Efficient Technical and Organizational Measures for Privacy-aware Campus Identity Management and Service Integration

Latifa Boursas*, Wolfgang Hommel**

*Munich Network Management Team, Munich University of Technology, Germany
boursas@tum.de

**Munich Network Management Team, Leibniz Supercomputing Center, Munich, Germany
hommel@lrz.de

Abstract

We present intermediate results of IntegraTUM, a project realized by the Technische Universität München (Munich University of Technology, TUM) and the Leibniz Supercomputing Center (LRZ) under partial funding by the German Research Foundation (DFG), which is intended as a prototype for the higher education institutions in Germany. At its core, a modern and scalable Identity & Access Management architecture has been developed to integrate both centralized and decentralized campus IT services and resources. We discuss the technical innovations as well as the organizational measures, and focus on the IT support of the university business processes as well as the user-friendly and privacy-enhancing interfaces for staff and students.

Keywords: Identity Management, Integrated Information Systems, University IT Strategy.

1. Introduction

Traditionally, the services and resources, which higher education institutions (HEIs) offer to their staff and students, have been developed and are being maintained independent from each other. This applies to both centralized services, such as the library, web portal and e-learning system, as well as services established decentrally, e.g. computer pools for students, file and e-mail servers, which are organized on a faculty or even chair level. This lack of coordination leads to redundant work, data inconsistencies, a higher total cost of ownership and deficiencies in usability. For example, students have to sign up for each service individually, remember many usernames and passwords, and have to update their personal and contact information in many places, e.g. if their home address or study course changes.

The Technische Universität München (University of Technology, Munich, TUM), Germany, does not have a single campus, but is spread across three main sites and further locations in Upper Bavaria, resulting in a multiplication of these drawbacks. Additionally, several central IT services are not being offered by a local IT department, but by the Leibniz Supercomputing Center (LRZ), which is the common computing center of both Munich universities and the other HEIs in the area of Munich.

Although many HEIs have undertaken IT recentralization and integration projects in the past few years, several of them provide only partial solutions to the underlying problems, because for example only technical aspects have been considered, the coverage area of solutions is limited to the central services, or the political and organizational support is missing.

In IntegraTUM [1], a project partly funded by the German Research Foundation (DFG), researchers, technicians as well as managers from both the centralized and the decentralized services are working together on a major reorganization of TUM's information and telecommunication infrastructure. 18 researchers have been newly employed for the project, in which also about 30 additional staff members from TUM and LRZ are tightly involved. The primary goal is to integrate the available IT resources, optimize them for the underlying business processes and thus provide seamless and user-friendly services to staff and students of the university [3]. Furthermore, the solutions are being designed with flexibility in mind, so the reuse of the developed methods and tools by other HEIs is explicitly encouraged.

In this article, we present selected intermediate results after two out of five years of project duration [2]. In section 2, we provide an overview of the systems which we are integrating. Section 3 gives insight into the work of the TUM's CIO/IO board, which has been institutionalized as a high level committee to enforce the realization of the necessary changes throughout the faculties and facilities. In section 4, we discuss our approach to campus-wide identity and access management, which – in contrast to other projects – uses a network of multiple directory services as backbone for identity and account data transportation and workflow support. In section 5, we discuss the applied privacy and data protection mechanisms, closely related to the role of the central university web portal as interface to the system and service administrators as well as to the end users, which is explained in section 6. A summary of our

achievements so far and an outlook to our next work items conclude this article.

2. Project and Service Overview

The project aims at the integration of existing centralized und decentralized services; at its core, a campus-wide central directory service is being developed as part of an Identity & Access Management (I&AM) system, which links all university members' digital identities. A thorough analysis of the underlying business processes and identity data workflows has been performed, followed by the specification of optimizations, which integrate the use of this directory service in all processes and software systems. Stakeholders from nine subprojects, spread across the whole university, have worked together to specify all requirements, e.g. regarding data models, approval workflows, protocols and synchronization frequencies.

