)

3.7.1 Objectives

The purpose of this workpackage is to deploy the provenance architecture in a realistic grid-based application environment. The application domain targeted is the tracking of patient medical record information and (more specifically) decision-making; the precise application scenario will be based on the process of organ/tissue transplant decision-making and the subsequent care regime.
Specific objectives for this reporting period:

1. Deploying provenance architecture: This task deploys provenance system in the organ transplant application and implements application specific support according to the specification D8.1.1. A partial deployment is an internal deliverable for the project.
2. Evaluating provenance architecture: In this application test bed, the provenance architecture is evaluated against the requirements identified by WP2.

3.7.2 Progress

Over the period of 01/03/06 to 30/11/06 , the focus in this workpackage has been on:

1. Development of the provenance-aware OTM application
    a. The provenance-aware OTM application has been implemented. The application models the services and user interfaces for the different actors in the system: Organ Transplant Authorities, Hospitals and Testing Laboratories. Due to the high amount of events and interactions needed to be recorded as p-assertions in OTM, this process involved quite some changes in the code. Code components were re-arranged to increase separation between the internal process control loop, the user interface and the provenance recording, to avoid impact of recording in the OTM execution.
    b. OTM interface to Provenance Store using the Client Side Library has been implemented. This interface demonstrates different provenance features, such as interaction, relationship and actor state p-assertions, usage of p-header, usage of view link, usage of object link, usage of time stamps in actor state p-assertions. In order to reduce the impact in OTM of changes in the Client Side Library during development, an intermediate layer between OTM components and the Provenance Client Side Library was created.
    c. Development of provenance-aware interface: an analysis of the events to be recorded about users in the user interface revealed that there were some important facts and causal relations that could not be automatically captured with the interface implemented in the CARREL@FIS project. The interface was re-arranged in order to guide user interaction with the system in a way that some causal relations could be inferred. Some other causal relations could not be automated, such as the elements that were the basis for a given medical decision. After several interactions with the end users and some prototyping, interface provides an easy way for users to “tick boxes” in a list of medical tests carried out to the patient. Each time a physician

The PROVENANCE project receives research funding from the European Commission’s Sixth Framework Programme

introduces a medical decision in the system, he has to "tick" the boxes associated to the tests that were relevant for his decision.

2. Development of the provenance-aware EHCR application
    a. EHCR Store is implemented with two web service interfaces for 1) remote applications using ENV13606 standard and 2) local medical applications like OTM.
    b. Several XML schemas were developed based on the XML Schema received from the convenor of the CEN TC251 Committee. These are the following: ENV 13606 schema for remote application interface, EHCR Store schema for local medical applications, EHCR Authentication schema for the EHCR Authentication service.
    c. EHCR Store interface to Provenance Store using the Client Side Library has been implemented. This interface demonstrates different provenance features, like interaction, relationship and actor state p-assertions, usage of p-header, usage of view link, usage of object link, usage of time stamps in actor state p-assertions.
    d. EHCR Authorisation service has been implemented. This service provides healthcare agent directory service as required by ENV 13606.
    e. A dependency checker has been implemented to examine the run-time environment of the system.
3. Integration of the OTM application and the EHCR Store.
    a. Design and implementation of the different phases of interaction between the OTM and the EHCR application: Login conversation, StartCase conversation, GetEHCR conversation, SetEHCR conversation, Logout conversation.
    b. Definition of the interaction p-assertions needed to properly record interaction between OTM and EHCR. Introduction of the interaction p-assertions' recording in OTM and EHCR applications. Usage of view links in interaction p-assertions, to record the OTM view and the EHCR view over the same interaction.
    c. Security: setup of the access control policies between OTM and EHCR. Usage of the EHCR Authorisation service by OTM. Encryption of all interactions between OTM and EHCR.
4. Integration of the OTM/EHCR applications with the tools.
    a. XSLT style sheet is implemented for the Tools to textually visualise a set of p-assertions.
    b. Sample p-assertions and sample p-queries were generated for the development and test of various provenance system components.
    c. Sample rules were generated for the development and test of the rule analysis tool.
5. Delivery and demonstration of a first draft of the OTM/EHCR demonstration system.
    a. This system shows the OTM / EHRC application running in conjunction with deployed provenance components including the IBM implementation of the provenance service, GT4, OGSA-DAI, eXist, Exo portal, Provenance Tool.
    b. Evaluation of the OTM/EHCR demonstration system based on a) user functional evaluation, and b) fulfilment of technical requirements identified in WP2.

