



Title: Three Month Report

Author: John Ibbotson

Editor: IBM United Kingdom Limited

Reviewers: Project partners

Identifier: None

Type: Internal Deliverable

Version: 1

Date: Tuesday, January 11, 2005

Status: Draft

Class: Confidential

Summary

This internal deliverable provides a brief report for the first three months of the Provenance project. Requested by the EU Project Officer, it describes the activities that have been taking place since the delayed start of the project on the 1st October 2004. Following an introduction, a brief review of each workpackage is provided.

Members of the PROVENANCE consortium:

IBM United Kingdom Limited	United Kingdom
University of Southampton	United Kingdom
University of Wales, Cardiff	United Kingdom
Deutsches Zentrum für Luft- und Raumfahrt s.V.	Germany
Universitat Politecnica de Catalunya	Spain
Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézet	Hungary

Foreword

At the request of the EU Project Officer, this document provides a brief report for the first three months of the Provenance project.

Copyright © 2004 by the PROVENANCE consortium

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

File name: GRID_PROVENANCE-GEN-M3-ThreeMonthReport-v0.sxw

Table of Contents

Introduction.....	5
Workpackage 1 – Management.....	5
Workpackage 2 – Requirements.....	5
Workpackage 3 – Architecture.....	6
Workpackage 9 – Implementation.....	8
Workpackage 10 – Collaboration.....	9

1 Introduction

The Provenance project started on the 1st October 2004. The contractual starting date was the 1st September 2004, but the delay was due to completion of the Consortium Agreement and subsequent contract signing. The project has proposed to slip two deliverables in workpackage 2 due to the delayed start. We expect to be able to meet the contractual deadlines for the other project deliverables.

The requirement workpackage (WP2) has completed its scheduled activities so far with the requirements questionnaire being completed and published on an external website. As part of its development, the questionnaire was completed internally and feedback used to modify the questions before making it publicly accessible. An early draft of the Organ Transplant Management (WP8) scenario was prepared as a way to identify their requirements. Full details of the WP2 activities are provided in a separate section.

At the kick-off meeting, it was decided to bring forward some of the activities in the Architecture workpackage (WP3). This was to provide a structure for the pre-prototype implementation (WP9) and to inform the later development of the architecture. Full details of the Architecture and Implementation workpackages are provided in later sections.

There have also been activities in the Management (WP1) and Collaboration (WP10) workpackages. No work has started on other workpackages.

2 Workpackage 1 – Management

The Provenance project kick-off meeting was held on 25-27 October 2004 at the IBM UK Laboratories. At this meeting, the working processes of the project were agreed together with a schedule for the first six months. Key points agreed were:

1. All project documents would be managed via a twiki. This has been created at <http://twiki.gridprovenance.org> with public and restricted areas. The restricted area is for project partners only. The restricted part includes areas for meetings, workpackages, deliverables and management.
2. Monthly management telcons are held with agendas published on the twiki. All meetings are minuted with actions tracked – again this is via the twiki.
3. Dates have been agreed for regular face to face meetings during the project lifetime. The next meeting will be in Barcelona at UPC in February 2005.
4. Standards were agreed for documents generated by the project. Where possible, they will conform to open standards and use the OpenOffice suite of tools.

3 Workpackage 2 – Requirements

Activities:

At the kick-off meeting, the mornings of 26 and 27 were dedicated to workpackage 2. All partners actively participated in the WP2 session. The results of the WP2 session at the kick-off meeting were:

1. a schedule for WP2 until the next face to face meeting was accepted,
2. initial work on requirements questionnaire was presented,
3. the content of the requirement questionnaire was discussed,
4. contact persons were assigned for each project to be interviewed.

The schedule of WP2 accepted at the kick-off meeting:

Draft 1 of questionnaire to include partner input		05/11/04
Draft 2		12/11/04
Internal final draft		19/11/04
UPC and DLR complete questionnaire		10/12/04
eDiamond and myGrid complete questionnaire		10/12/04
Publication of questionnaire		19/12/04
Identification of user requirements		31/01/05
User requirements doc (D2.1.1)	(31/01/05 in plan)	28/02/05
Internal draft of Software requirements		28/03/05
Software Requirements (D2.2.1)	(31/03/05 in plan)	28/04/05

Draft1 and Draft2 versions of the questionnaire were produced on time. The feedback from project partners was discussed in a telephone conference call on 18/11/04. As decided at the telephone conference, the internal final draft of the questionnaire was published on a password protected web page on 24/11/04. The on-line questionnaire was produced with an in-house developed form generator.

The internal final draft of the questionnaire was filled in by five projects: the TENT system by DLR, the Organ Transplant Management application by UPC, the eDiamond project by IBM, Combechem and MyGrid projects by Soton. The partners filling in the questionnaire also gave feedback on the questionnaire. The answers and the feedback from these projects were analysed and the modified final version of the questionnaire was published on 17/12/04.

The list of projects to be contacted for requirements input was collected. The first contact email is to be sent to them on the week of 20-23 December.

A template for the User Requirements Document was created. This template defines how to describe the user requirements collected in January 2005.