More than a dozen types of services, such as "a faculty's student computer pool", have been identified and will be connected to the I&AM system, either by directly accessing the central directory service or by being fed ("provisioned") by it. In the first stage of the realization, the following services will be integrated:

A) Central services, i.e.:

- The university library, which manages rather complex sets of rights to its various resources, e.g. access to selected volumes of scientific journals which is restricted to members of selected faculties.
- The e-learning system, which manages access to the content itself as well as various subservices, such as discussion forums for learning communities and file uploads for student homework.
- The university's web portal, myTUM [4], which on the one hand is the central information broker, e.g. for event announcements and students' tests results. On the other hand, it is a content management system for faculties and chairs which do not run their own web servers. Due to the importance of the information and services accessible through the web portal, timeliness and accuracy of the dynamic identity data synchronization pose very strict requirements on the central directory service.
- The student administration and human resources department supply the student and staff identity information as authoritative data sources. Additionally, a guest management system is being developed to handle guest professors and students, as well as conference attendees, who will receive temporary accounts, e.g. for personalized WLAN access.

Unfortunately, most of the software used for these central services does not support external I&AM systems directly, e.g. by providing an LDAP interface and supporting the correlation or linking of accounts. Thus, dedicated interfaces to these systems have to be implemented.

B) Decentralized services, i.e.:

- Faculties' computer pools and staff workstations. Faculty-level surveys [2] have shown that the Microsoft Windows and Linux operating systems are the most often used ones. Furthermore, while the number of machines running Solaris or UNIX derivates is decreasing, the number of Macintosh users increases. From the beginning, both Microsoft Active Directories as well as LDAP servers, such as OpenLDAP or Novell eDirectory, will be provisioned by the central directory service. Other user management methods, such as NIS, will deliberately not be supported centrally.
- File storage services. While the users' home directories are still typically stored on faculty- or chair-specific file servers, a central file server will provide additional features such as automated tape backups and long-term archiving. In the past, network connectivity issues often made users wary of central file servers. However, improvements in the network infrastructure, as well as the high costs for professional hardware and service administrators have led to an increased demand for centrally hosted home directories, which also allow browser-based access across the internet. Thus, this project focuses on the recentralization of file servers by providing seamless migration paths.
- E-mail servers. Much like to file servers, many faculties and chairs have set up their own e-mail servers in the past and are now facing the same cost issues. From the user perspective, multiple unrelated mailboxes for their @tum.de and @faculty.tum.de e-mail addresses had to be maintained. For students with minor fields of study, who also worked as assistants at yet other faculties, this often led to bad usability and technical problems such as cyclic e-mail forward chains. This project aims at the recentralization of e-mail servers, offering only a single mailbox for which an arbitrary number of faculty- or project-specific alias e-mail addresses can be configured, based in part on self services and on decentralized mail administrator interfaces.

In a further step, additional services such as the campus facility management and the telephone system will be added as data sources. Furthermore, additional target systems, such as the alumni administration and various standalone web-based applications, will also be integrated.

One important benefit of these centrally administered digital identities will be web single sign-on (WebSSO) that supports the campus-wide authentication and authorization for all connected services. Thus, users will have to remember only one username/password combination for all services and can switch between the web-based applications without re-authentication, hopefully resulting in a noticeable decrease of password-

related service help desk calls.

3. The CIO/IO Board

A committee, headed by the vice president of TUM as Chief Information Officer (CIO) with representatives (IOs) from each faculty and facility (e.g., library and administration), has been established to specify and approve the university's IT strategy and measures. The quarterly meetings are also attended by representatives of the employee committee and the LRZ.

The CIO/IO board has been granted decision-making power by the university's leadership, as well as the authority to issue directives to the facilities concerning all technical aspects of the TUM's information and communication infrastructure. Thus, discussions and decisions which previously required to bring together many people spread across the whole university, can now be made very efficiently, and also the time spent to implement agreed decisions has noticeably improved.

Reports about all major IT-related projects are also made to the CIO/IO board, resulting in a better coordination between them and the exploitation of synergies. Furthermore, to enhance the support for staff and students, a TUM-wide IT service support infrastructure based on ITIL concepts is currently being planned and realized, with service desks at the three major TUM campuses [2].