The following meetings were attended:

- Cardiff F2F February 22-24 2006
    - Security handling for OTM and EHCR applications
    - Examples of the kind of queries needed in OTM were presented
    - First tests of querying using Tools in the OTM application
    - Discussion on Documentation Styles.
- Barcelona developer meeting April 24-29 2006
    - Detailed plans for: Workflow Navigation, Using multiple Provenance Service with explicit giving Provenance Service, Create p-assertions with Client Side Library, Write XSLT (Short, Long), Using XSLT (Short, Long), Write Documentation Style Plug-In, Use Documentation Style Plug-In, Login, StartCase conversation, GetEHCR, SetEHCR conversation, Logout conversation, Security issues, P-query: all the cases of

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

the patient, P-query: all the places where EHCR of the patient has been updated, Using IBM Provenance Service, Refine the application specific content of p-assertions. Creating Documentation Style and DSP, Make p-assertions about all the communication, GUI, Run the Complete workflow of OTM

- Cologne F2F May 31-June 2 2006
  - EHCR application recording p-assertions
  - Query: displaying raw XML
  - Finding locations where related information is stored
  - Portal displaying ProvenanceStore content
- Southampton InteropF2F August 21-24 2006
  - OTM Demo with EHCR Store
  - Demo packaging
- Riva del Garda WP8 management meeting August 31 2006
  - Management issues in WP8 for the final project period: deliverables, test setup.
  - Detailed dissemination plan for WP8 results
- Cardiff InteropF2F3 October 9-11 2006
  - Demonstration infrastructure
  - Demonstration queries
  - Demonstration finalisation

3.7.3 Deviations

There are no deviations from the plan accepted after the extension of the project.

3.7.4 Deliverables

D8.2.1 Evaluation report (end of month 23)
D8.3.1 Final Deployment (end of month 23)

3.7.5 Milestones

M8.1.1 Application-specific provenance generation implemented (end of month 17)
M8.2.X Frequent deadlines set by the management to monitor progress of Code and Unit Test (CUT) (months 12 –23)

3.7.6 Knowledge Dissemination

In this reporting period, five presentations and papers were give based on this workpackage. The Final Project Documentation D1.2.1 and Exploitation Report D1.3.1 provide full details.

3.8 WP9: Implementation, Integration and Test

3.8.1 Objectives

In this final period of the project the objective of this workpackage was the completion of the reference implementation of the Provenance Architecture. In particular to:

1. Complete the Provenance Service implementation
2. Include Security and Scalability support for demonstration applications
3. Provide improved installation and configuration scripts
4. Respond to bug reports from application workpackages
5. Provide regular internal and external releases of the software

The priority was to support the development activities workpackages 6 and 7 so that there was a robust reference implementation available for their demonstration applications. The workpackage included

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

the development of the Client Side Library (CSL) by the University of Southampton. Implementations of the Security and Scalability functions were also included in the single reference implementation deliverable D9.3.3 though the activities related to these workpackages are reported elsewhere in this deliverable.

3.8.2 Progress

Following the first public release of the Provenance Service reference implementation at the end of February 2006, the objectives of this reporting period were to expand and complete the prototype including the security and scalability functions. The prototype was also to be integrated with the two demonstration applications from workpackages 7 and 8 together with the work being undertaken in the tools workpackage 6. This reporting period also included development of a Client Side Library (CSL) which provided application developers with a programming interface that hid the complexity of communicating with remote Provenance Services. The CSL development was undertaken by the University of Southampton with IBM being responsible for the server side software. All partners have access to the project CVS code repository. This allowed rapid access to bug fixes and new internal versions of the software.

The workpackage 9 timeline illustrates the activities undertaken in this reporting period. Following the software Beta 1 release in February 2006, two further beta releases were provided at the end of May and August 2006 which also included early versions of the tools and CSL. A final release of the software is scheduled for the end of the project in November 2006. All releases were via the project public website where uses could download the Provenance Service, CSL and Tools source code with documentation together with previous releases.

To support the integration of the software with the demonstration applications, a series of regular phone conferences were started in May 2006 involving the prototype and application development community. To accelerate the integration, these phone conferences were augmented with a set of face to face sessions starting at the DLR meeting in Cologne at the end of May. This was followed by two meetings at the University of Southampton at the end of July and August. Finally, another was held in Cardiff in October.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

