



Core User Directory Project

Request to seek confirmation of funding

Mike Fraser, Paul Jeffreys

29 June 2008

Ref: ODIT.40

The purpose of this document is to seek confirmation of funding from the Budget Sub-Committee of up to £123k for the Core User Directory Project, to record progress made in the first 6 months of the project and to present plans for the second half of the project.

Introduction

In October 2007, Council GPC gave its support for a project to plan, design and build an enterprise-wide Core User Directory (CUD). Availability of funding for the project, to be drawn from the Corporate part of the BSP five year plan, was confirmed by Michael Sibly and Val Johnston (email 5 November 2007). The Core User Directory Project lays the foundation for a future programme of activities to deliver an overall Identity and Access Management solution for the University of Oxford. The Core User Directory Project consists of two parts: a requirements gathering exercise to determine the needs of University administrators and ICT service providers in terms of their management and use of user attributes, and the deployment of a University Core User Directory. The first phase of the project has run from Jan 2008 to June 2008, and the report from the requirements analysis will be published in July.

The CUD Project is a short-term project, lasting only one calendar year (1 Jan 2008 – 31 Dec 2008), divided equally between requirements and implementation, and with an average of 1 FTE allocated to the project. The project plan focuses on what can be achieved in this period. The CUD Project is not, for example, intending to address the much wider spectrum of identity and access management issues. However, expert opinion suggests that having a CUD in place is a *sine qua non* of any coherent identity management programme.

The CUD Project is steered by the CUD Working Party which meets monthly and includes representation from across the collegiate University (<http://www.ict.ox.ac.uk/odit/projects/coreuser/>).

Overall Aim

The Core User Directory project aims to bring coherence to the way the University manages information relating to people. A Core User Directory (CUD) will hold sufficient details of individuals in order to uniquely identify them, and which can be used to assign a unique identifier for each record held. The role of the CUD is not to replace any of the existing databases but to enable more effective management of the data as it flows between different databases. The Project aims to make available a pilot University CUD by the end of Dec 2008. The pilot CUD will aggregate identity attributes from a number of authoritative sources and provide one or more interfaces to enable read-only lookup queries in the first instance.

Key Drivers

- Our administration processes that relate to people and their identities are complex, costly, potentially error-prone and do not enhance the experience of someone joining the University community. Errors to do with access to services are hard and time-consuming to troubleshoot and rectify.
- It is difficult, cumbersome, sometimes impossible, to join together identity attributes held in heterogeneous databases in order to provide key ICT services, whether centrally or within departments and colleges.
- Other Universities are significantly further ahead in resolving these problems than we are.

Benefits

- A clear understanding of key administrative processes and applications that manage or make use of user attributes, and how those processes could be streamlined and automated in some areas to simplify provisioning and de-provisioning of services to those individuals, given the use of suitable technology.
- A pilot Core User Directory comprising a minimum set of attributes and globally unique identifiers to serve as a hub for the future development of an Identity and Access Management framework for the University.

Outline Schedule

The schedule of activities from 1 Jan to 31 Dec is as follows (with an indication of work completed to date):

Requirements Gathering (Jan to Jun 2008, Completed)

- Month 1: Identify dedicated staff to undertake requirements gathering; specify detailed work-plan for first 6 months; set up interviews; gather information for initial use cases/scenarios.
- Months 2-4: Identify interviewees from amongst selected identity data providers and users. Create use cases/scenarios built on needs identified across the University.
- Month 5: Hold a facilitated workshop for the members of the CUD working party. Appraise the use cases/scenarios generated in months 2-4, extract requirements for the CUD (employing external expertise), and prepare for work in months 6-12.

Implementation Phase (Jul to Dec 2008, Proposed)

- Months 6-7: Agree internal partners comprising a mixture of enterprise and local data managers; and, where different, early adopter CUD users. Agree initial set of core attributes; develop attribute schema and associated provenance metadata schema; draft release and privacy policies. Understand the business processes associated with each attribute and model the flow of data into the CUD; identify reconciliation points.
- Months 8-9: Develop and document the processes (automated and human intervention) to reconcile data feeds consistent with the agreed authoritative sources of any given attribute. Design the system architecture for the pilot CUD, including database repository, input interfaces, output interfaces; database structure; data format standards etc.
- Months 10-12: Deploy backend CUD database repository. Deploy LDAP interfaces, building on Oak project. If resources permit, deploy secure but simple web-based ldap client to allow read-only lookups by authorised persons (e.g. Unit administrators). Pilot service with selected applications (e.g. Student System, OULS cardholders; OUCS Registration).

Deliverables

Outputs from the project include the following:

Requirements Gathering Phase

- A report that specifies use cases/scenarios of providers that manage data relating to people or require access to attributes relating to people, specifies work flows, determines which can be supported, and highlights deficiencies of the current arrangement [Completed].
- A report specifying detailed requirements from staff dealing with identity and an analysis of the implications for the CUD [Completed].
- A report from the facilitated workshop, which includes validation of the use cases [In Progress].

Implementation Phase

- Specification for a Core User Directory, in terms of its architecture, structure and the protocols it uses [months 7-9].
- Implementation of a pilot Core User Directory that will be tested as the hub for future identity and access management developments in the University [months 10-12].
- Interoperation of the implemented Core User Directory with a selection of key enterprise application databases [months 10-12].
- A report that describes the implementation of the CUD and its interoperation with a selection of enterprise databases [month 12].

Findings from Requirements Gathering Phase

The scope of the implementation phase has been defined by the findings of the requirements gathering phase, augmented by a workshop arranged for the CUD Working Party in May 2008.

The requirements gathering phase of the project comprised some dedicated effort to interview key stakeholders (data managers and users) and compile a set of use cases. During this phase a workshop was held which looked in more detail at selected use cases and discussed the principles and steps necessary to initiate the implementation phase.