4. Campus Identity & Access Management

An Identity & Access Management system, based on the common central directory services, is the backbone of the new IT infrastructure. Authoritative data sources such as the student administration and human resources administration systems feed these directory services, where the data is being aggregated, correlated and refined. Correlating data records from different sources, which is required in order to avoid multiple accounts per user, turned out to be rather difficult due to typing and transliteration errors in names and birth dates especially in foreign students' records. Thus, fuzzy-logic search mechanisms have to be implemented and integrated into data source systems such as SAP HR. Destination systems can then either directly access the central directory services or are being provisioned by it through so-called connectors.

Architecture

Unlike many other HEI identity management projects, the IntegraTUM directory service uses multiple LDAP servers which provide the data not only in one format, but in various, i.e. the data can be delivered to end systems in the required format and all necessary transformation steps are performed within the identity management system, resulting in transparency and more efficient change management for the connected services. Additional replication and clustering mechanisms provide the high availability required for the mission-critical directory components, such as the central authentication server.

A dedicated data schema has been developed for the backbone LDAP servers, because wide-spread standard LDAP objectclasses such as eduPerson [5] lack several attributes required for our purposes, and no standardised German or European alternative does exist yet. Thus, attribute mappings as well as syntactical conversions have to be performed when importing data from the source systems, as well as when exporting data to the provisioned services. Also, the internal structure of the LDAP server, the so-called Directory Information Tree (DIT), has been defined differently to most existing directory services. As can be seen in figure 1, we break with the traditional convention to store user objects as leaf objects, and instead turn user objects into container objects, so further objects representing the user addresses or student study course details can be stored beneath them. This design avoids problems encountered in previous directory services, in which only a limited number of addresses could be stored in the user object itself by using multiple address attributes (address1, address2, …), and also avoids the design flaw of turning an LDAP server into a pseudo-relational database by storing address objects in a separate branch of the DIT and linking user objects to their address objects.

Our implementation is based on Novell's eDirectory and

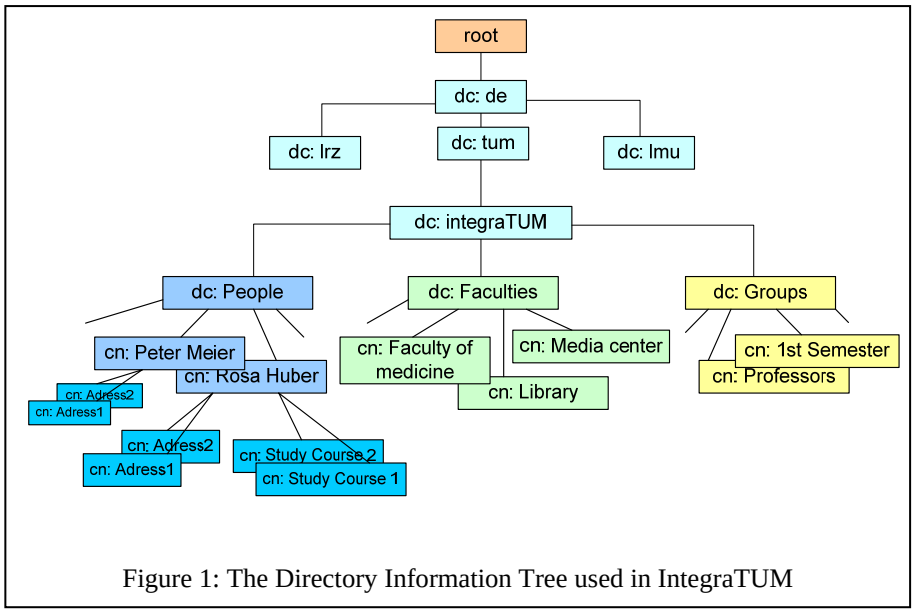

Figure 1: The Directory Information Tree used in IntegraTUM

Nsure Identity Manager software, which uses an XML-based internal data representation. Conversion rules can be applied within each service connector based on standard XSLT style sheets, which optionally can make use of Java code. This extension hook is also used to generate the users' login names and initial passwords.