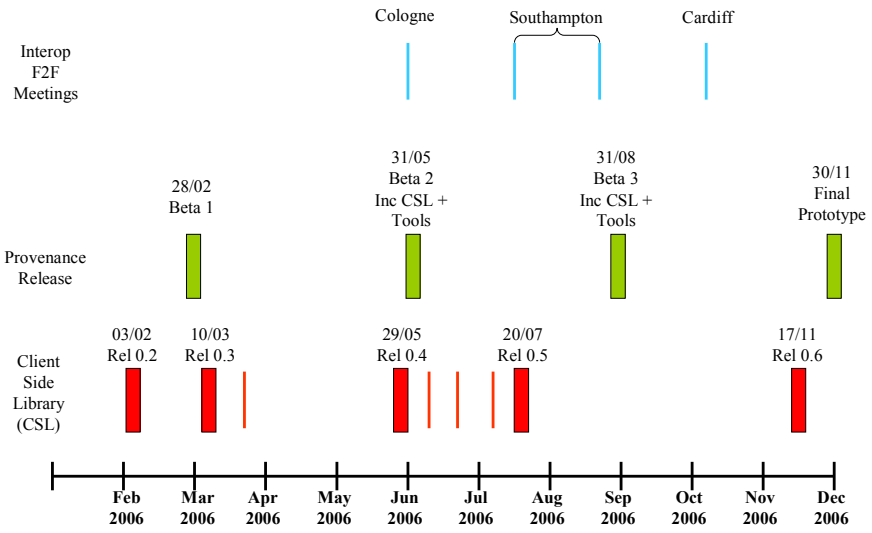

Figure 1: Workpackage 9 Development timeline

For the CSL, the design and implementation started at the beginning of 2006. For the initial version (v0.1), we generated stubs from the Provenance Service WSDL and constructed the implementation packages.

By 3rd February, we gave application developers our first release (v0.2). It includes only the generated Server API and an example to use it directly. We then implemented the basic Application API, which includes ProvenanceService, ViewRecord and XQuery. Necessary utilities and the relation helper were implemented as well in order to map the Application API to the Server API. By 10th March 2006, when we had released v0.3, we also added an ApplicationAPIExample and an XQueryExample. In order to fix several reported bugs and new application requirements, we made a minor release v0.3.1 on 24th March 2006.

The next release (v0.4) we made on 29th May 2006. It was implemented against an improved schema. This version was fully tested working with IBM WSRF-compatible provenance store. In the mean time, we implemented documentation style and the P-Header helper. PortFactory was introduced in order to simplify the usage of the CSL. After interaction with application developers, we released a minor version (v0.4.1) for fixed bugs and new requirements. V0.4.2 was released later with refined exception and better support documents provided. V0.4.3 was provided with new ProvenanceLink functions and JUnit testcases for CSL.

During the development period, it was our intent that changes to the Application API was minimized to avoid affecting the development of provenance-aware applications. Whilst this was desirable, changes could be made binary compatible and hence there would not be noticeable impact on application.

3.8.3 Deviations

For development, it was impractical to separate the implementation of the final prototype from other modules required for security and scalability. Therefore it was decided to include the implementation

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

activities for workpackages 4 and 5 into workpackage 9. The documentation of all the implementation activity is included in the workpackage deliverable D9.3.3.

3.8.4 Deliverables

D9.3.3 Final Prototype incorporating:
D4.3.1 Implementation of Security Specification
D5.3.1 Implementation of Scalability Specification

3.8.5 Milestones

No formal milestone in this period, but internal deadlines were set for code releases.

3.8.6 Knowledge Dissemination

No specific knowledge dissemination activities were undertaken for this workpackage.

3.9 WP10: Collaboration

Collaboration activities are reported in the final Collaboration Report deliverable PC3.

4 Consortium Management

This section describes the management of the consortium during the reporting period. It identifies the major activities during the final reporting period followed by issues identified during the reporting period and their resolution.

4.1 Management activities for months 19 - 27

The following management activities were completed during this reporting period:

1. At the annual project review in November 2005, the project requested an extension of three months to the end of November 2006. This was approved by the reviewers and project officer. To formalise the extension, the project coordinator sent a request letter to the project officer listing the changes to be introduced into the contract together with a justification of the changes. In addition, a new section 0 of the contract Description of Work (DoW) was added to provide details of the changes. In addition a revised CPF was provided to reflect the new completion date for the project. Confirmation of the extension was received by letter and email from the project officer on 31/05/2006.

   A copy of the request letter and section 0 of the DoW are included in Appendix 1 of this document.

2. The consortium analysed the reviewers' comments, implemented their suggestions and sent an extensive response to the project officer.

3. Following the reviewer's advice at the first annual review, the project adopted an iterative approach to managing the reference implementation and application integration. These activities were managed in three monthly cycles linked to completing subsets of the application functions. Regular telephone conferences were initiated between the developers leading to a series of face to face meetings in July and August hosted by the University of Southampton with a final meeting at Cardiff in October. This management process led to the successful completion of the reference implementation and integration with the two demonstration applications.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

4. Two face to face meetings were held during this reporting period hosted by partners in Cologne and Barcelona. The management issues discussed at these meetings were:
    a. Face to Face meeting held at DLR Cologne during May 31st to June 2nd 2006
        i. Payment for first reporting period
        ii. Update on implementation and application development including decision to initiate regular developer telcons and meetings
        iii. Update on Exploitation activities
        iv. Milestones for next period
    b. Face to Face meeting held at UPC Barcelona during November 6th to 7th 2006
        i. Discuss and agree proposed agenda and themes for final review in January 2007
        ii. Review the status of final deliverables and agree timetable for completion
        iii. Review and agree structure for project website and its lifetime following project completion
5. The project completed the all final deliverables due at month 27 of the project including financial and management information.
6. Following the comments from the reviewers, the project put in place an extra quality process, to check deliverables before shipping.
7. The management team initiated a review and restructuring of the project website at www.gridprovenance.org. This domain has been reserved and will be in existence for ten years following the end of the project.