Results:

5. Detailed work plan for WP2
6. Published questionnaire containing important aspects to be investigated when collecting the user requirements for the provenance architecture
7. First contact with other grid projects to involve them in the requirements capture process
8. Template for the User Requirements Document with a plan how to describe the user requirements.

Assessment of the progress with respect to the DoW:

Since the project started later than planned and WP2 produces the first deliverables of the project, the initial schedule had to be updated. The updated deadlines were met and WP2 has made enough progress to meet the updated deliverable deadlines.

4 Workpackage 3 – Architecture

The focus of the architecture workpackage during the first three months has been on the design of the pre-prototype architecture. The goals of the pre-prototype are:

1. to be a simple example to understand

Copyright © 2004 by the PROVENANCE consortium

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

2. to exhibit interesting examples of provenance use
3. to inform the future design of the architecture
4. to help build, setup and deploy a development environment for the Project.

We opted for a small example that shows how to bake a cake (Victoria Sponge). It consists of a simple workflow:

1. Whisk together the butter and sugar until light and creamy
2. Beat the eggs for a certain duration and add to the whisked sugar and butter
3. Fold the flour into the mixture and add the flavoring (vanilla, lemon)
4. Put the wet dough into the oven and bake for a given time at a given temperature

Despite its simplicity, this application leads to interesting questions that require the use of provenance in order to be answered. We have identified a preliminary list of these:

- Were the correct ingredients used at the correct step?
- Where was the longest time spent in for preparation?
- Why did the cake taste wrong? (answers can be related to Altitude, wrong temperature, eggs no beaten enough, proportion of ingredients)
- Was the proportion of ingredients right for the size of the cake?
- Did the baker follow the users' instructions (regardless of any claims from the baker)?
- Did the services follow the users' instructions (regardless of any claims from the services)?

Such questions can be answered if provenance is recorded, specifically consisting of:

- *interaction provenance*: the trace of all service interactions during workflow execution;
- *actor provenance*: a trace of internal service states at specific moments of execution.

After specifying the application's behaviour, we have designed its architecture. Specifically, we identified the services involved, their interactions (defined as sequence diagrams), and their interfaces (as wsdl files). The whole application data domain was defined as an XML schema. From a provenance viewpoint, we adopted (as indicated in the proposal) an existing implementation of a provenance service, designed as part of the pasoa project (www.pasoa.org), which is capable of recording actor and interaction provenance.

Implementation work has begun, split between IBM and Southampton. IBM are implementing the different services of the application, while Southampton focuses on the implementation of the provenance queries. In order to proceed with the later, without having an application available, a

Copyright © 2004 by the PROVENANCE consortium

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

dummy client simulated the whole cake baking process was created in order to submit all the necessary provenance information to the provenance service. A first provenance query verifying the baking temperature has already been implemented. Regular meetings have been held to coordinate the implementation and architecture activities.

5 Workpackage 9 – Implementation

The focus of the implementation workpackage in the first three months is two-fold.

- To put in place a development infrastructure that can be shared amongst the partners
- Development of a pre-prototype

The development infrastructure will include the following components:

- a. A CVS server to manage code developed by the partners
- b. A build server situated in IBM Hursley

The development code is to be managed in a CVS repository at the University of Wales, Cardiff. IBM does not allow access to internal systems except in extreme circumstances so this site was chosen to provide easy access to non-IBM partners. The machine in Cardiff has been configured and is being used by the IBM developers.

The development process to be used by the project will focus around daily builds of the code together with regression unit testing. Build scripts are being prepared which will pull the development code from the CVS server onto a build machine at IBM Hursley. The build process will then complete and run a set of regression tests on all modules. The regression tests will be written with JUnit. Following the build, a report will be prepared which can be accessed by all developers. The built system together with reports will then be pushed onto a system external to IBM for use by the development team.

To support the development process, coding and information standards have been prepared. A draft copy of the [HandBook](#) has been circulated to the project partners for comment.

To test the development process, a pre-prototype has been specified by the architecture workpackage team (see their draft report). During discussions between IBM and Southampton, we established a number of principles for the Provenance implementation:

1. The Provenance implementation would not impose a development environment on anyone that wanted to use the released code
2. The Provenance code releases would not include third party pre-requisite software. For IP reasons, users of the Provenance code will have to download pre-requisite software (such as Apache Tomcat and Axis) separately

Copyright © 2004 by the PROVENANCE consortium

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

3. The Provenance code will make use of Apache Ant build scripts to compile and deploy the system on an application server
4. The Provenance code will be developed using the Apache Tomcat application server and Apache Axis SOAP support
5. The Provenance code will be developed using the Java language to provide portability across systems

6 Workpackage 10 – Collaboration

The following activities have been completed to support the collaboration workpackage:

- a. We have produced our draft collaboration workplan that was delivered to the EU at the end of October 2004.
- b. Provenance (Luc Moreau) will attend the first meeting of the European Grid Standards Co-ordination Group, on January 28, 2005 at European Microsoft Innovation Centre in Aachen, Germany.
- c. The Provenance project will contribute a “Provenance” education session at the core grid summer school
- d. Provenance (John Ibbotson) has attended meetings of the key indicators working group with the EU in Brussels and has contributed to the working group report.