Unlike many other LDAP servers, including open source ones such as OpenLDAP, Novell eDirectory works event-driven. Thus, basic LDAP operations, such as adding, modifying, and deleting attributes or objects trigger events, which in turn are used to trigger the synchronization with the other LDAP servers in the backbone (as they use different data schemas, simple replication cannot be used) and to optionally initiate approval workflows or send signals to other systems which are not directly connected to the LDAP servers.

In order to avoid interdependencies among the systems, each service's connector to the central directory service is intended as its single interface to the outside world concerning all user-related information. Consequently, the directory backbone does not only broker information from the primary authoritative data sources, but also works as a user information transport medium between the connected services; for example, the user frontend for the e-mail service is running within the TUM's web portal, thus configuration changes, such as setting a new e-mail forwarding address, which are made in the web portal, have to be propagated to the e-mail system. Therefore, additional user attributes had to be added to the central directory service to facilitate this data exchange.

Figure 2 shows our directory backbone: At its core, a central meta-directory is used to synchronize the data of all connected directory servers, which we call satellite directories. Presently, three satellite directories are being implemented:

- The administration satellite directory aggregates and correlates the data from the authoritative sources, i.e. student administration, human resources department and the guest management system. Importing the data requires transformations and conversions, e.g. of date and telephone number formats, faculty and country names, and codes representing the persons' status, e.g. being a freshman or alumnus; an example is given in table 1. Furthermore, the default privacy policies are applied to set the appropriate LDAP access control lists.
  Automatically correlating data records from multiple source systems is error-prone, e.g. due to typing errors, if it is only based on matching attributes, such as names and birth dates. Due to the lack of a common key attribute in all systems this correlation has to be done semi-automatically once for the already acquired records. However, the underlying business processes in the human resources department are being adapted, so new employees who were at TUM previously, e.g. as students, are registered with their existing identifier.