4.2 Management Issues

There were no substantive management issues to resolve during this reporting period.

4.3 Project Resources

This section provides a justification of the resources in person months used by each partner throughout the project. The actual person months used by each partner are summarised in the following table. From the data in the table, the following resource issues are highlighted:

1. The project used 283.98 person months to complete the project compared with the 240 person months estimated in the contract negotiation. This was an additional 43.98 person months.
2. The University of Southampton consumed an excess of 11.5 person months in the Collaboration workpackage 10. This was due to the additional work contained in organising the IPAW06 workshop in May 2006. As a result of this workshop, the first Provenance Challenge was organised by Southampton which was held at GGF in Washington in September 2006. The organisation and participation in the challenge consumed further resources. These resources were not planned when the project was first proposed.
3. Workpackages 4 (Security) and 5 (Scalability) consumed less resources than expected in the original contract proposal. This was due to the choice of the Globus GT4 toolkit being used for implementation which required less work to add the Security and Scalability recommendations listed in the workpackage specifications. The project consumed 18.35 person months on these workpackages instead of the 42 person months planned in the project proposal. This under spend contributed to IBM's overall under spending on the project.
4. Workpackage 9 (Implementation) consumed 57.06 person months instead of the budgeted 42 giving an over run of 15.06 person months. This was due to the additional development of the Client Side Library which was not identified in the project Technical Annex. This library has been extremely useful in helping developers make their applications provenance aware. Also,

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Provenance Project Resources (in Person Months)								
		Totals	IBM	UoS	UWC	DLR	UPC	STA
Workpackage 1: Management	Actual period 1	8.02	7.07	0.25	0.25	0	0.25	0.2
	Actual period 2	9.6	5	3	0.25	0.5	0.25	0.6
	Project planned	14.6	12.1	0.5	0.5	0.5	0.5	0.5
	Project actual	17.62	12.07	3.25	0.5	0.5	0.5	0.8
Workpackage 2: Requirements	Actual period 1	29.54	4.14	4	3	6	4	8.4
	Actual period 2	0	0	0	0	0	0	0
	Project planned	36	6	6	6	6	6	6
	Project actual	29.54	4.14	4	3	6	4	8.4
Workpackage 3: Architecture	Actual period 1	9.11	2.11	7	0	0	0	0
	Actual period 2	25	1	24	0	0	0	0
	Project planned	22.8	9	13.8	0	0	0	0
	Project actual	34.11	3.11	31	0	0	0	0
Workpackage 4: Security	Actual period 1	6.35	3.35	3	0	0	0	0
	Actual period 2	4.5	2.5	2	0	0		0
	Project planned	21	12	9	0	0	0	0
	Project actual	10.85	5.85	5	0	0	0	0
Workpackage 5: Scalability	Actual period 1	6	3	3	0	0	0	0
	Actual period 2	1.5	1.5	0	0	0	0	0
	Project planned	21	12	9	0	0	0	0
	Project actual	7.5	4.5	3	0	0	0	0
Workpackage 6: Tools	Actual period 1	6	0	0	6	0	0	0
	Actual period 2	36.3	0	0	36.3	0		0
	Project planned	21.6	0	0	21.6	0	0	0
	Project actual	42.3	0	0	42.3	0	0	0
Workpackage 7: Aerospace	Actual period 1	6	0	0	0	6	0	0
	Actual period 2	12	0	0	0	12	0	0
	Project planned	18	0	0	0	18	0	0
	Project actual	18	0	0	0	18	0	0
Workpackage 8: OTM	Actual period 1	13.1	0	0	0	0	6.7	6.4
	Actual period 2	35.9	0	0	0	0	13.3	22.6
	Project planned	36	0	0	0	0	18	18
	Project actual	49	0	0	0	0	20	29
Workpackage 9: Implementation	Actual period 1	15.06	12.06	3	0	0	0	0
	Actual period 2	42	25	17	0	0	0	0
	Project planned	42	28.2	13.8	0	0	0	0
	Project actual	57.06	37.06	20	0	0	0	0
Workpackage 10: Collaboration	Actual period 1	2.5	0.2	0.5	0.5	0.5	0.5	0.3
	Actual period 2	15.5	0.3	13	0.5	0.5	0.5	0.7
	Project planned	7	1	2	1	1	1	1
	Project actual	18	0.5	13.5	1	1	1	1
	Total period 1	101.68	31.93	20.75	9.75	12.5	11.45	15.3
	Total period 2	182.3	35.3	59	37.05	13	14.05	23.9
	Planned total	240	80.3	54.1	29.1	25.5	25.5	25.5
	Actual total	283.98	67.23	79.75	46.8	25.5	25.5	39.2

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

the implementation of the Security and Scalability recommendations were included in this workpackage.