Requirements Gathering Report

Notable key findings from the requirements phase that potentially have a direct impact on the implementation phase include:

- Attributes tend to be passed from one system to another, making it unclear which system should be considered the authoritative source for any given attribute;
- OUCS already performs CUD-like activities such as reconciliation on behalf of other units, including within Central Administration;
- The CUD should be able to hold multiple status/affiliation data per identity record;
- Local units (departments and colleges) have to maintain databases about individuals with whom they have a relationship but the nature of whose role means they are not recorded in the enterprise systems.
- The University Card database is treated as an authoritative source of identity records, with the consequence that workarounds are required for individuals associated with the University but lacking a card status;
- For some processes the data is available but held in separate, unconnected systems, necessitating a time-consuming manual lookup process prone to error.

The Report recommended that for the implementation phase the project:

- establish the initial set of core user attributes, including the generation of a persistent, globally unique identifier (Core User ID);
- identify additional attributes, particularly attributes shared amongst various systems;
- identify the authoritative sources for each attribute;
- address the multiple status/affiliation issue.

Workshop

The CUD Project held a workshop facilitated by an external consultant in May 2008. The workshop looked in greater detail at two use cases, one from the perspective of a college and the other from the perspective of a central service provider. The key issues discussed at the workshop included:

- The operation of the CUD comprises two distinct activities: the 'cloud' activity involving data cleaning, reconciliation, etc; and the interfaces for data feeds in and out of the CUD. It was agreed at the workshop that for the pilot, at least, it would be sensible to release an LDAP interface.
- It was observed that for some attributes (e.g. Relating to status @ affiliation) colleges were one of the authoritative sources. The workshop agreed that the CUD pilot should directly address the multiple affiliation issue.
- The data 'cloud' activity, regardless of system deployed, would inevitably be a mixture of fuzzy logic and manual record checking.
- Above all the workshop was of the opinion that the implementation phase was a period of experimentation – that commencing on a solution was more important than a continued refinement of the CUD definition.

Implementation Phase Summary

In summary, both the requirements report and the workshop concluded that the implementation phase should:

- establish the initial set of core attributes, including a unique id;
- build on, and extend, the cloud activity undertaken by OUCS (Registration);
- build on, and extend, the provision of an LDAP service by OUCS (Oak);
- ensure that the pilot tests the reconciliation and provision of data from a representative; sample of data sources (examples given have included: selected colleges, HR/payroll, OSS, OULS, OUCS, Alumni);
- investigate (if not provide a solution for) the multiple status @ affiliation issue.

The workplan for the implementation phase of the project is given in Appendix A.

Budget

The staffing costs for each phase are calculated to allow for a mixture of payscale and consultancy-based daily rates. The project has completed the first phase of the project with 0.5 FTE allocated from both OUCS and BSP (in effect from March to June 2008). It is proposed that the underspend from the first phase is carried forward to the second phase. **The total request for the full 12 month project is: £122k.**

	Budget	Spend to Date (Jan-Jun, estimated)	Balance	Notes
Staff costs				
<i>Requirements Phase (Jan-Jun)</i>				
Business Analyst	37,500	21,300	16,200	1 FTE shared between OUCS and BSP
<i>Implementation Phase (Jul-Dec)</i>				
Project manager	14400	0	0	
Data Manager	31500	0	0	
Systems Developer	40500	0	0	
Technical author	6750	0	0	
Non-staff costs				
Workshops	10,000	3,000	7,000	One workshop in each phase with budget for venue and Facilitator costs (£5,000 each)
Totals	140,650	24,300		

Amount carried forward from phase 1 to phase 2			18,200	
Total requested	122,450			

Appendix A: Workpackages for Implementation Phase

Overview

The workplan to deliver the implementation phase of the project will deploy the following workpackages:-

Workpackages

Workpackage	Months (Jul-Dec 08)	Description	Estimated Resources (person days)
1. Project Management	Jul-Dec		22 days
2. Selection of initial data providers and users	Jul	Agree internal partners comprising a mixture of enterprise and local data managers; and, where different, early adopter CUD users.	10 days
3. Attribute schema, provenance, and release policy	Jul-Sept	Agree initial set of core attributes; develop attribute schema and associated provenance metadata schema; draft release and privacy policies.	25 days
4. Data flow design	Jul-Aug	Understand the business processes associated with each attribute and model the flow of data into the CUD; identify reconciliation points.	20 days
5. Data reconciliation	Sept-Dec	Develop and document the processes (automated and human intervention) to reconcile data feeds consistent with the agreed authoritative sources of any given attribute.	40 days
6. System Architecture	Aug-Sept	Design the system architecture for the pilot CUD, including database repository, import and export interfaces; database structure; data format standards etc.	15 days
7. Deployment of Database Repository	Oct	Deploy backend CUD database repository and data import scripts	30 days
8. Deployment of LDAP interface	Oct-Nov	Deploy LDAP interfaces, building on Oak project	20 days
9. Deployment of secure web interface	Oct-Nov	Secure but simple web-based ldap client to allow read-only lookups by authorised persons (e.g. Unit administrators). This workpackage is provisional.	tbc
10. Communications and support	Jul-Dec	Development of documentation; consulting and communicating with various stakeholder communities.	15 days
11. Evaluation	Sept-Dec	Software testing will be built into the deployment workpackages. This workpackage is designed to gather	10 days

	feedback from data providers and users alike. The results of the evaluation will determine subsequent phases of the project.	
Total days		207 (1 FTE at 6 months is 110 working days)

The resource allocation is currently provision and amounts to 1.88 FTE (excluding workpackage 9). The project has an underspend from the first phase which will be carried forward to contribute to the implementation phase.