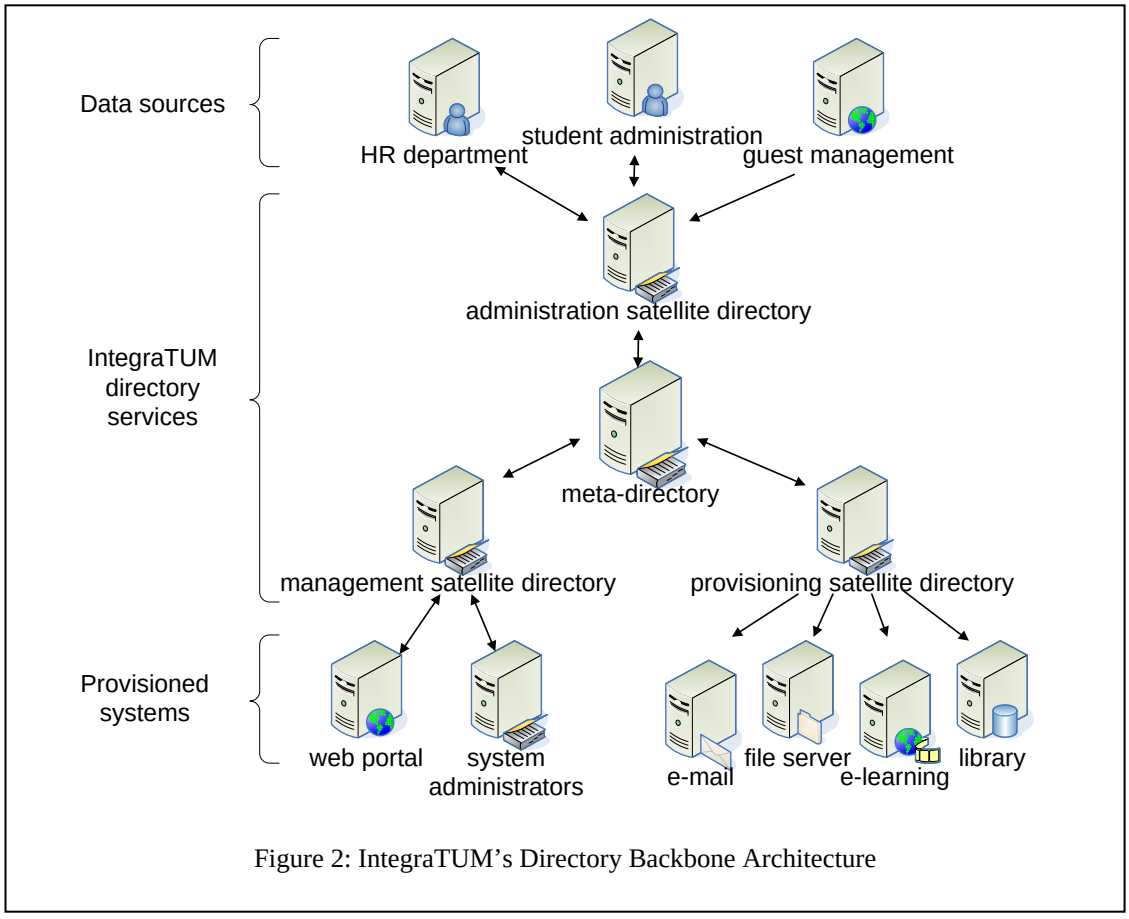

Figure 2: IntegraTUM's Directory Backbone Architecture

Retrieving data from the student administration and human resources software systems adds some complexity to the overall architecture: For security purposes, no access to these systems is possible from outside networks; thus, a so-called transfer server has been established in connection with the administration satellite directory, to which data exports are pushed from the source systems using an SSL-encrypted connection. Similarly, changes made to the data can be pulled from this transfer server and will be re-imported in the respective systems. Several sophisticated tools have been developed for validating data checksums and perform pre-processing steps for the LDAP import, which we also make available to other HEIs.

Service	Date format
Satellite Directory	2006-01-15
Student Administration	15 Jan 06
Guest management	January 15th, 2006

Table 1: Different date format in different systems

- The management satellite directory interacts with services which are allowed to change parts of the data, i.e. the central web portal which hosts the user self services and the faculties' systems administrators' services. As the administration is decentralized, and the fluctuation among administrators is considerably high, dynamic access control list management has to be implemented. To this extent, Novell's eDirectory has the advantage of storing directory service access control lists in the directory itself, as opposed to other products in which configuration files must be edited manually for each change. Plausibility checks and auditing mechanisms ensure the quality of the modified data.
- The provisioning satellite directory is the gateway to the other connected systems and services. For example, the e-learning system's internal user database is being fed through a connector attached to this directory, automating account creation, modification, and deletion while still leveraging its internal resource access control methods. Although some of the connected systems use LDAP servers as user databases, they typically require their own LDAP server products and data models. Thus, data still needs to be converted within the connectors, and also the rules concerning which objects have to be created where in the target system's DIT, known as placement policies, have to be adapted for each service.
  Again, the initial synchronization requires additional efforts, in order to correlate those accounts, which already exist in the target system, with their respective objects in the central directory services.

Combined with replication and a dedicated placement strategy for replicas, this architecture is very flexible, extensible, scalable and fault tolerant, e.g. if further services and a larger number of computer pools will be connected later.

Implementation

We are implementing the central LDAP server, which has the role of a meta-directory in identity management terms, using Novell eDirectory 8.8 on the SuSE Linux Enterprise Server 9 operating system [6]. The HEIs in the greater Munich area have a long tradition of using eDirectory, previously on Novell Netware, so extensive experience and know-how is already available. For synchronizing data between the connected system nearly in real time, Novell's Nsure Identity Manager 3 (formerly known as DirXML) is used. Besides extensive customizing options, it offers extension hooks for Java code.
Various connectors to systems such as SAP HR can be used almost out of the box: Figure 3, taken from Novell's web-based management front end, iManager, shows the various stages, which data from the source system passes through: After an event, e.g. about the registration of a new employee, is generated, so-called filtering, matching, mapping and placement policies can be applied in order to precisely specify which data fields shall be taken from SAP HR, under which conditions, how they shall be transformed, and where in the eDirectory DIT they shall be placed.