5. Workpackage 8 (OTM) has consumed 49 person months instead of the planned 36. This is explained by the complexity of the application and the work carried out by UPC and STA in evaluating and assessing the impact of the application with the medical user community.
6. Workpackage 6 (Tools) consumed 42.3 person months instead of the planned 21.6. This included development work using the eXo portal which was not anticipated in the initial project plan. Following feedback from the reviews at the 12 month review, Cardiff have collaborated with a number of other workflow development teams including ProActive, Cactus and Triana.
7. Generally, the overall person months used by the project have increased because of the extension to the project by 3 months. With approximately 7 people working on the project, this would have increased the resources used by 21 person months.

The following sections provide justifications for the major expense items for each partner during the final reporting period from month 13 to month 27 of the project.

4.3.1 IBM United Kingdom Limited, United Kingdom (IBM)

The major work items performed by IBM are as follows:

1. IBM led and coordinated Management workpackage 1 including:
    a. Planning and tracking partner activities.
    b. Chairing monthly management conference calls.
    c. Communications with EC Project Officer.
    d. Organisation and chairing of periodic face to face meetings between partners
    e. Preparation and delivery of project deliverables to EC Project Officer
    f. Coordination and preparation of periodic management and financial reports to EC Project Officer
    g. Distribution of payments from EC to project partners
2. IBM contributed to Architecture workpackage 3 by providing:
    a. Contribution to standardisation vision white paper
    b. Review of architectural and standards draft deliverables
3. IBM led and coordinated Security workpackage 4 including:
    a. Preparation of D4.2.1 Security specification version 2
    b. Implementation of Security specification as part of Implementation workpackage 9
4. IBM led and coordinated Scalability workpackage 5 including:
    a. Preparation of D5.2.1 Scalability specification version 2
    b. Implementation of Scalability specification as part of Implementation workpackage 9
5. IBM led and coordinated Implementation workpackage 9 including:
    a. Implementation of Provenance Store service including security and scalability features based on Globus GT4 toolkit
    b. Managing implementation with UoS and integration with tools and applications partners UWC, DLR, UPC and STA. This included regular phone conference calls and F2F meetings.
    c. Managing the release of the software through the project website. Releases to the external community were provided in February, August and November 2006
6. IBM contributed to Collaboration workpackage 10 by:
    a. Representing the project at the EU concertation meeting in Brussels during September 2006

The major cost items for UoS are:

1. Salaries for full time employees assigned to the project

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

2. Travel costs for:
    a. Provenance and EU meetings
        i. Budapest, Cardiff, Cologne and Barcelona F2F meetings
        ii. First annual project review in Brussels
        iii. Brussels concertation meeting (September 2006)
3. Support costs for:
    a. Project management including conference phone call facilities

4.3.2 University of Southampton, United Kingdom (UoS)

The major work items performed by UoS are as follows:

1. UoS contributed to workpackage 1 with the following:
    a. Participating in management face to face and telephone conference meetings.
    b. Preparing periodic management reports.
    c. Preparing financial statement.
    d. Previewing and quality checks on deliverables.
2. UoS led and coordinated the Architecture workpackage 3. The work undertaken in the last period has included:
    a. Creation of the Provenance Architecture deliverable including managing a significant review process amongst the partners
    b. Creation of the Open Specification deliverables as a set of 11 separate documents. This has included managing the review process amongst the partners and the wider research community
    c. Development of the Provenance Methodology and associated deliverables
3. UoS contributed to the Security workpackage 4 by providing:
    a. Contribution to the Security specification deliverables
    b. Globus GT4 specific skills for implementing the Security specification
4. UoS contributed to the Implementation workpackage 9 by providing:
    a. Development resources to design, code and test the Provenance Client Side Library (CSL)
    b. Hosting two F2F meetings in 2006 for the development teams in the project
5. UoS led and coordinated the Collaboration workpackage 10 which included:
    a. Jointly organising the IPAW'06 workshop in Chicago in May 2006
    b. Organising and participating in the Provenance Challenge which originated at IPAW'06 and was held at GGF in Washington in September 2006
    c. Both these activities have resulted in a significant amount of resources being expended in this workpackage

The major cost items for UoS are:

4. Salaries for:
    a. Full time employees
        i. Victor Tan
        ii. Steven Munroe
        iii. Sofia Tsasakou
        iv. Sheng Jiang
    b. Temporary employees
        i. Simon Miles
        ii. Paul Groth
        iii. Serena Raffin (Web Site)
    c. Summer students
        i. Andrej Kazakov
        ii. Thibaut Andrassi
5. Travel costs for:

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

   a. Meetings attendance
      i. Nice (grids at work),
      ii. AHM'05, AHM'06, Grid summer school
      iii. Athens GGF16
      iv. IPAW Chicago
      v. Boston, PASS Provenance workshop + IBM Watson
      vi. HPC'06
      vii. Provenance Challenge, Washington (GGF18)
      viii. SEM'06, Portland
   b. Provenance and EU meetings
      i. Budapest, Cardiff, Cologne and Barcelona F2F meetings
      ii. First annual project review in Brussels
      iii. Heathrow CT1 management meeting
6. Support costs for:
   a. Hosting integration meetings in Southampton
   b. Computers for Sofia Tsasakou, Steve Munroe, Sheng Jiang
   c. www.ipaw.info and www.gridprovenance.org web sites design
   d. CDROMs for tutorial

4.3.3 University of Wales, Cardiff, United Kingdom (UWC)

The major work items performed by UWC are as follows:

6. UWC contributed to workpackage 1 with the following:
   a. Participating in management face to face and telephone conference meetings.
   b. Preparing periodic management reports.
   c. Preparing financial statement.
   d. Previewing and quality checks on deliverables.
   e. Cardiff hosted the Provenance project meeting in February 2006 and the developer face-2-face meeting in October 2006.
7. UWC led and coordinated Work package 6 “Tools and Setup”. The work undertaken in the last period has included:
   a. Updating deliverable D6.1.1 based on comments from reviewers. This deliverable was then re-submitted.
   b. Specification, design and implementation of the Setup protocol, and identifying its relationship to other components in the Provenance System. Deliverable D6.1.2 was produced to cover the Setup protocol.
   c. Specification, design and implementation of the Configuration tool and identifying how it can be used alongside other components in the Provenance System. Deliverable D6.1.3 was produced to cover the configuration tool and the configuration process in general.
   d. Update to the Tool Suite to reflect requirements identified by the application users in WP7 and WP8. This also involved collaboration with developers involved in these WPs to help them install and deploy the tools at their local sites. Performance tests were also carried out to evaluate the overall user response from the portal. These are reported in deliverable D6.1.3.
   e. Development of a User, Administrator and Installation guide for the Tool Suite – connected via the Portal.
   f. Development of the rule engine and configuration of the engine based on requirements from WP7 and WP8. Performance tests were also undertaken to evaluate the scalability of the rule engine and the associated inference mechanism. These are reported in deliverable D6.1.3.
   g. Contribution to WP9 to support “Documentation Style” plug-in support in the Tool Suite.
8. UWC contributed to workpackage 10 on Collaboration including:

The PROVENANCE project receives research funding from the European Commission’s Sixth Framework Programme

a. Collaboration with the Triana development team to evaluate the inclusion of the Client Side Library in Triana, and subsequent access of data via the Tool Suite.
b. Collaboration with the Cactus team to analyse data generated by Cactus using the Tool Suite.

The major cost items for UWC are:
1. Salaries for:
    a. Vikas Deora
    b. Arnaud Contes
    c. Tom Goodale
2. Support costs for:
    a. Cardiff Project Meeting
    b. Cardiff Developers F2F Meeting
3. Travel costs for:
    a. Budapest, Cologne and Barcelona F2F meetings
    b. First annual project review in Brussels
    c. Project technical meetings in Barcelona and Southampton

Events Attended
1. IEEE CCGrid 2006/GridAsia 2006 (partial costs requested from Provenance project)
2. HPC ThaiGrid (partial costs requested from Provenance project)
3. DMC 5th Conference on "Autonomous Society", Cambridge
4. EDOC/SISS 2006, Hong Kong
5. IST 2006, Helsinki
6. Autonomous Adaptation workshop, Amsterdam
7. Grids@Work, Sophia Antipolis
8. Provenance Challenge, Open Grid Forum, Washington DC, 2006
9. Web Intelligence 2006, Hong Kong

Additional collaboration discussion meetings were held with Vrije University (Amsterdam), Bath University.

4.3.4 Deutsches Zentrum fur Luft- und Raumfahrt e.V., Germany (DLR)

The major work items performed by DLR are as follows:
1. DLR contributed to workpackage 1 with the following:
    a. participating in management face to face and telephone conference meetings
    b. preparing periodic management reports
    c. preparing financial statement
    d. reviewing deliverables
2. DLR led and coordinated workpackage 7 Application 1 - Aerospace including:
    a. Integration of Provenance architecture into the aerospace application.
    b. Preparation of demonstrator for evaluation purposes.
    c. Preparation of deliverable D7.2.1 for WP 7 (Aerospace Engineering Application).