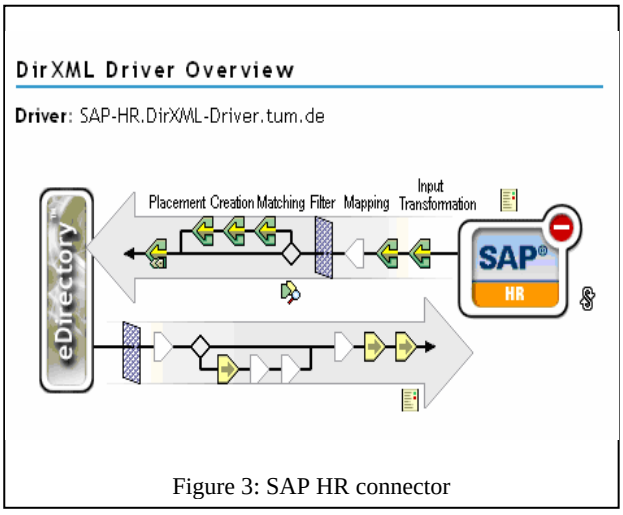

Figure 3: SAP HR connector

5. Privacy and Data Protection Mechanisms

The Identity & Access Management system described in the previous section aims at making user information available to the services and their administrators throughout the university. This obviously leads to privacy and data protection issues.

Furthermore, several services, such as mailing lists, e-learning and discussion communities, and file storage space for projects – as opposed to private home directories – are not being offered to individuals, but to arbitrary groups of persons.

Clearly, users must be able to control which services and other users are entitled to access which information about them. In general, trade-offs have to be considered between total anonymity, mandatory data required for service provisioning, and the use of optional features such as personalization and customization.

For example, course materials are typically made available to everyone, as long as access is restricted to university members, possibly further limited to students of certain study courses. However, for accessing these materials, users do not have to be personally identifiable.

On the other hand, participating in an e-learning course, with the option to take a test for the acquisition of study course relevant certificates, requires registration with the usual data, such as name and matriculation number. To ensure the correctness of this data and to avoid fake registrations, the identity information is retrieved from the central identity repository. The students must express their consent to this process in order to be able to use this service.

The TUM web portal, which is discussed in detail in the next section, provides a privacy control interface as part of its user self services. There, each user can specify which parts of her personal information are made available to whom. This task is facilitated by the pre-definition of groups such as “all service administrators” or even “everyone”.

The access control lists, which are derived from the user’s specifications, are stored in the central directory services. As all identity information flows through it, and the provisioned services and recipients are known, all privacy-relevant procedures and processes can easily be documented and the users can be given insight into how their personal data is being handled throughout the university.

This transparent strategy results both, in easier achievement of privacy and data protection law conformity, and in higher user acceptance, especially among the growing number of privacy-aware students of IT-related study courses, who have criticised the former procedures in the past.

Besides this user-centric privacy control, data protection is achieved by the exclusive use of encrypted network connections for the exchange of identity and account information. In most cases, this is achieved by using the SSL/TLS-enabled variants of the used protocols, such as LDAP+TLS or LDAPS instead of plain-text LDAP. Legacy software systems that only support plain-text protocols are connected using either SSL-encrypted tunnels (e.g., using the STunnel software or a SSH connection), or the whole machine is connected to a VPN using the IPSEC software, which is already used for campus-wide WLAN access.

6. The university web portal as interface to administrators and end users

In 2003, a university web portal called myTUM [4] has been established for about 20.000 students and 8.000 faculty members. It is based on the Zope/Plone content management system and features centrally managed information brokerage as well as web hosting for the faculties and chairs. The Leibniz Supercomputing Center hosts its highly-available LDAP user database and e-mail servers.
The web portal’s role is strengthened within the IntegraTUM project as it becomes the primary interface to both end users and administrators of all integrated services. Several new services are being implemented as part of the project:

- A guest administration interface, which can be used to add guest lecturers and students, as well as conference guests, e.g. for temporary, but personalized WLAN and file server access. The access control system controlling read and write operations to guest accounts does not only have to support decentralized administration, but also to delegate them. For example, guests may move from one chair to another, or a guest’s data, which was initially entered by a professor, later has to be edited by the secretary.
  Currently, guest accounts automatically expire at the end of each term. Workflows are in place to notify the respective guest administrators in time by e-mail, so the guest account can be renewed for another term.
  As with the other data sources, identity correlation is of major concern, so that guests are not registered multiple times. Due to the large number of users allowed to register guest users, protecting the privacy of other users which happen to have the same name as the new guest is more complex than in the HR department, which has full access to all employees’ data.