The major cost items for DLR are:
1. Salaries for:
    a. Guy K. Kloss
2. Support costs for:
    a. Cologne F2F meeting
3. Travel costs for:
    a. Budapest, Cardiff and Barcelona F2F meetings

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

b. First annual project review in Brussels
c. Project technical meetings in Cardiff and Southampton

4.3.5 Universitat Politecnica de Catalunya, Spain (UPC)

The major work items performed by UPC are as follows:

1. UPC contributed to workpackage 1 with the following:
    a. Participating in management face to face and telephone conference meetings
    b. Contributing to workpackage 8 periodic management reports
    c. Preparing UPC partner periodic management reports
    d. Preparing UPC partner financial statement
    e. Reviewing deliverables
2. UPC led and coordinated workpackage 8 Application 2 – Organ Transplant Management including:
    a. Development of 1) the provenance-aware OTM application, 2) the OTM interface layer with the Client Side Library, 3) the provenance-aware user
    b. Integration of the OTM application and the EHCR Store, including the interaction protocol and the setup of the access control policies between OTM and EHCR.
    c. Integration of the OTM/EHCR applications with the tools, by designing and testing p-assertions, analysis rules and XSLT stylesheets.
    d. Delivery and demonstration of a first draft of the OTM/EHCR demonstration system. Evaluation of the demonstration system was based on a) user functional evaluation, and b) fulfilment of technical requirements identified in WP2.
    e. Lead development of deliverables D8.2.1 and D8.3.1
3. UPC contributed to workpackage 10 with the following:
    a. Task 1: Exploitation of synergies / technical concertation:
    b. On-going collaboration (through UPC) with Hospital St. Pau, Barcelona, Spain
        1. Joint work on organ transplant management application.
        2. Regular interproject meetings with HSP lead Carrel FIS Project (Spanish-funded HealthCare Project).
    c. Ongoing collaboration (since October 2006) with the EU CONTRACT project
        1. The provenance architecture will be evaluated in the first phase of the project to determine if it can be used for the electronic contract monitoring mechanisms. This evaluation is expected to continue past the end of the Provenance project.
    d. Task 2: Joint fora for exchange and dissemination:
    e. UPC supported with its attendance the IPAW'06 workshop held in Chicago, in May 2006 http://www.sigmod.org/sigmod/record/issues/0609/sigmod-record.september2006.pdf
    f. Task 5: Collaboration on research inventory and roadmaps
    g. UPC and UoS have contributed (through UoS and UPC) to the AgentLink roadmap (coordinated by Michael Luck from Southampton)
    http://www.agentlink.org/roadmap/index.html
    M. Luck, P. McBurney, and O. Shehory and S. Willmott, "Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)", AgentLink, 2005. ISBN 085432 845 9.

The major cost items for UPC are:

1. Salaries for:
    a. Javier Vázquez
    b. Steve Willmott
    c. Sergio Alvarez
2. Support costs for:
    a. Barcelona F2F meeting
3. Travel costs for:
    a. Budapest, Cologne and Cardiff F2F meetings

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

b. First annual project review in Brussels
c. Project technical meetings in Cardiff and Southampton

The total person months used comes to 14.05 PM, which is slightly more than half of UPC's planned usage over the project. However this is in line with project planning since, as explained in the first reporting period,

1. A slightly increased load was expected in the second reporting period with increased activity in WP8.
2. In the first reporting period 4PM rather than 6PM were used in WP2 early in the project since UPC was able to complete tasks by drawing on existing expertise via the CARREL FIS project and save some resource for the second reporting period.

4.3.6 Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézet, Hungary (STA)

The major work items performed by STA are as follows:

1. STA contributed to workpackage 1 with the following:
    a. Participating in management face to face and telephone conference meetings
    b. Contributing to workpackage 8 periodic management reports
    c. Preparing SZTAKI partner periodic management reports
    d. Preparing SZTAKI partner financial statement
    e. Reviewing deliverables
2. STA worked on the Electronic Health Care Record Management part of the Organ Transplant Management (OTM) application of the project in workpackage 8. Activities carried out in workpackage 8 were:
    a. Several XML schema were developed based on the XML Schema received from the convenor of the CEN TC251 Committee. These are the following: ENV 13606 schema for remote application interface, EHCR Store schema for local medical applications, EHCR Authentication schema for the EHCR Authentication service.
    b. EHCR Store is implemented with two web service interfaces for 1) remote applications using ENV13606 standard and 2) local medical applications like OTM.
    c. Deployment and test of a full provenance environment including the IBM implementation of the provenance service, GT4, OGSA-DAI, eXist, Exo portal, Provenance Tool.
    d. EHCR Store interface to Provenance Store using the Client Side Library is implemented. This interface demonstrates different provenance features, like interaction, relationship and actor state p-assertions, usage of p-header, usage of view link, usage of object link, usage of time stamps in actor state p-assertions.
    e. Integration of the OTM application and the EHCR Store.
    f. EHCR Authorisation service is implemented. This service provides healthcare agent directory service as required by ENV 13606.
    g. XSLT style sheet is implemented for the Tools to textually visualise a set of p-assertions.
    h. A dependency checker is implemented to examine the run-time environment of the system.
    i. Sample p-assertions and sample p-queries were generated for the development and test of various provenance system components.
    j. Four publications were prepared (see dissemination).
    k. Contributions to Deliverables D8.2.1 and D8.3.1
3. STA contributed to workpackage 10 with the following:
    a. Joining the semantic grid activities of the collaboration between FP6 grid projects
    b. Studying the semantic grid descriptions of the collaboration between FP6 grid projects