- A group administration interface, which can be used to create groups consisting of individuals and other groups, and to apply for group-specific resources such as mailing lists and file server access.
  Several groups are being created and maintained automatically based on the users’ attributes, for example “all staff of faculty x”, “all students of study course y”. New groups are “owned” by their creator, who can appoint group administrators. Using these groups within other services can be limited to the group administrators, the group members, or arbitrary other groups, including “everyone”. Similarly, group administrators can specify who may send e-mails to the group’s mailing list, and whether those mailing lists shall be moderated or archived.

- The user self services, shown in figure 4, are revised, so changes made, e.g. to addresses, are not only stored in the web portal’s local user database but are also synchronized with the central directory services, which also propagate these changes back to the authoritative systems, e.g. in the student administration.

Due to the nature of the web portal software, these functions are not only available via web interface, but also

through XML-RPC, which can be used, e.g. by system administrators, within own programs and scripts to automate certain management batch tasks or to integrate this functionality into a faculty's own web site.

7. Summary and Outlook

In this article, we have shown the main challenges and our solutions regarding campus-wide service integration within the IntegraTUM project. A central Identity Management system, which is based on a backbone of directory services, supports the university's business processes and thus achieves a tight coupling of the connected systems and services, while it also provides better scalability and reduces costs for service administration and maintenance. User-friendly graphical interfaces for both end users and system administrators are being developed with a special focus on privacy management and group-specific applications such as mailing lists and file server access. We also described how organizational measures, such as the newly institutionalized CIO/IO board, complement the technical ones. Within the next 3 years, the coverage area of the Identity Management system will be improved: Presently, only two out of the 12 faculties prototypically participate in IntegraTUM, and additional data sources (e.g., the facility management database) and target systems (e.g., alumni administration) have to be connected.
Also, many business processes will have to be optimized to make use of the central directory services, within both the university's administration and the service providers.
Furthermore, a tighter integration of the TUM's Identity Management system with the other HEIs in the greater Munich area must be implemented. As an example, there are 15 study courses, such as medicine, in which students are enrolled in both Munich universities, LMU and TUM. By synchronizing the data between both university's Identity Management systems, data inconsistencies will be avoided and a unified login, i.e. a single username/password combination for all student-relevant systems at both universities, will be achieved.

Figure 4: myTUM Web Portal Screenshot

Acknowledgments

This work has been partly funded by the German Research Foundation (DFG) under contract WGI 554 975.

The authors thank the members of the IntegraTUM project team and the Munich Network Management Team for valuable comments on previous versions of this paper.

IntegraTUM is headed by the vice president and CIO of TUM, Prof. Dr. Arndt Bode (see http://portal.mytum.de/iuk/cio/).

The MNM-Team, directed by Prof. Dr. Heinz-Gerd Hegering, is a group of researchers from the University of Munich, the Munich University of Technology, and the Leibniz Supercomputing Center of the Bavarian Academy of Sciences. The web server of the MNM Team is located at http://www.mnm-team.org/.

References

[1] Arndt Bode et al., IntegraTUM Projektantrag, http://portal.tum.de/iuk/integratum/dokumente/index_html/CIO-TU_Muenchen.pdf, 2004

[2] Arndt Bode et al., IntegraTUM DFG-Bericht und Folgeantrag, http://portal.tum.de/iuk/integratum/dokumente/index_html/Folgeantrag.pdf, 2006

[3] Arndt Bode, IntegraTUM: Integriertes Informations-management an der TU München, Journal PIK - Praxis in der Informationsverarbeitung und Kommunikation, Volume 28 (2005) 3, pages 165-168, K. G. Saur Verlag, München 2005

[4] myTUM web portal, http://www.mytum.de

[5] eduPerson LDAP schema, http://middleware.internet2.edu/dir/schema/

[6] Novell eDirectory and Identity Management System, http://www.novell.com/products/edirectory/