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

The major cost items for STA are:

1. Salary costs for project personnel
2. Travel costs for:
    a. Cardiff, Cologne and Barcelona F2F meetings
    b. First annual project review in Brussels
    c. Project technical meetings in Barcelona, Cardiff and Southampton

No equipment purchases were made in this reporting period

Following the recommendations at the negotiation meeting, the travel costs were strongly kept at a low level by STA. With the extension of the project the person-month efforts are higher than planned, but the cost budget is kept within the limits.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

Appendix 1 – Project Extension Request Letter

REGISTERED WITH ACKNOWLEDGEMENT OF RECEIPT

Dear Mr. Gasós,

Subject: **Request for Amendment to Contract No. 511085**

Project "Enabling and Supporting Provenance in Grids for Complex Problems"

With reference to the above noted contract, I request to modify the contract as follows:

Extension of the duration
The duration specified in Article 4.2 of the contract is modified as follows:
New *duration*: 27 months

Modification of Annex I
Annex I - Description of work is modified.

The revised Annex I dated April 7th attached to this letter replaces any former version.

All other provisions of the contract and its annexes remain unchanged.

Yours sincerely,
……..
Authorised representative

Section 0 of modified DoW

Project Extension

At the Provenance annual project review held on November 14th 2005, the project consortium requested a three months extension to the project from August 31st 2006 to November 30th 2006. This was informally agreed by the Reviewers and the Project Officer. This additional section of the project description of work provides a justification for the request together with details of changes to the deliverable dates resulting from the extension.

Justification for the requested extension

The following justifications are provided for the requested extension:

1. The project was initially scheduled to start at the beginning of September 2004. Due to delays in signing the Consortium Agreement, project work did not start until the end of October 2004. This has resulted in a compression of timescales at the beginning of the project.
2. Some academic partners have experienced delays in recruiting research personnel to take part in the project.

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme

3. The project has been affected by the end of each reporting period coinciding with the holiday period of July and August. This has affected the quality of deliverables due to insufficient time to review them before delivery to the EC.
4. The original project plan required a large number of deliverables to be provided simultaneously at the end of the project. In particular the final version of the reference implementation, its installation in the two test applications and the production of evaluation reports. In order to phase the work for these deliverables, there would have been a considerable compaction of the critical path to produce these deliverables. By extending the project we are able to reschedule the deliverables that depend on the reference implementation and provide a better quality of results for the project.
5.

Changes to deliverables and schedule

On 23/06/2005, the Provenance project coordinator emailed the Project Officer with a request to modify the title and content of some contracted deliverables. This was agreed by the Project Officer by email on 29/06/2005. The details of the requested changes were:

Deliverable	Original Title	Revised Title
D4.1.1	Security Specification (Generation)	Security Specification version 1
D4.2.1	Security Specification (Access)	Security Specification version 2
D5.1.1	Scalability Specification (Generation)	Scalability Specification version 1
D5.2.1	Scalability Specification (Access)	Scalability Specification version 2

No changes were requested to the delivery dates for these four deliverables.

The requested changes to the schedule of deliverables resulting from the requested project extension are listed in the following table:

Deliverable	Title	Original Plan Date	Revised Plan Date
D1.1.4	Final Project Report	31/08/2006	30/11/2006
D1.2.1	Final Project Documentation	31/08/2006	30/11/2006
D1.3.1	Exploitation Strategy	31/08/2006	30/11/2006
D3.2.1	Standardisation Proposal	31/08/2006	30/11/2006
D4.3.1	Implementation of Security Specification	31/08/2006	30/11/2006
D5.3.1	Implementation of Scalability Specification	31/08/2006	30/11/2006
D7.2.1	Evaluation report	31/08/2006	30/11/2006
D7.3.1	Final deployment	31/08/2006	30/11/2006
D8.2.1	Evaluation report	31/08/2006	30/11/2006
D8.3.1	Final deployment	31/08/2006	30/11/2006
D9.3.3	Final Prototype	31/08/2006	30/11/2006

In addition, the consortium plans to provide an additional deliverable in the form of a *Provenance Methodology and Best Practice* document that will be delivered on 31/11/2006. All other deliverable dates are unchanged.

The content of the rest of the Annex 1 has not been modified. However, whenever there is a reference to month 23, other that the ones previously mentioned, they now refer to month 26 of the project

The PROVENANCE project receives research funding from the European Commission's Sixth Framework Programme